Erasmus
School of
Economics

Master Thesis – Economics and Business Economics
MSc. Urban, Port, and Transport Economics

Superstructures of Steel or Networks of Inefficiency: Analyzing Modal Accessibility Across Europe

Jan Stanisław Biegun
Student number: 537154

Supervisor: Dr. Michiel Gerritse
Second Assessor: Dr. Frank van Oort

Final Version Date: 11.10.2023
Word Count: 14,774

Abstract

This paper studies the market accessibility of 900 European regions to answer whether market access through road or rail provides greater contribution to these regions. The research is innovative in its development of a unique database by combining several sources, allowing for a cross-sectional study of accessibility in two instances. An origin destination matrix is used to investigate the influence of location capacity and origin potential on value added per employee at an industry level. This data is agglomerated and tested using a gravity model. The research finds evidence the railway accessibility is a more efficient productivity enhancer relative to motorway access. The relationship of accessibility on value added is indicated to be non-linear with increasing and diminishing returns to scale for road and rail respectively, coupled with evidence of complementarity between the two. The evidence matches theoretically defined concepts and research of literature, while exploring novel methods for accessibility studies and data construction.

Keywords: Market accessibility, Multimodality, Gravity models, Non-linear regression, European accessibility, Transportation economics, Road transportation, Railway transportation.

Acknowledgment: First, I would like to extend thanks to my supervisor Dr. Michiel Gerritse for the continued guidance and inspiration, both in the 7 months of this thesis process and over the last year and a half, having motivated my choice of the MSc in UPT. Secondly, thank you to my mom and dad, who financed the past four years of my education and who have been incredible examples of hard work, commitment for my whole life. Your care, love and support have carried me thus far and beyond. Likewise, to my brother Frank, you always challenged me to do my best as a true pedestal of excellence in writing, mój najważniejszy i największy kibic. I want to thank my girlfriend, Anhelina for pushing me to be the greatest version of myself and for the more than half a year of encouragement during my writing process, I cannot thank you enough. Finally, to all my friends in Rotterdam and beyond who through good times and bad, always held me on my feet to get to the finishing line.

Table of Contents

1. Introduction

Transportation is prerequisite to the economy, like the mycelium of fungi funnelling the growth of green forest canopies, it represents the veins without which the organs it feeds are void. Transportation is crucial across all spheres of human activity; one can cite the billions of Euros invested into Eastern European (EU) highways in the early 2000's which drove the unprecedented economic growth of Poland and Romania (Persyn, 2022). Even when looking at the most irrational human behaviour of war, during the ongoing Russian aggression in Ukraine, the strategies of Russian and Ukrainian forces rely on logistic railway hubs and highways (Latschan, 2022). The issue of transportation and infrastructure is foremost to nearly all sectors of human economic activity. An innate limitation of our modern systems; affecting environmental, geographic, and economic potentials can be resolved if the optimal choice and level of accessibility is defined and achieved. This is easier said than done, transportation is one of the most discussed topics on the EU agenda, with this cauldron of disunity being directly connected to half of the 2019-2024 EU priorities. In inexpedient dialogues, reducing road transportation is protested with critique to the variable reliability and efficiency of its more sustainable alternatives (Dailey, 2022). This thesis will focus on how the availability of rail and road infrastructure benefits different industries through their respective accessibilities, contributing to this discussion.

Most transportation in the world is done by road, rail, waterway, or air which in the context of continental transportation is mostly performed by the first 3 by throughput (Ford-Alexandraki et al., 2022). Air transportation has low capacity and is dedicated for time sensitive goods, while waterway transportation is geographically specific and almost entirely freight oriented. Therefore, the two crucial and indeed most universal forms of transportation infrastructure and mode are road and rail. The most definitive difference lays in rail maximizing efficient and theoretically uninterrupted bulk and crowd movement, meanwhile road infrastructure allows for the individual and direct flow of goods or commuters. Nevertheless, the comparison is not apple to oranges in this regard, with the purpose to move from point to point, the modes of transport allowed for increased economic growth by growing output (Parks, 2022). While this is certain in almost any context, it is crucial to understand what the differences are, and what are the discrete benefits of each modality. Transportation can be best defined in terms of derived demand, the value of the ability to access demand for goods and services at another location. Thereby implying that there is a value to access, or strictly an accessibility, which in term places a real value on the intangible derived demand at a location, describing its ability to connect displaced points of interest. Transportation theoretically is in demand because of an existing primary demand separated by displacement, but physically transportation is extant and geographically (un)available.

Accordingly, this research will aim to use answer the research question: **"To what extent do benefits from market accessibility differ between road and rail across European industries?".**

This question is answered with real data, by looking at the availability and operationality of railroads and highways to cities, industrial complexes, manufacturing clusters, and agglomerations. There are conceptual advantages to modality: the availability of superior country roads or high-speed rail improves market access for high income and high human capital industries, while slower rail or highway access is better suited for cargo freight. The extent of this will be studied using a non-linear regression design in a cross-sectional study between regions in Europe.

1.1 Relevance

Rail is as previously mentioned, a perpetually discussed article on the EU desk and abroad. As seen in **Figure 1.1** investment into infrastructure is hardly homogenous, with a consistent pattern of reduced spending in rail being moved to road development. France has moved to banning short haul domestic flights under 4 hours due to emissions (Limb, 2023) and educational institutions in the Netherlands outwardly do not reimburse business flights to destinations reachable by train within 6 hours (Niet, 2023). While both countries have exceptionally developed railway systems the pool of actors interested in a rail driven Europe is transnational. Looking at the broad three seas region of central Eastern Europe, there is a deluge of investments oriented at expanding an extensive rail network for freight and HSR across Ukraine, the Visegrad group and the Baltic States (CPK, 2023). The interest in rail solution is visible between both developed and developing countries.

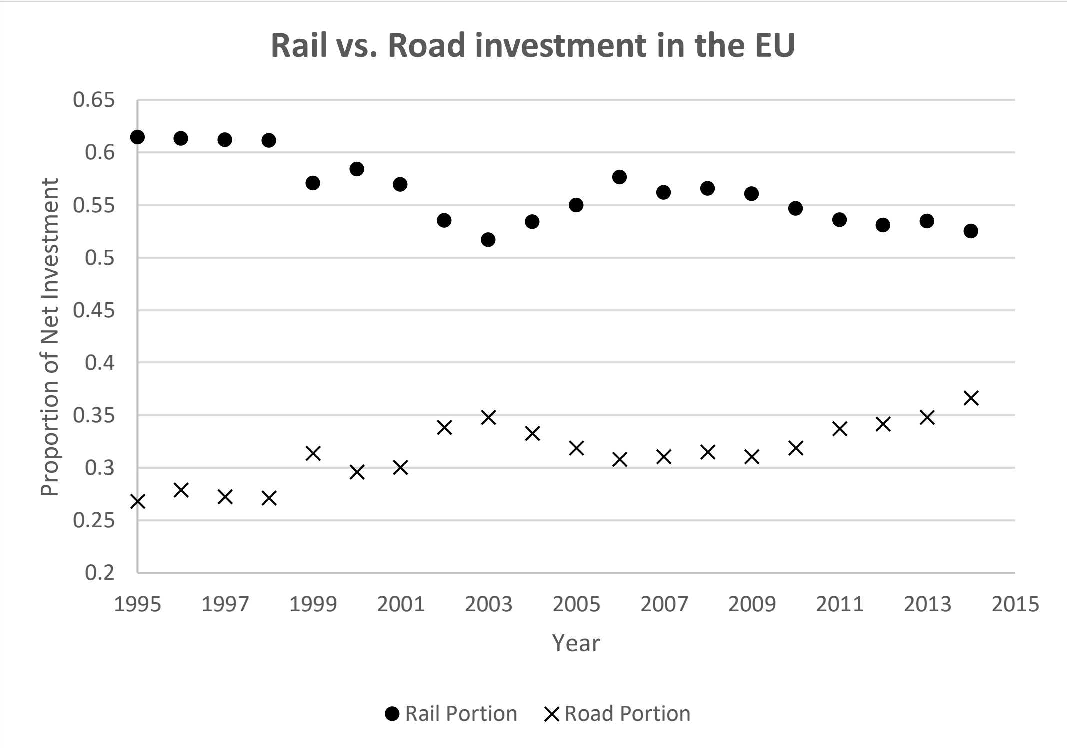


***Figure 1.1**: Rail and Road infrastructure spending in the EU as a portion of overall transportation investment.*

Despite the clear directions in many countries however, solutions are never this simple. In a country which is notoriously plagued by inefficient rail implementation, the German government has been in ongoing debate over shifting funding to road investments (Papatolios, 2023). The red tape bureaucracy in the German system amongst other problems means that roads offer superior real capacity then rail systems. Likewise, even European governments that stress environmental goals like Scotland are still prioritizing road travel instead of investing in their rail (Walton, 2022). While geographically this decision in a highland region is peculiar, low population densities are much better suited to road transport, due to the low traffic volumes.

The opacity of clear solutions is a clear obstacle for policy, strategy, business, and countless other fields due to the unresolved dialogues. Ironically, as seen in this prelude rail garners interest for its efficiency and sustainability compelling investment, even though it represents a much smaller volume of transportation in Europe. The implication of the former is that rail is a bad investment, a low return on investment despite its benefits. Such an answer would be hasty and unscientific, rather econometric analysis using market accessibility and real productivity statistics can numerate this problem and answer descriptively, what is the role of modal availability and potential in Europe. To answer this question best, this thesis focuses on Europe and studies market accessibility at a unimodal level to compare the discrepancies and forces at play across diversly specialized European clusters, providing insight for policy, and financial decision making by government and private entities.

1.2 Structure

This paper comprises several chapters, with an academic discussion on current and past literature that specify a context appropriate methodology to the problem which addresses the research question and embedded hypotheses. The theoretical review consists of chronological histories in model and theory development regarding modality, and market potential. The model development evolves out of this section adapting prominent studies to both the scope and availability of data in this work. A unique model is synthesized, and data is selected to best represent chosen variables empirically, which in turn justifies the chosen method. The results are analysed with literature and post-estimation to produce a reenforced and thoughtful conclusion. All findings are appropriately considered and evaluated to the highest relevant standard.

2. Theoretical Framework

One of the goals connecting urban and Transport economics is the ability to place value, and location in a quantifiable matric, physically defining the boundaries and extents of economic activity. One such topic relevant to the discussion from the introduction is rail capacity and integration for countries pivoting from automotive connectivity to railways. Otsuka et al. (2017), discussed the feasibility of developing intramodality in trans-European railway corridors with a specific focus on the Rotterdam Genoa Rhine-Alpine corridor. This paper focused on passenger and freight transport while contrasting de facto minimal share of European goods transport carried by rail despite its importance to the broader European economy. The paper listed priorities under 4 key factors for continued rail integration which are being financed by both local and EU stakeholders in the Rhine-

Alpine railway corridor. This brings two important pieces of information, firstly the immense potential of rail, and paradoxically its lack of integration.

Such dialogues, while qualitatively enriched tend to lack the substance of quantitative methods. “Relationship of Regional, Freight and Intermodal Market Access to Industry Location and Productivity” by Weisbrod & Goldberg (2022) gives a strong example of correct method design. A non-linear approach at estimating the relationships of market access on productivity and industry concentration allowed these authors to measure both efficiency and complementarity. Such non-linear approaches are currently preferred, as seen with the studies of Yi & Kim (2018) who like Weisbrod & Goldberg aimed to use adapted regressions. Past research used choice models and logistic regressions which leads to answers on modal choice or preference, but not necessarily efficiency. The newer paper looked at how labour markets, regional business and access to intermodal networks differently influenced industries. Its ability to categorize economic activity and adapt models similarly in both works serves as inspiration and a starting point.

One of the advantages of modern research is the present-day availability of information. The Weisbrod & Goldberg paper used open-source data and collected origin destination information from ticket services to establish a list of connections and general market access. Such a method is more feasible for most researchers and extends the definition of what tools one needs to perform superlative research in their field. While such a versatile adaptability deserves praise, when compared to the private and data intensive methods of Yi & Kim, the authors provide more precise estimates for comparing regional and local market accessibilities. Finally, the natural log-log regression used in both papers to study the elasticities of modal choice on concentrations inspired the research architecture for its context relevant ability to deal with skewed data and elasticities.

2.1 Location and Transportation

There is no simple answer to what the perfect modal splits are, 77.4% of European Cargo travels by road, as does over 80% of passenger transportation when measured by rail (Ford-Alexandraki et al., 2022) while as seen in **Figure 2.1** most vacation goers chose air travel. If context and purpose would be the sole drivers of modal choice, a consensus can form on optimal solutions, however location and geography also play a role. The Alpine regions of Europe are some of the most rail intensive in Europe, with Switzerland transporting over 70% of its freight cargo by rail, turning the previous statistic on freight around (European Rail Freight Association, 2021). This becomes more convoluted when other Alpine regions in Austria and France have lower values despite similar geographies, or when one considers the reliance on waterway travel for many coastal or riverine regions. There is far more to transportation choice, even if solely looking at the European context, then first meets the eye.

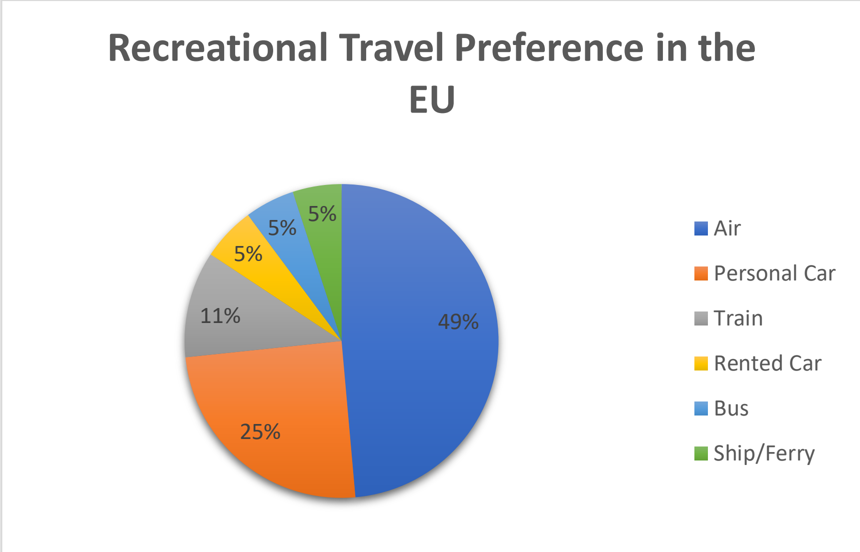


Figure 2.1: *European travellers' mode of transport for holidays in May 2023 according to Statista (2023).*

A European context constitutes specific requirements, but it will also share similarities with other regions. "Attenuation of agglomeration economies: Evidence from the universe of Chinese manufacturing firms" by Li et al. (2022) gives a unique view into the cardinal contributions of urban economics, namely, agglomeration economies: the observable superiority in cost and productive efficiency available and utilized in specific locations that both lead to and perpetuate the clustering of economic activity. Clusters have sizes and thus decay, also referred to as attenuation which indicates the sensitivity to distance and the discounting of value with said distance. This research found that attenuation is present to differing degrees between industries, hence it is crucial to treat sectors separately. If industries are sensitive to displacement or travel time, and transportation differs in time and cost, an adaptive solution is necessary.

2.1.1 Industry

An academic focus on manufacturing is due to a relatively straight forward relationship that upstream and downstream actors have in this industry. It is limiting to only focus on it and works like "The role of transalpine freight transport in a common European market: Analyses and empirical applications" by Reggiani et al. (1997) provided analysis on broad freight instead, while including topography which has to this point been ignored. This paper found further evidence for advantages to rail use, while commenting on the significant spill overs that access to such regions can provide for non-mountainous terrain. Despite less relevant logit and neural networks models being used, the two methodologies indicated that transalpine networks are economically feasible for improved connectivity and more efficient transport flows. These networks are more likely to ease issues of geography, but the models cannot answer to what extent. Such work has produced strong leaps in theoretical modelling for its time, however the previous more recent papers developed superior methodologies, which can better present relative advantages, not preferences.

The finding that topography influences preference because of rate of access implies that adverse geography can work as a type of border. This is argued on the basis of these findings' similarity to those of Lileeva & Trefler (2010) "Improved Access to Foreign Markets Raises Plant-Leve; Productivity… for Some Plants". This was an intervention study, looking at the liberalization of free trade across the US-Canada border in 1989. The model contributes theoretically, implementing discrete choice models with the Melitz cut off: a setting specific formula which describes rational choice to trade internationally under conditions. This transformation could be implemented in this study to better understand whether certain industries base thresholds or negative accessibilities when they are compared to more geographically adapted modalities. The concepts of market accessibility are further explored in the development of the theoretical framework as a key pillar and adaptation to the works of Weisbrod & Goldberg (2022). It is in this context that modality can be analysed independently while still allowing for a multilevel methodology. Additionally, there are further supplementing papers such as Yi & Kim (2018) or Otsuka et al. (2010) which would find ways to use some utility functions to better specify macroeconomic relationships.

2.1.2 Modal Choice

A definitive starting point to research into modality is ambiguous, as the choice of a form of transport vs. another is intrinsic to commuting and transportation choice, predating any modern economic theory. In his synthesis of contemporary research on the topic, Francois-Xavier de Donnea (1972) stated that in all models and observations, modal choice is a balance of monetary and time considerations. Both time and money have their scarcity and one can be traded for the other, or as one could colloquially say "time is money", this implied rate of exchange is the elasticity, and it is specific to each individual preference. Models' base utility on these parameters of time-cost of activity it could also be defined to represent loss of opportunity which is existent both in individual consumption and in company decision making. This innovation explained why rationally modal choice is almost never purely cost or time minimizing, rather there is a cobb-Douglas relationship, which necessitates an optimal split based on the utility exchange between the two.

Indeed, an academic focus on time value could be a reason for the stagnation and loss of market share in the European rail sector in the second half of the 20th century as theorized by Di Pietrantino & Pelkmans (2004). The authors explicitly prefaced their commentary on European policy "The Economics of EU Railway Reform" by aiming their future development to solve inefficiencies intrinsic to rail. Such a premise would be questionable in present academia, where rail is rather promoted and seen as a highly efficient tool that is more limited by the lack of policy and trans-regional frictions. Buehler (2011) compared the transportation behaviour of Germans and Americans, findings that there is a significant immaterial preference for car transportation in the US, which is only in part represented by transportation policy and costs. This behavioural bias to modal choice, where decision making is not measured as a prospect of experienced utility was likewise seen in freight transportation within Germany in the paper "The influences of behavioral biases, barriers, and facilitators on the willingness of forwarders' decision makers to modal shift from unimodal road freight transport to intermodal road–rail freight transport" by Elbert & Seikowsky (2017). If individuals faulter in their choices due to flawed perceptions, it is reasonable to assume that companies and governments might as well. For this reason, descriptive, not normative studies aid in separating the real from conjecture.

A step away from behavioural economics can reconnect to the topic of modal choice in the EU and how the internal market failures of rail stem from uncoordinated heterogeneity in ownership between European rail providers. The resultant inefficiencies in international transportation specifically inhibit freight transportation (Di Pietrantino & Pelkmans, 2004). This is a limitation that is likely to unmask itself in a real case study and is entirely an object of market failure. The same would be true for passenger transportation, which should be considered equally to freight, given that cross-border capable hubs can operate under nationally oriented management. Interestingly, despite mixed ownerships, subsidization and investment is relatively proportional which the authors of this work attributed to a relative inefficiency when comparing roadwork profitability. The final contribution is in the multi-product nature of railways, where the authors claimed networks effects to be present in the use of the costly infrastructure required for rail as there can be cross-subsidization between passenger and freight travel. This argument seemed to form as support publicly managed railways which was a key conclusion of the research. It also makes privy the treatment of both types of infrastructure as a market facilitator.

Much of the research on inter-, intra-, or unimodal transportation focuses on substitutability between rail and road, which is certainly not perfect. Intermodal competition in almost all these works of literature has been represented through elasticity, as there are clear operational differences in choosing transportation under different circumstances. For example, Westbrook & Buckley (1990) would find "vigorous competition" in the fresh fruits and vegetables market while Liu et al. (2019) found a much stronger argument for intermodal complementarity when studying mainland Chinese transportation corridors in heavy freight. Countless works have displayed this polarity; where certain industries posit an almost perfect substitution, thus warranting intense competition between modes of transportation, while other industries or geographies offer but a single solution to their most efficient modal split. While time and scope differ across the various authors studied, there are also important distinctions that certain authors fail to recognize.

2.2 Market Potential

Between the behavioural effects, ownership structure and topographical limitations, it is time to return to the most important factor from the beginning: the distance. A comparison of distances between road and rail is ambiguous in the more common research on intramodality, but naturally defining when one considers accessibility as unimodal. In this case, the levels of decay or loss of value at origin locations becomes the sole determiner which is also highly variable and dependent on the industry, location, scale, and context studied. The points of "The impact of distance on mode choice in freight transport" by Zgonc, Tekavčič, & Jakšič (2019) were that in a group of 15 studies crucial disparities between findings were usually due to the vast differences in discounting that was performed which would be calculated differently based on which geographic area and period was studied at which displacements. Indeed, this decay is a good way to introduce what is empirically the most important connection of theory to model, the market potential of a location. The conjecture on why certain findings prefer rail or road can be understood when one traces back how they discount the distance of locations, which makes it a driving force both literally in gravity models and figuratively in this thesis.

Market accessibility is an attribute of a location which defines its downstream markets and opportunities that are genuinely accessible; hence real access as one can think of. Given there is real access, there is also 'nominal' or potential market access. The terms are used interchangeably in academia, however for this thesis the terms are distinct, with the market potential outlining what is hypothetically achievable. The market potential of a location by no means refers to a location's capacity or magnitude, rather it is the endowment to singular connectivity, calculated with the magnitude of the downstream markets related to said industry. It is the frictionless potential, not accounting for the feasibility of an available trade/commuting volume between the origin and destination. This distinction will be important, as a destination might be optimally placed and theoretically in short distance to valuable origins, however low capacity can mean this source is hardly used. This distinction loosely borrows from the commentaries of Hochard & Barbier (2017) presenting robust normative solutions to issues of asymmetric market growth in developing accessibility. Their research manages to tie in aspect of geography, classic economics, and policy studies to present globally relevant solutions, through separating the potential access from theory to the real access.

With all the former in mind, three main conclusions can be established on modality and accessibility that allow to construct an economic model. Firstly, rail is in concept a superior mode of transportation, but its implementation and strategy often lead to inefficiency in operation despite good performance on paper. Secondly, the idea of market access and potential is principle in measuring the potency of modalities relative to one another. Lastly, distance, availability, and topography are the main drivers of disparities between modes. These assumptions mean that the model must both control for these important variables, and it must effectively estimate market accessibility vs. market potential.

2.2.1 Utility

Relationships in economics can be studied on several premises, the chosen of which is a utility model in most of the accessibility studies. In the paper "Relationship of Transportation Access and Connectivity to Local Economic Outcomes" by Alstadt, Weisbrod & Cutler (2012), the authors defined market access as the ability of transportation amenities to provide household and businesses with access to labour, material and/or customers. Vice versa this implies wages, product, and services, making the model directional. More importantly, the authors also provide three types of agglomeration premium that can be studied: employment premium or simply wage premium, output premium in unit markup or quality of service, and lastly synthesizing both into quality (value added) per employee. Effectively, the authors assume endogeneity of all outcomes on one another in three models which dependently on one another predict how big the three markets for labour, material and customers are met. When a model captures both effects, a non-linear trend is visible, suggesting that market access drives agglomeration economies asymmetrically with returns to scale.

A final theoretical contribution to three conclusions above is that both rail and road provide extensive market potential if implemented according to their industry preference. This should in turn be dependent on the decay parameters mentioned before. In "Long-Distance Passenger Rail Services in Europe: Market Access Models and Implications for Germany" by Beckers et al. (2009) the frictions in transportation availability and efficiency of operations between differing countries are outweighed by the cross-industry difference. In line with the authors recommendation, this is the reason why the thesis will focus on industry level differences ignoring the country level fixed effects as they are unsubstantiated in this context.

With all the following in mind several hypotheses will guide this research, firstly as was discussed by several authors, the relationship between rail and road transport is not one of direct competition hence choice is marginal not absolute.***H1a:** Rail offers greater benefits from market accessibility at a European level.* As has been indicated by most theoretical accessibility findings which look at cross-border freight. In line with this, ***H1b**: Road and Rail services are complementary within industry market access,* which is based off findings by Yi & Kim (2017) indicating that despite an absolute advantage of road in their case, rail had a comparative advantage in long distance.

Secondly, when looking at the normative discrepancies between studies on different industries mentioned previously ***H2**: rail is the preferred modality for high human capital and low value density industries, while road is superior for business that does not have these specific requirements.* This implies that high value service providers, or bulk production/commodity industries will benefit more from rail infrastructure while less specific manufacturing or entertainment services will align with road access. More directly, this hypothesis predicts that there should be advantages that are heterogenous by industry in intermodal market accessibility. Further from the recommendations of the Alstadt paper in respect to a need for industry level analyses the last hypothesis ***H3a**: Market access from road is most potent at short distances*. This hypothesis would likewise be in line with the Weisbrod & Goldberg (2022) local market thresholds. The final hypothesis to supplement this *is **H3b**: Geography has a significant effect on agglomeration development and preference*. This is to specify the effects that geography have on cluster development and thereby transportation.

2.3 Theoretical Model

This section will extract the assumptions and equations from works outlined previously, delving deeper into the mathematically derived constraints and utility equilibria to construct a credible and effective model. To start with, this work will develop on the Harris (1954) criterion of $MP_j = \sum_{k \in K} Y_k e^{-d_{jk}}$ for market potentials, a gravity model. On a theoretical level, the market potential of a location is a sum of location *k* incomes *Y* discounted by the distance *d* between locations *j* and *k*. The key variable to think of in this formula is the 'incomes' which for the region j is its respective GDP, although given the age of the model, and additionally the scope of this thesis to industry specific indicators across diverse regions, a value-added metric can be a superior variable. To elaborate, the assumption of homogeneous industries is dropped introducing a net number *I* of individual industries *i* with intermodal preference θ_i yielding $MP_j = \sum_{k \in K} \sum_{i=1}^{I} Y_{ik} e^{-\theta_i d_{jk}}$.

2.3.1 Individual consumption and preference

The following steps use Hanson's (2005) extension, which itself borrows from the work of Krugman (1991) and Helpman (1998) adapting the market potential function to work within a standard utility model. Firstly, utility $U = C_m^{\alpha} C_h^{1-\alpha}$ for all consumers spending their incomes on products manufactured C_m or housing C_h. As such industry level changes are constrained by consumer preference, and in a multi-industry model, C_m is transformed into C_t: the consumption of all defined tangible goods and services. Therefrom this non-housing variable is used identically within the utility function as it is simply an extension upon the assumption of only two consumables being available in the economy. The residual utility 'budget', which is a function of the alpha: cross-elasticity of consumption preferences, captures the maxima of optimal choice in the completed model.

The non-housing consumption can be defined by $(\mathbf{1})\ C_t = \sum_{i=1}^{I} (\omega_i * [\sum_f^F C_i^{\frac{\sigma-1}{\sigma}}]^{\frac{\sigma-1}{\sigma}})$ with σ as the elasticity of substitution between any pair of varieties *f* of all suppliers *F* in an industry i and ω_i as the contributed expense in the overall weighted basket of consumables C_t. Further, with this variable returned to the Cobb-Douglas utility function, there are increasing returns in production of each variety(company), where labour supply $L_{it} = a + bx_i$ is dependent on the quantity produced of all varieties in an industry. Effectively what the conditions developed imply is that all consumer utility is derived from a consumption of housing and tangible goods and services at a rate of labour supply defined by individual industry output. The condition assumes perfectly mobile and efficient labour market within and between industries. With the wage-labour limit the production potential of an origin *j* is:

$$(\mathbf{2})\ MP_j = \sum_k \sum_i^I \sum_f p_{fjk} C_{fjk} = \phi_j \sum_k \mu Y_k [\frac{\sigma}{\sigma - 1} b w_j e^{\tau d_{jk}}]^{1-\sigma} T_k^{\sigma-1}$$

This holds with the equilibrium condition that goods and services are priced at marginal cost $p_{ij} = \frac{\sigma}{\sigma-1} b w_j$. Since $w_j = x_i$, labour contribution is a function of the wage which is why it appears in the downstream origin. The real price in respect to transportation cost is $p_{fjk} = p_{fj}e^{\tau d_{jk}}$ where τ is the marginal rate of cost for transportation. In the case of services, the price of a delivered service being "transported" is zero as it does not apply, which in this case means that the real price is the same as nominal price for services in retail or corporate services as $e^0 = 1$. The market potential is an outcome of the decayed price time quantity (economic activities) of all individual destinations across all industries, adjusted for their utility contribution to the average consumer. The key question remaining being, how is this time and transportation preference calculated?

2.3.2 Cost of Time

In an industry level study, an industry modal preference θ_i should be considered (Yi & Kim, 2018). This is relevant for the extended Harris criterion in formula (1) a general decay function of $f(C_{ij})$ or in the vernacular used here $f(d_{jk})$. Travel time sensitivity or observed elasticity should capture the effect of preference. Intramodality is restricted, so $d \neq minf(d_{rail}d_{road}d_{air})$, rather distinct distance-time pairs must be included to find the optimal unimodal commute. Crucially, the unimodality is only to the extent of maximizing travel by mode. Rail travels hub-to-hub, hence in reality geographic information systems (GIS) data or equivalent is needed to find the intrazonal transportation times to hubs. More literally, while road time is the direct time that point to point travel times takes, rail time is $T_{jk}^{Rail} = \min(T_{js}^{Road} + T_{ss'}^{Rail} + T_{s'k}^{Road})$. Fortunately, such data as explored further is available serving as the best metric for real distance. Transportation hubs above can be stations or highway entrances indicated by s at origin and s' at destination.

Speed of travel is best standardized to have consistent intra-zonal commutes (Gutiérrez et al., 2011), this means that T_{js}^{Road} or $T_{s'k}^{Road}$ are such that $T = \frac{1}{2}\sqrt{\frac{A}{\pi}} \div v$ where *A* is the area in of a region studied and *v* is the average road speed in the region. While Yi & Kim used the maximum road speed, it is certainly more realistic to use a source from literature on average truck speeds in short haul as either through traffic/transit or technical reasons, lorries likely travel at lower average speeds. Lastly, it should be mentioned as a limitation, that this formula assumes that the proximity to hubs in a region is half of a radius of the region when imposed with circular geometry. This assumption accounts for industry and production location being incidental with hub development but likely is only significantly problematic if the regions studied will be notably non-circular, if region boundaries resemble long and slim polygons for example.

Finally, the measurement of output, per agglomeration is chosen to be the value added by rail or road, or the net value added in an industry region. Simply the advantage of a mode of transport is a comparison between the formulae: $VA_{road_k} = \eta_{road_k}\frac{VA_k}{road_k}\ and\ = VA_{rail_k} = \eta_{rail_k}\frac{VA_k}{rail_k}$ with the crucial variable, output elasticity of road/rail **(3)**$\eta_{road_k} = \frac{\partial VA_k}{\partial road_k}\frac{Road_k}{VA_k}$ and **(4)**$\eta_{rail_k} = \frac{\partial VA_k}{\partial rail_k}\frac{rail_k}{VA_k}$. Value added in a region is a linear function of all industries labour inputs, capital inputs, accessibility, and population density. These elasticity variables are implied on the assumption that the output in a region is tied to the discounted magnitude (value added) of origin downstream activities.

2.3.3. Complete Model

The value added in a region reflects its production value, making it directly a function of the producers input multiplied by the labour input and the reflection of downstream markets divided by their proximity. There is also the natural preference of industries, and a productivity of employees which further increases the labour input. The specified model can thus be adjusted to present a marginal value added of modality maximized with respect to modality:

$$AVA_j = \max(AVA_{road_j} + AVA_{rail_j})$$

$$(\mathbf{4}) VA_{road_j} = \eta_{road_{jk}} \sum\nolimits_k \sum_{i}^{I} \sum\nolimits_f p_{fjk} C_{fjk} = \phi_j \sum_k \mu Y_k [\frac{\sigma}{\sigma - 1} bw_j e^{\tau_{road_{jk}} d_{jk}}]^{1-\sigma} T_{road_{jk}}^{\sigma-1}$$

$$(\mathbf{5}) VA_{rail_j} = \eta_{rail_{jk}} \sum\nolimits_k \sum_{i}^{I} \sum\nolimits_f p_{fjk} C_{fjk} = \phi_j \sum_k \mu Y_k [\frac{\sigma}{\sigma - 1} bw_j e^{\tau_{rail_{jk}} d_{jk}}]^{1-\sigma} T_{rail_{jk}}^{\sigma-1}$$

Where in an efficiency maximizing region firms sort to modalities in, and the elasticity assumptions of (3) and (4) are introduced:

$$(\mathbf{6}) \overline{VA}_j = \max(\frac{\partial VA_k}{\partial T_{rail_{jk}}} \frac{T_{rail_{jk}}}{VA_k} \sum_k \mu Y_k [\frac{\sigma}{\sigma - 1} bw_j]^{1-\sigma} T_{rail_{jk}}^{\sigma-1}$$
$$+ \frac{\partial VA_k}{\partial T_{road_{jk}}} \frac{T_{road_{jk}}}{VA_k} \sum_k \mu Y_k [\frac{\sigma}{\sigma - 1} bw_j]^{1-\sigma} T_{road_{jk}}^{\sigma-1})$$

Considering a labour input, if all values are divided by their location specific endowment in the industry, then average value added can be more closely associated with the proximity to relevant markets.

$$(\mathbf{7}) \overline{AVA}_j = \max[(\eta_{rail_{jk}} \sum_k \mu Y_k [\frac{\sigma}{\sigma - 1} b]^{1-\sigma} T_{rail_{jk}}^{\sigma-1}) + (\eta_{road_{jk}} \sum_k \mu Y_k [\frac{\sigma}{\sigma - 1} b]^{1-\sigma} T_{road_{jk}}^{\sigma-1})]$$

$$with\ \sigma\ as\ industry\ level\ constant:\ \mu Y_k (\frac{1}{1 - \frac{1}{\sigma}})^{\sigma-1}\ b^{1-\sigma} T^{\sigma-1}$$

The labour wage limits allow for symmetry, the value added is elastic to the wage input which is in term a multiplicity of employment input. In effect, this operation reverses the studies found in literature that use a direct wage premium, instead capturing the effect in a broader employee output premium. This carries the advantage of being less sensitive to currency and purchasing parity, and better control for the disparity in the relative value of services. The formula (7) applies to orign-destination pairs, which following the summation produces a variable for market potential. In the initial formulation formula (2) indicates the value added to be the product of all potential downstream markets, however if labour volume is discounted, the output is a productivity measure. Lastly, this relationship is still naïve, as it constraints a frictionless world, wherein capacity is always at the maximum of potential. In reality $MA \leq MP$, with market access limited by modal capacity, and the values for b and μ as exogenous confounders for potential controls in a regression. The two variables have the same effect in both modalities; hence the model can be simplified by restricting them to confounders that were found in literature.

$$(\mathbf{8}) AVA_j = MA_{rail} + MA_{road} + \left[controls = \sum f(\mu, \phi, b)\right] + \varepsilon_j$$

One application for this function to adapt it to panel data to analyse the effect of policies as was done by Lileeva & Teffler (2010), however this method is vulnerable in larger settings where time lags, and exogenous shocks like recessions or political friction can influence the results. For this reason, the approach of Hanson (2007) was adapted in the logarithm above to fit a non-panel data set. A cross-sectional analysis is the alternative, and it was used effectively by the likes of Weisbord and Goldberg in similar research. The major advantage is that in analysing between subject variation, exogenous events have little effect, however control variables will need to capture all the observable differences between observations. This approach needs greater control for unobserved heterogeneity, and assumes that there is little idiosyncratic shock, which must be ensured by carefully choosing a less economically volatile year studied.

Table 2.1: *Necessary variables for research*

Variable	Type	Description	Database
AVA	Dependent	Value added per employee in an industry	Eurostat
MArail	Independent	Sum of weighted downstream industry VA at	Eurostat, ESPON,
MAroad	Independent	origin, discounted by travel and capacity.	University of Valencia
GDP(O/D)	Control	Gross domestic product adjusted for purchasing power parity for origin and destination	Eurostat
Age	Control	Mean age at the destination.	Eurostat
Density	Control	Population density per squared kilometre of Nuts3 regions for origin and destination.	Eurostat
Totalwater	Control	Total length of waterways in the Nuts3 region.	Eurostat
Topography	Control	Indicator of mountainous (>500m), or island geography.	European Commission

By using a cross sectional design, the model needs to choose variables based on reducing unobserved heterogeneity, so any endogenous collinearity or omitted confounders. The period studied should as previously mentioned not be subsequent of an abnormal economic event like a market crash, change in regional dynamic or another event. The choice of variables, the databases and regression design are discussed in the next section, however those variables already imperative to the literature review and theoretical modelling can be seen in **Table 2.1** above. Additionally, it is important to note that the relationship bases accessibility magnitude as a sum of inputs, and the proximity of outputs; as the upstream and downstream of activities, the function in economic terms will be an adapted translog production function (Yi & Kim, 2018). Lastly, concluding the literature review, is the fact that the dependent variable measuring agglomeration premium is the same, which allows for two different conditions to be tested.

3. Data

Synthesizing prior research, the research takes both modal, and industry level disparities in market access into account, hence an independent dataset needs to be constructed. First, transportation data on accessibility and consideration for all origin destination pairs between modalities. Further regional NUTS 3 level information with GDP per capita, population, wages, and employment statistics are needed to calculate the gravity of each point. Lastly, industry data to estimate the size, and monetary value of agglomerations will develop the extension for this work. Considering the diversity of this data, this will require several sources to establish a complete dataset and adaptation to make sure that the information is internally consistent. The region studied is Europe, however development and population density will surely diverge representativeness.

3.1 Origin-Destination data

The first category is the most challenging to quantify, and there are several ways this can be done with their own advantages and limitations. Firstly, the effect of distance can either be analysed through regional and local markets (Weisbrod & Goldberg, 2022) or by direct decayed parameters (Zgonc, Tekavčič, & Jakšič, 2019) the latter of which is superior for studies of larger regions. Decay parameters still have an implicit threshold which demarks preference if modal accessibilities are modelled simultaneously (Liu et al., 2019). Further, the data can either be Origin-Destination (OD), or factorial, indicating the endowment of a location, in this case the choice was to start with the former to calculate a unit for the latter. Important to keep in mind is that since the research does not start with an endowment (previously discussed as accessibility vs. potential), the OD data will need to be adjusted for capacity in this process.

The OD database used is ESPON dataset, which was developed by the University of Valencia in Spain, which is a complete dataset of all NUTS 3 regions. The authors used estimates for local average transportation times in calculating the time to commute between NUTS 3 region centroids through 4 potential modalities: road, rail, air transportation and multimodal. This data set provided over 2.2 million data points, and included regions outside of the EU, such as Turkey, Ukraine, and Switzerland. This data type, when used in the equation developed in the last section allows for the study to continue as a gravity model as clear distance relationships between nodes are available. The data uses average standard commercial traffic speeds and using OpenStreetMap to overlay the real travel distance on viable roads or rail lines, the assumption of standard speed which accounts for traffic and crossing is valid. This data has two limitations, firstly there it only gives a feasible commute not a throughput quantifiable route, this is resolved with inclusions of capacity later. A second limitation of this data is that it only has complete information for 2001, 2006, 2011 and 2014, the last data is later chosen due to its recency and normality.

Observable differences between modalities are already evident in **Figure 3.1** at this stage. The graph below shows the advantage of rail times to road times in transportation, likewise there is clear indication for air transport being the fastest unimodal connection seen in **Table 3.1** and multimodal solutions as the best overall possibility as it optimizes the three. A potential concern is that standard deviations are quite high in this data, which indicates that the data has a very big spread, however the variance is quite similar between road and rail.

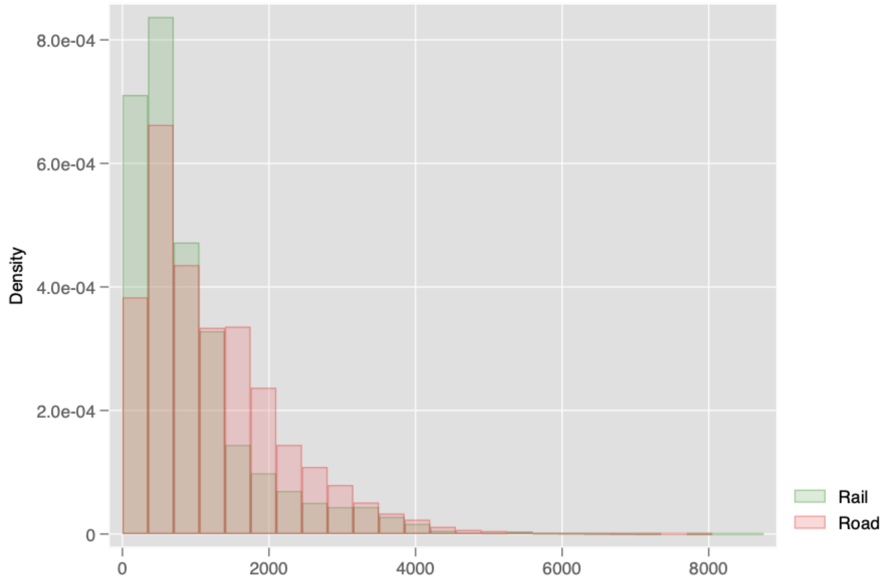


Figure 3.1: *Distribution of road and rail times between origin-destination pairs in the dataset.*

Table 3.1: *Descriptive statistics for modal proximities on Origin Destinations following initial data cleaning.*

Variable	Observations	Mean	Min	Max
Road travel	1,812,038	1,396.94	0	9,553
		(957.08)		
Rail travel	1,812,038	1001.02	0	16,100
		(1000.89)		
Air travel	1,812,038	538.69	0	1,229
		(166.74)		
Multimodal travel	1,812,038	509.24	0	1,229
		(185.86)		

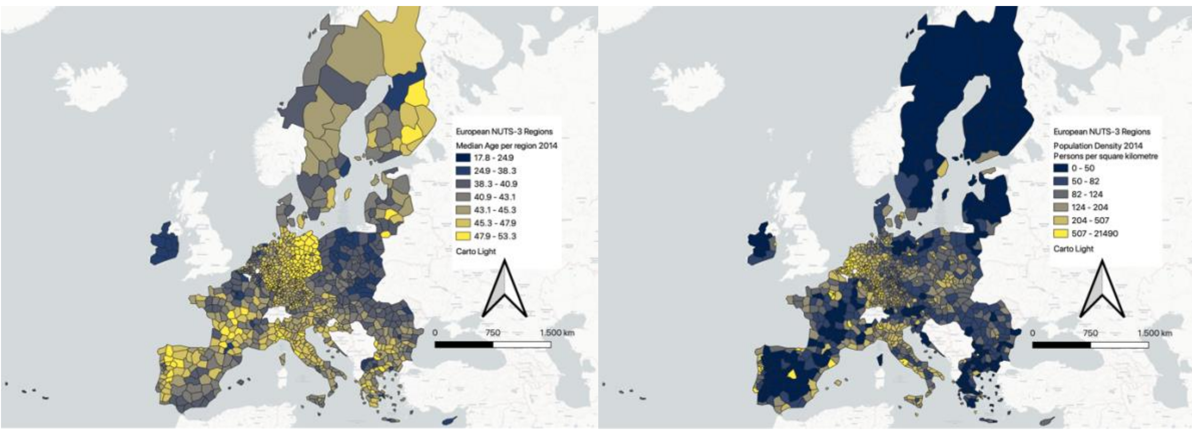
3.1.1 Regional Data

For regional data Eurostat: The European Commission's open-source data on countries and regions in Europe is used. This source can provide basic Nuts3 and Nuts2 region data from European countries on GDP per capita, employment, population, and value added. A superior data source is the Passport Euromonitor database, however it only provides national data. Eurostat data is transformed with the template of the European Commission ARDECO: over 10,000 micro regions of EU countries are available with population, employment, labour cost, domestic product, and capital formation data.

ARDECO itself is reaggregated in many of the variables of interest which is why the source data from Eurostat is itself transformed using the same methods to create a likewise robust source.

Data from Eurostat is provided at different levels ranging from country or EU wide down to the NUTS2 and NUTS 3 regions, with the latter being the smallest and comprising 150,000 and 800,000 inhabitants (Eurostat, 2023e). This can be somewhat deceiving as it is by no means a hard cut off, nevertheless regions above 1 million are still rare. The NUTS 3 region size is the most consolidated and smallest which is best for the model specified previously since market potential decays non-linearly. The data initially procured from the Eurostat database in NUTS 3 regions is employment (Eurostat, 2023a), GDP (Eurostat, 2023b), and population (Eurostat, 2023e). Data needed to be selected very carefully, as null observations would bias the regression or be omitted. For this reason, more generalized variables were key as if missing data are not consistent between variables, then more regions are likely to be omitted.

Value-added is the first metric included in the database: Eurostat (2023f) publishes estimates per NUTS 3 region of value added per industry as categorized by their nomenclature of economic activity (NACE). These values are crucial as they allow to estimate the efficiency per region, which is the dependent variable, and to establish the mass of an origin for the independent. The next step is adding employment values to find value added per worker. Introducing this variable reduces the 2,292,543 observations to 1,812,038 observations, which mainly isolates information from Turkey, Ukraine, Russia, and large parts of the Balkans. This leaves just above 1,200 NUTS 3 regions in the dataset. A value that drops to 1154 when GDP, population density, median age, and region size are included. Thus, as seen in **Figure 3.2 and 3.3** a still mostly complete picture of Europe can be reproduced, and the data can be visually confirmed if one sees for example the population density trends peaking in yellow on cities like Paris, Madrid, Rome, or the Randstad region.



Figures 3.2 & 3.3: *Maps showing the median age of the population and population density across available regions in the database, constructed using Eurostat (2023d) and Eurostat (2023e) data. Opensource shapefile from the European Commission.*

3.1.2 Industry Data

Eurostat data for gross value added per region is ordered by the statistical classification of economic activities in the European community (NACE) code, which allows to find the exact contribution and relative size of each industry in a region. With 21 NACE codes, and 16 in the data source, it is imperative to group these activities into their industries. This can be done by adapting the categories of Weisbrod & Goldberg (2022). Thus, six agglomeration varieties are formed as seen in **Table 3.1** below, the NACE codes, reasonings for grouping and factors for the industries to cluster are listed. Additionally, based on the literature, within the cluster type is also the intensity of the activity which can help answer hypothesis 2 later. Due to constraints in availability of average value-added data, missing industries are dropped, to focus on completeness of regions in the final dataset.

Table 3.1: Cluster categories for different NACE codes provided by the European Commission.

Cluster Type (Intensity type)	NACE Codes	Reasoning for grouping	Cluster Factors (Localization Economies)
Raw Material Commodity: (Low Value Density)	A(Agriculture) B_E (Mining and Water supply)	Bulk cargo with low value density, almost exclusively produced on scale as an intermediate good.	Likely to occur around rail or waterway access with sufficient space for infrastructure.
Manufacturing (Regular Industry)	C(Manufacturing)	This is a very large and diverse category which is grouped by the European Commission.	Likely to locate nearest to highly accessible routes and highways to optimize delivery of inputs and product.
Construction (Regular Industry)	F(Construction)	These operations are likely to produce the most inputs on average as it is relevant for all industries regardless of geography.	Likely to flourish in all available agglomerations. Higher values in metropolitan areas, edge cities or industrial clusters.
Business Services (High Human Capital)	K-N (Finance and Insurance, Administration and supporting activities)	Purely service companies, likely to use or even share the same office style infrastructure and interact within the same branches.	Likely to settle in central business districts, in large cities with sufficient complementary services around and opportunity for fast short-distance transport and spill overs.
Consumer Tourism (High Human Capital)	R-U (Art, entertainment, and foreign oriented businesses)	This is a consumption industry; hence tourism, entertainment, and general leisure activities are operating together.	Likely to form within cities which have high creativity output or historical/natura; beauty with access to diverse transportation options.

In the 1,328 regions, and of the 16 NACE categories provided, the amount of datapoints per NACE code ranged from 628 for real estate to 1,302 for public administration and defence in 2014. This is due to publishing and data laws in different countries, nevertheless the average number of observations per category is 1,000. Further descriptive statistics on this variable can be found in **Table 1** of the appendix. An issue is that certain industries are difficult to classify, real estate activities for example, make relatively low percentages of economic activity but have an extensive influence on all other activities. While this is explored further, it is not a great source of error, as predictive classification is not relevant to the research question. Going forward, the activities from **Table 3.1** were used to develop the descriptive **Table 3.2** below. The data indicate a foreseeable trend, where corporate services like finance, and administration have the greatest value added while manufacturing is lower, due to few European economies relying on basic manufacturing anymore.

Table 3.2: *Categorized NACE activities from Eurostat with value added expressed in millions of Euros.*

Industry	Observations	Mean	Standard deviation	Range
Commodity	1,153	1,977.12	(2,415.10)	26,770-11
Manufacturing	1,153	466.62	(666.93)	8,076-2
Construction	1,153	1,498.99	(2,001.46)	22,360-3
Corporate Service	1,153	2,444.75	(5,636.12)	86,184-27
Consumer Tourism	1,153	2,142.09	(4,484.85)	57,044-37

3.1.3 Market Accessibility

With the data from the previous two subsections, market accessibility can now be adequately studied. The Hansen criterion weighs market potential as the mass of the origin decayed by the time required to access a location. Rather than using GDP, a more adaptive measure for origin mass would be the "industry world input-output tables" published by the University of Groningen (Timmer et al., 2015). Within the context of the completed model from the previous section, the weights capture the θ_i. The information provided was agglomerated by adding all European countries results and merging the sub-industries into their main NACE classification. The tables were filtered to look specifically at the industries from **Table 3.1** finally producing the data seen in **Table 3.2** below and in **Table 2** of the appendix. In both types of markets notable relationships (>15% contribution) are highlighted in red.

Table 3.2: *Output table for the selected industries*

Output	ABE	C	F	KNJ	GIR	Total
ABE	19.9%	67.3%	2.5%	3.7%	6.5%	100.0%
C	4.6%	68.1%	11.1%	5.9%	10.4%	100.0%
F	2.7%	8.8%	54.6%	24.6%	9.3%	100.0%
KNJ	3.3%	15.8%	5.6%	57.2%	18.1%	100.0%
GIR	2.8%	21.2%	7.5%	43.5%	25.0%	100.0%
Intensity	33.3%	181.2%	81.3%	135.0%	69.3%	

Effectively, data for output contribution is given for each industry cluster by each industry cluster, all the values for one industries output percentage are used as weights which are multiplies by the value added of these industries at each origin, which produces a value for downstream market. With the mass calculated, there is still the matter of the decay denominator which is crucial in credibly estimating market access. The time differences between rail and road transportation are likely to differ the most at the beginning of trips where short distances are more available by road, and on the long distance where the point-to-point connectivity can better maximize benefits of rail efficiency. When mass is divided by undiscounted travel times, this relationship can be seen in **Figure 3.4** below.

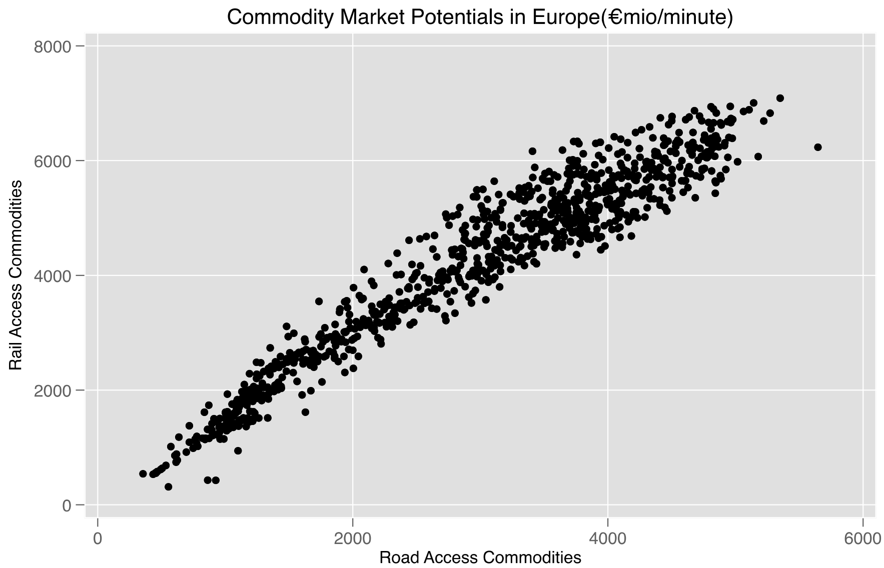


Figure 3.4: *Scatter plot showing the market accessibility by road and rail transport across European NUTS 3 Regions. Variables represent the value added at the sum of all origins in a year per minute of displacement.*

3.2 Decay Parameters

Market potential is calculated through three variables, the mass described above, the travel time and most importantly the rate of decay (Gutiérrez et al., 2011). Adapting the method of Reggiani, Bucci & Russo (2012), an auxiliary regression on the location mass and commuting/trade flows can be analysed while controlling for the time taken between these origins and destinations, and their appeal. While there is actual trade flow data at a NUTS 2 level for rail transportation, the necessary disaggregation would likely make the decay coefficients inaccurate, and road would still be missing. This is a significant issue, and as Tiefelsdorf (2003) explored, one of the two reasons for spatial relation misspecification when using gravity models, while the other of heterogeneity in distance variable estimation was addressed by consistent and valid dataset on OD’s. Therefore, several potential estimations are used: migration flows, figures from literature, and contemporary use of input-output tables. This enables future sensitivity analyses on decay parameters.

Two potential forms for decay factors are tested, which despite lower confidence in validity give credible values to test decay parameter sensitivity later. The first is the use of literature values, namely from Rosik, Stępniak, & Komornicki (2015), who analyzed over a dozen works on decay parameters in Europe outlining a credible range for trans regional decay parameters between 0.0005 and 0.29. They relate how greater disparities can also be valid as with 0.25 and 0.7 for rail and road decay in Zgonc, Tekavčič, & Jakšič (2019) when purely long freight industries are discussed. The second form for decay parameters is calculated, by using Eurostat (2015) data on migration flows. There is robust evidence that migration flows are predicted directly by commuting flows (Shuai, 2012), hence conducting a regression in the form $\ln(Migflow_j) = \ln(Migflow) + traveltime + \ln\,(Capacity)$ using migration data from post 2014, produces coefficients of 0.0008 and 0.0018 for rail and road. The travel time coefficient in this regression is an elasticity for distance, however this method assumes the relationship between commuting and migration from Shuai (2012) to hold. With two pairs of parameters for sensitivity analysis, the next method is the one to be used as true values.

3.2.1 Trade Flow Estimation

The European Join Research Centre (JRC) data provides a close estimation of trade flow relationships. The JRC produces input output tables like previously, but on a NUTS 2 level for net trade. Already having a NUTS 3 OD matrix, this data could be used by equally weighing population density and GDP to find which NUTS 3 region is the economic centre for their NUTS 2 region. Once this was done on all NUTS2 region origin and destination pair, then there would be a unique NUTS 3 OD pair for each NUTS 2 OD pair. This process can be seen in **Figure 3.5** below, with Centre Val de Loire (FRB0) and Toscana (ITI1), where the NUTS 3 regions representing the two greater regions are Loiret (FRB06) with the city of Orleans, and Firenze (ITI14) with the city of Florence. Thus, the travel times ITI1-FRB0 are approximated to FRB06-ITI14.

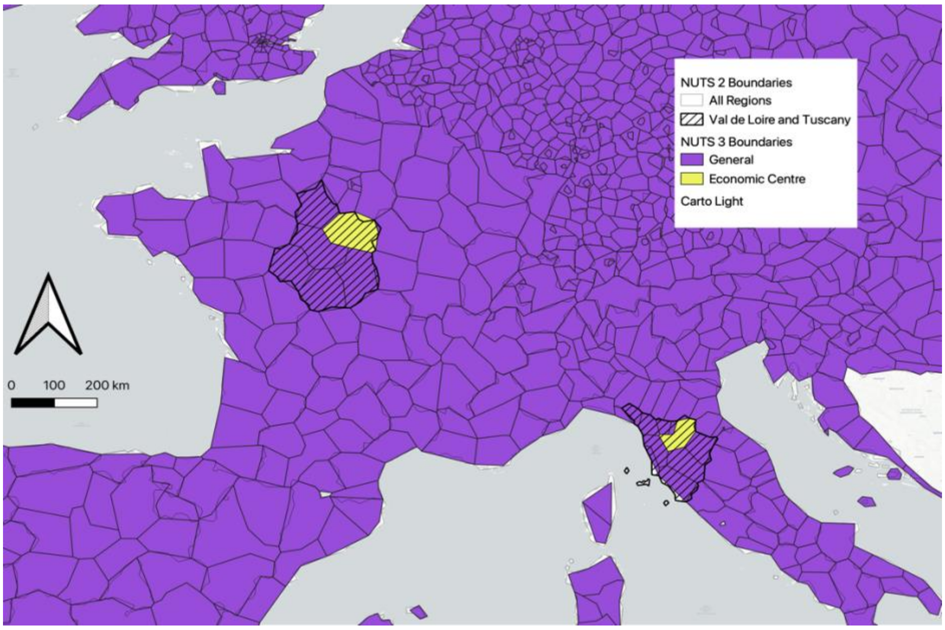


Figure 3.5: *Visual representation of the NUTS conversion using the example of FRB0 and ITI1, the chosen NUTS 3 regions both are the largest cities with the greatest GDP: Orleans and Florence.*

In **Table 3.3** results of the regression, following the formulation $\ln(Tradeflow_j) = \ln(VA_k) + modetime + \ln(GDP_k) + \ln\,(GDP_j)$, indicate quite small and consistent values for decay parameters. The method described above was performed, leading to 24,964 OD pairs, which represent an average of 25 potential origins per destination. Overall, there seems to be very little variation, with all industries having slightly more sensitive time elasticities in road demand at around -0.0015 and less sensitive rail demands at around -0.00145. All the results are highly significant at a 99.9% confidence. The R^2 of all the results is above 0.8, meaning that these variables explain trade flows quite well, however as the aim of model is not predictive this is unimportant.

Table 3.3: *Decay parameter regression by industry.*

Variables	Commodities	Construction	Manufacturing	Corporate	Retail
Rail	$-1.42\text{x}10^{-3}$***	$-1.40\text{x}10^{-3}$***	$-1.45\text{x}10^{-3}$***	$-1.45\text{x}10^{-3}$***	$-1.46\text{x}10^{-3}$***
	$(1.78\text{x}10^{-5})$	$(1.78\text{x}10^{-5})$	$(1.78\text{x}10^{-5})$	$(1.78\text{x}10^{-5})$	$(1.78\text{x}10^{-5})$
ln(VA)	0.250***	0.257***	0.124***	0.332***	0.054***
	(0.017)	(0.014)	(0.021)	(0.022)	(0.021)
ln(GDPDes)	0.483***	0.456***	0.783***	0.313***	0.627***
	(0.017)	(0.016)	(0.020)	(0.026)	(0.023)
ln(GDPOg)	0.625***	0.627***	0.622***	0.622***	0.622***
	(0.011)	(0.011)	(0.011)	(0.011)	(0.011)
Road	$-1.48\text{x}10^{-3}$***	$-1.46\text{x}10^{-3}$***	$-1.50\text{x}10^{-3}$***	$-1.50\text{x}10^{-3}$***	$-1.52\text{x}10^{-3}$***
	$(1.39\text{x}10^{-5})$	$(1.41\text{x}10^{-5})$	$(1.39\text{x}10^{-5})$	$(1.39\text{x}10^{-5})$	$(1.40\text{x}10^{-5})$
ln(VA)	0.252***	0.190***	0.022***	0.293***	0.113***
	(0.016)	(0.013)	(0.019)	(0.020)	(0.019)
ln(GDPDes)	0.494***	0.528***	0.676***	0.370***	0.582***
	(0.016)	(0.015)	(0.019)	(0.025)	(0.022)
ln(GDPOg)	0.638***	0.639***	0.636***	0.636***	0.635***
	(0.010)	(0.010)	(0.010)	(0.010)	(0.010)
Observations	24964	24964	24964	24964	24964

3.3 Methodology

With every necessary piece of data accessible, the methods used to answer the research question will be logarithmic regression based on the translog production outlined in theory (Yi & Kim, 2017), this offers two distinct advantages. Firstly, the logarithmic model allows for market access and value added to be expressed in terms of elasticities, or simply the sensitivity and degree of change in one on the other. Secondly, the data for many of the variables is highly skewed as seen with the distance times previously, this is also the case for population densities and value added as the study covers a very large and diverse area and deals with clusters. The data could be imputed; however, the asymmetric distribution makes this approach less viable. By construction and due to the skewedness being theoretically consistent, a log transformation is optimal.

When using cross-sectional data, there are several considerations that must be first accounted for. If one takes the adapted Harris formulation $MP_j = \sum_k \frac{MP_k}{T^{\eta_{road}}} + \frac{MP_k}{T^{\eta_{rail}}} + \cdots \varepsilon_j$, the log of both sides forms the function $\ln(MP_j) = 2 \times \sum \ln(MP_k) - (\eta_{road} + \eta_{rail}) \ln \sum(T_{jk}) \ldots \varepsilon_j$ for each origin k. In this formula T is a displacement metric, and the η variables are the modal elasticities: the degree to which changes in distance impact the change in market potential along an OD pair. A greater elasticity means that market potential is more sensitive to distance, which seen previously is true for road.

The relationship described above can be extended from previous modelling, remember that independent variable $AVA_{ij} = \frac{VA_{ij}}{employee_{ij}}$ for average value added per employee. The term MA is a sum of all potential origin markets adjusted for modal capacity (Eurostat) and discounted by travel time, making it an adaptation of MP: $(\mathbf{9}) MA_{ji} = \sum_{k=K}^{K} \frac{Cap_k * \sum_{i=I}^{I} \sum \omega_i inputs_k}{T^{\eta_i}} = Cap_k \times MP_i$ for either modality. An interaction is included to capture the influence the differing modes of transport which were outlined in the log form of the Hansen criterion. Theoretically this is justifiable by the expected complementary or substitutive effects between modalities. Since market access is dependent on a mass that is industry specific, this interaction should also differ across industries.

$$(\mathbf{10}) \ln(AVA_j) = \beta_0 + \eta_{rail} \ln(railMA_k) + \eta_{road} \ln(roadMA) + \beta_1 \ln(railMA) * \ln(roadMA) + \beta_2 \ln(density) + \beta_3 \ln(Odensity) + \beta_4 \ln(GDP) + \beta_5 \text{age} + \varepsilon_j$$

Several iterations of the regression will be performed, a purely linear form, and in line with the research of Yi & Kim (2018) a formula with squared values for market accessibility to address potential non-linearity. This will be done by squaring the logarithms in the formula above in the case of the former, as there is a possibility that the relationship between independent and dependent variables is non-linear with decreasing or increasing returns to scale. This is to control for potential differences in the preference that transportation has depending on the distance that needs to be covered, and the relevant thresholds at which one mode would be substituted by another. Considering that transportation is a derived demand, the interaction term captures the mathematically and theoretically likely existence of multimodal spill overs and economies of scale in infrastructure development.

3.3.1. Feedback Simultaneity

When analysing the relationship between areas of valuable productive efficiency and availability of transportation, a key factor is potential feedback simultaneity. Namely, while it is assumed in this thesis that superior infrastructure, connecting relevant market for a given industry leads to increased value added per employee, the opposite could also be argued. If an efficient market develops in one region, then investments into the transportation network could be a response to increase market accessibility for demand markets. A common test for whether there is a degree reverse causality is by using an instrumental variable. In this case, the instrument tested is a variable that is used to estimate the market accessibility in the standard formula and treated as a potentially independent variable that can mediate the relationship of theoretical market access and value added in a region.

While instrumental variables are tested further on to supplement the main findings, feedback simultaneity is minimized through other means. The choice in this research to distinguish market potential and market accessibility as theoretical market possibility vs. a market scaled to what is available to use means that the independent variables are more dimensional. Capacity and potential are quite unlikely to be caused by productivity premia, but it is even less likely that both would be driven by this variable. This stems from formula (**9**).

Capacity was measured from the Eurostat NUTS 2 data for modal network densities by NUTS 2 region is useful once more. Data is published on the length of waterways, and the length per 1000 square km of motorways and railways, the former of the three introduces a useful control variable for the commodities sector, while the other two are used to scale potential to real accessibility. The density of networks is very likely to measure the capacity (Hochard & Barbier, 2017). Individually the effect of this variable on AVA is likely to be low, as the extent of railway development is not likely to improve how much value is produced per employee. Instead, as a derived demand this variable is likely to remove the friction in transportation across markets, i.e., mediate through the market accessibility variable while maintaining the exclusion restriction as there should be no direct effect on the outcome.

To understand the further application of the data from Eurostat (2023c), it is important to note that for road transportation only motorways were chosen, and likewise only freight capable railways were used. This is due to the restrictions on freight traffic (Engel, 2010). As for motorway choice, commuting for service industries between the relatively large NUTS 3 regions, it is unrealistic to assume regular roads would be used due to speed and trucking limits. Given that the independent variable is a mix of much differently constructed variables, with low chances of reverse causality for either of them, mitigates the likelihood of feedback simultaneity.

3.4 Descriptive Statistics

Finally, in **Table 3.4** below, the data on regression variables are aggregated where necessary and presented. Value added and market potential have been averaged across the five industries to give a broader picture of the regions studied. It can further be also seen, that following the integration and refitting of the Origin-Destination matrix, the final number of observations available for regression is at 982 European NUTS-3 regions. The high Standard Deviation on all variables except for Age is in line with the previous justifications for logarithmic transformation in the model refinement. Finally, all values and ranges are reasonable, with maxima like 210 bn Euro GDP responding to Paris, or 11,538 km of traversable waterways for the Finish region of Etelä-Savo.

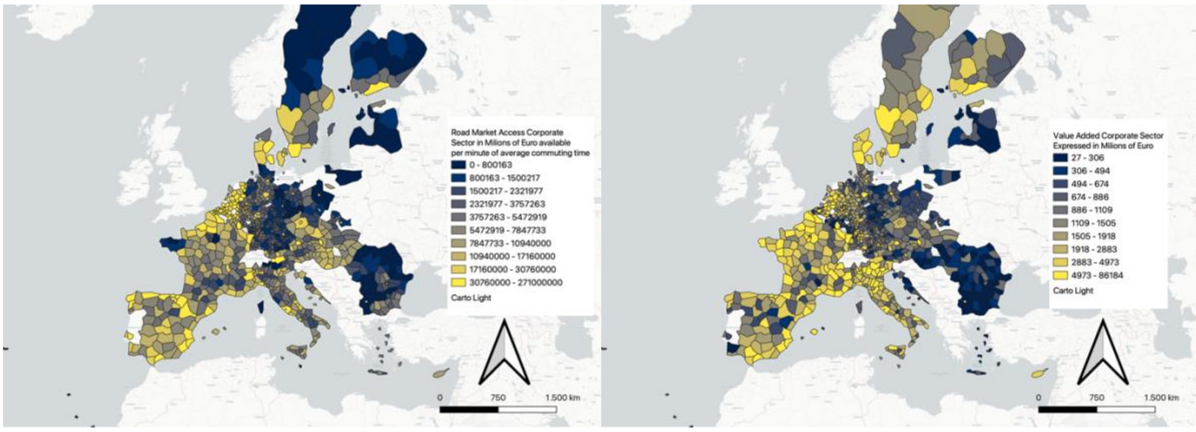
Table 3.4: *Descriptive statistics for the complete model regression*

Variable	N	Mean	SD	Range
Value Added per Employee in Euro	982	272,614.20	97,545.10	42,392 - 695,360
Average Market Potential of the Density of Available Motorways per Km^2	1,137	5.74	10.73	0 – 126
Average Market Potential of the Density of Available Railways per Km^2	1,137	13.18	27.05	0 - 497
GDP in Millions of Euros	982	10,707.75	18,193.81	192 – 210,140
Area in Km^2	982	3,556.33	6,420.69	14 – 105,208
Average Age	982	44.12	3.35	33 - 54
Navigable Waterways Length Km	1,193	215.86	932.20	0 – 11,538
Population Density – Persons per Km^2	982	420.88	1,064.50	2 – 21,490

The results of the research are split into a preliminary model and then one with geographic/topographic indicators. This extension is likely to provide additional accuracy and precision to the model if it is robust and correctly defined, hence for this reason the two previously mentioned variables of waterway access seen in the table above, and geography are included. The geography of the regions studied can be seen in **Figure 2** of the appendix, where over a quarter of all European regions studied are either mountainous or islands. Further 662 regions out of 1,193 have some level of river or canal pathways, indicating over half of European regions although many of these regions are very sparsely populated, and work as transit. These variables will be useful later in the results when the geographic control variables are added to improve accuracy of predictors.

4. Results

With Data and variables described, there are some valuable findings that can be assessed visually. In **Figure 4.1 & 4.2** below the real market access by road and value added for the corporate/financial sector are seen. Firstly, one can see that these values somewhat align, which indicates the weak but still present correlation between market access and output of 0.4. Further, it also shows that the accessibility metric was well developed as areas around service-based economies like in the Benelux or in southern Scandinavia the regions have higher market access values. Those are areas with high rail density and close to large agglomerations. Secondly, the issue of size can be seen, as Germany is mostly blue despite being the largest economy. This is a scaling issue, with German NUTS regions being relatively small hence the productivity is relatively lower seen in **Figure 4.2**. In **Figure 3** of the appendix an unadjusted AVA map shows how further the area and population density of a region can inflate the statistics as seen in Northern Scandinavia.



Figures 4.1 & 4.2: *Maps showing the road market accessibility (left) and the value added (right) locally in European region.*

4.1 Initial Results

The initial models can help specify the significance and need for adjustments to the regression design. To this end, three models can be developed, looking at first, a naïve regression, then a relationship including basic control variables, followed by a permissibility of interaction between market access ($\beta_1 \neq 0$), and lastly testing for non-linearity (exponential η_{road} & η_{rail}). The model progression can be seen in **Table 4.2** below, and identical table can be found for retail in **Table 3** of the appendix. The first positive identifier from the regression is that most variables are highly significant at a 99.9% level. Secondly, the results are consistent in controls, apart from population density at the origin in Model 4. In fact, Model 4 deviates significantly from past observations, which despite its intentional inclusion sheds new light on the interpretation of the variables of interest. The formulation for the retail sector in the appendix section reflects almost precisely exact changes in variables and their corresponding coefficients. Interactions are not significant outside of model 4, and further there the model indicates that there are significant rates of change in the effect of modal market access across industries. These returns to scale are justified in literature as can be seen in Gutiérrez et al. (2011), additionally the high mean variance in factors is expected due to the repeated presence of independent variables. The mean VIF when merging the variable influence is around 4.83 which is just below the threshold for significant collinearity in variables that would bias the coefficient. Additionally, this model has the best AIC, BIC and R^2 values suggesting the greatest predictive power, also when considering over/under specification and best fit respectively.

The first hypotheses relate to baseline model specification, it is clear in the first three models' that rail is an inferior market enabler relative to road. However, when a non-linear specification is introduced, this relationship changes around. These models indicate that regions with 1% better rail accessibility keeping all else constant, are associated with 6.7, 5.9 and 6.3% lower value added per employee in manufacturing at 99.9% confidence. Interestingly, in this case the interaction is also not significant in model 3, hence there is insufficient evidence for complementarity despite the positive effect. What is more relevant is that the relatively better fitting polynomial model with more constant residuals with natural distribution reverses this relationship. Such a model however is much more difficult to interpret, hence an example region is best used. If Sankt Pöteln (AT123) is used with a lnMA of 15.71 and 16.97 for road and rail respectively, a 1% increase in rail leads to a 1.99% increase in AVA for commodities. A 10% increase leads to 21.69% increase in AVA, meanwhile if the same two increases are done for road transportation, the values are 1.74% and 18.74% ceterus paribus.

The model indicates the following: rail accessibility is more potent at improving AVA, and in both modalities, there are diminishing effects, however a positive interaction term captures how the complementarity of rail and road make both accessibilities a net positive. The model highlights both modalities importance as regions which would completely lack accessibility by road, would have very low value added per employee. The non-linear model captures more of the relationship, and it is better at describing the necessity for both modalities, which explains the awkward fit of the previous three models. As seen in **Figure 1** of the appendix, this model has heteroscedasticity issues, but its residuals are relatively more consolidated, making the model less sensitive to the asymmetric kurtosis of certain controls such as GDP and age exhibiting outlier propensity. This fit issue is expected even with a log transformation; hence the model residual normalizing indicates the best fit.

Table 4.2: *Model progression using the commodities sector.*

ln(Value Added per Employee) Commodities Sector	(Model 1) Basic	(Model 2) Control	(Model 3) Interaction	(Model 4) Non-linear
ln(road accessibility)	0.099***	0.064***	0.056***	-0.150***
	(0.010)	(0.009)	(0.018)	(0.032)
ln(rail accessibility	-0.067***	-0.059***	-0.063***	0.156***
	(0.012)	(0.010)	(0.013)	(0.027)
ln(rail)*ln(road)			0.001	0.003**
			(0.001)	(0.001)
$\ln(\text{road access})^2$				0.011***
				(0.002)
ln(rail access)2				-0.013***
				(0.002)
ln(density)		0.072***	0.069***	0.044**
		(0.017)	(0.017)	(0.018)
ln(Origin density)		-59.028**	-67.510**	70.300**
		(27.251)	(29.350)	(28.905)
ln(GDP)		0.342***	0.339***	0.321***
		(0.026)	(0.029)	(0.027)
age		0.062***	0.062***	0.045***
		(0.007)	(0.007)	(0.007)
Constant	10.501***	361.467**	412.806**	-418.869**
	(0.142)	(164.964)	(177.046)	(174.369)
Observations	973	973	973	973
R-squared	0.215	0.477	0.477	0.520
AIC	1,877	1,488	1,419	1,411
BIC	1,892	1,517	1,531	1,454
Mean VIF	1.32	1.27	4.10	17.32

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

4.1.1. Industry Analysis

Model 4 captures the full theoretical relationship with non-linearities, as it captures the pattern found in theory for diminishing returns and for complementarity. Not including the interaction term changes the coefficients significantly and leads to greater heteroscedasticity meaning that in polynomial modelling the complementarity of modalities is captured. The model has the highest R^2 value relative to the other models, which despite being an average metric for validity, gives a good indication for the fit of models relative to each other. Model 4 tested in all five industries for 2014 can be seen in **Table 4.3** below, the model represents value added with the assumption of non-linearity and interaction between the modes of access. There are several points of interest, firstly on the congruency in the results and secondly with the magnitude and direction of the control variables.

The population density of origins has a relatively strong and negative effect on manufacturing, corporate services, and retail but a positive effect on construction and commodities. This indicates that these industries differ from the others, which is likely considering that a sector like commodities with agriculture, mining, etc. is certainly going to occur in areas of low density with inputs from higher density areas where downstream markets for these goods are located. As previously, the metrics are based on roughly the same masses, and travel times which are based on distance, with the only completely independent variable of capacity separating the accessibility indices; hence there is a baseline average VIF greater than 5. This is however not a problem as it is implicit to the formulation, and all controls are still low (<2).

Several decay parameters were tested, for example using literature values of 0.25 and 0.7 provided results that were less collinear and had visually isotropic residual plots. Nevertheless, in **Table 5** of the appendix, using these values provides similar results in coefficient direction to those below. It would be hasty to validate the coefficients from Zgonc, Tekavčič, & Jakšič (2019), as this similarity could just reflect that capacity plays a much higher role in determining the relationship then the decay parameter. In the appendix **Table 6** values calculated from migration, 0.18 and 0.08 were tested and once more the results are comparable. The magnitude of the coefficients is likely most accurate with the model in 4.2, as the decay parameters were specifically adapted to this setting. It can already be stated that the model is not very sensitive to changes in the independent variables stemming from discounting. However, the differences in post estimators shows that the model could still be improperly specified.

Figure 4.3: Regression of Market Accessibility on value added across five industries.

ln(Value Added per Employee)	Manufacture ln(VA/emp)	Corporate ln(VA/emp)	Construction ln(VA/emp)	Commodities ln(VA/emp)	Retail ln(VA/emp)
ln(road accessibility)	-0.113***	-0.082***	-0.158***	-0.150***	-0.107***
	(0.024)	(0.020)	(0.027)	(0.032)	(0.023)
ln(rail accessibility	0.126***	0.184***	0.198***	0.156***	0.095***
	(0.022)	(0.018)	(0.024)	(0.027)	(0.021)
ln(rail)*ln(road)	0.002	0.000	0.002*	0.003**	0.001
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
ln(road access)^2	0.009***	0.007***	0.011***	0.011***	0.008***
	(0.001)	(0.001)	(0.001)	(0.002)	(0.001)
ln(rail access)2	-0.011***	-0.012***	-0.015***	-0.013***	-0.008***
	(0.001)	(0.001)	(0.001)	(0.002)	(0.001)
ln(density)	-0.016	-0.053***	0.035**	0.044**	-0.009
	(0.013)	(0.010)	(0.015)	(0.018)	(0.012)
ln(Origin density)	-169.547***	-387.281***	161.749***	70.300**	-80.515***
	(24.037)	(20.710)	(25.661)	(28.905)	(21.347)
ln(GDP)	0.232***	0.161***	0.280***	0.321***	0.270***
	(0.021)	(0.014)	(0.023)	(0.027)	(0.020)
age	0.025***	0.022***	0.027***	0.045***	0.018***
	(0.005)	(0.004)	(0.006)	(0.007)	(0.004)
Constant	1,032.056***	2,349.224***	-970.285***	-418.869**	493.941***
	(145.191)	(125.111)	(154.339)	(174.369)	(128.517)
Observations	973	972	972	973	972
R-squared	0.433	0.399	0.484	0.520	0.479

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

As discussed, the implication that improvements in road as a modality diminish the value added per worker in the regions studied, is incorrect. The value added, or more strictly within the model, output elasticity per employee is decreasing and increasing in returns to scale in the chosen modality, indicates distinct advantages based on the levels of accessibility. It implies that minor relative improvements into rail improve the value added in all industries. For road, large differences in accessibility diminish in their adverse effects, a possible reason for this is that roads need a minimum accessibility to really facilitate trade. It is possible that a motorway needs to be developed sufficiently to allow trade in the first place, as an insufficient size or capacity might deter users. This can be seen like the transportation equivalent of the Melitz cut off, where same as a market efficient area needs to produce at a certain threshold to engage in international trade, motorways need a certain level of capacity for users to consider them. These conclusions can be justified by the unique construction of market accessibilities and the non-linear relationships implicit to the study. Naïve regressions on market access seen in **Table 3** of the appendix give very hazy results, with no robust indication of association between value added and accessibility.

4.2. Combined Models

While the previous section created a minimum bias model, the following extension will aim to improve the predictor accuracy by introducing geographical considerations for trade. Transportation is a way of addressing issues with the natural landscapes in which and between which economies appear and flourish, once more using **Figure 2** of the appendix, over a quarter of the regions studied are geographically challenging. Thus, two more metrics were introduced, topography and waterway access. In theory the former should be a limiter, while the latter is a facilitator of trade flows and market access. As seen in **Table 4.5** below the issue is more complex, where in most cases, adverse geography is associated with lower value added, however in the case of the corporate sector, locating in mountainous regions is associated with 33.8% higher value added ceterus paribus. Islands, like with corporate financial sectors are positively associated with retail value added, which makes sense for its connotation with tourism. Nevertheless, islands have a low representation in the data hence the evidence could be circumstantial. Lastly, coefficients between **Table 4.5** and **Table 4.3** overlap standard errors at a 95% confidence across the board, indicating that there is no significant change in association. This means simply that the relationship between value added, and accessibility is unaffected by geography, it solely helps better predict the levels of value added. Interestingly waterway development has close to no effect on the value added in regions, with the only highly significant values for commodities and construction.

Table 4.5: *Final result: polynomial log-log regression of market access on the AVA of all five industry types.*

ln(Value Added per Employee)	Manufacture ln(VA/emp)	Corporate ln(VA/emp)	Construction ln(VA/emp)	Commodities ln(VA/emp)	Retail ln(VA/emp)
ln(road accessibility)	-0.116***	-0.091***	-0.166***	-0.157***	-0.121***
	(0.024)	(0.020)	(0.027)	(0.031)	(0.022)
ln(rail accessibility)	0.133***	0.196***	0.217***	0.172***	0.118***
	(0.023)	(0.018)	(0.024)	(0.028)	(0.022)
ln(rail)*ln(road)	0.002	0.001**	0.002*	0.003*	0.002*
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Topography					
Island	-0.021	0.338***	-0.083	-0.068	0.309***
	(0.108)	(0.064)	(0.104)	(0.128)	(0.081)
Mountainous	-0.061*	0.040	-0.101***	-0.159***	-0.072**
	(0.035)	(0.025)	(0.038)	(0.043)	(0.033)
Total Waterway Length	0.009	-0.007*	0.026***	0.023***	0.009*
	(0.006)	(0.004)	(0.006)	(0.007)	(0.005)
ln(road access)^2	0.009***	0.006***	0.012***	0.012***	0.008***
	(0.001)	(0.001)	(0.001)	(0.002)	(0.001)
ln(rail access)2	-0.011***	-0.012***	-0.016***	-0.015***	-0.009***
	(0.001)	(0.001)	(0.001)	(0.002)	(0.001)
ln(density)	-0.017	-0.050***	0.039**	0.042**	-0.008
	(0.014)	(0.011)	(0.016)	(0.020)	(0.014)
ln(Origin density)	-158.707***	-409.588***	176.728***	99.099***	-82.651***
	(24.848)	(21.974)	(27.097)	(30.258)	(21.881)
ln(GDP)	0.231***	0.170***	0.275***	0.318***	0.276***
	(0.021)	(0.014)	(0.023)	(0.027)	(0.020)
age	0.024***	0.025***	0.025***	0.044***	0.020***
	(0.005)	(0.004)	(0.006)	(0.007)	(0.005)
Constant	966.595***	2,483.713***	-1,060.676***	-592.772***	506.609***
	(150.375)	(132.450)	(163.369)	(182.711)	(131.840)
Observations	973	972	972	973	972
R-squared	0.439	0.424	0.506	0.539	0.495
Geographic Dummies	Yes	Yes	Yes	Yes	Yes
Country Dummies	No	No	No	No	No

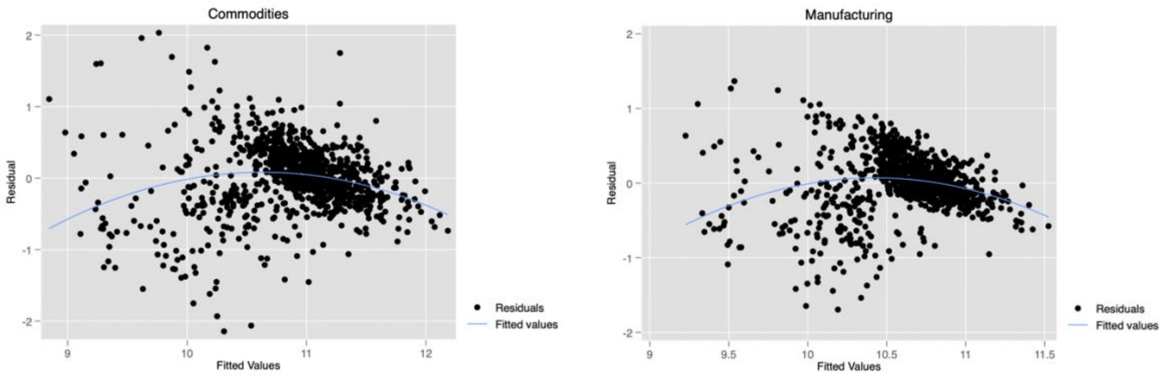
Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

The model indicates that once capacity is considered as a constraint to the market potential of a destination, roads are a superior mode of transportation for all five industries at shorter distances/higher values of market accessibility. The interaction term had indicated a low degree of complementarity at a 95% level with corporate services, and at a 90% confidence with Commodities and retail. There was a concern of collinearity between the population density and market accessibility variables and endogeneity with the topography variable however seen in **Figure 4 & 5** of the appendix the former was visually unsubstantiated and the VIF for topography was very low. Given the coefficients, it can be concluded that for example, a 5% increase in rail market accessibility for the commodities sector increases the value added by 0.71% while a 10% increases value added by 1.42%. The rate of change reduces but always remains positive, which does not mean that the less efficient road access is useless, as the full extent of the effect of both accessibilities is also driven by a positive interaction between modalities, which indicates complementarity.

The interpretation is multi-pronged, a limitation of the model's construction as the present index is both a metric of distance and mass, but the crucial theoretical contribution is that rail is preferred to road. A strong driver of market access is the distance, so what this result indicates is that potentially at very high accessibilities, and small distances the road transportation exhibits greater properties which the model cannot calculate. This study looks at relatively large distances, with a previously seen mean of over 1,000 minutes for travel for both modes of transportation. Markets that are on average more than 16 hours away skew the coefficients from what a shorter distance study would. 'Local markets' of <2 hours which are highly car driven as seen in Yi & Kim (2018) or Weisbrod & Goldberg (2022) are rare in this dataset. Th result opposes some principles of induced demand, which was not included in the utility of the function however it is not necessarily a wrong conclusion if the relationship is driven to a greater extent by distance/potential and not capacity.

4.2.1. Heteroskedasticity and Non-linearity

The five regressions are statistically significant and align with theoretically embedded relationships. While looking at the VIF the regressions likewise fit within a rule of thumb of <5 to declare low correlation between variables, although this only provides one view on the robustness of the results. Certain assumptions or robustness can be omitted, namely normality due to sample size, and auto-correlation due to the cross section. A key measure here, is the previously mentioned residuals, which should visually indicate non-linearities in the linear regression of model 3, and trends in the data. This gives a final reinforcement for the choice of model 4, and more importantly gives another view of the model's data estimation before non-linear heteroscedasticity tests are performed on model 4.



Figures 4.3 & 4.4: *Residual plots of the capacity constrained models on the commodities and manufacturing sectors respectively.*

The Commodities and Manufacturing sectors most explicitly expresses heteroscedasticity in the model with a downwards trend in residuals. Interestingly, the problem with the preliminary model, and seen once more in **Figure 4.3 & 4.4** above is the degree of clustering and trend downwards in residuals past the 10.5 threshold. Firstly, a pronounced non-linear the trend for fitted values is visible, secondly despite a log formulation the data still visually clusters implying abnormality or outliers. Interestingly in **Figure 4.3** where this regression had a highly significant water access variable, there is less clustering, or more homogenous residuals than in manufacturing. The waterway access variable was insignificant in the manufacturing regression, hence there could be a third unobserved variable which confounds that regression.

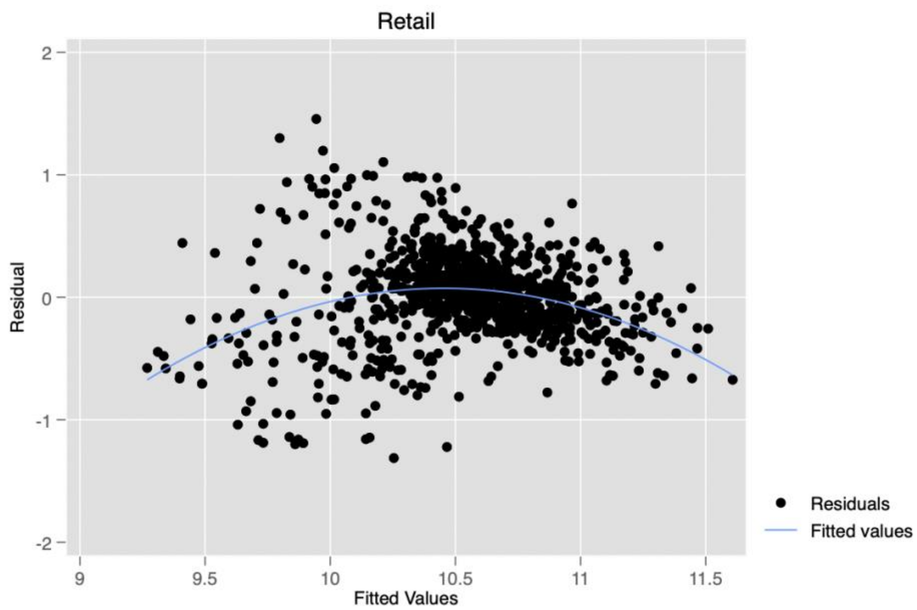


Figure 4.5: *Residual plot of the Retail sector.*

In **Figure 4.5** above, the retail sector is faced with a similar issue of heteroscedasticity to the previous two industries studied, however it is less clear whether there is a relationship between the predictors and the error term. All five industries studied have objective differences in the types of business they perform, however, retail is the most business-to-customer which could justify it having the most divergent trends. Heteroscedasticity can also imply that there are outliers in the three industries of retail, commodities, and manufacturing. The first two are limited geographically to points of natural beauty/interest or resource availability to flourish, while in manufacturing there may be an issue with the use of value added (Marschinski & Turegano, 2019). The metric is critiqued internally and externally by the EU as European manufacturing is quite different to manufacturing on other continents due to both its maturity and local economies.

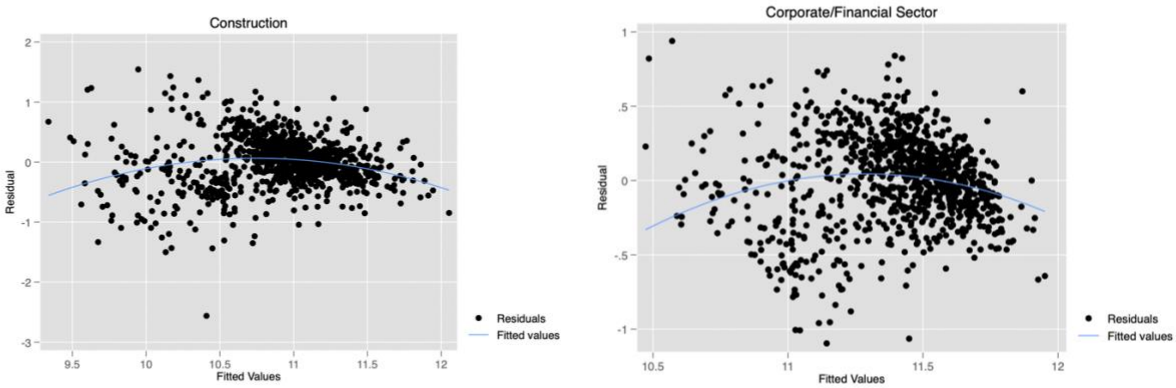


Figure 4.6 & 4.7: *Residual plots for the Construction and Corporate sectors respectively.*

Lastly in the cases of the construction and corporate sectors seen in **Figure 4.6 & 4.7**, there are no drastic patterns, with very minor coning, and no strong trends in the estimated vs. observed predictors. Once more there are pronounced non-linearities and the corporate/financial sector is visually the most homoscedastic. While construction is very region specific, skill, education, or material availability are less likely to impact the model to the same extent as commodities or manufacturing. In the corporate/financial sector, the regression seems to have the most consistent error terms which implies the model fits this industry best. This industry is the least likely to be influenced by geography, and it is very dependent on its upstream markets as it is purely a service business, hence it could be argued that there is a relatively minimal extent of OVB.

4.2.2. Variable Postestimation

With the Non-linear choice justified on several grounds, it is key to address the distribution of the results, measured heteroscedasticity, and the sensitivity. Firstly, already seen before, error terms are clustered to differing extents between the industries but overall, there is insufficient homoscedasticity to claim independence in error terms. This is an issue, and it does not significantly resolve with differing data and parameters. This is the first and most significant model limitation of this research, it implies that there are as previously mentioned patterns and relationships between the error and model predictors.

Using Lavene's test categorizing variables on a country and a geographic level, all p-values are highly significant, likewise based on kurtosis estimates, following the logarithmic transformation of the variable the data does not follow a normal distribution. Paramount during the regression and model development, the traditional imputation methods such as winsorizing proved futile. The reasons why are understood when the distribution of outcome variables is visualized as seen in **Figure 4.8** The variables do not follow a normal distribution, rather the variable distribution itself is non-linear which is not a fault of the outliers. Indeed, when an independent variable is squared as seen in **Figure 6 & 7** of the appendix, the distribution normalizes, which makes sense considering the already established robustness of this transformation. Nevertheless, despite positive two-way ANOVA results, the credibility of post estimators cannot be guaranteed as two of the assumptions for the regression are not fulfilled. Nevertheless, the results convincingly indicate a non-linearity, now further supported by the distribution curves, the coefficients are reasonable and will be tested in one more way further in the report despite problematic post-estimators.

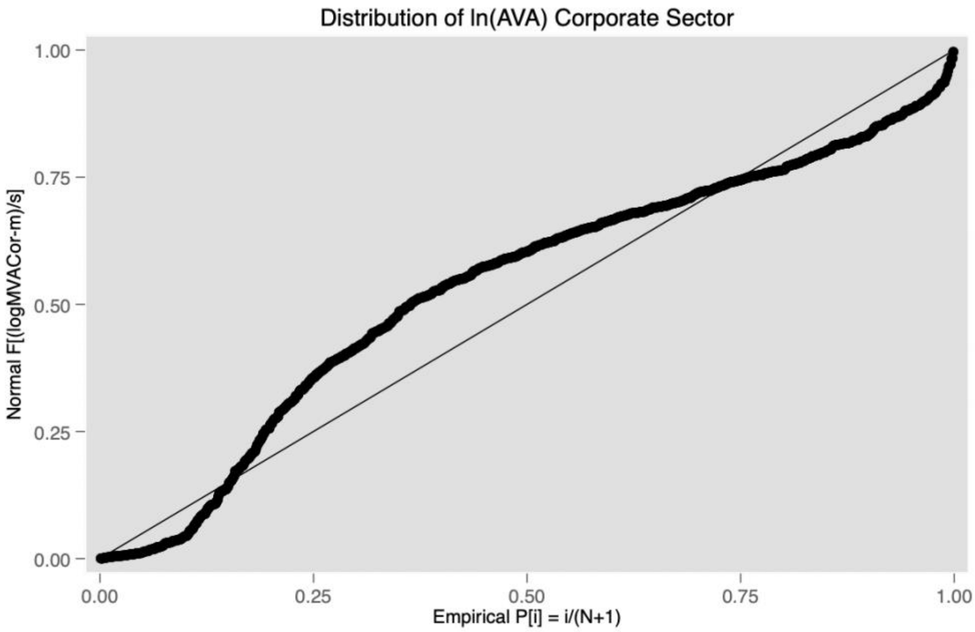


Figure 4.8: *Quantile-Quantile plot of the outcome variable (value added per employee) in the corporate sector relative to a normal distribution.*

To additionally comment on sensitivity, the independent variable was constructed considering both accessibility theories and capacity constraints. A crucial part of the analysis was in the development of decay parameters, and those seemed to mainly affect the significance of coefficients as seen in **Table 5 and 6** and minorly the error terms as discussed in reference to **Figure 1**. The independent variable is novel in its formulation, and thus there is no standard approach to testing its robustness. It could be tested whether there is an extent of relationship between capacity and theoretical market access, which in the case of no relationship would imply that indeed either capacity or potential drives the value-added relationship. Borrowing from alternative uses of instrumental variables (IV) discussed by

Pokropek (2016), in **Table 7** of the appendix, a fictitious IV regression is performed, where instead of a variable MA, there is a variable MP which is instrumented by capacity. The variables are completely uncorrelated, hence this faux IV is performed to test the relationship between MP and capacity better understanding how MA behaved in the final model above. The estimation showed that despite a lack of correlation, there is some evidence for predictive power of capacity through market potential, although for rail this is only supported by an F-test, not the Wald X^2. Lasty, when instrumented by capacity and geographic dummies, there was a significant association for rail accessibility.

4.2.4. Change of Year

The final robustness test for the model is to trace back the study testing the same relationships at a different point in time. In this case where a descriptive measure of elasticities is performed, a robust model should produce the same coefficients for the variables of interest and controls at a different date. With inter-modal market accessibility, there can be degrees of innovation in the long run that could change the influence or preference for modes of transport, however this is unlikely to be visible unless the studies would be several decades apart. As such, the 2011 origin-destination matrix was chosen to build the dataset, the two other available OD matrices were from 2006 and 2001 however the EU had fewer countries in those years then in 2014 which would lower the representativeness of this operation.

Table 4.7: *Final model reproduced with 2011 data.*

VARIABLES	Manufacture ln(VA/emp)	Corporate ln(VA/emp)	Construction ln(VA/emp)	Commodities ln(VA/emp)
ln(road accessibility)	-0.138***	-0.074	-0.174***	-0.205***
	(0.042)	(0.049)	(0.042)	(0.041)
ln(rail accessibility	0.036	0.071	0.066	0.059
	(0.051)	(0.064)	(0.051)	(0.047)
ln(rail)*ln(road)	0.001	-0.000	0.001*	0.002**
	(0.001)	(0.001)	(0.001)	(0.001)
Topography				
Island	-0.678***	-0.082	-0.324**	-0.637***
	(0.190)	(0.169)	(0.157)	(0.175)
Mountainous	-0.145**	0.007	-0.037	-0.203***
	(0.070)	(0.070)	(0.061)	(0.065)
Total Waterway Length	0.012	-0.017*	-0.001	0.014
	(0.009)	(0.010)	(0.009)	(0.009)
ln(road access)^2	0.011***	0.006**	0.013***	0.016***
	(0.002)	(0.003)	(0.002)	(0.002)
ln(rail access)2	-0.004	-0.005	-0.006**	-0.007**
	(0.003)	(0.003)	(0.003)	(0.003)
ln(density)	0.042	0.036	-0.001	0.023
	(0.039)	(0.038)	(0.034)	(0.040)
ln(GDP)	0.151***	0.092***	0.138***	0.158***
	(0.028)	(0.027)	(0.024)	(0.027)
age	-0.004	0.005	-0.008	0.011*
	(0.006)	(0.007)	(0.006)	(0.006)

Table 4.7 continued

Constant	9.235***	9.941***	9.486***	8.541***
	(0.500)	(0.505)	(0.424)	(0.480)
Observations	973	972	973	974
R-squared	0.254	0.078	0.242	0.352

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

Tracing back the data provides some interesting results which reinforce the accuracy of the model while evaluating its precision. Firstly, there is a greater variance in factors across the board, crossing the conservative threshold of 5 however as previously it is a lesser concern. Secondly all coefficients have a greater standard error, which means that data estimation is more spread out, further making many of the coefficients no longer significant. Comparing **Table 4.7** alongside **Table 4.5** indicates this disparity specifically in rail, density, and age coefficients. This is a major issue, however seeing how the direction of all coefficients is in line with previous findings there could be greater statistical noise in 2011. The 2011 date was much closer to the 2008 housing market crash which would have influenced companies across industries, making the VA metric less representative. Performing a comparison of 2001 and 2006 could confirm this, as the dot com boom would be more felt in 2001 then 2006, although likely only in corporate and manufacturing sectors. The most important finding is that every coefficient, within its respective standard deviation is in the same 95% confidence range of the previous findings. There is insufficient evidence to indicate that the coefficients are different between 2011 and 2014, which supports the robustness and accuracy of this model in measuring the relationships of interest.

A very clear divergence from the original results, is that in 2011 the retail sector is missing, this was due to lacking value-added data from this year in this NACE category, with less than 650 regions publishing data on retail. This is simply a symptom of database limitations, as using industry specific value-added metrics is dependent on intergovernmental institutions procuring information from different local and federal bodies that in term would need to manage private company data. The fact that 4 industry groups can consistently have complete data to the extent that the results differ in sample by 1 observation (973 vs. 972) is an impressive achievement in its own regard.

This test for model validity provides some mixed conclusions. On one side, the magnitudes and relative forces outlined are in line with the original model, but there is also less confidence in the value of these coefficients. Changing the point in time of a study is a key indicator of reliability for cross-sectional studies, it can determine the model's credibility in accounting for time variant confounders that are by construction exogenous. This is crucial for directions of the relationships, as a change in relationship would indicate that the model is inconsistent, and thereby lacks predictive power. However, the results above show models that have similar findings, albeit at a lower significance in certain variables. Up to 2014, the investment into infrastructure had been shifting from rail to road as was seen in **Figure 1.1**, which either directly or following time lags could mean that there were real differences in infrastructure capacity in 2014, that were less evident in 2011.

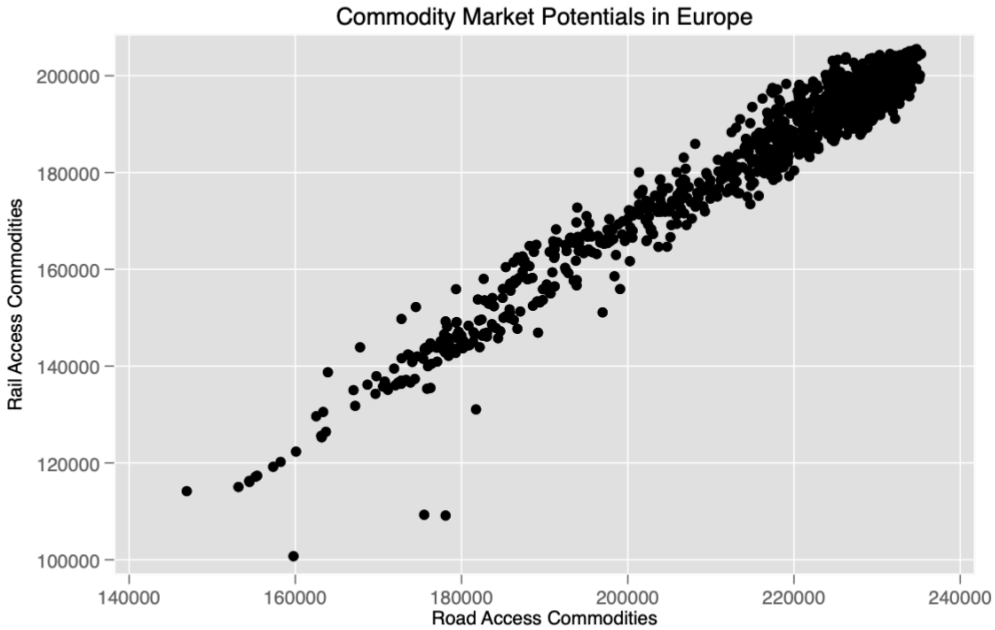


Figure 4.9: *Scatter plot showing the market accessibility by road and rail transport across European NUTS 3 Regions in 2011 in 1,000 of Euro per minute.*

A potential explanation for the low confidence of results can be derived from **Figure 4.9**. Visually it is clear when comparing to **Figure 3.4**, the same scatterplot from 2014, one scatter is linear the other not. Further, there is less variance in the results, and the overall size of market potentials is significantly lower. The lesser variance implies a greater collinearity as a stronger correlation is present which explains the higher VIF scores. The lower market potentials are reflective of market growth, since when referring to the CME group market cap which indexes financial derivatives in commodities, the market cap for this sector grew by around 35% in the three years between 2011 and 2014 (Macro Trends LCC, 2023). Overall, the model still produces consistent, but insignificant results that support the observations from the initial model.

5. Conclusion

This study aimed to answer how European regions benefit from market accessibility depending on the availability of specific modalities and for distinct industries. Evidence points to rail being a more powerful driver of market potential, however it is hindered by diminishing returns to scale. Road transportation is associated with greater increases to average value added at greater values of market access if all else is kept constant, however within the European context it is inefficient relative to rail. Market accessibility between modes of transport is found to be complementary, with a positive cross elasticity, confirming the first hypothesis group H1a and H1b. Additionally, this thesis indicates that the unimodal relationships of market accessibility and productivity are non-linear, supporting the theoretical findings of several authors. The study looked at transportation unimodally to develop an understanding of the intermodal differences between road and rail transportation. This allowed for individual efficiencies, complementarity, and interaction to be better understood across a large geographical area. A strict examination of the model and results will be followed by the contributions and future developments of this research.

While the first hypothesis was proven correct with significant and positive associations between rail market access and average value added alongside positive interaction in some industries, the models found insufficient evidence to disprove the second hypothesis. There were no significant differences between industries in the research, with the same direction and preference for rail across all five industries studied. This is likely a limitation by construction as the model prioritized measuring market accessibility at an individual industry level but did not measure industries simultaneously or in a single regression. With the null not rejected, and industries assumed to have no significant difference in direction for market accessibility potency, the second sub-hypothesis of H1b should be kept in mind. A noticeable industry level difference was in manufacturing where there was insufficient evidence to indicate complementarity between modes of access. While a possible problem with the fit of the model, as this interaction was mostly positive and in Yi & Kim (2018) in a study of Korean manufacturing. There is also the factor of European manufacturing being a small low value industry, which both makes the lack of evidence both logical and expected.

More importantly, the same non-linearity is found in the European model as was in the Korean research, and as was found in several works outlined by Zgonc, Tekavčič, & Jakšič (2019)/ Firstly, by formulation and as indicated in the results, this confirms the first geographic hypothesis that returns to scale are mainly driven by distance accepting H3a. As for the last hypothesis H3b, the extension partly addressed physical geography, with robust findings that adverse geography is associated with more productive regions. Those in term were regions with greater accessibility, and although those areas also had greater market access, there was no evidence for a preference to road or rail aside from specific methodological context where geography would be assumed to be a driver of potential. This means that there is insufficient evidence to reject the null of H3b.

The research was posed with significant challenges which would later force limitations, however it also brings several contributions to the current study of intermodal market accessibility. Firstly, this research took advanced data from the University of Valencia, and using publicly available data was able to compound a database of over 900 regions in Europe with industry data and market accessibility data at differing levels and across two years. The framework for using, adjusting for changes in nomenclature, and conversion can be reproduced on a myriad of European studies by following the methodology. Further, several methods of (dis)aggregating data, converting matrices, and transforming I-O and trade flows tables were used in diverse applications; including estimating downstream markets and calculating decay parameters for market potential. A wide range of modelling techniques was used, testing production, efficiencies, and intermarket relationships were estimated using both linear and polynomial log-log regression.

Besides the aforementioned literature, several other theories and works were expanded upon. Firstly, there is sufficient evidence to indicate complementarity in several industries, which with freight would align to the findings of Liu et al. (2019). The complementarity is not evident in the 2011 model, but given ample literature and significant results in 2014, an argument can reasonably made to extend this finding to the European setting. As Gutiérrez et al. (2011) discussed the role of value-added is more complex then can be seen, when looking through a simple scope of market potential rail is a stronger driver. The diminishing returns to scale in accessibility indicate that uni-modal solutions are inefficient, as relationships between modalities are neither transitive nor linear. A possible contribution from earlier, is the work of Lileeva & Teffler (2010) in estimating the Melitz cut off, the results fit into a logarithmic model as there is a predictably large number of regions that support domestic markets and do not export. Therein, a low market accessibility leads to low influence on value added which is captured in the constant of the regression, as a sensible cut off would be difficult to determine and would funnel down the regression into highly productive regions.

The final contribution is that this research has extended to look at three dimensions, differences by mode, industry, and region. As was indicated in metric comparisons to Rosik, Stępniak, & Komornicki (2015) studies usually scale their region size to their area of interest. This thesis has used small intra-regional NUTS 3 regions on a continental level of analysis. Hence aside from the non-linearities, by virtue of scope where distances analysed are very far apart, rail transportation is the logically superior mode of transportation. The local markets make up a small sample of the population, which is insufficient to seriously impact the model.

Regarding the limitations of this paper, they can be summarised to the assumptions, data availability, and choice of model. Firstly, the travel travel times used were produced by assuming circular region geometry. This made it easier to determine the points of origin and destination between NUTS 3 regions, however it is not very representative since businesses do not have to locate at the centre. Likewise, regions are not always circular in shape as is the case with Italy and Norway. Two other assumptions were in the efficiency of labour markets and prices of goods being equal to marginal cost. On this basis, value added could be fully attributed to superior accessibility and agglomeration economies, however it is certain that there are various frictions and market failures. Effectively, it is not credible that the market is perfectly mobile for firms and workers, and no industry faces a perfectly competitive market. While the issue of circularity harms the precision of results, the latter two assumptions likely inhibit the results accuracy, leading to downward bias as the frictions in competition weaken the influence of market access on industry performance. Future research should base measurements on real prices, flows and displacements perhaps by focusing on country level research where a complete database could be established.

With that being said, data availability was a major hindrance to the extent and precision of control variables available for research. Eurostat is an extensive and useful database for recording data across very diverse regions from over 27 countries, however data will inadvertently be missing, or non-existent in certain regards. Select approximations can be accepted, hence a robust model was able to estimate the relationships studied, however in the future more attention is needed to enable precise and accessible European studies. Select countries such as Denmark, Ireland, Spain, or the Netherlands are meticulous in documentation, but that is not true for all. A future extension on the paper would be to introduce more complex database methods to go further beyond the available Eurostat repository. This limitation for the research inhibited the scope that the model could address as construction was limited, nevertheless the research question could be better approached if individual country data would be standardized and manually constructed. Additionally, a panel data set could be developed if O-D matrices were available on an annual basis with enough subsequent time periods.

Lastly from the major limitations are the methodological concerns. While there was a vast array of logistic, multilevel and event study methods that could be used to address the question, the choice of a logarithmic regression was most optimal for the scope of the research question while satisfying the above conditions. The model could not capture discrete firm differences which are likely to drive a large amount of variation. This on top of a limited dimension, which does not consider success and conception probabilities, levels of effects, or time variance limits the internal and external validity. Due to the multivariable nature of market access, value added, and decay parameters, the model was satisfactory in describing a European setting but likely would falter in other regions. This is due to structural differences drastically influencing model predictive power as was seen in the 2011 results. Despite this, these errors were by construction, since the method was chosen to control for internal validity by being verifiable in its interactions, mediation and omitted variable bias.

Overall, this thesis found evidence for greater potency of railway market access in Europe when it comes to improving the value added per employee across five industries. Excluding manufacturing, the research also indicated minor but significant complementarity between road and rail market access. The research is valid and could use further developments of advanced data prospecting, and modelling. Accessibility was indicated to be driven by displacement and capacity, and the in-depth analysis of decay parameters offers a robust way to measure potential. Crucially, the paper sheds light onto the challenges of analysing a vast area such as Europe, and future research should focus on intramodality to calculate the thresholds and combinations that benefit regional market efficiency.

Bibliography

Alstadt, B. B., Weisbrod, G., & Cutler, D. (2012). Relationship of Transportation Access and Connectivity to Local Economic Outcomes. *Transportation Research Record*, *2297*(1), 154–162. https://doi.org/10.3141/2297-19

Beckers, T., Haunerland, F., Hirschhauser, C., & Walter, M. (2009). The Future for Interurban Passenger Transport. In *Long-Distance Passenger Rail Services in Europe: Market Access Models and Implications for Germany* (18th ed., pp. 287–310). OECD. https://doi.org/10.1787/9782982102688-en

Buehler, R. (2011). Determinants of transport mode choice: a comparison of Germany and the USA. *Journal of Transport Geography*, *19*(4), 644–657. https://doi.org/10.1016/j.jtrangeo.2010.07.005

Centralny Port Komunikacyjny (CPK). (2023, March 17). *High-speed rail to boost economic development in Central and Eastern Europe*. Global Railway Review. Retrieved March 29, 2023, from https://www.globalrailwayreview.com/article/142463/high-speed-rail-boost-economic-development-central-eastern-europe/

Dailey, E. (2022, October 21). Connecting the hinterlands: challenges faced by rail and barge | Project Cargo Journal. *Project Cargo Journal*. Retrieved October 7, 2023, from https://www.projectcargojournal.com/modalities/2022/10/21/connecting-the-hinterlands-challenges-faced-by-rail-and-barge/?gdpr=deny

De Bartolome, C. A., & Ross, S. L. (2003). Equilibria with local governments and commuting: income sorting vs income mixing. *Journal of Urban Economics*, *54*(1), 1–20. https://doi.org/10.1016/s0094-1190(03)00021-4

De Donnea, F. (1972). Consumer behaviour, transport mode choice and value of time: Some micro-economic models. *Regional and Urban Economics*, *1*(4), 355–382. https://doi.org/10.1016/0034-3331(72)90025-5

Di Pietrantonio, L., & Pelkmans, J. (2004). The Economics of EU Railway Reform. *Competition and Regulation in Network Industries*, *5*(3–4), 295–346. https://doi.org/10.1177/178359170400500304

Engel, A. W. V. D., van den. (2010). Driving Restrictions for Heavy Goods Vehicles in the European Union. In *transport.ec.europa.eu* (No. DM28, 0/36). European Commission. Retrieved July 15, 2023, from https://transport.ec.europa.eu/system/files/2016-09/2010_12_driving_restrictions_for_heavy_goods_vehicles_in_the_european_union.pdf

European Environment Agency. (2021, May 11). *Investment in transport infrastructure*. Retrieved May 27, 2023, from https://www.eea.europa.eu/data-and-maps/indicators/infrastructure-investments/assessment-3

European Rail Freight Association. (2021, September 28). *Achieving modal Shift: The Swiss Alpine Example - ERFA - European Rail Freight Association*. erfarail.eu. Retrieved September 7, 2023, from https://erfarail.eu/news/press-release-achieving-modal-shift-the-swiss-alpine-example#:~:text=The%20first%20half%20of%202021,sure%20the%20highest%20in%20Europe.

Eurostat. (2015, December 15). *Net migration into NUTS 3 regions 2001-2011 | Data | European Structural and Investment Funds*. https://cohesiondata.ec.europa.eu/Other-Eurostat/Net-Migration-Into-NUTS-3-Regions-2001-2011/74hm-qims

Eurostat. (2023a, February). *Employment (thousand persons) by NUTS 3 regions*. ec.europa.eu. Retrieved April 6, 2023, from https://ec.europa.eu/eurostat/databrowser/view/NAMA_10R_3EMPERS__custom_6995369/default/table?lang=en

Eurostat. (2023b, February). *Gross domestic product (GDP) at current market prices by NUTS 3 regions*. ec.europa.eu. Retrieved April 5, 2023, from https://ec.europa.eu/eurostat/databrowser/view/NAMA_10R_3GDP__custom_6995574/default/table?lang=en

Eurostat. (2023c, February). *Inland transport infrastructure at regional level*. ec.europa.eu. Retrieved June 30, 2023, from https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Inland_transport_infrastructure_at_regional_level

Eurostat. (2023d, February). *Population on 1 January by age group, sex and NUTS 3 region*. ec.europa.eu. Retrieved June 6, 2023, from https://ec.europa.eu/eurostat/databrowser/view/DEMO_R_PJANGRP3/default/table?lang=en

Eurostat. (2023e, February). *Population structure indicators by NUTS 3 region*. ec.europa.eu. Retrieved May 3, 2023, from https://ec.europa.eu/eurostat/databrowser/view/demo_r_pjanind3/default/table?lang=en

Eurostat. (2023f, February 21). *Gross value added at basic prices by NUTS 3 regions*. data.europa.eu. Retrieved April 3, 2023, from https://data.europa.eu/data/datasets/vhcfyrau2sc2fmn0pneyuw?locale=en

Ford-Alexandraki, E., Xenellis, G., Corselli-Nordblad, L., & Jere, N. (Eds.). (2022). Key figures on European transport. In *Eurostat*. European Union. https://doi.org/10.2785/322262

Fujita, M., Krugman, P., & Venables, A. J. (1999). The Spatial Economy. In *The MIT Press eBooks*. https://doi.org/10.7551/mitpress/6389.001.0001

Gutiérrez, J. C., Condeço-Melhorado, A., López, E. S., & Monzon, A. (2011). Evaluating the European added value of TEN-T projects: a methodological proposal based on spatial spillovers, accessibility and GIS. *Journal of Transport Geography*, *19*(4), 840–850. https://doi.org/10.1016/j.jtrangeo.2010.10.011

Hochard, J., & Barbier, E. B. (2017). Market Accessibility and Economic Growth: Insights from a New Dimension of Inequality. *World Development*, *97*, 279–297. https://doi.org/10.1016/j.worlddev.2017.04.018

Latschan, T. (2022, May 7). Ukraine: Will the railroad decide the war? *dw.com*. https://www.dw.com/en/ukraine-will-the-railroad-be-what-decides-the-war/a-61714831

Li, J., Li, L., & Liu, S. (2022). Attenuation of agglomeration economies: Evidence from the universe of Chinese manufacturing firms. *Journal of Urban Economics*, *130*, 103458. https://doi.org/10.1016/j.jue.2022.103458

Lileeva, A. L., & Treffler, D. T. (2010). Improved Access to Foreign Markets Raises Plant-Leve; Productivity. . . for Some Plants. *The Quarterly Journal of Economics*, *125*(3), 1051–1099. https://www.jstor.org/stable/27867506

Limb, L. (2023, April 4). *It's official: France bans short haul domestic flights in favour of train travel*. Euronews. Retrieved April 4, 2023, from https://www.euronews.com/green/2022/12/02/is-france-banning-private-jets-everything-we-know-from-a-week-of-green-transport-proposals

Liu, D., Deng, Z., Sun, Q., Wang, Y., & Wang, Y. (2019). Design and Freight Corridor-Fleet Size Choice in Collaborative Intermodal Transportation Network Considering Economies of Scale. *Sustainability*, *11*(4), 990. https://doi.org/10.3390/su11040990

MacroTrends LLC. (2023). *CME Group Market Cap 2010-2023 | CME*. MacroTrends. Retrieved July 23, 2023, from https://www.macrotrends.net/stocks/charts/CME/cme-group/market-cap

Marschinski, R., David, M. T., & Turegano, M. D. (2019). Reassessing the Decline of EU Manufacturing: A Global Value Chain Analysis. *Publication Office of the European Union*. https://doi.org/10.2760/30611

Niet, I. (2023, February 15). *Low carbon travel: all universities in action? - Low Carbon Travel Initiative*. Low Carbon Travel Initiative. Retrieved April 2, 2023, from http://lowcarboninitiative.nl/2023/02/15/low-carbon-travel-all-universities-in-action/

Otsuka, A., Goto, M., & Sueyoshi, T. (2010). Industrial agglomeration effects in Japan: Productive efficiency, market access, and public fiscal transfer. *Papers in Regional Science*, *89*(4), 819–840. https://doi.org/10.1111/j.1435-5957.2010.00286.x

Papatolios, N., & Papatolios, N. (2023, February 9). *Road or rail investments? German rail associations think this debate has gone too far | RailFreight.com*. RailFreight.com. Retrieved March 22, 2023, from https://www.railfreight.com/policy/2023/02/10/road-or-rail-investments-german-rail-associations-think-this-debate-has-gone-too-far/

Persyn, D. (2022). *The ripple effects of large-scale transport infrastructure investment*. http://hdl.handle.net/10419/265239

Reggiani, A., Bucci, P., & Russo, G. (2012). Accessibility and Network Structures in the German Commuting. *Networks and Spatial Economics*, *11*(4), 621–641. https://doi.org/10.1007/s11067-010-9149-0

Rosik, P., Stępniak, M., & Komornicki, T. (2015). The decade of the big push to roads in Poland: Impact on improvement in accessibility and territorial cohesion from a policy perspective. *Transport Policy*, *37*, 134–146. https://doi.org/10.1016/j.tranpol.2014.10.007

Shuai, X. (2012). Does commuting lead to migration? *Journal of Regional Analysis & Policy*, *42*(3). https://scholarship.richmond.edu/spcs-faculty-publications/2/

Statista. (2023, July 19). *Preferred mode of transport for European travelers 2023*. Retrieved September 7, 2023, from https://www.statista.com/statistics/1325231/preferred-mode-of-transport-for-european-travelers/

Timmer, M. P., Dietzenbacher, E., Los, B., Stehrer, R., & De Vries, G. J. (2015). An Illustrated User Guide to the World Input-Output Database: the Case of Global Automotive

Production. *Review of International Economics*, *23*(3), 575–605. https://doi.org/10.1111/roie.12178

University of Valencia. (2020, April 21). *Indicator: Road travel time between centroids of NUTS-3 regions | ESPON Database Portal*. ESPON. Retrieved March 30, 2023, from https://database.espon.eu/indicator/1542/#metadata-download

Walton, S. (2022, December 15). *Freight figures but not top priority in Scottish transport plans | RailFreight.com*. RailFreight.com. Retrieved March 30, 2023, from https://www.railfreight.com/railfreight/2022/12/16/freight-figures-but-not-top-priority-in-scottish-transport-plans/

Weisbrod, G., & Goldberg, J. (2022). Relationship of Regional, Freight, and Intermodal Market Access to Industry Location and Productivity. *Transportation Research Record*, 036119812211313. https://doi.org/10.1177/03611981221131306

Westbrook, M. D., & Buckley, P. F. (1990). Flexible Functional Forms and Regularity: Assessing the Competitive Relationship between Truck and Rail Transportation. *The Review of Economics and Statistics*, *72*(4), 623. https://doi.org/10.2307/2109602

Yi, Y., & Kim, E. (2018). Spatial economic impact of road and railroad accessibility on manufacturing output: Inter-modal relationship between road and railroad. *Journal of Transport Geography*, *66*, 144–153. https://doi.org/10.1016/j.jtrangeo.2017.11.017

Zgonc, B., Tekavčič, M., & Jakšič, M. (2019). The impact of distance on mode choice in freight transport. *European Transport Research Review, 11*(1). https://doi.org/10.1186/s12544-019-0346-8

Appendix

Table 1: Reported value added by NACE category across NUTS 3 regions.

NACE Code	Observations	Frequency
A	1292	0.97289157
B-E	1290	0.97138554
C	1296	0.97590361
F	1296	0.97590361
G-I	721	0.54292169
G-J	1296	0.97590361
J	725	0.54593373
K	726	0.54668675
K-N	1296	0.97590361
L	684	0.51506024
M-N	726	0.54668675
O-Q	732	0.55120482
O-U	1302	0.98042169
R-U	726	0.54668675
TOTAL	**1328**	**0.758821**

Table 2: Weighted average input table component from the data studied.

Input	ABE	C	F	KNJ	GIR	Intensity
ABE	33.0%	13.3%	1.8%	1.2%	3.9%	53.3%
C	38.5%	68.4%	41.1%	10.0%	31.8%	189.8%
F	3.9%	1.5%	35.1%	7.2%	4.9%	52.5%
KNJ	16.0%	9.1%	12.1%	55.5%	31.9%	124.6%
GIR	8.6%	7.6%	10.0%	26.2%	27.4%	79.8%
Total	100.0%	100.0%	100.0%	100.0%	100.0%	

Table 3: Model development using the retail sector.

ln(Value Added per Employee)	(Model 1)	(Model 2)	(Model 3)	(Model 4)
Commodities Sector	Basic	Control	Interaction	Non-linear
ln(MA_Road)	0.062***	0.038***	0.038***	-0.107***
	(0.007)	(0.006)	(0.011)	(0.023)
ln(MA_rail)	-0.033***	-0.033***	-0.033***	0.095***
	(0.007)	(0.007)	(0.009)	(0.021)
ln(MA_Road)* ln(MA_rail)			0.000	0.001
			(0.001)	(0.001)
$\ln(MA\_rail)^2$				0.008***
				(0.001)
$\ln(MA\_rail)^2$				-0.008***
				(0.001)
ln(density)		0.015	0.015	-0.009
		(0.011)	(0.012)	(0.012)
ln(OriginDensity)		-138.261***	-138.398***	-80.515***
		(20.801)	(21.405)	(21.347)
ln(GDP)		0.286***	0.286***	0.270***
		(0.019)	(0.021)	(0.020)
Age		0.026***	0.026***	0.018***
		(0.004)	(0.004)	(0.004)
Constant	10.157***	842.225***	843.050***	493.941***
	(0.081)	(125.038)	(129.242)	(128.517)
Observations	972	972	972	972
R-squared	0.163	0.443	0.443	0.479
Mean VIF	1.31	1.27	4.10	17.50

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

Table 4: *Naïve regressions of market access by mode on value added in manufacturing.*

VARIABLES	(Naïve 1)	(Naïve 2)
log(Market_Access_Manufacturing_rail)	0.005	
	(0.003)	
log(Market_Access_Manufacturing_road)		0.052***
		(0.007)
Constant	10.682***	9.870***
	(0.051)	(0.099)
Observations	973	973
R-squared	0.000	0.143

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 5: Initial regression with migration approximated decay values 0.25 and 0.70

VARIABLES	Manufacture ln(VA/emp)	Corporate ln(VA/emp)	Construction ln(VA/emp)	Commodities ln(VA/emp)	Retail ln(VA/emp)
ln(road accessibility)	-0.100***	-0.070***	-0.167***	-0.147***	-0.096***
	(0.028)	(0.022)	(0.031)	(0.036)	(0.026)
ln(rail accessibility	0.107***	0.175***	0.185***	0.134***	0.078***
	(0.022)	(0.017)	(0.023)	(0.026)	(0.021)
ln(rail)*ln(road)	0.002	-0.000	0.003	0.004*	0.001
	(0.002)	(0.001)	(0.002)	(0.002)	(0.002)
ln(road access)^2	0.014***	0.010***	0.018***	0.018***	0.013***
	(0.002)	(0.001)	(0.002)	(0.002)	(0.002)
ln(rail access)2	-0.012***	-0.013***	-0.016***	-0.015***	-0.008***
	(0.002)	(0.001)	(0.002)	(0.002)	(0.001)
ln(density)	-0.051***	-0.071***	-0.007	-0.005	-0.039***
	(0.014)	(0.011)	(0.016)	(0.019)	(0.013)
ln(Origin density)	-151.309***	-377.682***	188.626***	101.495***	-62.528***
	(23.978)	(20.457)	(26.282)	(28.615)	(21.754)
ln(GDP)	0.219***	0.156***	0.267***	0.305***	0.259***
	(0.020)	(0.013)	(0.022)	(0.025)	(0.019)
age	0.021***	0.020***	0.022***	0.040***	0.014***
	(0.005)	(0.004)	(0.006)	(0.007)	(0.004)
Constant	922.351***	2,291.472***	-1,132.101***	-606.679***	385.697***
	(145.014)	(123.653)	(158.262)	(173.104)	(131.114)
Observations	973	972	972	973	972
R-squared	0.469	0.415	0.519	0.554	0.506

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

Table 6: Initial regression with migration approximated decay values 0.08 and 0.18

VARIABLES	Manufacture ln(VA/emp)	Corporate ln(VA/emp)	Construction ln(VA/emp)	Commodities ln(VA/emp)	Retail ln(VA/emp)
ln(road accessibility)	-0.115***	-0.081***	-0.165***	-0.155***	-0.108***
	(0.025)	(0.020)	(0.028)	(0.033)	(0.024)
ln(rail accessibility	0.122***	0.183***	0.196***	0.152***	0.091***
	(0.022)	(0.018)	(0.023)	(0.027)	(0.021)
ln(rail)*ln(road)	0.002	0.000	0.002*	0.003**	0.001
	(0.001)	(0.001)	(0.001)	(0.002)	(0.001)
ln(road access)^2	0.010***	0.007***	0.012***	0.013***	0.009***
	(0.001)	(0.001)	(0.002)	(0.002)	(0.001)
ln(rail access)2	-0.011***	-0.012***	-0.015***	-0.014***	-0.008***
	(0.001)	(0.001)	(0.001)	(0.002)	(0.001)
ln(density)	-0.024*	-0.057***	0.026*	0.034*	-0.015
	(0.013)	(0.010)	(0.015)	(0.018)	(0.013)
ln(Origin density)	-169.380***	-387.090***	162.956***	71.862**	-79.304***
	(24.579)	(20.306)	(26.115)	(28.540)	(21.394)
ln(GDP)	0.229***	0.160***	0.277***	0.317***	0.267***
	(0.021)	(0.014)	(0.023)	(0.026)	(0.020)
age	0.024***	0.022***	0.026***	0.044***	0.017***
	(0.005)	(0.004)	(0.006)	(0.007)	(0.004)
Constant	1,031.146***	2,348.113***	-977.459***	-428.170**	486.713***
	(148.048)	(122.589)	(157.543)	(172.281)	(128.952)
Observations	973	972	972	973	972
R-squared	0.440	0.402	0.490	0.526	0.484

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

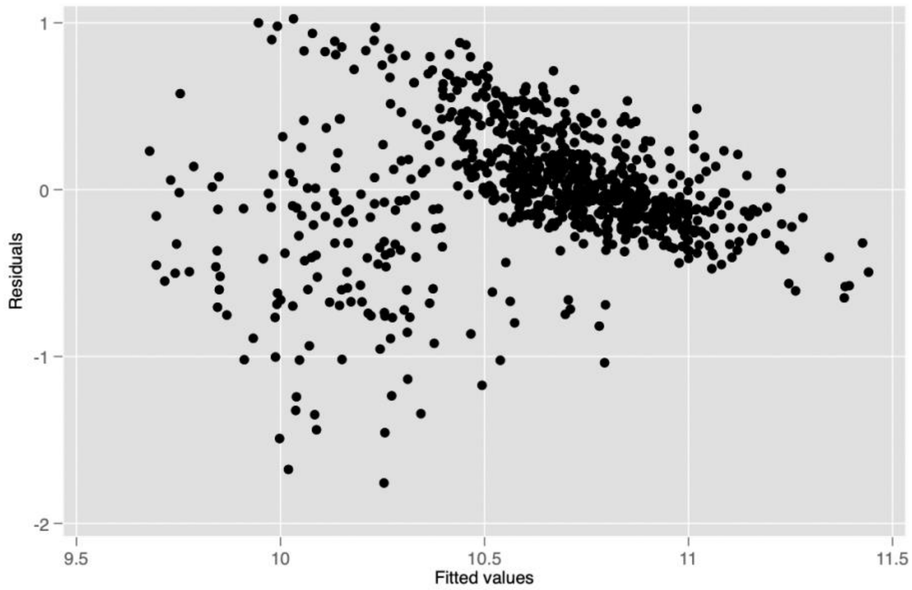


Figure 1: *Bias and heteroscedasticity evident from the initial regression.*

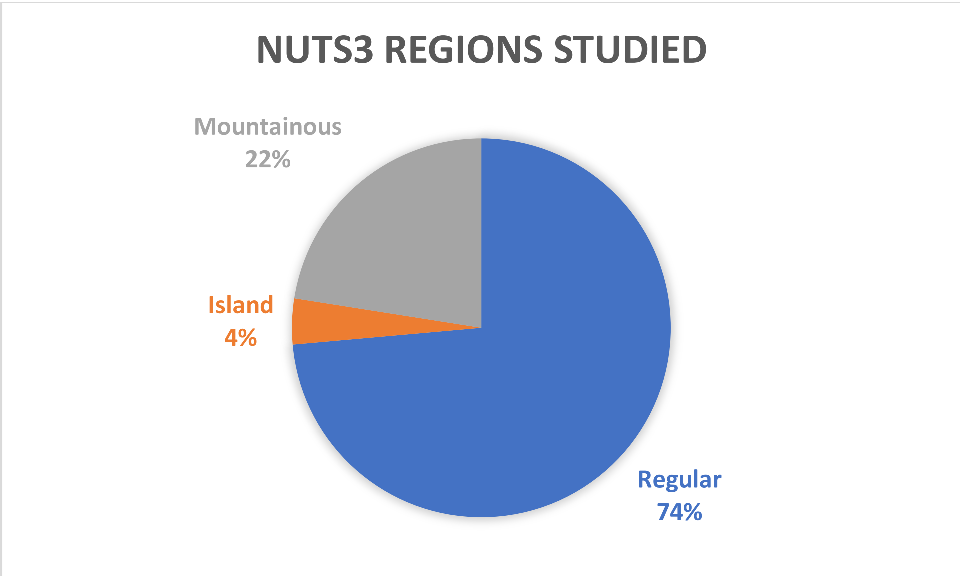


Figure 2: *Proportion of the 974 regions that are either mountainous (n=219) or isolated by sea (38)*

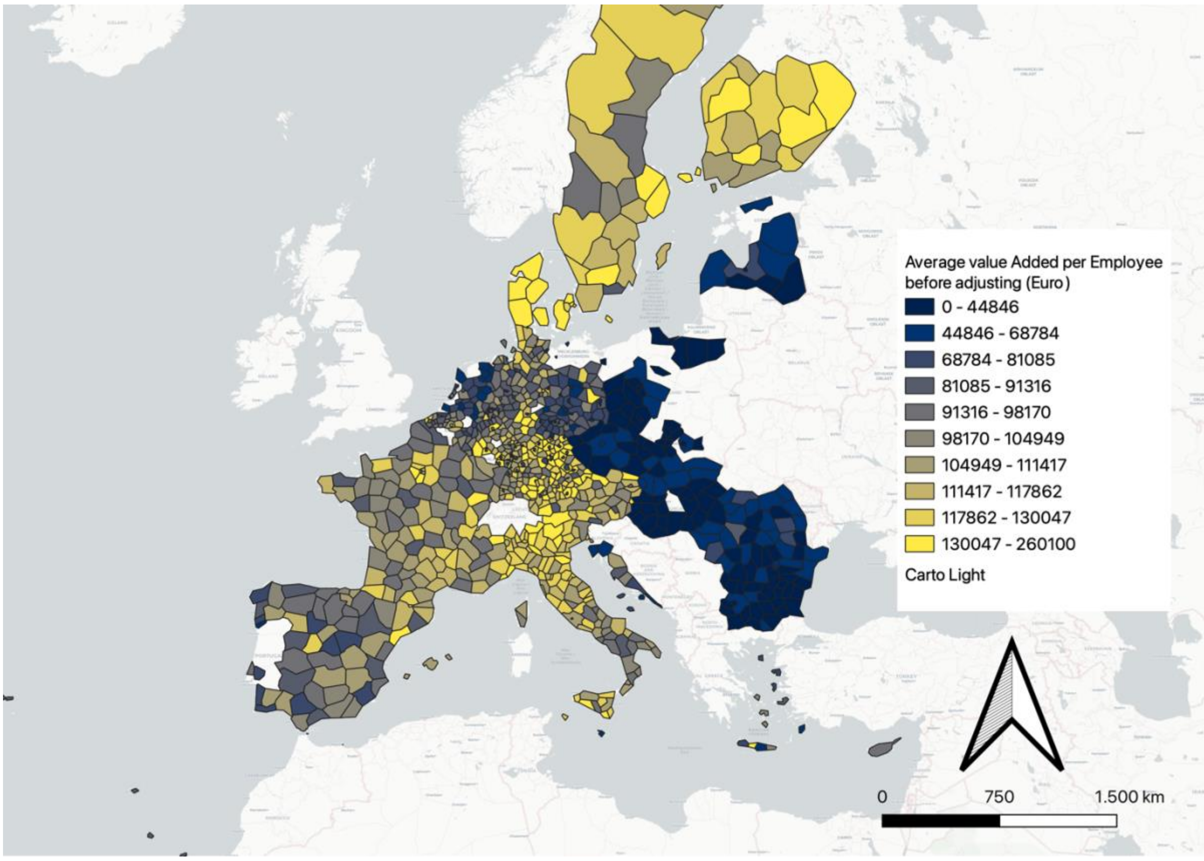


Figure 3: *Average value added per employee in the corporate/financial sector across different European regions.*

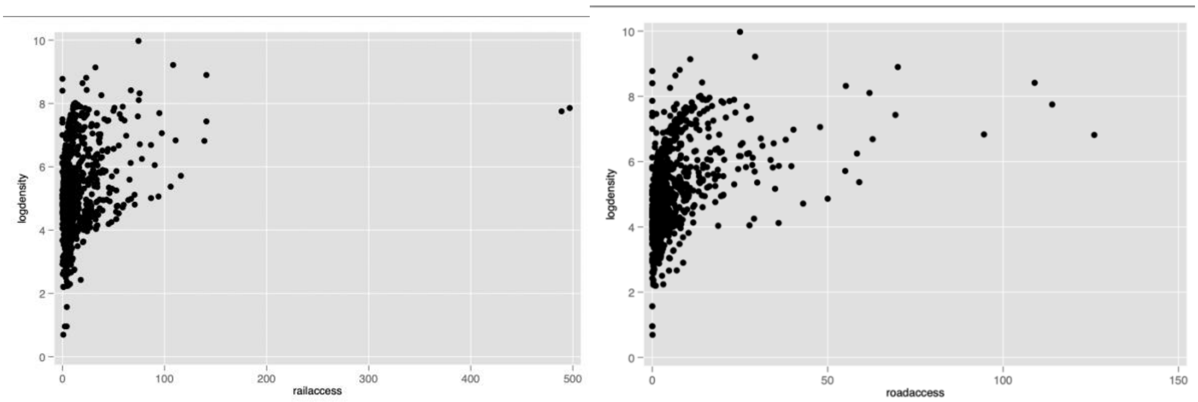
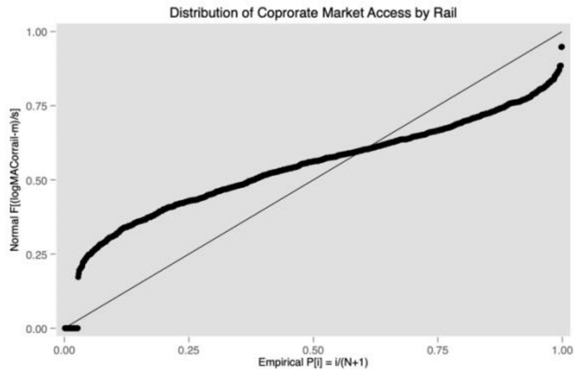


Figure 4 & 5: *relationship between the population density and accessibility of locations.*

Table 7: *Validity and endogeneity tests on the instrumental variables introduced. Sargan X^2 performed with geographical dummy variable as IV component.*

Variable	Test	Value
Log(Rail Accessibility)	Adjusted R^2	0.65**
	F-test	5.18***
	Wald Chi-squared	Not satisfied, below critical value: 19.93
	Sargan Chi squared	5.43**
Log(Road Accessibility)	Adjusted R^2	0.64***
	F-test	57.70***
	Wald Chi-squared	Satisfied, above critical value: 19.93*
	Sargan Chi squared	0.16



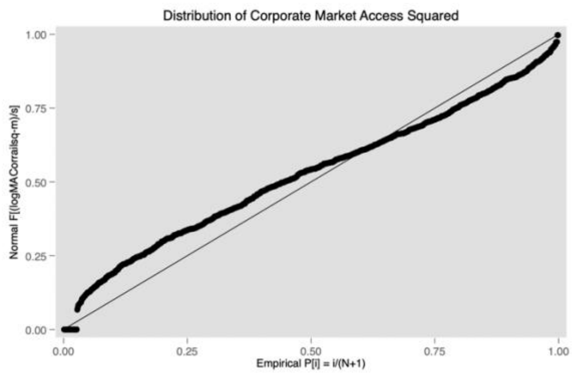


Figure 6 & 7: *Distribution Q-Q plots for corporate market access and corporate market access squared respectively.*