SPATIAL COMPETITION BASED ON CUSTOMERS' CHOICE HISTORIES:
A STUDY OF TRADE FLOWS

by

Mona Kashiha

A dissertation submitted to the faculty of
The University of North Carolina at Charlotte
in partial fulfillment of the requirements
for the degree of Doctor of Philosophy in
Geography and Urban Regional Analysis

Charlotte

2014

Approved by:

____
Dr. Jean-Claude Thill

____
Dr. Harrison Campbell

____
Dr. Wenwu Tang

____
Dr. Craig A. Depken, II

ABSTRACT

MONA KASHIHA. Spatial competition based on customers' choice histories: a study of trade flows. (Under the direction of DR. JEAN-CLAUDE THILL)

In a world of differentiated firms, customers with the same observable characteristics reveal different choice behaviors. This research contributes to the modeling of competition in a market characterized by spatially differentiated firms and significant consumers' taste heterogeneity. With recognition of the fact that the degree of competition depends on the extent to which a representative customer switches among firms, this research proposes an emprical framework for quantifying competitive interactions. The framework is built on the Latent Class Logit and customers' choice histories that allow estimation of heterogenous preferences. Furthermore, relying on unique data of geographical distribution of demand, this framework particularly emphasizes the spatial dimension of competition and provides a disaggregated measure of regional contestability, which is often ignored in the previous studies.

The applicability of the proposed framework is tested on shipper-level data from a business trade dataset to model choices for seaports, as critical nodes in logistics networks that support global trade. We estimate how distance, crossing a national border and port characteristics influence port choice across shippers of different size. The results suggest significant heterogeneity in preferences that gives rise to the intensity of competition. Segments of customers that are willing to cross borders and travel broad geographical distances toward efficient ports generate international competitive forces and encourage ports to improve efficiency, while customers that avoid borders and put large negative

value on distance support the local monopoly of the closest ports with no incentive for efficiency improvement.

ACKNOWLEDGMENTS

I would like to express my gratitude first and foremost to my mentor, Professor Jean-Claude Thill, for his continuous support and intellectual contribution to this dissertation. Jean-Claude provided me with a scientific environment based on academic freedom and gave me the opportunity to try new ideas and follow new paths. I also greatly thank Professor Craig A. Depken, II for his invaluable support and encouragement during the final year of my PhD. I appreciate very much his guidance on the economic aspects of my research, empirical analysis and publications. I thank my committee members, Dr. Harrison Campbell, Dr. Wenwu Tang, Dr. Jennifer Troyer and Dr. Rob McGregor for their invaluable comments and critiques on my research and dissertation.

I have been very fortunate to share my PhD life with wonderful friends and great colleagues. I would like to express my deepest appreciation to Dr. Elizabeth Delmelle, Dr. Diep Dao, Dr. Jae Soen Son, Yuhong Zhou, Kailas Venkitasubramanian, Ran Tao, Yasaman Kamyab, Sheelan Vaez and Kaveh Daneshvar, for their love, encouragement, support, humor, and wisdom. On the personal note, I am very thankful to my parent and my sister for their unconditional love and support.

Finally, I am thankful for the financial support from the Graduate Assistant Support Plan (GASP) from UNC Charlotte as well as fellowship from BELK College of Business.

TABLE OF CONTENTS

CHAPTER 1: INTRODUCTION 1

1.1. Statement of Research 2

1.2. Application of the Model: Shipping Industry and Seaports 4

1.2.1. Modeling Consumer Choices and Port Competition 6

1.3. Structure of the Dissertation 8

CHAPTER 2: LITERATURE REVIEW 10

2.1. Literature Review: Spatial Competition 11

2.1.1. Theory of Spatial Competitions 11

2.1.2. Empirical Modeling of Spatial Competition 14

2.2. Port Systems 17

2.2.1. Evolution of Port Competition 18

2.2.2. Methodological Review 20

2.2.3. Port Choice Framework 22

2.3. Literature Review: International Trade 24

2.3.1. Distance and Geography 26

2.3.2. Border 28

2.3.3. Infrastructures 29

2.3.4. Modality 31

2.3.5. Port Performance and Efficiency 32

2.3.6. Data type, Sampling, and Geographical Setting 34

CHAPTER 3: RESEARCH QUESTIONS 35

3.1. Probabilistic Modeling of Spatial Competition 36

CHAPTER 4: DATA PREPARATION 38

4.1. Data Source 38

4.1.1. Geographic Information Retrieval 39

4.1.2. Identification of Place Names 41

4.1.3. Full-Text Search Capabilities 42

4.1.4. Web-mining 43

4.1.5. Disambiguating Place Names 43

4.1.6. Geocoding 44

4.1.7. Verification and Market Intelligence 45

4.1.8. Fuzzy Matching 47

4.1.9. Final Dataset 48

4.2. Port Efficiency: Stochastic Frontier Analysis 53

4.2.1. Methodology 55

4.2.2. Model Specification 57

4.2.3. Ports Technical Efficiency 58

CHAPTER 5: MODELING FRAMEWORK 63

5.1. A Framework for Choice Theory 64

5.1.1. Choice Set Definition 66

5.1.2. Competition Measurement 68

5.2. Modeling Customer Choices 70

5.2.1. Standard Logit Model 70

5.2.2. Latent Class Logit 73

CHAPTER 6: RESULTS 76

6.1. Description of Data 76

6.2. Model Specification 81

6.3. Empirical Results 85

6.3.1. Market Segmentation 89

6.3.2. Competitive Interactions 93

6.4. Simulations 102

6.4.1. Border Dissolution 102

6.4.2. Improving Efficiency 105

6.4.3. Reducing Distance to Ports 112

6.4.4. Structural Change 120

CHAPTER 7: CONCLUSIONS 122

REFERENCES 128

CHAPTER 1: INTRODUCTION

As a result of advanced communication and transportation technologies, geographical barriers have lost potency in separating people, businesses, regions and cultures. These entities now readily interact with each other in response to various intrinsic attributes, while not being merely subservient to geography anymore. This has led to a world of differentiated firms offering products and services to diverse people, while transportation costs add spatial differentiation to the firms dispersed across space (Sloev, Ushchev, & Thisse, 2013).

Consumers distribute their purchases between firms depending on how they evaluate transport costs and other firm characteristics and depending on what range of product varieties is available in the marketplace. Hence, with the changing role of distance which has been likened the "death of distance" by Cairncross (2001), the exclusive market of firms has been increasingly penetrated by competing firms and inter-firm competition has intensified (Cairncross, 2001). In other words, consumers enjoy diversity by purchasing from different firms, and the degree of competition between two firms depends on the extent to which customers switch among these firms.

Although there is a large body of literature on spatial competition, research is mainly based on theoretical discussions, or empirical studies with a limited access to disaggregated spatial demand, which is so crucial to apprehend competitive forces. The existing empirical literature seems to ignore the heterogenous valuation systems of decision makers and underestimates the complex nature of choice behavior by observing single choices of individuals.

This study proposes an empirical framework for spatial modeling of competition between firms in a market characterized by significant consumer taste heterogeneity. This analysis relies on consumers' choice set, consumers and firms' characteristics.

1.1.Statement of Research

This research contributes to the spatial modeling of competition by examining disaggregated choices over a certain period of time, which includes repeated decisions. Our analysis aims at the probabilistic modeling of competitive interactions between spatially differentiated firms in a market characterized by significant taste heterogeneity, whereby an interaction flow space is transformed into a competition space.

We infer the feasible choice set of individuals based on observed choice patterns of decision makers. There is rich variation in consumer choices that can be associated to diverse valuation systems, unknown to the analysts. We claim that two choice alternatives compete only if some switching behavior is observed; in other words, only if both choice options coexist in an individual's feasible choice set. Imputed choice sets are composed of selected alternatives and choice probabilities are associated to each of them.

In fact, competition can be justified only in the presence of variations among the preferences of decision makers. Recently, scholars have recognized the fact that the degree of competition between differentiated products will depend on the degree to which preferences vary in the population (Hastings et al., 2005).

Therefore, the observation of repeated choices is central to competition modeling. Analyzing choice sets, which consist of the repeated choices made by the same consumer over a certain period of time, reflects the consumers' preferences for characteristics of the differentiated firms. With only a single choice it is difficult to distinguish if unexpected behavior is due to a strong preference for certain attributes or is only due to the unobservable factors (Hastings et al., 2005).

This necessitates the integration of concepts of spatial competition and random coefficient discrete choice models for modelling competition under spatial differentiation. Hence, this research employs a Latent Class Logit (LCL) model that endogenously segments the markets into customers with homogenous preferences.

The construct of choice sets allows preference estimation and market segmentation simultaneously. On the one hand, the market is partitioned into consumer segments with similar preferences based on observed choice sets as well as consumer and supplier characteristics. Segments of consumers whose preferences restrict choice to a single supplier constitute captive hinterlands. Conversely, the preference structure of other consumers may be conducive to multiple firms being selected at different occasions; collectively, these consumers form contestable hinterlands. In fact, contestable hinterlands create severe competition by switching between alternatives and represent the market that those alternatives compete over. On the other side of the analysis, competition between two alternatives, generated through market contestability, is quantified using choice probabilities of individuals' choice sets. Simulation allows the calculation of elasticity of demand for each firm with respect to firms' strategies.

The proposed framework can be used in any spatial choice context involving repeated choices where 1) the firms are differentiated, because of their inherent characteristics including geography and other product characteristics, 2) and consumers, whose geographical locations are known, have heterogeneous preferences for differentiated products, 3) and consumer behavior cannot be predicted with certainty.

The applicability of the proposed framework is tested on shipping route choice behavior. This research uses firm-level data from a business trade dataset to model choices for seaports, as critical nodes in logistics networks that support global trade. We estimate preferences for port characteristics (location, efficiency, etc.) and reveal competition between them.

1.2. Application of the Model: Shipping Industry and Seaports

The global economy has experienced tremendous expansion under the combined effect of the liberalization of international trade and the continuing decline in transportation costs, which have opened an era of deep and far-reaching globalization. The increasing intensification of the volume of trade and of its geographic dispersion has necessitated the tremendous expansion of the shipping industry over recent decades. Ports are nodal points of the worldwide maritime networks that interconnect continental inland logistics sub-networks. The intensification of long-haul trade routes has reinforced the critical role of seaports, as gateways to economic spaces and as nodes on the global deep-sea liner shipping networks (T. E. Notteboom & Rodrigue, 2005; J. L. Tongzon & Sawant, 2007). Port competition is the focus of this research as a case study. Traditionally, competition between ports has been related to the exclusive dominance of ports over a region (Slack, 1985), due to the immovable geographic location of the ports and to the concentration of productions (Y. Chang & Lee, 2007; Cullinane & Song, 2006). As a result of containerization and of the emergence of intermodal rail and barge corridors, exclusive hinterlands are now invaded by competing ports and inter-port competition has become a core topic of transport economics (Meersman, Van De Voorde, & Vanelslander, 2010; Slack, 1985).

There is a large body of theoretical and empirical literature that studies port competition from diverse perspectives, such as port performance and efficiency, merging and alliances, ownership and privatization, and port selection factors (Y. Chang & Lee, 2007). We believe that understanding competition requires sound micro-level models of underlying port choice decisions made by port users as manifested by freight flows. In their conceptual paper, Notteboom and Rodrigue (2005) argued that "inland distribution becomes of foremost importance in port competition". Hence, our work is more in line with this body of work that looks at port competitiveness based on port selection factors. However, except for very few papers that provide measures of competition, with a single aggregate market share (M. B. Malchow & Kanafani, 2004; S. J. Veldman & Bückmann, 2003), this body of literature mainly focuses on understanding determining factors and on formulating port selection behaviors (Slack 1985, Murphy and Daley 1994; Nir 2003, Tiwari et al 2003; Blongin and Wilson 2006; Tongzon 2007; Veldman et al 2011). The data used in these studies are survey-based, or aggregated, and represent small samples of the entire population. The somewhat inconsistent results of these studies of determinants of commodity flow distribution are evidence of the complexity of a situation that cannot fit into traditional logit models.

Our proposed framework perfectly matches this spatial choice situation since ports are differentiated by their location, different levels of facilities, services, prices, and different levels of efficiency. Also, they serve a wide range of consumers dispersed across the geographic space, including carriers, forwarders, small and large shippers and so on, with unobservable taste heterogeneity (unknown to the analyst).

1.2.1. Modeling Consumer Choices and Port Competition

We model competition by measuring disaggregated choices imputed from a rich and disaggregated hinterland-port-foreland commodity distribution database. This research focuses on modeling port competition in Europe, which is recognized as a region of fierce competitive in terms of ports interactions (Veldman & Bückmann, 2003). A large and unique database of containerized door-to-door freight shipments from European countries and bound for the United States, compiled from the 2006 PIERS Intelligence database, is used. In this research, the decision making unit refers to "shippers" that indiscriminately represent a wide range of consumer of shipping services, including manufacturers, carriers and freight forwarders, who ship from a production location in Europe to a final destination in the United States. Data contains repeated decisions of each shipper. Data on repeated choices render the understanding of choice behavior much more precise and complete.

There is significant variation in customers' choices; variations across choice sets of customers dispersed across space can largely be explained by the geographical characteristics. However, there is significant variation within choice sets as approximately fifty percent of shippers in the sample data set ship via multiple ports. These variations are caused by unobservable or difficult-to-measure variables and allow us to estimate variations in preferences, as an underlying force of competition.

This research employs a random coefficient logit model with a discrete mixing distribution to explain heterogeneity in preferences over shipper and port characteristics by segmenting shippers into classes with homogenous preferences. The empirical analysis estimates how distance, border crossings, and port characteristics influence the choice of which port to ship from and how these preferences differ across segments of shippers. Efficiency is a port characteristic that has been the focus of a number of studies of port competition. However, there are no universal port-level efficiency indicators or even a standardized measurement available. This dissertation utilizes Stochastic Frontier Analysis (SFA) to estimate ports' relative efficiency.

The LCL model accommodates a wide range of behaviors, avoids the well-known limitations associated with the standard logit model, and has been extensively used in the industrial organization and marketing literature. The standard logit model, which is popular in the port-choice literature, cannot capture heterogeneous preferences and treats repeated choices by the same decision-makers as independent (cross-sectional) observations. Another well-known limitation of the logit model is the Independence of Irrelevant Alternatives (IIA), which imposes restrictions on the relationship between spatial choices.

The results show that consumers are heterogeneous in their valuation of proximity, border crossing, and ports' characteristics. All customers value distance negatively, avoid border strongly and put positive weight on ports efficiency. While the smallest shippers have a strong aversion to crossing borders and are not willing to ship further to reach more efficient ports, the largest shippers cross border readily and put less weight on distance to efficient ports. Competition is rather uneven between ports and its intensity varies based on characteristics of each segment.

With demand estimated with the proposed model, we are able to track changes to the competitive relationships of ports in the entire system in response to strategic positions taken by port authorities, shippers, and shipping lines. Particularly, simulations of customers' responses to improvements in ports' transportation service and efficiency enables port authorities to identify the potential consumers to be particularly targeted for marketing and price discrimination. Also, they can find where their current customers are in danger of being attracted to other competitors as conveyed by customers' degree of contestability.

Specifically, we show that Italian, French and British ports are advantaged by the barrier imposed by national borders, while Dutch, Belgian, Spanish and German ports would greatly benefit if the borders were removed. Also, we find that demand elasticities are small in areas close to the ports and increase with distance from the ports, but beyond some distance, demand is no longer responsive. More interestingly our results demonstrate that demand of efficient ports is much more elastic to efficiency improvement policies than demand of inefficient ports. In other words, efficient ports obtain more market share in response to increases to efficiency level than inefficient ports. While efficiency-related strategies put efficient ports in more competitive positions, inefficient ports have a local monopoly over neighboring customers that do not have high preference for efficiency.

1.3.Structure of the Dissertation

The dissertation is organized as follows: The next chapter is devoted to the literature review consisting of three sections. First, it reviews the theoretical and empirical literature on spatial competition. In the second section, it investigates port-related studies to understand the nature of the shipping industry and port systems. Then, it reviews literature on international trade to identify influential determining factors of trade flows. Chapter 3 states the research questions of this research. Then, in chapter 4 we present the data used in this research and briefly review the techniques used for preparing the disaggregated container flow database. This chapter continues with a presentation of the process of estimation of port efficiency. Chapter 5 proposes a methodological framework for spatial modeling of competition. It is followed by discussions on the modeling results in chapter 6. Chapter 7 concludes.

CHAPTER 2: LITERATURE REVIEW

This chapter reviews the literature that builds the fundamental concepts of this dissertation. It is hard to fit this research into one body of literature as it benefits from concepts rooted in different disciplines. Figure 2.1 summarizes the literature reviewed in this chapter and specifies the position of our research in relation to them. It is worth noting that Figure 1 does not necessarily cover all the disciplines involved in topics of Spatial Competition, Port Systems, and International Trade, but depicts those that have been reviewed in this study.

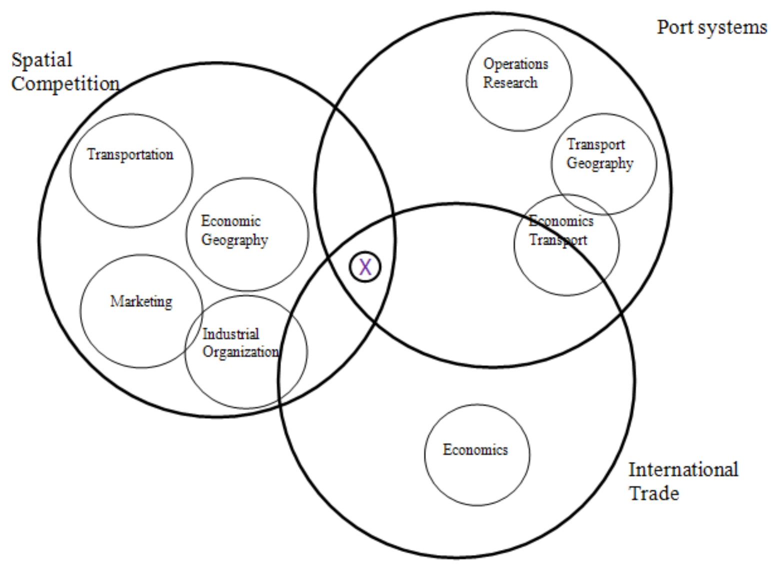


Figure 2.1: Diagram of the literature reviewed in this chapter and position of this research (X) in relation to other disciplines.

This chapter starts with foundations of spatial competition modeling, as the theoretical framework of this research and then reviews the empirical literature on competition modeling in a wide range of industries in order to derive an empirical framework for modeling competition. The next section reviews studies of port systems and particularly port choice papers to build a comprehensive understanding of port industry as an application of this study. The last section reviews papers in international trade to specify the underlying factors of commodity movement. There is a rich established literature on international trade that shares the same context with shipping industry, as the choice of shipping routes is an indispensable step of international trade.

2.1. Literature Review: Spatial Competition

2.1.1. Theory of Spatial Competitions

Spatial competition, as the basic concept of the space-economy, forms the background of this research. Spatial competition models are partial-equilibrium models with oligopolistic competition, which allow researchers to introduce the spatial dimension in the modeling of interactions between economic agents dispersed across space (Combes, Mayer, & Thisse, 2008). Its main purpose is to model location choices made by economic agents maximizing their profits.

Similar to the notion of “intervening opportunities” and “competing destinations” in spatial interaction modeling (Roy & Thill, 2003; Wilson, 1970) and “multilateral resistance” in Anderson-Wincoop’s gravity model (Anderson & Wincoop 2003), the concept of spatial dependence is fundamentally involved in the theoretical foundations of spatial competition. The optimal location of an agent depends on the location chosen by the other agents with which it interacts.

Spatial competition modeling was initiated by Hotelling (1929), who demonstrated that firms agglomerate at the center of a linear, bounded market, with homogenous products and exogenous price (Hoteling, 1929). In his framework, consumers are assumed to buy a single unit of the product and must overcome a linear transportation cost in order to buy the product. His idea became accepted unquestionably among economists as the ‘Principle of Minimum Differentiation’ until 50 years later, when some scientists started to question the model, mainly due to the simplified assumptions made for mathematical convenience (Biscaia & Mota, 2013). His framework was reformulated with more realistic assumptions, such as a quadratic transportation function instead of linear (d’Aspremont, Gabszewicz, & Thisse, 1997), price maker firms instead of exogenous price (d’Aspremont, Gabszewicz, & Thisse, 1997), circular instead of linear market (Ben-Akiva, De Palma, & Thisse, 1989), differentiated instead of homogeneous products (S. Anderson et al., 1988; Irmen & Thisse, 1998), and multi-dimensional product space instead of geographical space (Irmen & Thisse, 1998b; Thill, 1992)

However, until the appearance of the influential paper by de Palma et al. (1985), all of Hoteling’s successors modeled firms’ behavior on the basis of a deterministic consumer utility function and delineated mutually exclusive firms’ markets (De Palma, Ginsburgh, Papageorgiou, & Thisse, 1985). That is, market boundaries between two firms were identified as the locations of consumers who are indifferent between buying from either firm. In recognition of the random taste heterogeneity of consumers, which are not observable a priori, de Palma et al. (1985) introduced a stochastic term to the framework.

de Palma et. al (1985) argued that in a market characterized by differentiated products, due to the inherent attributes of firms, consumers would have different preferences for products, due to their own specific system of valuation. In such a market, in response to the decreasing effect of location as a factor of differentiation, consumers evaluate other characteristics of firms according to their preferences and distribute their demand between firms. In fact, as soon as transportation costs relative to the degree of market differentiation decrease, consumers stop patronizing a single firm, and the hinterlands of firms start to overlap (Combes et al., 2008). Since firms do not have enough information on consumer preferences to formulate their utility with certainty, a random utility variable is assumed, which reflects the lack of information on the part of firms (analyst).

Interestingly, Anderson et al. (1988, 1989, and 1992) demonstrated the linkage between the random utility framework and product differentiation theory (S. Anderson et al., 1988; S. Anderson, Palma, & Thisse, 1989; S. Anderson & Palma, 1992; Ottaviano & Thisse, 2011). The "representative consumer" model (Dixit & Stiglitz, 1977), as a product differentiation model, assumes that taste heterogeneity of consumers can be described by the choices made by a single individual with variety-seeking behavior (Anderson & Palma, 1992). The variety-seeking behavior justifies individual's tendency to avoid the boredom generated by the repeated consumption of the same variety (Adamowicz & Swait, 2013; Combes et al., 2008). In other words, the behavior of consumers having preference for variety, faced with differentiated products, can be described by a purchasing probability (or frequency). In the same way, random utility theory assumes that individual taste heterogeneity is not completely observable and describes individual choice behavior using a probability function.

2.1.2. Empirical Modeling of Spatial Competition

Competition intensity is measured differently depending on its definition and on characteristics of the industries under consideration. Despite the centrality of the concept of competition in economic analysis, it is often measured inadequately using aggregate market concentration indices such as market share, Herfindahl index, price cost margins and so on (Boone, Griffih, & Harrison, 2004; Boone, 2008). These measures could give an incorrect view of competition, as they do not take the consumers' choice, their locations and their preferences for differentiated products into account (Hastings et al. 2005), mainly because data on disaggregated choice may not be readily available. Unobservable or difficult-to-measure characteristics of products and consumers make the analysis of competition even more complicated (Goettler & Ronald, 2001).

This research particularly reviews literature that models competition based on disaggregated choices from different disciplines, such as industrial organization, marketing and transportation research. While some of these studies deal with spatial choice (spatially differentiated alternatives) such as schools, airports, seaports and shopping centers, the geographical space is not involved in other studies that model competition between different brands of product, TV shows, political parties, and mode of transportation. However, concepts of spatial competition are used in a large number of research without spatial alternatives. These studies conceptualize consumers' preferences for alternative with distance between consumers and (non-spatial) alternatives and employ spatial competition theories.

The early empirical research that specifically builds on the concepts of spatial competition, Thill and Rushton (1992) identify the intensity of market contestability based on the relative location of consumers and market places. At any consumer location, their "space-price" competitive model calculates the differences in delivered price for the two best market places as a degree of rivalry (Thill & Rushton, 1992).

Hastings et al. (2005) deal with spatial choices and integrate concepts of spatial competition to the discrete choice analysis in a paper on school competition. They use parental ranking of the top three choices of school in order to identify the preference parameters leading to more severe competition between schools. In the context of the airline industry, Hess and Polak (2005) analyze airport competition and use a mixed logit model. They found significant heterogeneity in preference for access time (Hess & Polak, 2005). Estimating both conditional and mixed logit models, Ishii et al. (2009) found that non-price characteristics like access time, delays, and flight frequency can strongly affect airports' market shares (Ishii, Jun, & Van Dender, 2009). Although a considerable number of research studies competition in the airline industry, very few integrate the concept of market segmentation into competition modeling (Wen & Lai, 2010). Also, while emphasizing taste heterogeneity, they do not explore individuals' choice histories.

The literature discussed above has recognized the fact that the degree of competition between alternatives will depend on degree of variation in consumers' preferences for characteristics of the alternatives. They estimated advanced discrete choice models to accommodate taste heterogeneity. However, the empirical research on port competition does not acknowledge heterogeneity in preferences despite the wide range of port users including shippers, carriers, and forwarders that are characterized by different market size. They often rely on logit and nested logit models and except for very few papers that provide measures of competition, with a single aggregate market share

(Malchow and Kanafani 2004; Veldman & Bückmann, 2003), this body of literature mainly focuses on understanding determining factors and on formulating port selection behaviors (Slack 1985; Murphy and Daley 1994; Tiwari et al 2003; Blongin and Wilson 2006; Tongzon 2007; Veldman et al. 2011; Nir, Lin, & Liang, 2003). We will review this body of literature more thoroughly in the next section.

In the context of non-spatial choice, the established literature in marketing research has focused on analyzing market segmentation and competitive market structures, by studying the pattern of brand switching. Grover and Srinivasan (1987) argued that the same brand choice probability can provide a basis for market segmentation as well as competitive structures of the product market. They use the joint probability of choosing two different brands by the same consumer to infer the extent of market competiveness. They analyzed patterns of brand switching using observed choice probabilities, but without modeling the underlying determinants of switching behavior (Grover & Srinivasan, 1987). Kamakura and Russell (1989) proposed a latent class logit model to characterize market structure, emphasizing the destabilizing role of marketing variables (e.g., price) as underlying factor of switching behavior (Kamakura & Russell, 1989).

Goettler and Shachar (2001) presented an empirical study of spatial competition estimating a discrete choice model in the network television industry and generated an attribute space for products (television shows) with unobservable characteristics. Distance in this space captures the extent to which two shows compete for the same viewers. Similarly, the political science literature uses spatial competition concepts under the assumptions that citizens will vote for the candidate whose policy position is closest to their own views (Jessee, 2009). Poole and Rosenthal used choice histories to estimate the parameters of the utility function and the location of the choice options and the decision makers (Poole & Rosenthal, 1985). Central to the discussion of product competition is a measure of similarity of choices between two products that is revealed by the number of consumers who purchase both products (mutual or joint consumers). This means that two products consumed by many of the same individuals are in more severe competition.

There are numerous transportation related studies on intramodal and intermodal competition that employ discrete choice models to understand how a change in one of the alternatives will affect market shares (Behrens & Pels, 2012; Hess & Polak, 2005; Ishii et al., 2009; Ortúzar & Simonetti, 2008; Wen & Lai, 2010). The multinomial logit model, which has been traditionally used in transportation research, is substituted by more flexible and advanced discrete choice models. Behrens and Pels (2009) analyze inter- and intra-modal competition between high speed rail and air transport using nested and mixed logit models. Wen and Lai (2010) employed a latent class model and stated preferences to reveal passengers' preferences for international airlines and to segment the market.

2.2. Port Systems

Regarding the importance of ports as gateway to global market and the reliance of the national economies on ports efficiency and competitive power, the topic has been the focus of research in a wide range of disciplines such as geography and regional planning, economics, operational research, management, and environmental studies, and so on (Woo, Pettit, Kwak, & Beresford, 2011).

Port competition constitutes a main research theme of the literature on transportation economics (Meersman et al., 2010; Slack, 1985). The research that has focused on port competition has been categorized by Chang and Lee (2007) based on five aspects: 1) competitive policy, 2) cooperation, merging and alliances, 3) governance, ownership and privatization, 4) performance, productivity and efficiency measurement, and 5) port selection factors (Chang & Lee, 2007).

We survey the literature belonging to all five categories. Theoretical discussions and conceptual frameworks proposed in the first three categories aid us to gain comprehensive insights into complex logistics networks. We believe that understanding competition requires sound micro-level models of underlying port choice decisions made by port users as manifested by freight flows. In their conceptual paper, Notteboom and Rodrigue (2005) argued that "inland distribution becomes of foremost importance in port competition". Hence, methodologically, our approach to competition is closer to the last strand of literature. Also, we examine methods and factors discussed in the fourth group to identify key competitive factors.

2.2.1. Evolution of Port Competition

Containerization has been considered as one of the most important technological revolutions in the transport sector of the last decades (Clark, Dollar, & Micco, 2004; Hummels, 2001). Using standardized containers, cargos travel long haul and transfer from one transport mode to the other without being unpacked or repacked (Hummels, 2007). Also, deployment of larger container vessels requires fewer ports of call along shipping lines' service route (Notteboom 2009; Malchow and Kanafani 2001). As a result, containerization or unitization of general cargo (Slack, 1985) has led to large cost reduction, cargo transshipment facilitation, specialization power decay, world trade growth, and severe inter-port competition (Clark, Dollar, & Micco, 2004; Cullinane & Song, 2006; Luo & Grigalunas, 2003; Notteboom, 2009; Zhang, 2009).

Traditionally, competition between ports was related to the exclusive dominance of ports over a region (Slack, 1985). Due to the immovable geographical location of the ports, concentration of productions, and monopoly of ports, inter-port competition was not a major concern (Chang & Lee, 2007; Cullinane & Song, 2006). As a result of containerization and of the emergence of intermodal rail and barge corridors, exclusive hinterlands are penetrated by competing ports and inter-port competition has become a central topic of transport economics (Meersman et al., 2010; Slack, 1985). Also, the highly fragmented maritime transport chain has been reformed to a fully integrated global supply chain through the increasing incorporation of shipping lines, terminal operator companies, port authorities, and transport service providers to improve the quality of services and reduce costs. Therefore, yesterday's competition between individual ports has been transformed into competition between supply chains, particularly in the Western world (Magala & Sammons, 2008; Robinson, 2002).

Interestingly, besides extensive discussions on port competition, there is literature that remarks on the interest of port authorities in cooperation with other ports. Ports in the same region cannot offer any meaningful distance advantage over others; however, they can take advantage of cooperative agreements among ports of a region to increase their competitive position vis-à-vis other port regions (De Langen, 2007; Heaver, Meersman, & Van De Voorde, 2001).

2.2.2. Methodological Review

The inception of containerization and ensuing fierce port competition has encouraged scholars to better understand the underlying factors of competition. Ports competition as an important subject in transportation economics has been investigated through different methodologies that basically can be divided into quantitative studies and conceptual studies. Earlier we referred to some of the available descriptive literature, which expresses the complex nature of logistic systems and inter-port competition, without being concerned about modeling aspects and practical solutions. Conversely, quantitative contributions simplify the situation with straightforward assumptions. Early quantitative research mainly bases the analysis of data driven from survey and interviews. Slack, Murphy and Daley, Brooks, Song and Yeo, Tongzon and Sawant, De Langen are among those who focus on interviews and stated preferences of shippers, carriers, forwarders and experts (Slack 1985; Murphy and Daley 1994; Brooks 1990; M. R. Brooks 1995; Song and Yeo 2004; J. L. Tongzon and Sawant 2007; De Langen 2007).

Slack (1985) shifts the focus of research on cost-based inter-port competition towards a service-oriented view, since he found evidence that non-cost factors such as reliability, speed, quality of service, and source of information are important considerations (Slack, 1985). The statistical results of his interviews with Northern American/ European shippers and forwarders suggest that price and service considerations of land and ocean carrier are more influential than port attributes such as port infrastructure and other port-related economies. He notes the relative importance of cost and service consideration in relation to the size of the agency. Price is more important for the smaller companies, while larger companies seek to avoid congestion and associated delays along container flows.

Murphy and Daley find shipment information, and loss and damage performance are the most important factors (Murphy & Daley, 1994).

In an extension of this approach Song and Yeo (2004) investigate the competitiveness of Chinese ports based on a series of surveys on a group of experts using the Analytic Hierarchy Process (AHP). The interviewees were asked to identify competing criteria and also to indicate the relative importance of each factor (Song & Yeo, 2004). The most important identified factors are cargo volume, port facility, port location, and service levels that are quantified simply based on an attribute of each port.

De Langen surveyed port's selection criteria and the competitiveness of ports from the perspective of shipper's and forwarders in contestable hinterlands of Austria (De Langen, 2007). The sample collected from 23 Austrian shippers and forwarders shows that shippers have a less price-sensitive demand, and are more concerned with reliable and damage-free handling. Although the sample size is small for any certain conclusion, he explains the price-elastic behavior of forwarders by the fact that one of their capabilities is to purchase transport services cheaply for large volumes of cargo, while transport costs are only a fraction of overall shippers' costs that it even may pass to their customers. These findings are consistent with Tongzon' results that show shippers' concern with indirect cost, such as unreliability, damage and adverse reputation effects (J. L. Tongzon, 1995).

However, the finding from interviews and surveys markedly differs from results based on revealed preferences and policies have to be derived from players' actions rather than their statements (Tongzon & Sawant, 2007). Brooks importantly remarks on the differences between important and determinant factors, since a factor might be perceived important by interviewees, but not differ across alternatives (Brooks 1990; Brooks 1995).

The next sub-section reviews some of the recognized quantitative research on port competition.

2.2.3. Port Choice Framework

Traditionally, discrete choice modeling has been proposed to estimate demand for freight and passenger in transport economics. It has been initiated in port selection studies with Malchow and Kanafani about a decade ago by employing a logit framework (Malchow & Kanafani 2001).

Veldman and Buckmann (2003) implement a conditional logit framework to model choice among West European container ports (Veldman & Bückmann, 2003). Their model is intended to explain port's routing choice that maximizes the shipper's utility. Port's routing is a combination of the carrier, the port and the mode of transportation, while explanatory variables include transportation cost, transit time, frequency of service, and indicators of quality of service. In fact, they assume that a shipper or receiver has the choice of four container ports in the Antwerp-Hamburg range, 25 shipping lines, and three modes of inland transport (road, rail or waterway). Tiwari et al. proposes a logit model to estimate the underlying factors that affect the shipper's selection of a port and a carrier in China (Tiwari, Itoh, & Doi, 2003). Their logit model's choice set is composed of the combination of three shipping lines and five ports.

Malchow and Kanafani (2001) stand out by modeling containerized commodity distribution from a carrier perspective. They claim that the selection of ports has shifted from the shipper to the carrier, and utilize a logit framework to study port choice for export from the U.S. to eight foreign countries (Malchow & Kanafani 2001; Malchow & Kanafani 2004). The utility of a port for a shipment is defined as a function of five factors including oceanic distance, land distance, frequency of ship sailings, vessel capacity, and probability of being the last visited port. Surprisingly, they consider the cross-functional integration between carriers and ports by pointing to the fact that the carrier's intermodal transfer process at each port could vary with the carrier. They estimate port-specific constants varying by carrier to take care of unobserved factors that could influence each carrier's selection of a port.

The logit model has been utilized in the majority of the research on port choice that has been reviewed. Under the multinomial logit formulation, the disturbance components are assumed to be independently distributed. This structural restriction, Independence of Irrelevant Alternatives (IIA), generally does not hold true in port competition environments in which unobservable characteristics associated with neighboring ports are perceived to be correlated. While most scholars who have applied this model acknowledge the restrictions imposed by this assumption, no real action is taken to circumvent its unfortunate consequences (Malchow & Kanafani 2001; Tiwari, Itoh, and Doi 2003; Veldman and Bückmann 2003).

The only empirical implementation for relaxing this restrictions is the nested logit model on Spanish container trade by Veldman (2011) who test a two-level choice function, where the higher level sets apart Mediterranean and Atlantic coast ports, while the lower level handles specific ports within each of the two high-level clusters (Veldman, Garcia-alonso, & Vallejo-Pinto, 2011). However, structure of spatial alternatives hardly fits into a hierarchical setting. That is, splitting spatial entities that have been spread over a geographical area into clear-cut hierarchical divisions is not always conceptually meaningful. The structural restriction of the conditional logit encourages Blonigen and

Wilson to employ a gravity model instead (Blonigen & Wilson 2006). Also, Ferrari et al. used a gravity model to measure container traffic diversion from Ligurian ports to the main Italian and European competitors (Ferrari, Parola, & Gattorna, 2011).

Luo and Grigalunas (2003) uniquely estimate port demand by simulating the container transportation process through a multi-modal transportation network. The core of their simulation model is a Dijkstra algorithm that selects the least-cost route composed of ports, rail, highway, and international shipping lines (Luo & Grigalunas, 2003). Port demand is estimated by the frequency that the focus port is located along selected least-cost routes. It is seen that the estimated demand often differs from the actual port throughput. It can be argued that operations research is a suitable solution to the port choice problem under the assumption of a deterministic choice behavior, where a unique port minimizes the total cost. But, such an assumption contradicts the observed behavior of shippers that ship via more than one port and shapes the phenomenon of inter-port competition. However, this research interestingly discusses the role of competing ports and shows through sensitivity analysis that the service area of ports and cross-price demand curve may be affected by policy and charges at other ports.

2.3. Literature Review: International Trade

In addition to studies reviewed in section 2.2, commodity distribution has been extensively studied in the literature of international trade. This body of research often relies on gravity models and studies trade patterns between nations or regions, while some studies have focused attention on ports as intermediate nodes on international trade routes. Reviewing studies of international trade helps to extend our insight into this considerable body of literature that shares context with port systems. Therefore, this chapter continues with underlying factors of trade discussed in these studies along with factors found to be influential on ports competitiveness.

Baher and Venables summarize the determinants of international trade through the formulation below. In this formulation, transportation costs shape trade and is shaped by a variety of other factors such as distance, geography, infrastructure quality, transport technology, fuel cost and so on (Behar & Venables, 2011).

$$\textit{Trade} = \mathrm{F}\{\textit{income, policy, cultural affinity, transport costs} = \mathrm{f}(\textit{distance, geography, infrastructure, trade facilitation, technology, fuel costs})\}$$

With the reduction of countries' tariff and nontariff barriers to trade, due to development strategies of global economy (Clark et al., 2004), transport costs are the remaining trade barriers that require attention. Different variables in addition to distance are incorporated into the transportation cost function, including dummies to control for countries with common border, common language, common currency, common colonizer, the existence of a free trade area, and the condition of being landlocked countries to study trade between nations (Anderson & Wincoop 2003; Brun et al. 2005; Hummels 1999). Brun et al. add several other factors, such as an index of infrastructure quality and oil cost, into the standard specification of transportation costs (Brun et al., 2005). Clark et al. put emphasis on the effect of insurance on transportation cost: the more expensive the product is, the higher the associated insurance and therefore the transportation cost is (Clark et al., 2004).

2.3.1. Distance and Geography

Quite naturally, geography, and particularly distance, is one of the most studied components of transportation costs (Clark et al., 2004). Distance is a very important factor and, in most studies, distance serves as an approximation for transportation costs. Notteboom's notion that immediate regions compose the backbone of port's cargo basis makes direct reference to the distance effect (Notteboom 2009). According to this author, around 40% of Antwerp's customers are located in a radius of 50km of the port. Besides distance, other geographic characteristics such as having a common border, being landlocked, and country area (Behar & Venables, 2011) have been found influential.

Research on international trade studies the elasticity of costs with respect to distance and the distance elasticity in response to different modes as well as over time. Distance is decomposed into oceanic and inland distances by Malchow and Kanafani due to their different cost and reliability (Malchow & Kanafani, 2004). These authors found that the marginal rate of substitution between inland and oceanic transit correlates positively with the commodity value. Hummles quantifies the effects of distance on transportation costs for different modes, and found values of 0.46, 0.39, 0.275, and 0.22, respectively for air, rail, road and sea transportation (Hummels, 2001).

The significant negative effect of distance on trade interactions is obvious, but it is observed to strictly increase rather than decrease over time. This contradicts the perception of "death of distance" in the current wave of globalization (Brun et al., 2005), and this distance puzzle has attracted attention in gravity-based international trade studies. Various researchers, including Brun et al. and Cao and et al., have obtained paradoxical results and evidence of distance decay over time (Brun et al., 2005; Cao, Mamoulis, & Cheung, 2005).

On the other hand, studies reviewed by Leamer and Levinsohn, Burn, and Disdier and Head failed to observe a decreasing distance coefficient over time (Brun et al., 2005; Disdier & Head, 2008; Leamer & Levinsohn, 1995). Anderson-Wincoop's new non-linear gravity model implies that bilateral trade is homogenous of degree zero in trade costs (Anderson & Wincoop 2003). Then, they express that "The invariance of trade to uniform decreases in trade costs may offer a clue as to why the usual gravity model estimation has not found trade becoming less sensitive to distance over time". Disdier and Head conclude that their findings based on the systematic analysis of 1467 distance coefficient estimates in 103 papers "represent a challenge for those who believe that technological change has revolutionized the world economy causing the impact of spatial separation to decline or disappear" (Disdier & Head, 2008).

However, it is worth mentioning that distance captures all variations caused by accessibility, transportation infrastructure, transportation time, costs and so on. In fact, distance is a very important factor and serves as a proxy for transportation costs in studies of trade pattern (Clark et al., 2004). However, Glaeser and Kohlhase (2004) report that "80% of all shipments (again by value) occur in industries where transport costs are less than 4% of total value" (Glaeser & Kohlhase, 2004). Grossman (1998) claims that estimated distance effects are about an order of magnitude too large to be explained by shipping costs. He speculates that the reason why distance matters so much is the lack of familiarity or cultural differences (Grossman, 1998).

Furthermore, distance may reflect shipping time as Hummels estimates a one percent reduction of export probability in response to each additional day that a product is delayed prior to being shipped (Hummels, 2001). Therefore physical distance may mirror transportation infrastructure such as road, rail, and barge services that facilitate overcoming the physical distance in relation to time.

2.3.2. Border

One of the spatial dimensions that has caught the attention of economists is national borders (Combes et al., 2008). With the reduction of countries' tariff and nontariff barriers to trade, due to development strategies of global economy (Clark et al., 2004), the remaining significant border effect is puzzling. A large body of empirical literature investigates the sizeable negative impact of border on trade. The border puzzle was introduced by McCallum (1995), who found that the U.S.-Canada border caused the Canadian provinces to trade 22 times more with each other than with U.S. states (Mccallum, 1995). Further research confirmed the substantial home bias in trade between integrated and culturally homogenous countries, where a free trade agreement is supposed to ease the movement of trade. For instance, Helliwell (1996, 1998) confirms McCallum's original finding, by analyzing trade between the US and Canada using data on the post-NAFTA period, but shows that trade fell by a factor of twelve (Helliwell, 1996, 1998). European Union national borders have also been studied extensively (Chen, 2004; Nitsch, 2000; Turrini & van Ypersele, 2010). Chen's results imply that a EU country trades about six times more with itself than with a foreign EU country. The border effect in Nitsch's paper indicates that a border reduces trade by a factor of 6.8 between European countries. Also analysis of trade among OECD countries, by Turrini & Ypersele (2010), shows that the absence of borders raises trade by a factor of 9.59.

The frictional effect of a border has largely been explored using international trade data, as well as sub-national trade, there is not yet a consensus view on its underlying causes

(Turrini & Van Ypersele, 2010). The significance of trade barriers in analysis implemented at the regional level does not support many of the suggested hypotheses. Turrini and Ypersele (2010) summarized the existing papers investigating the border effect. Different factors such as exchange rate volatility, non-tariff barriers and regulation differences across countries, informational barriers and the role of commercial networks, weak institutions and widespread corruption are suggested as determinants of border effects.

Turrini and Ypersele suggested that the border matters because of the differences in legal systems across the OECD countries. After controlling for country-specific factors, distance, the presence of a common border and a common language, variables capturing the impact of heterogeneous judicial systems have a significant impact on international trade. They found that trade between two countries with “identical legal procedures to refund an unpaid check” is 70% higher than with a “fully differentiated procedure”.

Some studies find relationships between a border effect and the degree of substitution between domestic and imported goods. For instance, Chen’s (2004) results in the EU indicate that bulk commodities like concrete, stone, concrete products or mortars suffer the greatest border effect. Moreover, a border effect can arise endogenously and as a result of spatial clustering of firms. In order to reduce transportation costs, intermediate and final goods procedures tend to agglomerate (Chen, 2004; Hillberry & Hummels, 2005; Wolf, 1997).

2.3.3. Infrastructures

In addition to physical distance and borders, infrastructure frequently comes into consideration, which includes manmade constructions such as road, rail, and port services that facilitate overcoming the physical distance in relation to time. Some studies have applied indices of road, rail, ports infrastructure and connectivity across countries to account for their impact on transportation costs and trade (Brun et al., 2005). Measuring transit time is a good indicator of service related impacts since the fact that shipments wait long time in ports for the arrival of the next ship or standing still at borders has a negative effect on trade interactions. Hummels estimates a one percent reduction of export probability in response to each additional day that a product is delayed prior to being shipped (Hummels, 2001). This amount of reduction is assessed to be equivalent to distancing by about 70 km on average from trading partners. With the same rational, Veldman et al., Lou and Grigalunas include transit time besides transit costs to the port demand analysis (Veldman & Bückmann 2003; Luo & Grigalunas 2003).

Export documentation and customs handling can slow down and negatively affect cargo movement. The large cross-country variations in transportation time are exemplified by Djankov et al.: they show that moving an export container from the factory in Baghdad to the nearest port and loading the cargo onto a ship takes 105 days. It takes 93 days from Almaty (Kazakhstan), while it only takes 5 days from Copenhagen, 6 days from Berlin, and 20 days from Shanghai (Djankov, Freund, & Pham, 2010).

Recently, rail and barge services have brought fundamental changes to hinterland accessibility and inter-port competition, which used to largely rely on trucks and road haulage (Notteboom 2009). In the next subsection, we review the role of multimodality on cargo flow. Also, the infrastructure of ports draws significant attention in economic and transportation research as an aspect of port competitiveness. This section ends by discussion on port efficiency.

2.3.4. Modality

Thill and Lim refer to intermodalism as a “powerful force of change in the spatial economic organization of places and spaces” (Jean-Claude Thill & Lim, 2010). Modal choice has gained attention in recent port selection modeling work as an influential component of supply chain competition (Magala & Sammons, 2008; Veldman & Bückmann, 2003). Notteboom remarks on two sides of intermodal development’s impacts. On the one hand, it helps port hinterland expansion and port competitiveness (Notteboom 2009).

Zhang investigates port competition in relation to the development of alternate intermodal transportation (Zhang, 2009). He finds that port output responds differently to the development of rail and waterway corridor capacity versus development of road capacity. While the former will improve the port’s output, and reduce the rival port’s output, the latter increases delay and also induces greater commuter traffic, i.e. traffic other than the seaport cargo, and in total may lead to a decrease of port’s output. For instance, congested roads in Los Angeles area can inhibit business at the port of Los Angeles and Long Beach and substantially impact supply chains throughout the country (Blonigen & Wilson, 2007). However, other research has conversely argued that the growing congestion in the hinterland of large ports have not necessarily shifted traffic towards smaller and less congested ports located in proximity of large ports (Joint Transport Research Centre, 2009). This is supported by the fact that the share of traffic handled by Northern European ports has remained stable between 1975 and 2007 and their growing share of the Mediterranean market. Also, the share of Los Angeles/Long

Beach in the west coast container traffic has remained at about 70% over the last two decades in spite of substantial congestion problems.

On the other hand, intermodality increases regional accessibility to more than one port. This dimension is emphasized in Thill and Lim's paper (Jean-Claude Thill & Lim, 2010). They measure regional accessibility across the United States at the zip code level with respect to both intermodal and highway networks, and then the accessibility gains are evaluated by comparing two accessibility measures. They illustrate how the spatial pattern of accessibility gains varies across the country and how remote regions can benefit from less expensive access to new foreign markets.

2.3.5. Port Performance and Efficiency

Port efficiency is recognized as one of the factors of port competition. Clark et al. find that an improvement in port efficiency from the 25th to the 75th percentile reduces shipping costs by more than 12% or the equivalent of 5000 miles in distance. For instance, they estimate that transportation costs in countries like China, Indonesia or Mexico would drop by 10% if they improve their port efficiency to the levels of France or Sweden (Clark et al., 2004).

Despite the availability of a huge body of literature dealing with port efficiency, there is not a common agreement on the significance of their effects and even a standardized measurement. While Tongzon asks shipping lines to rank port efficiency as one aspect of port's attributes that they consider in port choice(Tongzon & Sawant, 2007), Ching Tang et al. define a combined measure of efficiency from a large number of attributes such as port charges, annual port calls, water depth, trade volume, and availability of inter-modal facilities (Tang, Low, & Lam, 2008). Clark et al. relate port efficiency to activities like customs requirements, policy, and management, besides the physical infrastructure of ports (Clark et al., 2004).

Based on the 1999 Global Competitiveness Report (GCR) of the World Economic Forum, Clark notes that North America and Europe are ranked as having the most efficient ports, followed by the Middle East, and East Asia and the Pacific. The lowest ranking is associated with ports of regions like Latin America, South Asia, and West Africa (Clark et al., 2004). Besides the measure of port efficiency reported by GCR, some economists have designed a consistent measure of port efficiency.

For instance, Blonigen and Wilson measure port efficiency using U.S. census data on import charges. Their analysis is based on the premise that higher costs of getting the cargo to the docks and unloaded is a result of port inefficiency, which contributes to a rise in import charges. Tang et al. apply Factor Analysis to identify key dimensions of port attributes and ends up with three factors that are interpreted as port efficiency, scale economies, and the convenience in using the port, respectively. While trade volume, turnaround time, and port charges heavily load on the port efficiency component, port traffic, number of port calls and drought variables load on scale economies, and annual operating hours of port and the availability of intermodal transport facilities load on the third factor (Tang et al., 2008). In recent years, approaches such as stochastic frontier analysis (SFA) and data envelopment analysis (DEA) have been widely used to measure the degree of relative efficiency in port industry. However, we cannot utilize their estimates of port efficiency because their samples are not comprehensive and do not include all the ports we used in this study. We used SFA to estimate technical efficiency for the ports involved in this research. Chapter 4 will discuss it in more detail.

2.3.6. Data type, Sampling, and Geographical Setting

The literature that attempts to model commodity flows uses datasets with different levels of aggregation and completeness. Among the papers reviewed here, the most disaggregated data is used by Malchow and Kanafani on individual shipments, who takes a sample of 4843 shipments from Port Import Export Reporting Service (PIERS)(Malchow & Kanafani 2001). Also, Tiwari uses a sample of 1033 shipments whose shippers were asked about the choice of carrier and port (Tiwari et al., 2003). These two datasets include information like origin, destination, and volume of shipment in TEUs, port and carrier and are most comparable with the dataset used in this study. For simulation of container trade between U.S. and foreign countries, Luo and Grigalunas indicate that the best data source is PIERS, but it is very expensive and not available to their research (Luo & Grigalunas, 2003).

The data used in other port choice studies often come from more aggregated sources. Regions constitute the unit of analysis in the study conducted by Veldman et al. on West European Container ports. The data was compiled from national transportation and port statistics and covers 33 regions in the Netherlands, Germany and Belgium (Veldman & Bückmann 2003). In some cases, the data exhibit very limited geographic disaggregation. For instance, Blonigen and Wilson use aggregated data of annual import volumes at various U.S. ports, with no information on hinterlands and forelands. The volume of trade between states and ports is estimated (but not observed) by the gross state product weighted by the distance between the port and the state (Blonigen & Wilson 2006).

CHAPTER 3: RESEARCH QUESTIONS

The extant literature not only underscores the need for adopting a new framework to model spatial competition in a market with significant heterogenity in preferences but also points to several unresolved issues with existing approaches, especially dearth of data, simplification of reality, and lack of attention to the spatial organization of markets.

Although there is a large body of literature on spatial competition, research is mainly based on theoretical discussions, or empirical studies with a limited access to the spatial distribution of dissagregated demand as crucial competitive force. The existing empirical literature seems to ignore the heterogenous valuation systems of decision makers and to underestimate the complex nature of choice behavior by observing single choices of individuals.

A major contribution of this research is therefore to address these issues and to propose an approach for quantifying competitive interactions between spatially differentiated firms at the micro level. The proposed research methodology contributes to the methodological body of literature in economic geography due to the uniqueness of data, pervasive exploratory data analysis, and novelty of the methodology.

The central goal of the analysis is to model interactions of competitive nature between firms that can be imputed from heterogeneous preferences for firms' attributes. We first model the individual choice set based on the spatial structure of flows from customers' locations to suppliers (ports in our case study). Subsequently, we simulate demand elasticities and competitive interactions in response to firms' strategies.

This framework is not domain-specific and is seen to be potentially applicable to a variety of domains of application dealing with significant alternative substitutability, particularly with spatial choices. Port competition is the focus of this research as a case study. As a result of containerization and of the emergence of intermodal rail and barge corridors, exclusive hinterlands are now invaded by competing ports and inter-port competition has become a core topic of transport economics (Meersman et al., 2010; Slack, 1985). There is a large body of literature that studies port competition from the perspective of port selection. However, existing studies overly generalize individuals' decisions in traditional logit model and provide basic measures of competition with a single aggregate market share (Malchow and Kanafani 2004, Veldman et al. 2003).

3.1. Probabilistic Modeling of Spatial Competition

The exploratory analysis of origin-port-destination flows reveals large variations that can be ascribed to the diversity of shipping decisions of consumers, even after constraining for the origin and destination of the shipments. That is, the persistence of variations, after controlling for the geographic origin and the geographic destination, indicates that port choice factors hitherto identified as the most influential in shaping long-haul international freight distribution systems, including port characteristics such as port efficiency, infrastructure, congestion connectivity and invariant origin-destination variables such as distance, transportation infrastructure, and accessibilities, do not suffice to explain choice behavior. The switching pattern between ports by shippers constitutes the choice set. This analytical construct originally coined in choice theory forms the basis for identifying the distribution of preferences in population, market segmentation and measuring competition.

This research proposes an empirical framework for modeling spatial choices in a market characterized by significant heterogeneity in preferences. In this framework, observable explanatory factors can estimate individuals' feasible choice sets, whose elements effectively compete over the same market. To account for taste heterogeneity, we estimate separate models for consumers that have different system of valuation by analyzing their choice histories. This research employs a Latent Class Logit framework that allows market segmentation and preference estimation simultaneously. With this approach, we endogenously cluster consumers into homogenous segments with similar preferences.

With choice probability estimates from the proposed framework, we calculate the elasticity of demand for each port in response to firms' strategies and decision variables, such as transportation costs and port efficiency policies. This dissertation answers the following specific research questions:

1.1. How can the decision maker's choice set be defined?

1.2. What are underlying determinants of choice behavior?

1.3. How do preferences for underlying factors vary across segments of populations?

1.4. How can spatial competition be measured between firms?

1.5. How does demand respond to changes to underlying factors of choice? How does market share changes in response to these changes?

CHAPTER 4: DATA PREPARATION

In the previous chapters, a huge emphasis was placed on the unique nature of the dataset that is applied in this research. What makes such detailed data absent from existing research is related to its price and the difficult and time-consuming process of cleaning and verification. The preparation of an accurate, reliable and rich spatial database from a large dataset of localities with a variety of addressing system is a time-consuming process that requires the involvement of different database capabilities, business and human intelligence. In this chapter we discuss the data sources and the steps undertaken to prepare this data for our analysis.

Also, our analysis needs a complementary source of information containing ports efficiency. As discussed earlier, we could not find a universal measure of efficiency that covers all the ports we used in this study. Therefore, we estimate port efficiency utilizing SFA method. This chapter is followed by presenting this method and our results.

4.1. Data Source

Freight shipment flows analyzed in this study are part of a dataset of door-to-door U.S.-bound waterborne containerized export shipments with European origins. The dataset includes all waterborne containerized U.S. imports during the month of October 2006. The data is comprised of information submitted through both the U.S. Customs and Border Protection Automated Manifest System (AMS) and manifests submitted at the ports. These original documents have been corrected, cross-referenced, and improved by data supplier

Port Import Export Reporting Service (PIERS) and distributed to users as the PIERS Trade Intelligence data product (PIERS, 2007). This data source provides appropriate attributes of the shipment of each bill of lading, including shipper company name and address, commodity description and type based on the Harmonized Commodity Description and Coding System (HS), quantity (twenty-foot-equivalent-unit [TEU], estimated value, weight), carrier, forwarding port, pre-carrier location, U.S. port of entry, and U.S. consignee and its address (if the shipment is not in transit to a third country).

Development of geographic information retrieval, machine-learning algorithms and manual methods for verification of data items such as shipment origins (by country and locality), coordinates of shipment origin, and categorization of shipment movement typology, as well as other improvements in support of these processes, has been undertaken in house. Data preparation has implemented full text search queries and fuzzy matching algorithms to extract and geocode shipper origin localities from shipper company names and addresses. The unique data cleaning process employed for this dataset includes corrections based upon verification of physical origins (production locations) of shipments.

This research applies some functions and techniques available in SQL Server 2008 to determine potential geocodable places in textual addresses and detect groups of items that are similar on the basis of fuzzy-sets considerations.

4.1.1. Geographic Information Retrieval

The large volume of unstructured data with some locational content that is available in a modern society can be turned into useful geospatial information, but only if locational content can be extracted. Examples of such data sources are unstructured, non-geocoded dataset of localities, often collected for commercial and government applications all over the world, with a variety of addressing systems. Furthermore, real data are dirty due to inconsistencies, misspellings, truncations, omission, null fields, unexpected abbreviations, and so on. Finding an appropriate geocodable locality from textual address data in huge datasets requires considerable resources for cleaning, standardizing, and extraction of useful information from existing data.

To extract spatial information from databases, such as the U.S. waterborne imports data used in this research, we borrow techniques from research that concerns the retrieval of information from documents with geographic-related data. Geographic information retrieval (GIR) has been gaining attention with the increasing growth of immense but implicit geographical information on the World Wide Web (Larson 1996; McCurley 2001; Borges et al. 2007; Jones and Purves 2008). GIR concerns the retrieval of information from documents with geographic-related data. It involves the extraction, semantic analysis and indexing of spatial locations from unstructured or semi-structured textual data by integrating aspects of databases, human-computer interaction, GIS, and IR (Borges et al., 2007; Larson, 1996).

The main purpose of the literature on GIR is to use the geographic context and location semantic of web pages in response to user queries (Markowetz, Brinkhoff, and Seeger 2008; Gan et al. 2008), or enhancing spatial databases by converting extracted geospatial data from web pages for later use e.g. in urban GIS (Borges et al., 2007). Several stages are recognized in this process and they are now discussed in relation to our own research.

1- Identification of geographic references in the form of place names from textual descriptions with little or no structuring
2- Disambiguation of place names
3- Determination of correct location
4- Development of methods to validate the performance of GIR

4.1.2. Identification of Place Names

In order to examine commodity distribution at a high spatial resolution, identifying physical origin (production location, for instance) of shipments is necessary. To unambiguously reference geographic information, access to external data sources of place names and hierarchy between them is necessary (Ahlers & Boll, 2008). A common approach consists in using Gazetteers, which contain extensive information about geographic entities and their relations (Hill 2000; Stokes et al. 2007). We utilize two reference shape files with over 3,000,000 place names across the world namely, USGS's Geographic Names Information System (GNIS) and Europa Technologies' Global Insight data product.

McCurley differentiates between the processes of geoparsing and geocoding, where the former is the process of recognizing geographic context with analyzing text while the latter refers to the process of assigning geographic coordinates (McCurley, 2001). Due to the lack of a standard formatting of addresses across countries and also variations on abbreviations, punctuations, suffixes, line break and others, the geoparsing process is not a trivial task (Borges et al., 2007).

Borges et al. divide address into three parts: basic (street type, name and building number), complement (neighborhood name), and location identifier (postal code, phone number, city and state, country). The complement part is optional and the sequence of components varies across countries (Borges et al., 2007). The location identifier is mainly the part of interest in this research because it is consistent with the city-level spatial resolution of the analysis. SQL-like queries and full-text search are performed on the shippers' address to identify the city, province (state) and country of product source, which are discussed in next section. Postal codes contain precise locational information; also area code of phone numbers may refer to spatial localities. However, their differentiation from each other and also from other numeric data is sometimes hardly possible. Hence, postal codes are examined only when there is no useful information extracted from another part of the address. Phone numbers that are accompanied by “TEL” are examined only if locational information is not identified in the postal code.

4.1.3. Full-Text Search Capabilities

Two tables are involved in the geoparsing process; one is the PIERS data with the shippers’ address fields (source table) and the other is the place name table with their spatial coordinates (reference table). A relation between two tables can be defined if the source table contains keywords of the reference table. The LIKE predicate evaluates whether a string expression contains a string pattern, but its efficiency decreases drastically with large source and reference tables such as those involved in this research. Full-text search capabilities have been integrated to major relational database management systems (DBMSs) to handle textual indexing and searching (Liu, Yu, Meng, & Chowdhury, 2006).

The SQL Server database application used in this research provides full-text search functions and predicates that significantly outperforms LIKE queries in the context of place name identification. Also, they allow the specification of queries to find keywords that appear near each other. This search condition enables us to test the relative position of address components, such as the proximity of city and province names. Other advantages of full-text search methods include relevance-based ranking that indicates the relevance of returned data with the search condition (Hamilton & Nayak, 2001), accent insensitive searches that return the same for either São Paulo or Sao Paulo, for instance, and ignorance of whitespaces, dashes that returns same for Ho Chi Minh and Hochiminh and also Minato-ku and minatoku.

4.1.4. Web-mining

The addresses in the PIERS dataset are not always complete and contain the location identifier component of shipper addresses. Searching for incomplete addresses and the Shipper company names on the Web may lead to finding complete addresses. Although human intelligence plays a significant role in detecting geographical location, this is supplemented by a web search application developed in C#.Net to automatically extract geographical location from company's web pages.

However, geographical information extraction requires further cleaning on identified place names to ensure an unambiguous and valid location. For instance, multiple locations may be identified from a single address. The spatial association between different locations helps to detect outliers. For instance in the case of extracted location names such as "Shanghai", "China", and "CH", and "Paris", that form one address string, the last one likely happened due to textual ambiguities and is excluded as an outlier.

4.1.5. Disambiguating Place Names

Once a place name is detected in a text field, determination of the actual entity that the name refers to is the next stage. Textual ambiguities of geographical entities are divided into two tasks (Ahlers & Boll, 2008). Nongeo-/geo-ambiguity happens when a place shares name with other entities, such as Washington, which is a name for both person and place. The use of one same term for naming different places is called geo-/geo-ambiguity such as London in the U.K. and London in Canada.

Contextual information is required to filter coincidental names. In order to handle geo-/nongeo-ambiguity, we take the position of keywords into account by giving higher priority to the keywords at the end of a text string. The observed pattern in our database (Figure 4.1) makes the information on right-hand side of the textual address more related to the city-level spatial resolution of our application.

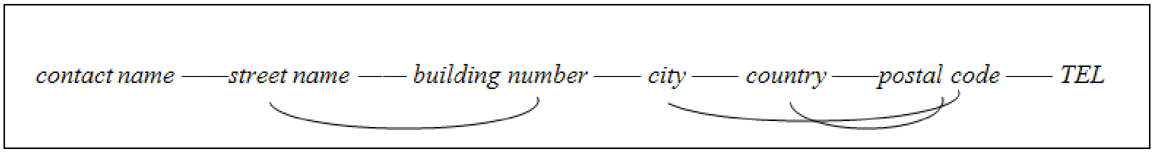


Figure 4.1: The textual address usually starts with contact name and is followed by building number and street names and then city and country names (or country's abbreviations and FIPS codes), postal codes and phone numbers. The curve lines depict the possibility of replacements in address components.

Searching for other related place names, especially in the vicinity of the first place name's position along the string, is our approach to geo-/geo-disambiguation. The related place names are mainly the first place's parent region, which include the province (state) and country names, and sometimes, postal codes.

4.1.6. Geocoding

The coordinates of localities in reference files are assigned to the identified place names. Geocoding allows the mapping of extracted information, which then enables us to visually check the validity of the results. The fact that shipments are not generally shipped via ports that are relatively far from the shipment origin is used to detect possible errors by scanning origin-destination maps. Figure 4.2 represents the visualization of the distance to a specific port produced by the spatial extension of SQL Server. This information is used in some parts of the verification task in conjunction with information on container feeder routes.

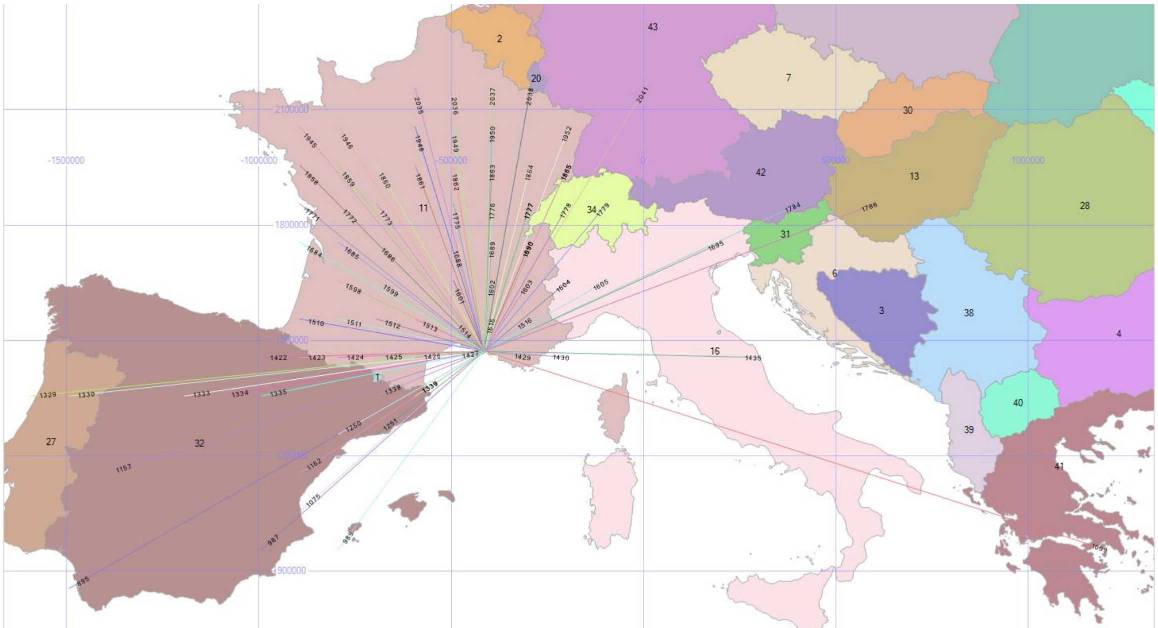


Figure 4.2: Spatial SQL Server representation of production locations shipping via seaport of Fos for map-based verification

4.1.7. Verification and Market Intelligence

The role of human-centered evaluation has been mentioned during the explanation of other stages of the adopted GIR framework. In fact, market intelligence is central to this process, as other pieces of information in the shipment record and corporate websites may incorporate useful evidence on the geographic location of firms. Also, the automatic part of the GIR process may end up with a large number of shipment records left either without unambiguous and sensible location or with duplicate potential locations. The methodology that is undertaken for the implementation of GIR is summarized in Figure 4.3.

Market intelligence is based on the logistical realities of global commodity chains that are connected with production locations for companies with multiple facilities. Production/warehouse facility locations are determined from corporate websites and from evidence within shipment records - such as in the case of a "precarrier location" data field matched with a particular production facility (precarrier location being the point at which a carrier takes possession of the cargo, as indicated in the shipment record). Carrier services are also used during this process.

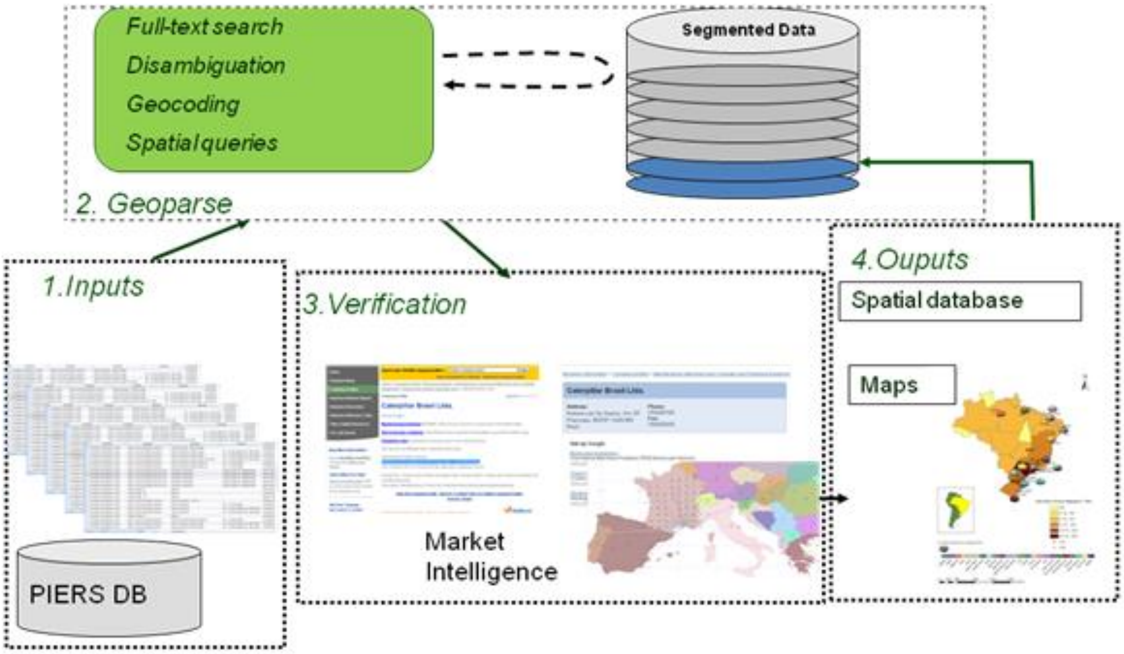


Figure 4.3: Different stages of implementing geographical information retrieval

Market intelligence is a slow process of knowledge building that justifies developing a gradual knowledge base to learn from market intelligence's outcome. Machine learning techniques allow the incorporation of knowledge into the GIR process while minimizing the need for human interaction. Therefore, the raw dataset has been segmented and a spiral workflow is implemented, as depicted in Figure 4.4.

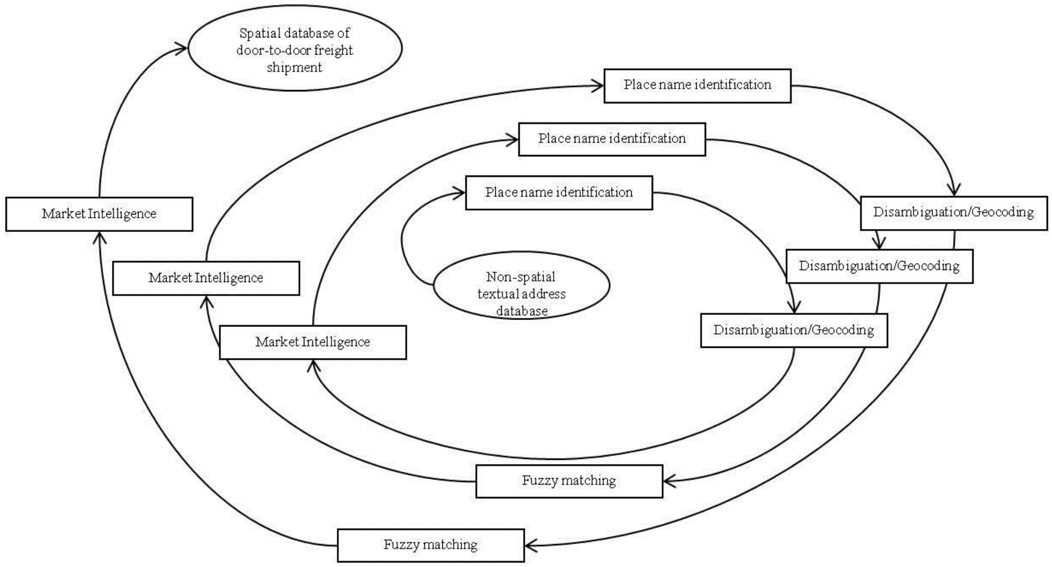


Figure4.4: Implemented spiral workflow for GIR

4.1.8. Fuzzy Matching

Regarding the significant amount of time and energy spent on data verification, a prudent alternative is to match the unseen data against verified records. However, due to the different spellings, abbreviations, truncations, and omissions, two records referring to the same locations may not exactly match. A robust matching method is required to be resilient to small differences and to find unverified records that are reasonably close to the verified tuples (Chaudhuri, Ganjam, Ganti, & Motwani, 2003). Fuzzy matching is an instance-based learning operation that determines which tuple of the training set (verified table) is the closest to an unknown test set (unverified table) (Witten & Frank 2005).

Instance-based learners define the notion of closeness by a distance function such as the Euclidean distance or string edit distance. In SQL Server Integration Service (SSIS), the notion of closeness between records is measured by a similarity function. The fuzzy matching similarity function tokenizes data using string’s delimiters and measures the cost of transforming one record’s into others. Transformation operations include replacement, insertion and deletion of tokens. However, the cost of transformation of tokens varies with respect to the token's importance. The importance of tokens decreases with their frequency in a document: for instance, in a string on a BMW company, BMW is more informative and should weigh more. Inverse document frequency (IDF), which is a well-known weight in the information retrieval literature, is utilized in the similarity function (Chaudhuri et al., 2003).

Records matched with similarity score higher than a pre-specified threshold are likely valid and the rest requires involvement of market intelligence. Moreover, the remaining records are grouped to very similar records and only one representative from each group needs human interaction for verification.

4.1.9. Final Dataset

The combination of the implementation of automated machine-learning algorithms and of the market intelligence method enables us to compile a unique, consistent and comprehensive database of U.S.-bound containerized door-to-door freight shipments with multiple locational and business attributes. It thus allows the direct study of the hinterland-port-foreland triptych (Charlier, 1992; Robinson, 1970). Records where relevant attributes are suppressed are excluded from the dataset. The October 2006 timeframe is important because it provides a baseline preceding the trade downturn and restructuring associated with the Great Recession of 2008.

Data processing led to the exclusion of 12,169 bill-of-lading records due to suppressed or incomplete information. The resulting usable dataset contains about 107,000 bill-of-lading records, which collectively represents U.S.-bound traffic of 202,837 TEUs. This cross-North-Atlantic trade is estimated to represent about 13.4% of U.S. containerized imports (Carluer, 2008) and 21.1% of European containerized exports (Amerini, 2010). The following section presents data with some descriptive statistics and visualization of the commodity flows. Figure 4.5 presents visualization of the commodity flows from geocoded origins to the 22 major European ports[1]. It should be noted that certain European ports may play a significant role in trade with the United States, but as terminal points of a feeder service only and not as forwarding ports. This is precisely the reason for St. Petersburg, Russia, or Constanta, Romania to not appear on this list. This research focuses on a sample of this data. Chapter 8 presents descriptive statistics of the sample data used in the analysis.

[1] This Figure is extracted the book chapter: Kashiha, M., and Thill J.-C. "The Functional Space of Major European Forwarding Ports: Study of Competition for Trade Bound to the United States," in A. Verhetsel, T. Vanoutrive (editors), Smart Transport Networks: Decision Making, Sustainability, and Market Structure, Edward Elgar (2013).

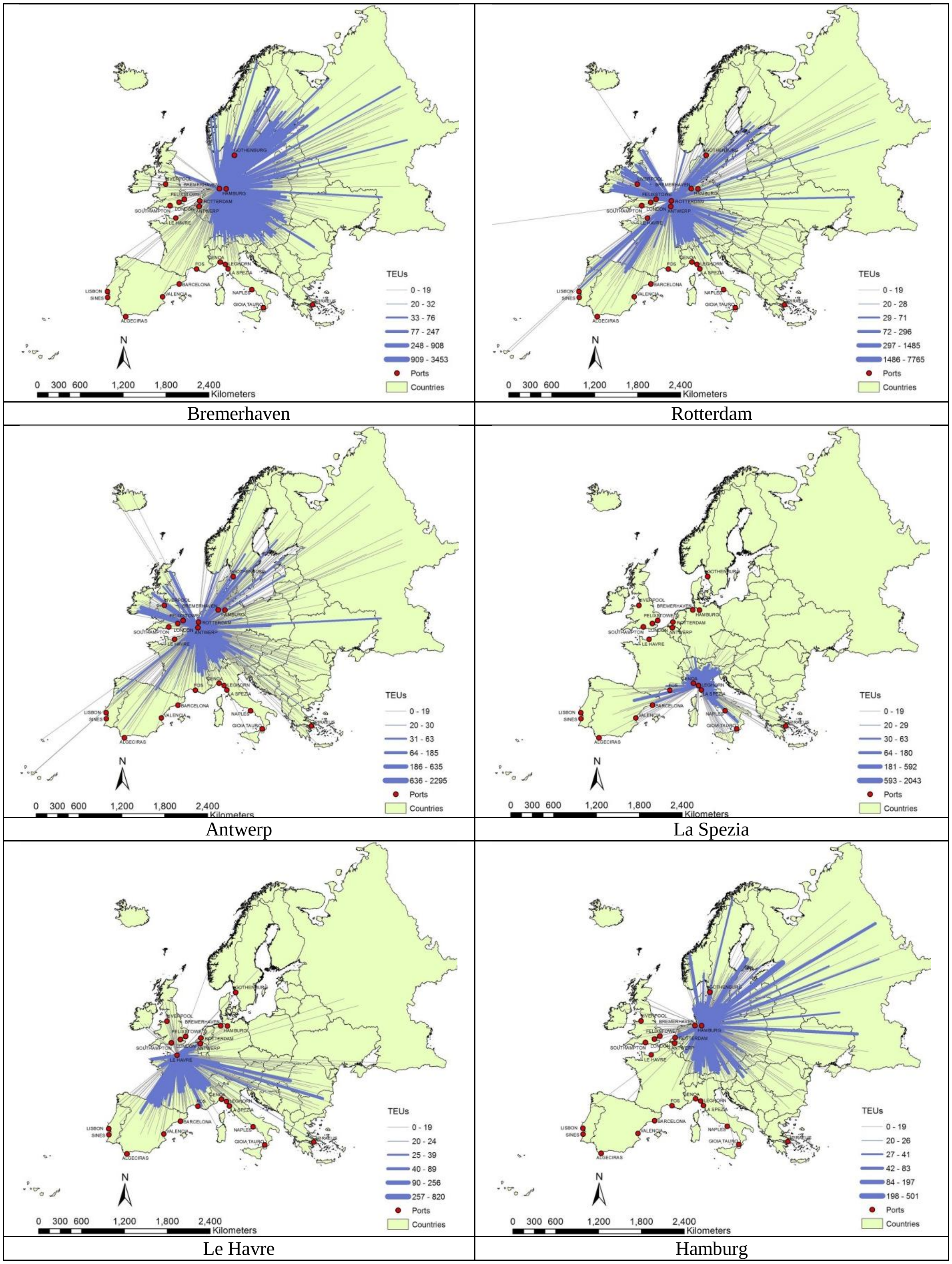


Figure 4.5: Hinterlands of major forwarding European ports

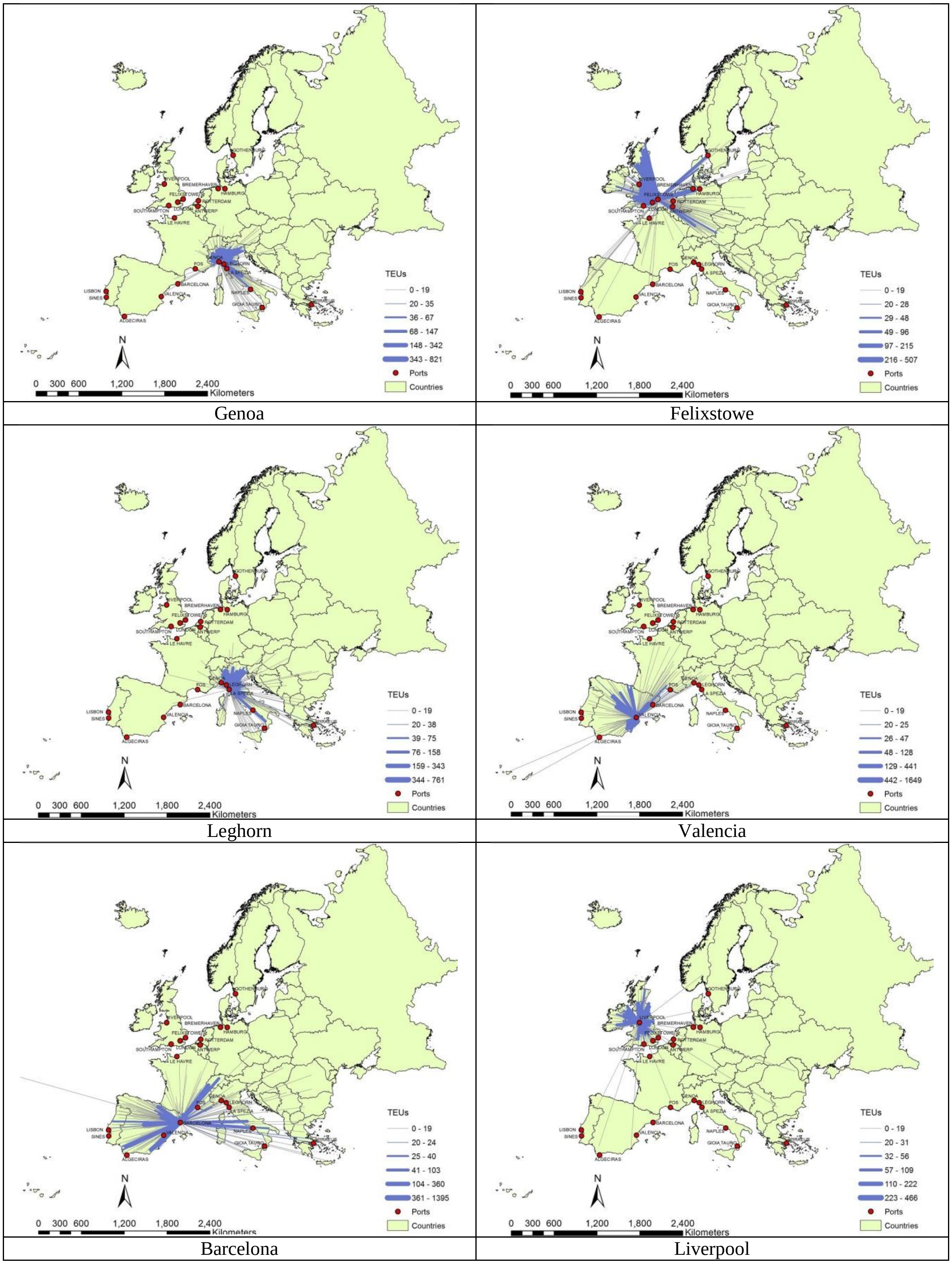


Figure 4.5: (continued)

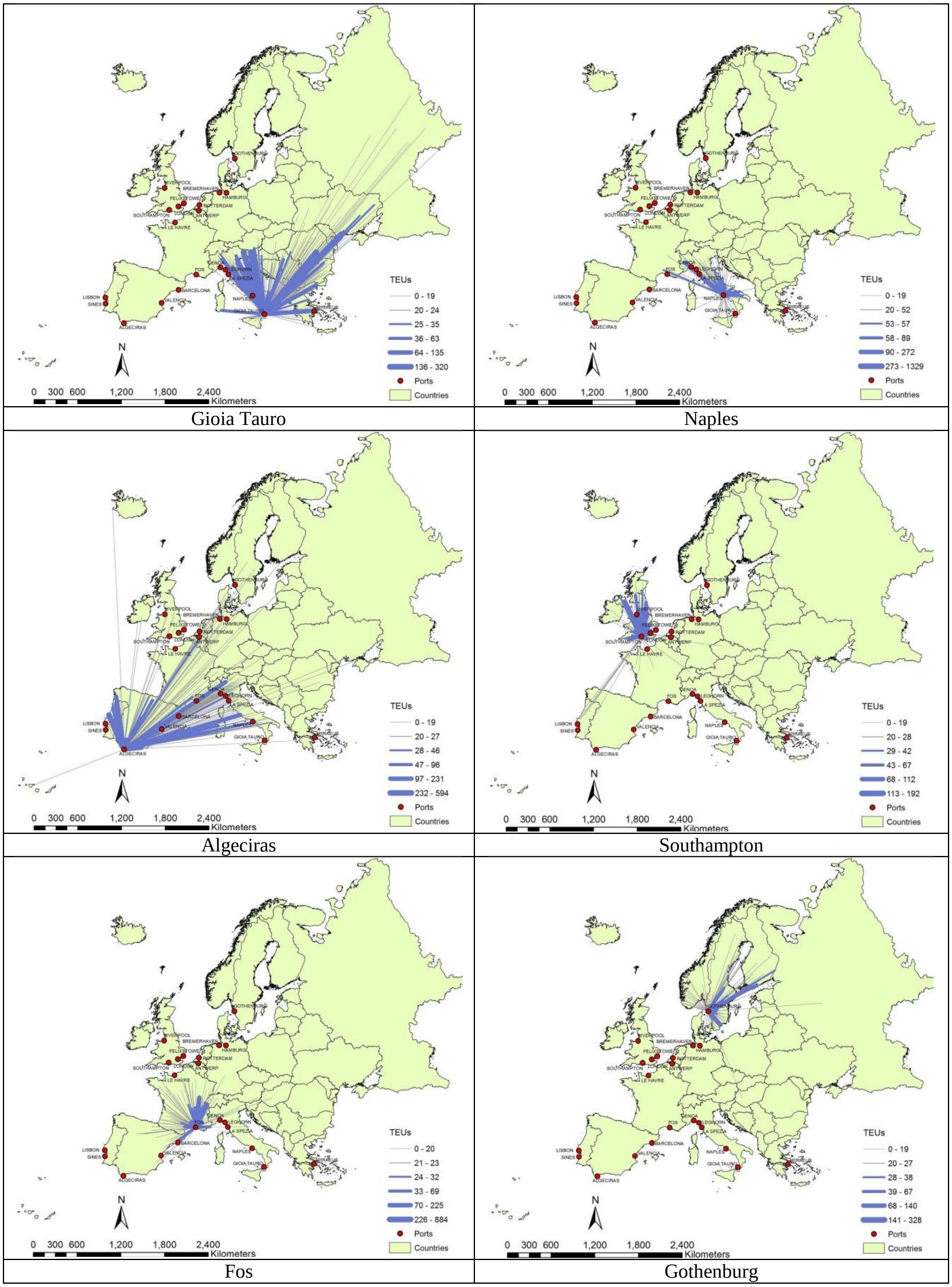


Figure 4.5: (continued)

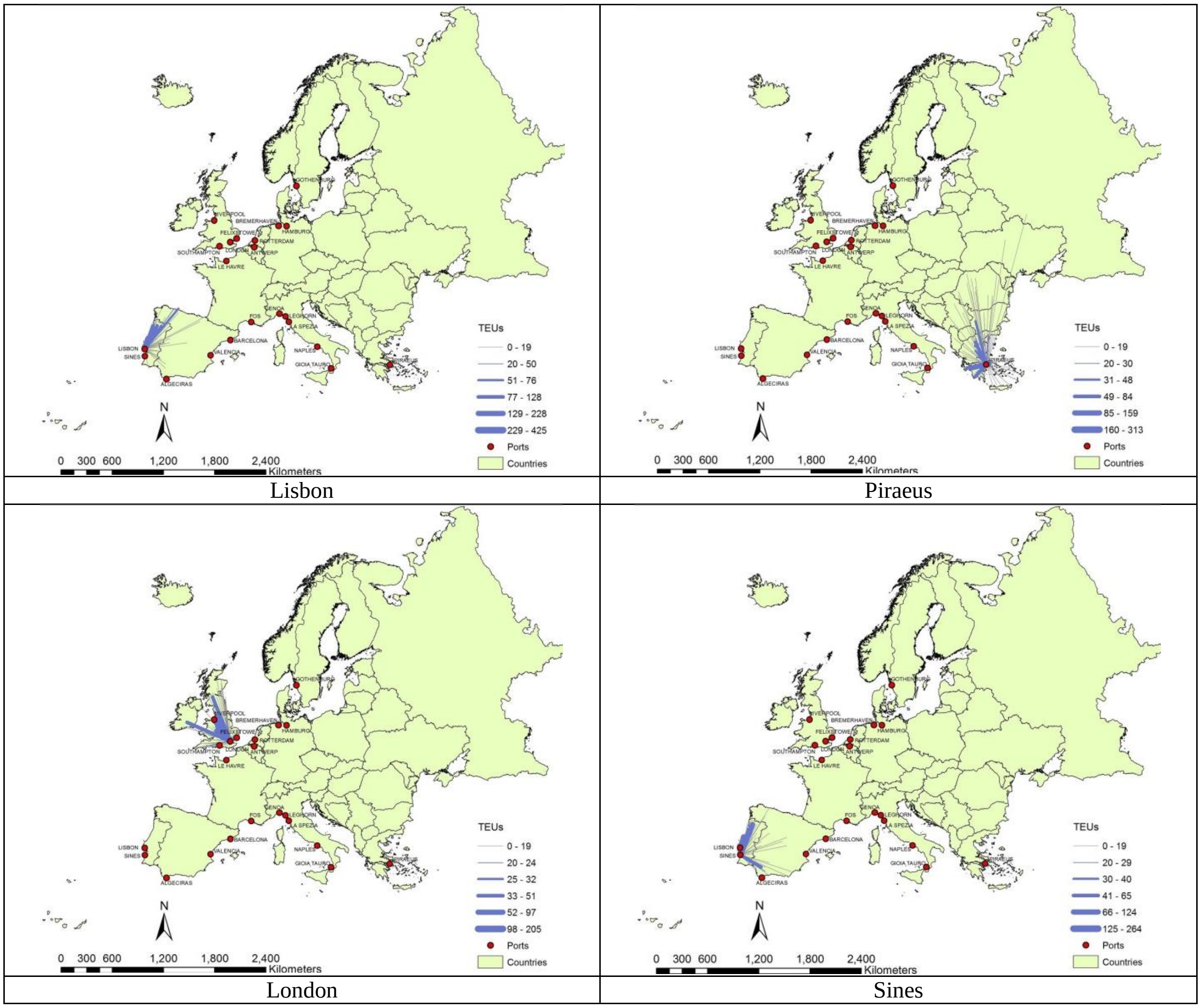


Figure 4.5: (continued)

4.2. Port Efficiency: Stochastic Frontier Analysis

Despite the availability of a huge body of literature dealing with port efficiency, there are no universal port-level efficiency indicators or even a standardized measurement available. Besides the port efficiency indicators published by the Global Competitiveness Report (GCR) at the national level, some economists have designed consistent measures of port efficiency suitable to international/national comparative analysis through Data Envelopment Analysis (Barros, 2003; Turner, Windle, & Dresner, 2004; Wang & Cullinane, 2006) and Stochastic Frontier Analysis (Coto-Millan, Banos-Pino, & Rodriguez-Alvarez, 2000; Cullinane & Song, 2006; Cullinane, Wang, Song, & Ji, 2006; J. Tongzon, 2001; Wanke, Barbastefano, & Hijjar, 2011), multivariate analysis (J. L. Tongzon 1995; L. C. Tang, Low, and Lam 2008, Sanchez 2003), and regression models (Clark et al. 2004; Blonigen & Wilson 2007), to name a few. However, as summarized in a survey of empirical research on port efficiency by Gonzalez and Trujillo (2009), literature on efficiency in the port industry is relatively new compared to studies conducted on other transport sectors such as railway, airport, and airline sectors (Gonzalez & Trujillo, 2009).

Blonigen and Wilson (2007) measure port efficiency using U.S. census data on import charges. Import charges can be decomposed into costs of loading at foreign ports and unloading at the U.S. ports, and costs of port-to-port transportation. The first costs are replaced by dummy variables for foreign and U.S. ports, respectively, which approximate the foreign and U.S. ports relative inefficiencies. Their analysis is based on the premise that higher costs of getting the cargo to the docks and unloaded is a result of port inefficiency, which contributes to a rise in import charges. While this logic can be criticized on the ground that omitted inland transportation infrastructures, which is correlated to port efficiency, may bias estimates (Blonigen & Wilson 2007). Tang et al. apply Factor Analysis (FA) to identify key dimensions of port attributes and end up with three factors that are interpreted as port efficiency, scale economies, and the convenience in using the port, respectively. While trade volume, turnaround time, and port charges heavily load on port efficiency component, the port traffic, number of port calls and drought variables load on scale economies, and annual operating hours of port and the availability of intermodal transport facilities load on the third factor (Tang et al., 2008). Sanchez analyzed ports efficiency based on a survey of Latin American common users of ports. Principal component analysis (PCA) has been used to group different variables on port efficiency into three factors: time inefficiency, productivity and stay per vessel (Sánchez et al., 2003).

In recent years, approaches such as stochastic frontier analysis (SFA) and data envelopment analysis (DEA) have been widely used to measure the degree of relative efficiency in port industry, in which the terminal or ports are considered as the production unit. This research utilizes SFA, which is an econometric method, allowing for random shock measurements. Under this method, the output of efficient units lies on the production frontier function, while inefficient units lie below the frontier (Cullinane & Song, 2006).

4.2.1. Methodology

The theoretical definition of a production function, expressing the maximum output that can be obtained from a specific set of inputs with fixed technology, is not sufficient for empirical work. The observed input-output values show that firms do not produce the maximum feasible output for the inputs involved, given the technology available. To bridge the gap between theory and empirical work, Farrell (1957) proposed the so-called frontier production functions to characterize discrepancies between outputs of firms with the same input bundles or to explain why firms operate below the frontier production (Aigner, Lovell, & Schmidt, 1977; Battese & Coelli, 1993). A measure of the technical efficiency of firm TE_i with output y and input x is given by

$$TE_i = \frac{y_i}{f(x_i;\beta)}, \qquad (1)$$

where $f(x_I; \beta)$ is the production frontier, which assumes a parametric function between production inputs and outputs, and β is a vector of parameters coefficient to be estimated. Here a cross-sectional data on a set of N inputs is assumed to produce a single output for each firm i. Equation 1, however, cannot distinguish production efficiency from other sources of random shocks and environmental variations in the error term. As an alternative, the stochastic frontier model has been suggested, which imposes a more logical error structure. Under the stochastic frontier model, equation 1 is rewritten as

$$TE_i = \frac{y_i}{f(x_i;\beta).\exp\{v_i\}}, \qquad (2)$$

where $f(x_i;\beta).\exp\{v_i\}$ is the stochastic production frontier and consists of two parts: a deterministic part $f(x_i;\beta)$ and a stochastic part $\exp\{v_i\}$, which captures the effect of random shocks on each firm (Kumbhakar, 2003). Two different research groups introduced the stochastic production frontier model at the same time; Aigner, Lovell, and Schmidt (1977) and Meeusen and Van den Broeck (1977). The log-linear Cobb-Douglas function has been considered as a basic production function between inputs and outputs (Kumbhakar, 2003). Particularly, the production frontier function used in studies of port efficiency (S. Chang, 1978; Cullinane & Song, 2006; Wanke et al., 2011) takes a log-linear Cobb-Douglas form. Then equation 2 can be written as

$$\ln y_i = \beta_0 + \textstyle\sum_n \beta_n \ln x_{ni} + v_i - u_i \qquad (3)$$

where v_i is the symmetric stochastic component, and asymmetric u_i is the inefficiency component, of the error term, ε_i. Our objective is to obtain estimates of input coefficients, β_n, and of the technical efficiency of each firm. The latter estimation requires separate estimates of v_i and u_i from ε_i. Thus distributional assumptions on the two error components are necessary (Kumbhakar, 2003). The noise component v_i is assumed to be identically and independently distributed as $N\,(0, \sigma_v^2)$, and independently of u_i. Different one-sided distributions (such as half-normal, exponential, gamma distribution, etc) have been proposed, of which the half-normal is employed more frequently. A nonnegative distribution of u_i reflects the fact that production units operate below the frontier.

Given the density function of u and v, the joint density function for u and ε is simply calculated. The density function of ε can be derived by integrating u out of $f(u, \varepsilon)$

$$f(\varepsilon) = \frac{2}{\sigma} f^*\left(\frac{\varepsilon}{\sigma}\right)[1 - F^*(\epsilon\lambda\sigma^{-1})], \qquad (4)$$

where $\sigma^2 = \sigma_u^2 + \sigma_v^2$, $\lambda = \sigma_u/\sigma_v$, and $f^*(\cdot)$ and $F^*(\cdot)$ are the standard normal density and distribution functions.

The log-likelihood function, using equation 4, can be maximized with respect to λ, and σ^2 to obtain maximum likelihood estimates of these parameters. However, the estimation procedures yield $\varepsilon_i = u_i - v_i$, rather than u_i. The next step is to extract the information on u_i from the conditional distribution of u_i given ε_i as shown by Jondrow et al (1982).

$$\hat{u} = E[u_i|\varepsilon_i] = \frac{\sigma\lambda}{(1+\lambda^2)}\left[\frac{f^*\left(\frac{\varepsilon_i\lambda}{\sigma}\right)}{1-F^*\left(\frac{\varepsilon_i\lambda}{\sigma}\right)} - \frac{\varepsilon_i\lambda}{\sigma}\right] \quad (5)$$

where u_i follows half-normal distribution. They suggest that the mean of this distribution, $E[u_i|\varepsilon_i]$, is an appropriate estimator for u_i. The estimate of the technical efficiency can be calculated from (Kumbhakar, 2003).

$$TE_i = \exp\{-\hat{u}\} \quad (6)$$

The conditional distribution of u_i given ε_i, under other forms of distributions (such as exponential, truncated normal) has been formulated and employed (Aigner et al., 1977; Kumbhakar, 2003). This research also estimates relative efficiency under three different distributional assumptions.

4.2.2. Model Specification

In order to define the port production function, and its input, and output, we follow Cullinane and Song (2006). Similar to the functional form chosen in most of the related studies, they assume a log-linear Cobb-Douglas production function (Gonzalez and Trujillo, 2000). Focusing on container terminals, they argue that the production output should be measured in terms of the number of containers handled (TEUs), instead of total tonnage. Their justification is based on the fact that the production inputs for moving any container, regardless of their size and weight, is roughly the same. In terms of input, they rely on certain key physical characteristics of terminals, including terminal berth length, terminal area, and the number of cargo handling equipment pieces. The labor factor has been omitted from the input set, under the assumption that there is high degree of correlation between the physical quantities of capital and labor factor in production (Notteboom, Coeck, & Broeck, 2000; J. Tongzon & Heng, 2005; Turner et al., 2004). We extract the same set of variables from the Containerized International Yearbook 2008, for 128 European terminals (aggregated to 60 ports). Although purpose of this study is to obtain port efficiency for 22 major European ports that are actively involved in trade with the United States, the analysis is conducted on larger sample to ensure the consistency and robustness of the results. Descriptive statistics for the data are illustrated in Table 4.1. The estimation of port efficiency scores is conducted in the Stata statistical software, which relies on Jandrow et al.'s (1982) formula (Belotti, Daidone, Ilardi, & Atella, 2012).

Table 4.1: Descriptive statistics of the variables for European ports included in analysis

Variables	Mean	S.D.	Min	Max
Port throughput	1182107.051	1960985.838	20120	9654508
Port facilities	86.40677966	107.187898	4	595
Port area	887320.1186	1331482.458	32000	7018911
Berth length	2168.457627	2441.099324	200	11995

4.2.3. Ports Technical Efficiency

However, before estimating a stochastic production frontier, we check the presence of technical inefficiency in the data. In the presence of technical inefficiency, the error term, $\varepsilon_i = u_i - v_i$ is expected to be negatively skewed. Thus a symmetrical distribution of the OLS residuals means that u_i=0, while a negative skewness of residuals is evidence of technical inefficiency, u_i>0.

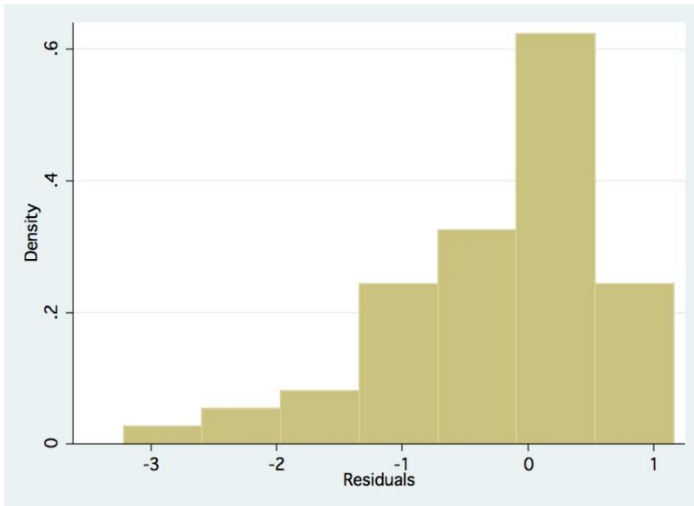


Figure 4.6: Histogram of the OLS residuals

Figure 4.6 confirms that the OLS error term is negatively skewed. So we proceed with the implementation of the SFA on port cross-sectional data. It is worth noting that before SFA estimation, we aggregated the terminals to the port level, in order to obtain relative efficiency for the 22 ports involved in the port choice analysis. Information on terminals' throughput is sometimes missing from the Containerization Yearbook. In such cases we aggregate input variables of terminals (terminal berth length, terminal area, and the number of cargo handling equipment pieces) within the ports. For the output variable we used the port throughput reported in the Containerization Yearbook. Table 4.2 represents the OLS and SFA estimations for each of the three distributions of the inefficiency term.

Table 4.2: Estimation of technical efficiency

Variables	OLS	MLE		
		Half-normal	Exponential	Truncated Normal
Constant	6.006*** (1.53)	6.216*** (1.24)	5.814*** (1.11)	5.982*** (1.19)
ln (facilities)	.534*** (.17)	.470*** (.14)	.437*** (.12)	.450*** (.13)
ln (area)	.341*	.263*	.281**	.269*

	(.17)	(.15)	(.140)	(.14)
ln (length)	.099 (.21)	.337** (.17)	.364** (.17)	.362** (.17)
λ		4.00	2.40	
σ_v^2		.219	.240	.051
σ_u^2		.875	.576	1.785
Sigma2				1.836
gamma				.972
Log-likelihood		-39.89	-39.82	-39.54
R^2	.737			

*** significant at 1% level
**significant at 5% level
*significant at 10% level
Figures in parentheses indicate standard error

The goodness of fit of the OLS model implies that the four input variables explain around 74% of the variation in port throughput. The MLE results indicate that the parameter estimates are statistically significant at the 10% level and their signs conform the a priori expectations, where extra physical equipment is expected to increase port throughput. The results also show that the coefficient estimates are very similar under the three alternative distributions, but relatively different from OLS estimates. The disagreement between OLS and MLE estimates is explained by the relative variability of the one-sided and symmetric error terms, as computed by λ. The results suggest that λ is significantly different from 0, meaning that one-sided errors become the dominant source of random variation in the data (Aigner et al. 1977).

Although the averages of the efficiency estimates obtained from the three alternative distributions are slightly different, the ranking of port efficiencies turns out to be robust and not sensitive to the distributional assumptions. Spearman rank correlation coefficients between the three different efficiency estimates are over 0.97, revealing highly correlated results. Under all three distribution forms, the port of Algeciras is ranked the highest; it is followed by Naples, Rotterdam, Felixstowe, Hamburg, La Spezia, Liverpool and Valencia (with almost same orders across each distributional form). The container ports of Genoa. Leghorn, Le Havre, Lisbon, and

Gothenburg are estimated to be the most inefficient ports. Figure 4.7 depicts the relative efficiency estimates of 22 European container ports, under the assumption of half-normal.

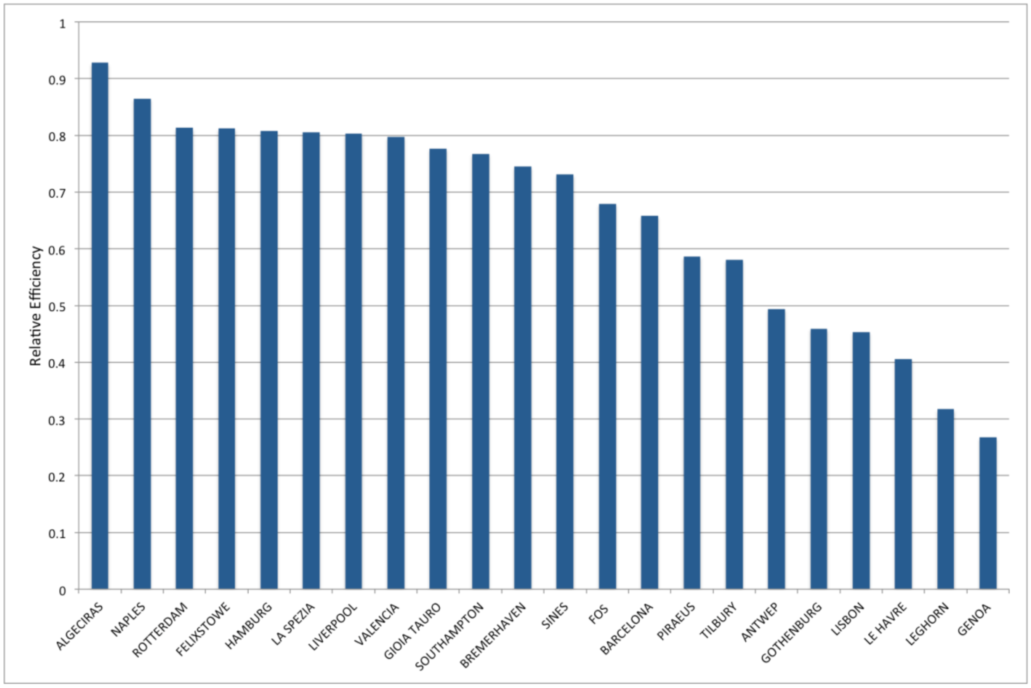


Figure 4.7: Relative port efficiency estimates for 22 European ports (Half-Normal distribution)

Examination of the efficiency indices reveals that larger ports are expected to operate more efficiently, depicted in Figure 4.8. That is, there is a positive correlation between port throughput and efficiency estimates ($\rho = 0.44$). This is in line with other studies that find a positive relationship between port size and efficiency level (Cullinane & Song, 2006; Gonzalez & Trujillo, 2009; J. Tongzon & Heng, 2005; J. L. Tongzon, 2009). These studies show that the larger a container port is in production scale, the more likely it is associated with higher technical efficiency score. However, there is research that reaches an opposite conclusion such as Coto-Millan et al. who examined Spanish ports. They applied stochastic frontier cost function and concluded that the relatively larger ports are more economically inefficient (Coto-Millan et al, 2010).

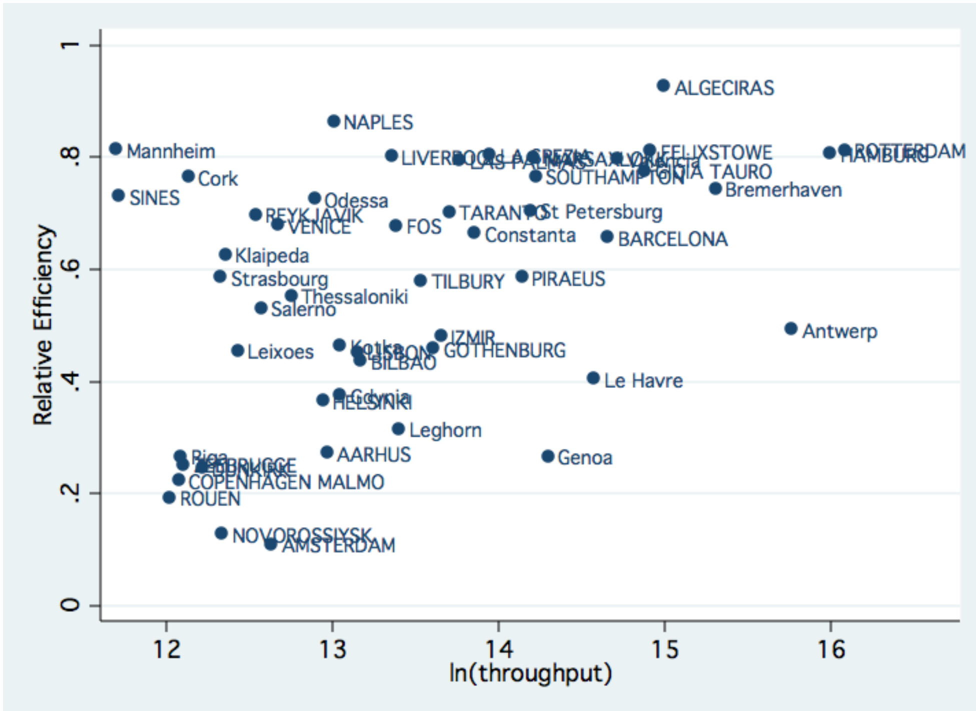


Figure 4.8: The relationship between ports' relative efficiency and throughput

CHAPTER 5: MODELING FRAMEWORK

This chapter is devoted to building a methodological framework for modeling spatial competition between firms at a fine level of disaggregation. This chapter contributes to the existing literature on competition modeling by measuring the probabilistic aspect of choice in a market characterized by significant preference heterogeneity. The developed probabilistic spatial choice framework allows us to transform spatial interaction flows into the space of competition.

In this section we discuss the empirical framework employed to model competitive interactions between suppliers and market contestability by structural estimation of market segments' preferences for firm attributes. A Latent Class Logit (LCL) model, known as a discrete random coefficient logit, is the basis of the framework that accommodates taste heterogeneity of the population and simultaneously estimates underlying preference parameters and segments the market. This approach relies on the individual's choice set consisting of choices made by each individual over a certain period of time.

Methodologies related to answering our main research questions are discussed in this chapter under two main sections. First, we start with a definition of choice sets, and market captivity/contestability that are the central concepts in this research. Then, we propose a measure for quantifying competitive interactions based on individual choices. Second, we model customers' choices as a function of the characteristics describing firms and customers. We specifically articulate the limitations of the standard logit model and justify the necessity for using a latent class logit modeling approach.

5.1. A Framework for Choice Theory

The decision to ship via a port boils down to a discrete choice from a set of alternative ports. The hypothetical decision-maker faces different costs and services associated with each possible port. The decision maker evaluates suitability of available alternatives based on his/her best knowledge. However, he/she may not be aware of all the possibilities. Available information is structured in the form of decision rules to arrive at a choice among a set of alternatives. According to Ben-Akiva and Lerman (1985) a theory of choice is composed of the following elements:

a. Decision maker

In this research, the decision making unit is referred to as "shippers". It indiscriminately represents a group of shipping agents, including shippers, carriers and freight forwarders, who ship from a certain production location.

b. Alternatives/choice set

In this research, 22 European major ports collectively form the universal set of competing alternatives. The individual's choice set, as a subset of the universal set, contains alternatives feasible to the individual. Contrary to the prevalent approaches that set deterministic constraints, such as physical availability, monetary resources, time availability, information constraint, and so on, to identify the feasibility of an alternative (Swait & Ben-Akiva, 1987; J.-C. Thill, 1992), our choice set is formed of observed choice distribution. With the second objective, we attempt at modeling choice set.

c. Attributes of alternative/ taste variations among decision makers

The attractiveness of ports such as port performance, connectivity, location and so on, is expressed as attributes of alternatives. Also, taste variations of individuals may lead them to evaluate attributes differently and require that parameter estimates vary across decision makers.

d. Decision rule

This research defines decision rules with a set of “if-then” rules that describe the mechanism of information processing towards a unique decision.

Rational behavior is the basic axiom of microeconomic consumer theory and individual choice theory. Consumers are rational in the sense that they have consistent preferences. In other words, placed under identical situations, an individual will repeat the same choice. Or, two or more individuals with the identical choice sets, attributes, and socioeconomic characteristics, will select the same alternative (M Ben-Akiva & Lerman, 1985; McFadden, 1981).

But, in practice individuals are seen not to choose the same alternative in repetitions of the same choice circumstances. Behavioral inconsistencies are taken to be due to the inability of individuals to discriminate perfectly, or to the inability of the analyst to fully measure the context of the choice (M Ben-Akiva & Lerman, 1985; McFadden, 1981). Probabilistic choice theory is developed to accommodate this observed probabilistic behavior.

The literature reviewed in chapter 2 put a great emphasis on geographical distance, transportation infrastructures and accessibilities, port efficiencies and so on, as driving factors of port selection. Hence, individuals with the same locational characteristics are expected to reveal unique preference behavior. However, our exploratory data analysis reveals large behavioral inconsistencies at least from the observational perspective of the analyst. Such alternative switching can be justified through unobservable taste heterogeneity of decision makers, timing and capacity considerations, and other underlying determinants that cannot be explained by time-invariant elements (Kamakura, Russell, 1989).

5.1.1. Choice Set Definition

We quantify competition by measuring disaggregated choices over a certain period of time, which includes repeated decisions. We claim that two alternatives compete only if both options coexist in an individual's feasible choice set. Choice sets are composed of selected alternatives and the choice probabilities are associated to them in a choice vector.

To determine the choice set, let us assume that there are J alternatives and I individuals. J alternatives form the universal choice set, A^U. The choice probability, p_{ij} , is denoted by the probability that individual i chooses alternative j (or market share). a_i is a choice vector of individual i that is an n-component vector of choice probabilities, p_{ij} . It follows that

$$a_i = \{p_{ij} \mid p_{ij} \geq 0 \wedge \sum_{j=1}^{J} p_{ij} = 1\}, \qquad (5.1)$$

and $A_i$is the choice set of individual i that

$$A_i = \{j \mid p_{ij} > 0 \wedge \sum_{j=1}^{J} p_{ij} = 1\}, \qquad (5.2)$$

Since we are interested in modeling competition based on observed choices (besides the estimated choices) we define the observed choice probability as the ratio of instances where alternative j is chosen by individual i, N_{ij}, to the total number of choices made by individual i, N_i,

$$p_{ij} = \frac{N_{ij}}{N_i}, \quad (5.3)$$

$$N_i = \sum_{j=1}^{J} N_{ij}, \quad (5.4)$$

Certain individuals ship via a single port, which therefore collectively constitute captive hinterlands. Other individuals may ship via multiple ports, which collectively form contestable hinterlands (Zhang 2009, De Langen 2007, Notteboom 2009). This definition of hinterlands is consistent with Grover and Srinivasan's loyal and switching market segments and Novshek and Sonnenschein's definition of marginal consumers (Grover & Srinivasan, 1987; Novshek & Sonnenschein, 1979). We use the terms captive and contestable following De Langen who made a distinction between captive and contestable hinterlands. According to him, a captive hinterland is composed of all regions where one firm has a substantial competitive advantage. Contestable hinterlands consist of all regions where there is no single firm with a clear cost advantage over competing firms. Consequently, various firms will have a share of the market (De Langen 2007).

However, in economics, contestable market specifically refers to the concept proposed in Baumol (1982) where it characterizes a market structure in which firms behave competitively, irrespective of being in a monopoly, oligopoly or perfect competition. The critical characteristic of a contestable market is its vulnerability to hit-and-run entry that enforces firms to behave in a competitive manner (Baumol, 1982).

The concept of contestability in this research refers to a similar market structure, but from a demand perspective. In other words, contestable demand does not guarantee firm's profit and readily switches to the other firm as soon as it offers higher utility.

A captive market j contains individuals with a unit choice vector that always choose alternative j. It can formally be defined by

$\exists! \ p_{ij} \in a_i \ | p_{ij} = 1 \wedge p_{ij} \times p_{ik} = 0 \ for \ j \neq k$ (5.5)

Based on this definition, J distinct vectors can be identified, which represent loyal individuals to alternative j, a^j

$\forall a^j, a^k \ | a^j \bullet a^k = 0 \ for \ j \neq k.$ (5.6)

Accordingly, the choice vector of a contestable market follows as

$\exists \ p_{ij}, p_{ik} \in a_k \ | p_{ij} \times p_{ik} \neq 0$ for $j \neq k,$ (5.7)

In order to take into account the effect of the choice probabilities distribution, a Herfindahl, H, index weight individuals when delineating captive and contestable hinterlands,

$H_i = \sum_j^J p_{ij}^2.$ (5.8)

Equations 5.5 to 5.8 are used to define indexes in Tables 6.4 and 6.7.

5.1.2. Competition Measurement

In order to clarify how we characterize competition in this research, let us consider a market with two choice alternatives, j and k, each of which with a market share of 50%. It is hard to say if j and k really compete, unless information on who selects j and who selects k is available. If people who select j are likely to select k and vice versa, j and k may compete; otherwise if people who select j are not seen to select k, and vice versa, no rivalry exists between j and i. Accordingly, the approach proposed here assesses the magnitude of competition between k and j across individuals who are likely to substitute j for k. Hence, in contrast to approaches that measure competition on the basis of a single aggregate market share index, which masks whether two alternatives share the same market or not, our approach measures competition ensuring that both alternatives coexist in the individual choice set.

In order to model and quantify competitive interaction at a certain location, the feasible choice set must be identified. Once the choice set is determined, fluctuations between choice probabilities reveal the magnitude of competition. This research computes degrees of competition between two alternatives within individual' choice sets locally. The local measure of competition can be accumulated to a global rate at which two alternatives compete.

Within contestable market i, the degree of competition between alternatives k and j, s_{ijk}, is proportional to the joint probability of choosing i and j. The joint probability states the probability of choosing alternative j on one occasion and alternative k on another occasion (Grover and Srinivasan 1987). The justification for using a joint probability is that we expect the two alternatives with highly correlated choice probabilities to draw proportionally more from each other's markets than from alternatives with which they are uncorrelated or negatively correlated.

$$s_{ijk} \propto p_{ik} \cdot p_{jk}, \qquad (5.9)$$

While S_{jk} measures global competitive interaction between alternatives j and k, weighted by total choices made by individual i,

$$S_{jk} = \sum_{i=1}^{I} N_i s_{ijk}. \qquad (5.10)$$

Interactions of a competitive nature between alternatives can be visualized by a network structure. In order to spatialize competition, a connectivity matrix, $S,$ is created that shapes the topology of a competition network. In this network, nodes and edges represent alternatives and competitive interactions respectively. Two nodes are connected if $s_{jk} \neq 0$, and the magnitude of s_{jk} determines the weights of the edges.

This chapter continues with a discussion on modeling customers' choices. The standard logit model, which is popular in the port-choice literature, cannot capture heterogeneous preferences and treats repeated choices by the same decision-makers as independent (cross-sectional) observations. Another well-known limitation associated with the logit model is the Independence of Irrelevant Alternatives (IIA), which imposes restrictions on the relationship between choices. In the following section we briefly overview limitations of the standard logit model, and why it does not satisfy our objective.

5.2. Modeling Customer Choices

5.2.1. Standard Logit Model

Let us consider an individual n who is presented with J alternatives. Under the random utility framework the individual's utility from alternative i is $U_{ni} = V_{ni} + \varepsilon_{ni}$. The first component, V_{ni}, is deterministic, i.e. it is known to the analyst using observable variables x_{ni}. The component ε_{ni} is stochastic, not observed, and assumed to have a density of $f(\varepsilon)$. Individual n chooses the alternative that gives the highest utility. Since ε_{ni} is random, individual choice cannot be predicted with certainty, and should be estimated probabilistically. What differentiates the various discrete choice models is the distributional assumption for the random component, ε_{ni}. The Logit model assumes that ε_{ni} are independently, and identically Gumbel distributed

$$P_{ni} = \frac{\exp(\beta' x_{ni})}{\sum_j \exp(\beta' x_{nj})}$$

in which v_{ni} is a linear function of observable variables, $v_{ni} = \beta x_{ni}$.

The independence of the error terms is a very strong assumption and implies that the unobserved component of utility for one alternative is not correlated to the unobserved portion of utility for another alternative. This restrictive assumption can be violated by taste heterogeneity, substitution patterns, and repeated choices, as discussed by (Train, 2003).

a. Taste heterogeneity:

The recognition that people are different and place different values on the attribute of each alternative requires an empirical model that can handle taste heterogeneity (Train 2003). Thus, there might be fundamental differences in preferences of individuals under the same choice circumstances. While conventional logit models can capture taste variations in a deterministic way by estimating separate coefficients for mutually exclusive groups or by interacting observable characteristics, any correlation of preference with unobserved factors limits the appropriateness of the logit. For example, the shipping industry is composed of a variety of decision makers all with different tastes for transportation costs, transportation time, and quality of services. Their tastes for each of these factors will vary depending on contractual relationships, and variables such as reputation effects, or personal connections, which are not observable by the analyst.

b. Substitution pattern:

The main concern about the logit model is its assumption of independence from irrelevant alternatives (IIA), which is another consequence of the assumption of uncorrelated errors. The cross-elasticity of Logit is represented by

$$E_{iz_{nj}} = -\beta_z z_{nj} P_{nj},$$

where E_{iznj} only depends on alternative j, and alternative i does not enter the formula. It means that an improvement in the attributes of one alternative reduces the probabilities for all the other alternatives by the same percentage.

However, the pattern of substitution among alternatives has important implications for competition. For example, if one supplier improves an attribute of a product, it can expect to be interested in a prediction of the percentage of the market they can draw customers away from competitors. However, if a logit is used to model customer choice, this substitution pattern is not modeled. The behavioral weakness of this assumption is clearly seen in port competition because unobservable characteristics associated with neighboring ports are likely to be correlated. It is obvious that the probability ratio of individuals choosing between two ports depends on the availability or attributes of the other port. Also an improvement in the attribute of one port is not expected to reduce the probabilities for all other ports by the same percentage.

c. Repeated choice

Another restrictive aspect of the uncorrelated errors assumption is the case of repeated choices made by the same individual. It is reasonable to expect the random portion of utility for each individual to be correlated over different choices. Hence the logit assumption is violated if data with repeated choice is treated as purely cross-sectional data (Train 2003). If the data contain repeated choices then the estimation process should allow unobserved factors to be correlated over time.

As discussed above, modeling the choice of which port to use needs to rely on a model that allows for correlated errors and relaxes the IIA property, since the data contain repeated choices with significant unobserved taste heterogeneity.

Computational development and ease of simulation allows estimation of more realistic behavioral models using more complicated methods, which were previously unapproachable due to computational complexity (Train 2003). One established solution to the limitations of logit is to allow for variation in preferences across individuals. This entails the decomposition of the error term into two components; the first component can be correlated over alternatives and captures unobservable taste heterogeneity by allowing for random preferences, while the rest is an identically and independently distributed error. Let

$$U_{ni} = \beta_n x_{ni} + \varepsilon_{ni} \qquad (5.11)$$

The probability of individual n choosing alternative i is a weighted average of the logit formula evaluated at different values of β, with the weights given by the density f (β). The density of random preferences can be expressed either as a continuous or discrete distribution. The former leads to the mixed logit, while the latter leads to the latent class logit model that underlies this research (Train 2003).

5.2.2. Latent Class Logit

To model competition we identify consumers with homogenous preferences by observing individuals' choice histories. Therefore, we use a discrete choice model that accounts for taste heterogeneity as well as segmenting the market into segments with similar choice sets. Given the limitations of the standard logit model, we employ a latent class logit (LCL) model, which has been popular in the marketing literature. This method relaxes the IIA restriction and specifies preferences to vary across finite classes of decision makers.

Let us take an LCL model with S classes, which can have S different vectors of preferences, $\beta_1, \ldots \beta_s$. Let $P_n(i|\beta_s)$ provide the probability of choosing alternative i conditional on individual n falling into class s. The LCL probability is the unconditional choice probability that individual n chooses alternative i. It is given by the weighted average of conditional probabilities, with the weights given by the class membership probabilities (Hess, Ben-akiva, & Walker, 2011)

$$P_n(i \mid \beta_1, \ldots, \beta_S) = \sum_{s=1}^{S} \pi_{ns} P_n(i \mid \beta_s), \tag{5.12}$$

This specification can be simply generalized to situations of repeated choices by calculating the weight summation of the product of logit probabilities, one for each time period, instead of only a single logit probability:

$$L_n(j_{n1}, \ldots, j_{nT_n} \mid \beta_1 \ldots \beta_S) = \sum_{s=1}^{S} \pi_{ns} \left(\prod_{t=1}^{T_n} P_n(j_{nt} \mid \beta_s) \right). \tag{5.13}$$

Class allocation probability π_{ns} represents the prior probability of individual n being in class s. If no prior information on class allocation is available, the model assigns individuals to segments only based on their choice histories. In such case, model estimation starts with assuming that probability of being in each class is same for all individuals and equals to one divided by number of classes. However, the LCL model also allows us to link the class allocation (prior probability) to some of the alternative-invariant characteristics of individuals. A multinomial logit model can be used to relate this probability to the alternative-invariant (socio-demographic) characteristics of the decision makers (Greene & Hensher, 2003; Hess et al., 2011; Train, 2003).

$$\pi_{ns} = \frac{exp(\theta_s z_s)}{\sum_{s=1}^{S} exp(\theta_s z_s)} \tag{5.14}$$

After estimation of β_s, the probability of individual n being in class s conditional upon the observed choice history (predicted choice probabilities) is obtained as a posterior probabilities in a Bayesian fashion. Assume that j_{n1}, j_{n2},…, J_{nTn} is the choice history of individual n, represented by H_n, then

$$\pi_{s|H_n} = \frac{L_n(H_n|s)\pi_{ns}}{\sum_{s=1}^{S} L_n(H_n|s)\pi_{ns}} \qquad (5.15)$$

The choice of the number of segments, s, is important in LCL modeling. It is common to choose the optimal number of latent classes by examining information criteria such as the BIC (Bayesian Information Criterion) or the AIC (Akaike Information Criterion) (Train 2008).

The latent class logit model accommodates taste heterogeneity and leads to realistic substitution patterns by relaxing the IIA assumption, which is crucial to understand competition in this research and particularly, for simulated scenarios discussed in the next chapter.

CHAPTER 6: RESULTS

6.1. Description of Data

The freight shipment flows analyzed in this study are part of a dataset of door-to-door U.S.-bound waterborne containerized export shipments with European origins via 22 major European ports for the period of July 2006 to June 2007. Here, our focus is on shipments made by shippers that made three or more shipments decisions (and therefore three port choices) during the month of October 2006. The first column in Table 6.1 represents number of shippers, number of shipments and total TEUs associated with this data. A total of 40,965 shipments[2] meet this condition. They amounted to 81,234 TEUs and were made by 4,974 distinct shippers (last column of Table 6.1). Shippers that shipped less than three shipments (second column of Table 6.1) are excluded from our analysis because with small number of choices it is difficult to distinguish if unexpected behavior is due to a strong preference for certain attributes or it is only due to the unobservable factors. This excludes the shipments made by 25,467 shippers in October 2006. These shippers collectively accounted for 33,743 TEUs.

Not only do we exclude small shippers that ship less than 3 shipments in a month, we also exclude very large shippers that handle more than 600 TEUs (third column in Table 6.1) in a month on the ground that these shippers often have different behavior.

[2] Shipment may range from 0.01 TEUs (twenty-foot-equivalent-unit) to 78 TEUs, with an average of 6.27 TEUs and a standard deviation of 8.93.

They often use one or two ports (with which they have contract) for a large number of shipments and there is not much variation in their choice sets. Then, we selected a random sample of 4,974 among 7,968 shippers to reduce the computational complexity of the LCL model.

Table 6.2 reports the countries of origin and their aggregate level of shipments (number of shipments) included in the analysis.

Table 6.1: Descriptive statistics of different samples of data

	Population	Shippers less than 3 shipments	Shippers more than 600 TEUs	Sample included in analysis
Shippers	33,435	25,467	15	4,974
Shipments	101,269	30,636	3,842	40,965
TEUs	188,642	33,743	23,265	81,234

Table 6.2: Countries of origin included in the analysis

Countries	Number of shipments (Percentage of Total Shipments)	Countries	Number of shipments (Percentage of Total Shipments)
Italy	9,754 (23.68%)	Poland	811 (1.97%)
Germany	8,916 (21.64%)	Ireland	618 (1.50%)
United Kingdom	4,018 (9.75%)	Finland	606 (1.47%)
France	3,838 (9.32%)	Czech Republic	482 (1.17%)
Spain	2,781 (6.75%)	Portugal	371 (0.90%)
Netherlands	1,881 (4.57%)	Norway	292 (0.71%)
Belgium	1,457 (3.54%)	Greece	260 (0.63%)
Sweden	1,099 (2.67%)	Russia	217 (0.53%)
Denmark	1,010 (2.45%)	Hungary	190 (0.46%)
Switzerland	903 (2.19%)	Slovakia	175 (0.42%)
Austria	823 (2.00%)	Slovenia	116 (0.28%)

Table 6.3 reports the types of commodities and their aggregated level of shipments (in TEUs). For the purpose of examining the commodity composition of freight traffic of data, we collapsed classes of the Commodity Description and Coding System (United States International Trade Commission, 2007), termed “HS codes” hereafter, into seventeen categories. Commodity categories are comprised of groups of similar commodities. As indicated by Table 6.3, U.S. imports from Europe are dominated

by Prepared Foods (18.08%), Machinery and Mechanical Appliances (13.24%), Chemical Products (9.23%), Base Metals and Articles Thereof (8.60%).

Table 6.3: Commodity categories included in the Analysis

Commodity Category	Number of shipments (Percentage of Total Shipments)	Commodity Category	Number of shipments (Percentage of Total Shipments)
Prepared Foodstuffs	7356 (18.08%)	General Merchandise	1409 (3.46%)
Machinery & Mechanical Appliances	5388 (13.24%)	Wood Pulp Products	1328 (3.26%)
Articles Of Stone, Plaster, Cement, Asbestos	4557 (11.20%)	Textiles & Textile Articles	1049 (2.58%)
Chemical Products	3757 (9.23%)	Mineral Products	991 (2.44%)
Base Metals & Articles Thereof	3500 (8.60%)	Wood & Wood Products	836 (2.05%)
Handicrafts, Pearls, semi/Precious stones, metals	2809 (6.90%)	Instruments - Measuring, Musical, arms & ammunitions	463 (1.14%)
Plastics & Rubber	2574 (6.33%)	Footwear, Headgear	176 (0.43%)
Transportation Equipment	2192 (5.39%)	Hides & Skins	149 (0.37%)
Vegetables, Animals & their Products	2150 (5.28%)		

Table 6.3 reports various characteristics of the 22 ports used in the analysis. The first column reports each port's count and share of shipments during the sample period of one month. The statistics reported in Table 6.4 identify Bremerhaven as the most frequently selected port (23% of total decisions include Bremerhaven as the final port choice), followed by Rotterdam (12.5%), Antwerp (11.7%), La Spezia (9.6 %), Le Havre (5.8%), Genoa (5.5%), and Hamburg (5.5%). No other port is selected in more than 5% of total choices. Columns 2 and 3 report the basic composition of the market in terms of

captive (loyal) and contestable (switching) customers. Among shippers in the sample, 2,721 were captive, as they consistently shipped via the same port. The remaining shippers (2,253) ship via multiple ports and contribute to the creation of contestable hinterlands. Table 6.4 shows that Valencia, Lisbon, Piraeus, Barcelona, Gioia Tauro, Liverpool, Naples are the ports with the largest share of captive consumers; this can be associated with their remote geographical location within the European space, as measured by the remoteness index[3].

Furthermore, the large ports of Bremerhaven and Le Havre, which are surrounded by competing ports, also have markets that are largely captive. This may reflect the contractual relationships between ports and shippers that are unobservable in data. Customers of the large ports of Leghorn, Hamburg, Genoa, Rotterdam, La Spezia, on the other hand, switch between ports frequently.

Relative port efficiency is measured by a port efficiency index calculated using Stochastic Frontier Analysis based on information on port throughput and port infrastructure and equipment (As discussed in Chapter 6). The Oceanic connectivity of a certain port (to the U.S.) is measured by the number of vessels providing service between the European port and all U.S. ports. We use the 2008 Containerization Yearbook for measuring port efficiency and oceanic connectivity. The land connectivity of ports is represented by the total length of highways, rail and highways in a 100 km buffer around each port, collected from Europea Technologies' Global Insight.

---

[3] Remoteness index defined as the weighted distance to all other ports, where weight is size of ports (Brun et al. 2005).

$$\mathrm{R_i} = \sum_{j}^{22} \mathrm{w_j}\, \mathrm{d_{ij}} \text{ where } \mathrm{w_j} = \frac{\mathrm{s\,j}}{\sum_{\mathrm{j}} \mathrm{s_j}}$$

Table 6.4: Shares of contestable and captive consumers for each port

Ports	Number of shipments (Percentage of Total Decisions)	Captive consumers (Percentage of Total Shippers through the Port)	Contestable consumers (Percentage of Total Shippers through the Port)	Remoteness Index
BREMERHAVEN	9981 (23.1%)	4503 (0.45)	5478 (0.55)	381
ROTTERDAM	5411 (12.5%)	1489 (0.27)	3922 (0.73)	323
ANTWERP	5075 (11.7%)	1788 (0.35)	3287 (0.65)	319
LA SPEZIA	4155 (9.6%)	1168 (0.28)	2987 (0.72)	524
LE HAVRE	2501 (5.8%)	1354 (0.54)	1147 (0.46)	422
GENOA	2382 (5.5%)	649 (0.27)	1733 (0.73)	505
HAMBURG	2361 (5.5%)	478 (0.20)	1883 (0.80)	413
LEGHORN	1949 (4.5%)	308 (0.16)	1641 (0.84)	558
FELIXSTOWE	1725 (4.0%)	705 (0.41)	1020 (0.60)	388
VALENCIA	1261 (2.9%)	815 (0.65)	446 (0.35)	812
BARCELONA	1056 (2.4%)	516 (0.49)	540 (0.51)	672
LIVERPOOL	1045 (2.4%)	492 (0.47)	553 (0.53)	551
NAPLES	750 (1.7%)	349 (0.46)	401 (0.54)	778
SOUTHAMPTON	729 (1.7%)	291 (0.40)	438 (0.60)	460
ALGECIRAS	639 (1.5%)	217 (0.34)	422 (0.66)	1113
GIOIA TAURO	638 (1.5%)	308 (0.48)	330 (0.52)	946
FOS	425 (1.0%)	144 (0.34)	281 (0.66)	549
GOTHENBURG	370 (0.9%)	122 (0.33)	248 (0.67)	665
TILBURY	299 (0.7%)	47 (0.16)	252 (0.84)	410
PIRAEUS	216 (0.5%)	120 (0.56)	96 (0.44)	1218
LISBON	210 (0.5%)	130 (0.62)	80 (0.38)	1088
SINES	47 (0.1%)	11 (0.23)	36 (0.77)	1115

6.2. Model Specification

In the previous chapter we discussed the necessity of employing a random coefficient logit model, with discrete mixing distribution, instead of a standard logit, although the latter is more popular in the shipping industry literature. To the best of our knowledge this is the first time this model is applied to port choice behavior and inter-port competition. The dataset covers shippers that make at least three choices over study period. Our empirical model uses these choices, along with data describing each shipper and port, to estimate the preferences for port characteristics and simultaneously classify shippers based on their preferences. This section starts with model specifications and continues with the modeling results.

Our empirical model uses repeated choices made by shippers, along with data on each shipper and ports attributes, to estimate the preferences for port characteristics and to segment the market into classes with similar preferences. The LCL performs behavioral market segmentation based on preferences for distance and port characteristics.

The model specification includes several variables to capture the importance of proximity and transport costs. This includes the distance from the production origin to the various ports (*distance*); an indicator that takes a value of one if the port is the closest port from the shipment origin, and zero otherwise (*home port*); an indicator that takes a value of one if a border is crossed and zero otherwise (*bordercross*). We also include the oceanic distance to the U.S. to test if the distance to U.S ports affects the choice of port (*oceanic distance*). In addition to location, other attributes that describe characteristics of ports are tested in the model, namely efficiency (*efficiency*), land connectivity (*landaccess*), and oceanic connectivity (*connectivity*)[4]. We also include interactions between distance and efficiency to test the effect of port efficiency on shippers' valuation of travel cost/time.

[4] Also, our setup allows for including price differentials in terms of transportation services, ports and carriers. But we do not have access to price data.

To control for the observed heterogeneity across shippers, the LCL model endogenously estimates different preferences for segments of homogenous shippers. The LCL model accepts alternative-invariant variables as a priori probability for class allocation. In this study, the size of the shipper (*shipper size*) as measured by the total TEUs recorded in the dataset is used for this purpose on the ground that large and small shippers have different choice behaviors (Brooks, 1995; Slack 1985; Cahoon & Notteboom, 2008;). Slack's statistical results of interviews with Northern American/ European shippers and forwarders show the relative importance of cost and service considerations in relation to the size of the companies. Price is more important for the smaller shippers, while larger shippers put more weight on service considerations. Similarly, Brooks find that cost is more important for small shippers, while large shippers and forwarders base their choice on service quality (Slack 1985). Also, other survey-based studies find that market is certainly not homogenous in its requirements for port choice since they are evaluated differently by different consumer groups. De Langen represents differences between shippers and freight forwarders and explains the price-elastic behavior of forwarders by the fact that one of their capabilities is to purchase transport services cheaply for large volumes of cargo, while transport costs are only a fraction of overall shippers' costs that it even may pass to their customers (De Langen 2007). Also, Cahoon and Notteboom differentiate choice behavior between shippers, forwarders and shipping lines.

Figure 6.1 shows the distribution of shippers based on total TEUs reported in the data. The distribution is extremely skewed towards smaller shippers.

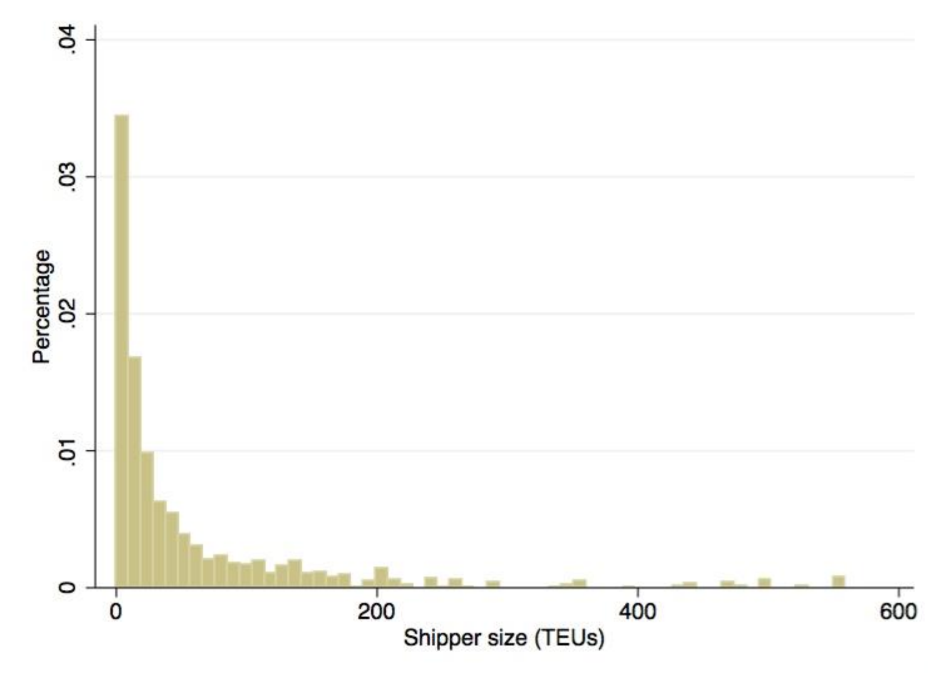


Figure 6.1: Distribution of shippers by TEUs in October 2006.

The model has been estimated for different numbers of classes. The BIC (Bayesian Information Criterion) and AIC (Akaike Information Criterion) values determine that five is the proper number of classes.

Table 6.5 reports the estimates of the LCL model for five segments[5] of shippers. The coefficient estimates differ significantly across the five segments, while within each segment preferences are relatively homogenous.

[5] This research uses the terms ‘class’ and ‘segment’ interchangeably.

Table 6.5: LCL segmentation, parameter estimates and standard errors (in parentheses). All parameters are statistically significant at 5%. The stared values are insignificant.

	Conditional logit	Latent Class Logit				
		Segment 1	Segment 2	Segment 3	Segment 4	Segment 5
Distance	-0.008 (0.0001)	-0.015 (0.001)	-0.014 (0.0005)	-0.004 (0.0003)	-0.013 (0.002)	-0.031 (0.001)
Homeport	0.384 (0.015)	-0.802 (0.056)	-1.806 (0.101)	-0.750 (0.47)	0.185 (0.87)	1.425 (0.051)
Bordercross	-1.549 (0.018)	-2.779 (0.168)	-6.882 (0.149)	-2.704 (0.053)	1.138 (0.075)	-1.293 (0.10)
Distance to US	-0.001 (0.000)	-0.002 (0.0002)	-0.001 (0.0002)	-0.001 (0.0001)	-0.001 (0.0002)	-0.005 (0.0004)
Efficiency	0.340 (0.051)	-0.512 (0.402)	6.477 (0.226)	-6.046 (0.176)	-0.260 (0.358)	0.616 (0.188)
Distance X efficiency	0.003 (0.0002)	0.023 (0.001)	-0.016 (0.0008)	0.008 (0.0004)	0.032 (0.002)	0.021 (0.0007)
Connectivity	0.027 (0.0002)	0.001 (0.0009)	0.107 (0.002)	0.034 (0.001)	0.028 (0.001)	0.057 (0.001)
Land access	0.158 (0.023)	1.203 (0.141)	-5.192 (0.167)	1.204 (0.066)	-1.389 (0.095)	2.904 (0.168)
Shipper size		-0.004 (0.002)	-0.006 (0.002)	0.006 (0.005)	0.008 (0.001)	0.000
Constant		-0.307 (0.066)	0.038 (0.050)	-0.681 (0.65)	-1.286 (0.071)	0.000
Class share		0.205	0.282	0.140	0.084	0.288
LL	-54948	-38311				
Observations	901,230	901,230				
Shippers		4,974				

6.3. Empirical Results

Looking at the model fit, the Latent Class Logit model is found to easily outperform the conditional logit model (log-likelihood (LL) of -38311 and -54948, respectively) thanks to the 42 additional parameters. The conditional logit model is estimated using 8 parameters associated to the variables listed in Table 6.5. On the other hand, the LCL is estimated by 8 parameters in each of the five segments identified in the final specification (8x5); also membership functions include 2 parameters in each segment (2x5). Table 6.6 compares prediction accuracy of the conditional logit model with LCL at the port level. Overall, the LCL predicts selected ports correctly at the rate of 70%, which is much higher than the conditional logit (54%).

Table 6.6: Prediction accuracy of conditional logit and LCL

Ports	Prediction Accuracy Conditional logit	Prediction Accuracy LCL	Ports	Prediction Accuracy Conditional logit	Prediction Accuracy LCL
BREMERHAVEN	90%	81%	LIVERPOOL	80%	87%
ROTTERDAM	39%	63%	NAPLES	84%	87%
ANTWERP	29%	70%	SOUTHAMPTON	19%	23%
LA SPEZIA	58%	80%	ALGECIRAS	29%	53%
LE HAVRE	69%	89%	GIOIA TAURO	8%	30%
GENOA	46%	60%	FOS	71%	70%
HAMBURG	0%	43%	GOTHENBURG	74%	60%
LEGHORN	32%	50%	TILBURY	0%	21%
FELIXSTOWE	35%	68%	PIRAEUS	94%	88%
VALENCIA	86%	87%	LISBON	32%	74%
BARCELONA	45%	77%	SINES	6%	45%
Total	**54%**	**70%**			

More detailed discussion of the parameter estimation results follows.

Proximity

To capture the importance of proximity and transportation costs on port choice, the specification includes distance from the origin of the shipments to each port (measured in miles). Given our prior expectations, preference for proximity would decrease as port efficiency increases. Hence, we include the interaction of distance with port efficiency in the model. Each efficiency value is subtracted from the sample mean before inclusion in the interaction term. Thus the interaction coefficient measures the impact of distance to a port with an average efficiency value.

Shippers in each group place a negative value on distance, but the effect is different for each segment. Each additional mile of distance reduces the odds of choosing a port by approximately 0.4% for segment 3 and approximately 3.05% for segment 5. Except for segment 2 (smallest shippers), other segments evaluate distance more positively for reaching more efficient ports. In other words, shippers differentiate between equidistant ports on the basis of the level of efficiency, with an inefficient port treated as if it were farther than an efficient port.

The variable of the closest port (home port) tests the impact of port's geographic monopoly over its hinterland, which used to be the most important factor of port choice. However, this variable may pick the non-linearity of distance that reflects added preferences that shippers have for closeness of ports due to the travel costs and time. The preference for the homeport is higher for larger shippers and lower for smaller shippers. This may reflect the influence of preferences that are not measured here such as discounts, reputation and peer effects.

All segments of shippers place negative weight on oceanic distance, but always smaller than the weight on land distance. This makes sense regarding the lower rate of overseas shipping compared to land transportation (Malchow & Kanafani, 2004).

Border

As discussed, there is a large body of literature that discusses the friction caused by borders and the puzzle of border effects that remain in integrated and free trade regimes, such as the European Union or NAFTA. We include an indicator if a border must be crossed to reach a port. The results in Figure 6.5 indicate that shippers often perceive a border crossing as very negative, no matter what segment they belong to, but that the impact of a border is smallest (positive) for larger shippers (segment 4) and largest for smaller shippers (segment 2). For a shipment with average value, a border is a strong barrier for the smallest group but its effect decreases with the value of shipment.

It is worth noting that, except for countries whose ports are involved in our analysis, the model considers the rest of the countries as landlocked, because they have no choice but to cross a border to reach any of these 22 ports. Shippers in these countries have no effect on the estimation of the border effect, since this variable is always equal to one for shipments originating in these countries.

Our model estimates that shippers have a strong aversion to crossing a border which most likely reflects the strong effect of cultural, language, and information barriers. In the presence of such barriers, one can argue that what is measured by distance, domestic or international, is also affected by these barriers. In other words, the impacts of common polity, common language, and common culture, or imperfect information, which are missing from existing models of trade, largely affect borders and distance estimates. In fact, while distance measures continuous effects of familiarity and information decay, borders measure any discontinuity in these factors.

Ports attractiveness

In addition to location, other attributes that describe characteristics of ports are efficiency, land connectivity, and oceanic connectivity. The conditional logit model provides an average estimate of preferences that shippers put on port attractiveness. The first column of Table 6.5 shows the positive preferences for efficiency, accessibility and connectivity.

Also, the LCL estimates positive preference for connectivity in each segment. Except for segments 2 and 4 that value land accessibility negatively, the rest of the segments have high preference for higher accessibility. The loyalty of segment 2 could justify the negative valuation of accessibility, but it is harder to justify it for segment 4.

We discussed efficiency earlier in relation to its interaction with distance. The main efficiency variable (without interaction with distance) is insignificant for two segments. Segment 2, which is the only segment that are not willing to travel further to reach more efficient ports, places positive value on the main efficiency variable.

Shipper size

As discussed in Chapter 5, the LCL model accepts alternative-invariant variables for class allocation, as a priori probability, to relate unobservable heterogeneity to the characteristics of the decision makers. In this study, the size of the shipper is introduced for class allocation. Indicated by coefficient of *shipper size* in Table 6.5, segment 4, contains larger shippers. They are the only shippers that cross border easily, which leads to most intense competition. The average size of shippers decreases for segments 3, 5, 1, and 2 respectively. Segment 2, the most captive segment, consists of the smallest shippers.

6.3.1. Market Segmentation

The spatial distribution of customers belonging in each segment is represented by small dots in the maps of Figure 6.5. It is worth mentioning that each dot may be representative of multiple shippers with the same spatial and behavioral characteristics. Also shippers with the same spatial characteristics may fall in different segments if they are behaviorally different. Table 6.7 reports the descriptive statistics for each segment to better characterize each group in terms of contestability/captivity, shipper size, commodity distribution, and port market shares.

Table 6.7: Descriptive statistics of segments

	Segment 1	Segment 2	Segment 3	Segment 4	Segment 5
Avg shipper size std	16 32	13 26	18 47	25 45	17 31
Avg value std	98,077 481,576	88,209 200,811	94,658 224,227	132,709 335,935	104,775 365,713
Avg TEUs std	2.85 9.69	1.76 2.23	2.21 3.55	2.56 4.64	2.08 3.20
Contestability: # of shippers (%)	669 (79%)	368 (23%)	446 (69%)	241 (62%)	529 (36%)
Captivity: # of shippers (%)	188 (21%)	1261 (77%)	198 (31%)	150 (38%)	924 (64%)
Avg shippers choice set diversity (HHI)	0.65	0.90	0.68	0.72	0.86
Tradeoff Between distance and border	119	551	606	-87	43
Avg distance	195	251	345	287	142
Prob crossing border	0.19	0.28	0.47	0.72	0.31

Segments consist of different shares of customers as indicated by *class share* in Table 6.5; while segments 5, 2 and 1 have shares of 29%, 28%, and 21% respectively, segments 3 and 4 are composed of smaller numbers of shippers (14% and 8%, respectively). Shippers in Segment 2 are typically smaller than shippers in other segments; it is the only segment that is not willing to ship further to reach more efficient ports. They strongly avoid borders and as the trade-off value in Table 6.7 shows they are willing to travel approximately 551 more miles to avoid a border. This segment forms the most captive markets, where about 77% of shippers (Table 6.7) ship via one single port. Rows 4, 5 and 6 of Table 6.7 reports on the degree of contestability/captivity of segments as explained by equations 5.5 to 5.8. Row 4 reports the number of shippers that always ship via one single port, while row 5 describes the number of shippers that switch between multiple ports. Row 6 reports the average diversity in shipper's choice sets by means of Herfindhal Index.

After segment 1, segment 5 constitutes the second largest share of captive markets. This segment values distance the most negatively and has largest preference for the home port (closest port). As implied by the trade-off value in Table 6.7, these shippers leave the home country relatively easily to reach their closest port, located in another country.

Segments 1, 3, and 4 form the most contestable markets, with 62% to 79% of shippers switching between ports. Segment 4 is the least reluctant to crossing borders and segment 3 is the least reluctant to traversing distance. The insensitivity to distance can partly be due to shipping via transshipment hubs (e.g. Algeciras and Gioio Tauro) and using barge corridors to reach them. Shipper size coefficient in Table 6.5 shows that segments 3 and 4 are usually larger shippers. These segments have diverse choice sets that reflect underlying preferences of the segments and intensify competition. While choice sets of segments 1 and 3 are composed of domestic ports and raise domestic competition, shippers in segment 4 lead to international competition by choosing ports outside their home country.

We are interested to know if the revealed behavior of shippers is affected by the types of commodity they ship. Figure 6.2 depicts the commodity profile of each segment. Overall, commodity profiles do not exhibit very discriminating patterns. This emphasizes again on the commodity specialization decay with unitization (containerization) of cargo.

However, some noticeable pattern is that *Prepared Foodstuffs* dominate shipments in segment 3 and *Chemical Products* followed by *Prepared Foodstuffs* dominate shipments in segment 4. We should notice that the prevailing share of these two categories does not imply that they are shipped less frequently in other segments. In other words, the smaller size of segments 3 and 4 (14% and 8% of class share) gave rise to the share of these two commodities while number of shipments is not necessarily more than other segments. However, it is still important to know why these two categories are significantly dominant in segments 3 and 4. It may therefore suggest that the switching behavior of these two segments results from the high perishability of products shipped, which have to be shipped in a timely manner and cannot just sit and wait for a better time to ship. Commodity profiles of segments 1, 2 and 5 are composed of *Machinery and Mechanical Appliances* and *Articles of Stone, Plaster, Cement, and Asbestos,* which are relatively less time sensitive, in addition to *Prepared Foodstuffs*. Therefore we can conclude that time-sensitive commodities could increase competitive interaction between ports. Because shippers may need to consider ports over a large geographic area and cross borders to be able to ship them on a timely manner.

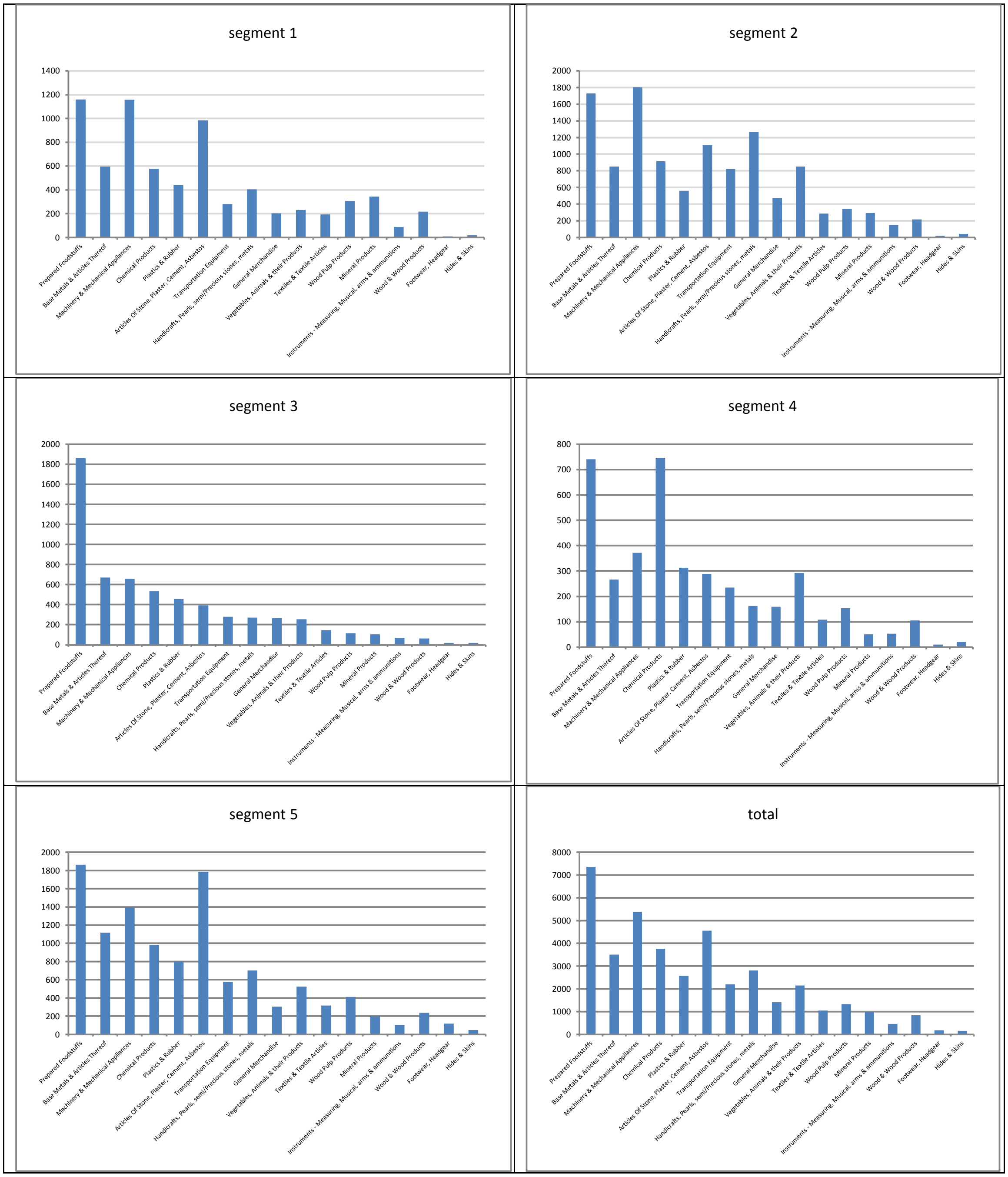


Figure 6.2: Commodity composition of segments. Vertical axis represents the number of shipments

6.3.2. Competitive Interactions

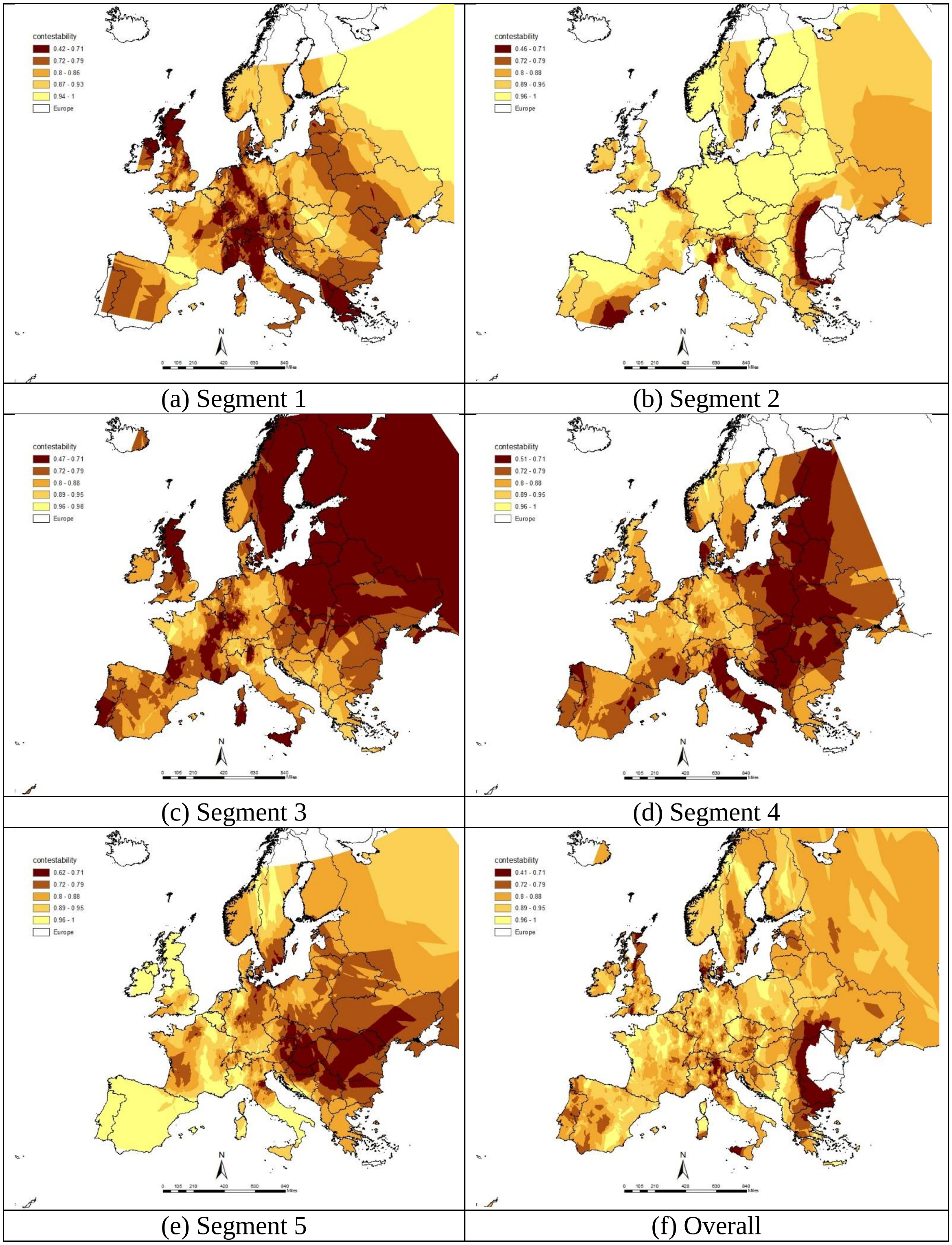
Examining characteristics of customers across segments reveals interesting properties of the model in that it can segment the customers into contestable and captive markets. In other words, the uniqueness of this research is its ability to measure disaggregated contestability at the customer-level. Figures 6.3 and 6.4 demonstrate the contestability of local markets and its variability across the geographic space by calculating ports diversity in customer choice set (HHI) as discussed in chapter 5. Figure 6.3 depicts contestability at shipment origins measured on the basis of the observed behavior of shippers. On the other hand, Figure 6.4 shows the same concepts, but based on predicted behavior modeled by the LCL.

Figures 6.3 and 6.4 average contestability values when multiple businesses are located at the same origin. In order to create a continuous contestability surface, a kriging interpolation technique is employed. More contestable areas, which are associated with a smaller Herfindhal index, are represented in darker colors. As the Herfindal index increases, areas become more captive and lighter on the map. It should be pointed out that interpolated contestability values cannot be calculated for some peripheral regions of Europe in because no data are observed (modeled) in these areas. For the sake of comparison, the same range of classification is applied to individual segments and to all shipments.

Figures 6.3 and 6.4 show that the segments of customers whose choices are restricted to a single supplier constitute captive hinterlands. Conversely, the preference structure of other customers may be conductive to using the services of multiple firms. Customers that value distance very negatively and strongly avoid borders are tied to one single choice and constitute captive markets (esp. segments 2, 5). However, customers whose demand is inelastic to distance or cross the border readily have diverse choice sets (segments 1, 3, 4). Figure 6.3f depicts overall regional contestability, irrespective of segments of customers.

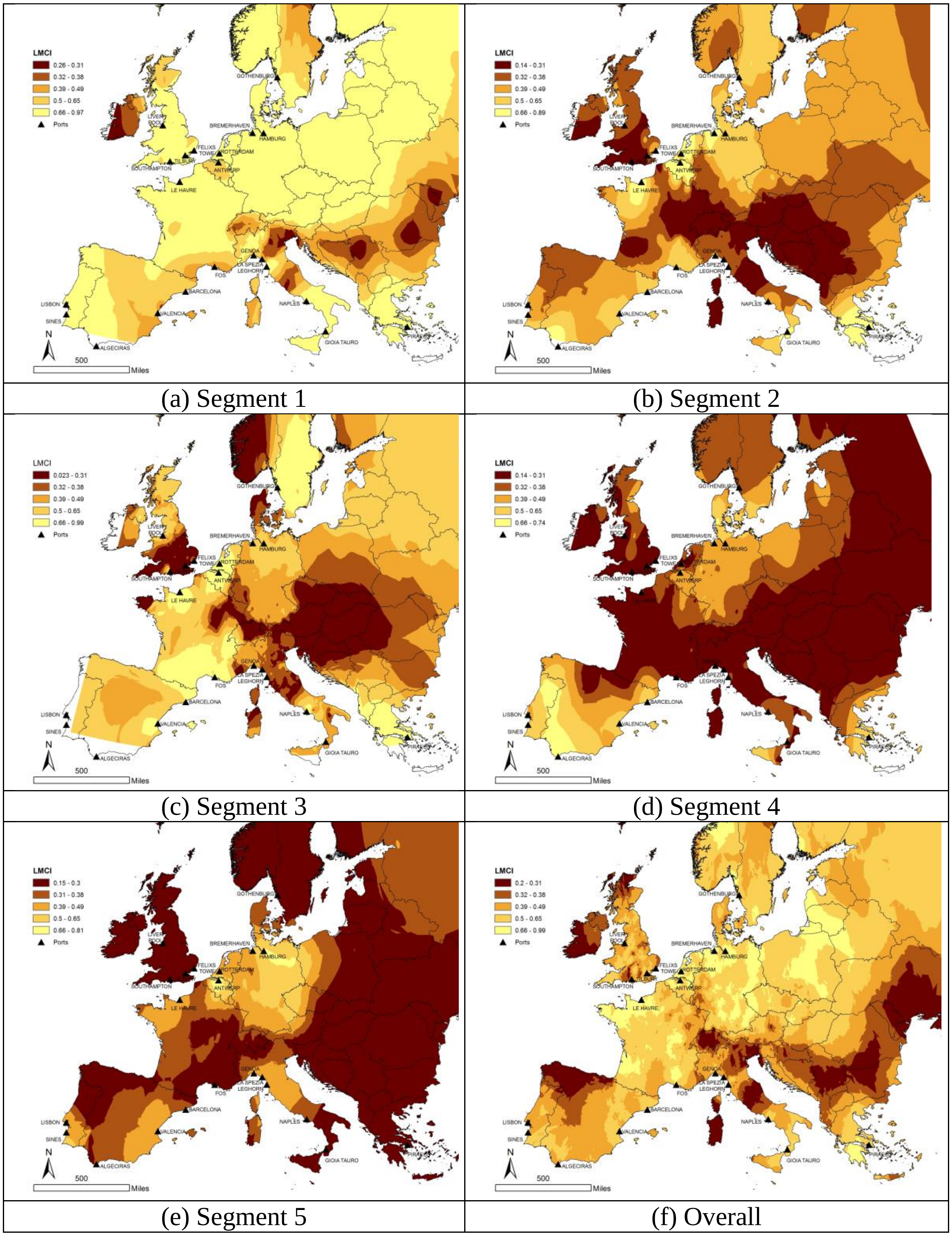
The general trend in Figure 6.3 is that British and Italian shippers are highly contestable due to the availability of a large number of domestic ports. Also, a large degree of contestability is seen in landlocked countries. However, it is worth noting that large contestable areas in Eastern Europe are interpolated based on very few, but contestable (switching) customers.

However, Figure 6.4 (based on modeled choices) presents a much wider range of contestability, between 0 to 1 (maps legend), compared to Figure 6.3 (based on observed choices) that ranges between 0.4 to 1. Furthermore, the surface generated based on modeled choice sets is very smooth compared to the sharp patterns that are seen in the surface of observed choice sets. This is due to the fact that our observed data is only a snapshot of the reality that captures limited variation happening during a specific time period. Hence, deriving feasible choice sets from available data may underestimate the variability in substitution alternatives. On the other hand, modeling competitive interactions will account for possible choices that may not be captured in the timeframe of the dataset.



(a) Segment 1 | (b) Segment 2
(c) Segment 3 | (d) Segment 4
(e) Segment 5 | (f) Overall

Figure 6.3: Observed choice set and disaggregated contestability



(a) Segment 1 | (b) Segment 2

(c) Segment 3 | (d) Segment 4

(e) Segment 5 | (f) Overall

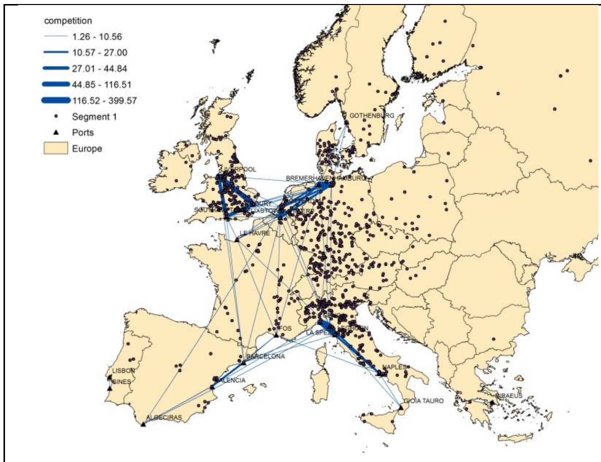
Figure 6.4: Modeled choice set and disaggregated contestability

After partitioning customers into different segments, with different degrees of contestability, we examine the structure of competitive interaction between firms. As specified in Chapter 5 competition intensity between firms is calculated by the joint probability that customer i chooses firm j on one occasion and firm k on another occasion. Networks in Figure 6.5 depict competition between ports, where the edges depict the intensity of competition measured on the basis of the observed behavior of mutual customers of ports.

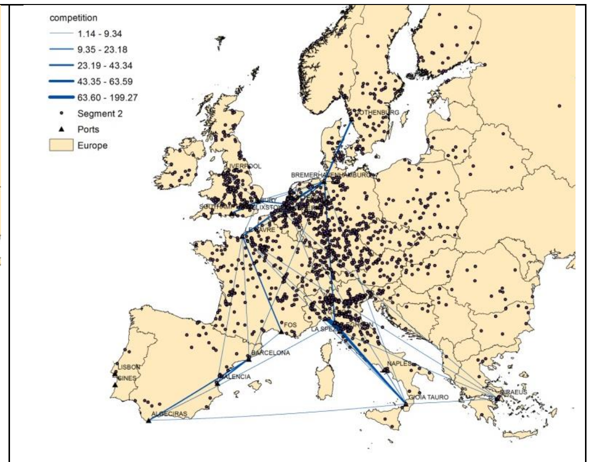
Figure 6.6 shows the same concepts, but based on predicted behavior of mutual customers of ports modeled by the LCL. Besides Table 6.6, comparing Figures 6.5 and 6.6 on a segment basis shows that our model reproduces competitive interactions pretty well. The structure of networks in Figures 6.5 and 6.6 reveals hierarchical clustering between ports. At first glance, we can see two main clusters (communities): North European ports and Mediterranean ports, where the intensity of interactions is high within each community, but relatively lower between the two communities. Furthermore, each community is composed of smaller communities. Clusters of British and Le Havre-Hamburg range ports are identifiable in North European ports. Also, interactions within Italian ports are much more severe than between Italian and Spanish ports.

Similar to the discrepancy discussed between Figures 6.3 and 6.4, Figure 6.5 (based on modeled choices) presents a denser network, implying a higher degree of competition, compared to Figure 6.6 (based on observed choices). It is again related to larger variations in modeled choice sets that are not limited to the observed choices.

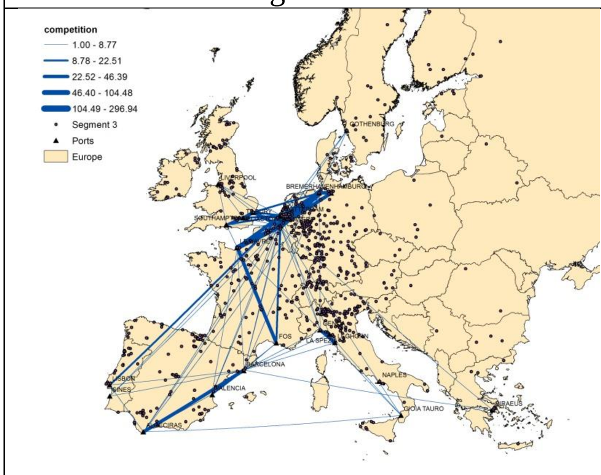
Segment 1 creates the most intense competitive interaction (in terms of size) but often between neighboring ports. Segments 3 and 4 are the most competitive in terms of diversity of choice sets reflected by the length of the edges. We can see that edges are relatively longer in segments 3 and 4. The more distant two competing ports are located, the longer is the edges. According to the results in Table 6.5, the most inelastic to the distance (i.e. segment 3) and most open to crossing a border (i.e. segment 4) have the lengthiest networks. Conversely, the more elastic to the distance a segment is (i.e segments 2, 5), the shorter is the length of the edges.



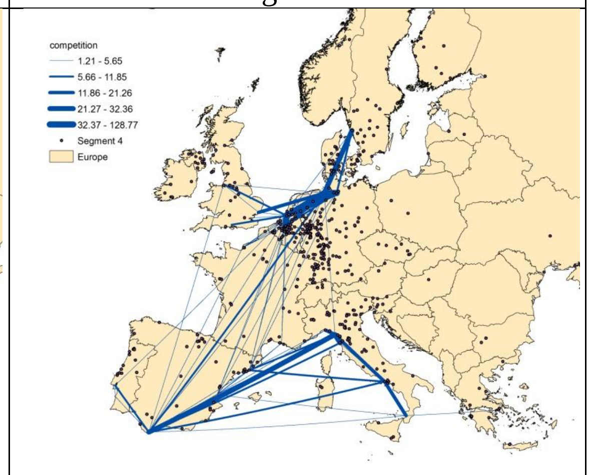
Segment 1



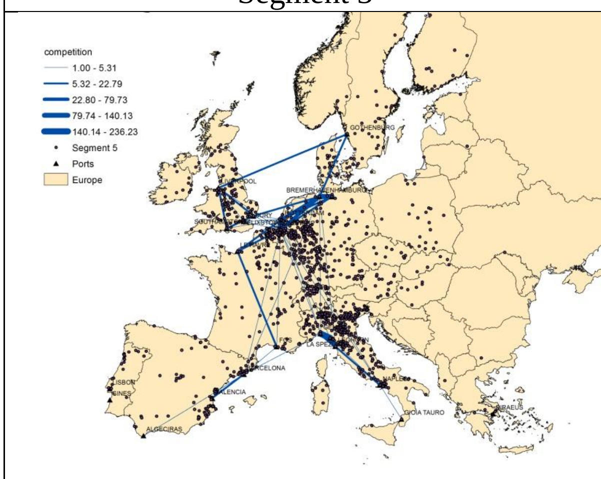
Segment 2



Segment 3



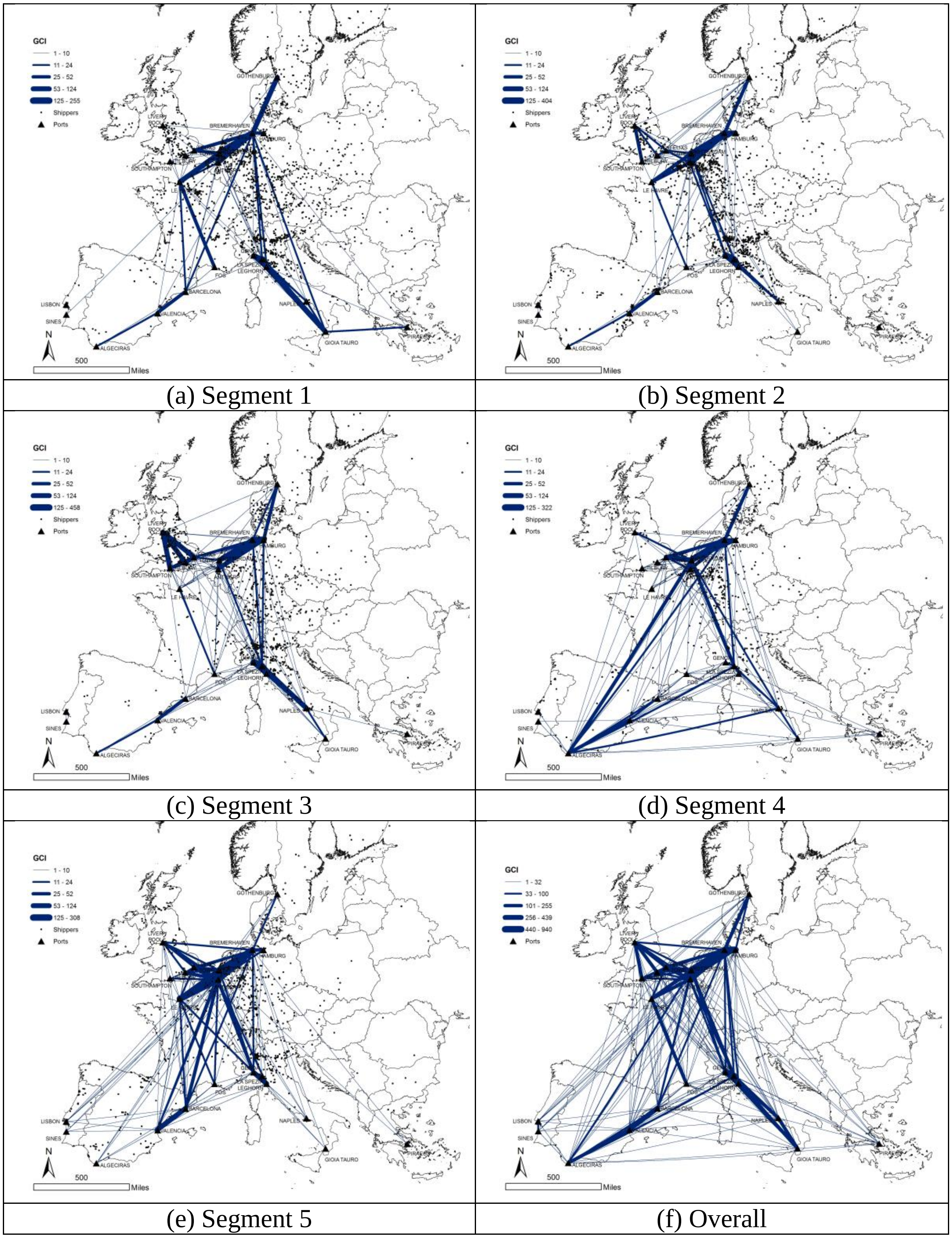
Segment 4



Segment 5

Overall

Figure 6.5: Observed choice sets and competition



(a) Segment 1 | (b) Segment 2

(c) Segment 3 | (d) Segment 4

(e) Segment 5 | (f) Overall

Figure 6.6: Modeled choice sets and competition

6.4. Simulations

So far in the discussion of results we have focused primarily on model estimates, segment characterization and competitive forces on the basis of observed port attributes. However, the proposed framework built on the LCL model that relaxes the IIA assumption allows us also to model substitution patterns in response to different policy changes. Several scenarios are presented hereunder that clearly show how structure of competition responds to changes to the underlying forces.

6.4.1. Border Dissolution

As noted earlier, national borders are an important focus of empirical studies of international trade patterns. This study shows that within a free-trade zone like the European Union border effects still exist. Although there is no consensus on its underlying causes, it could be associated to several factors such as language, business culture, or informational barriers. With the first scenario we are interested in teasing out what would happen if borders (and its underlying causes) were completely removed. In order to shed some light on this question, we change the border variable to zero, simulate the responses for all shippers, and aggregate these responses (market share) for each port over all segments. Then, the changes in share associated by the border dissolution are represented in Figure 6.7 on a port-by-port basis.

Le Havre and the Italian ports of La Spezia, Leghorn and Genoa as well as the port of Felixstowe that are all relatively large ports, would lose a large number of customers if borders were removed. Conversely, this assumption works to the advantage of the ports of Rotterdam, Antwerp, Barcelona, Bremerhaven, and Southampton. The difference in responses to this assumption is associated to variations of border effects across countries.

The LCL model in Table 6.5 includes one border variable that accounts for any national border that is crossed. However, in order to explain variations in changes to port market shares in Figure 6.7, we are interested to know how border effects vary across countries. To test if the preference for border crossing varies over countries, indicator variables for the top seven countries of origin, which are Italy (24%), Germany (22%), United Kingdom (10%), France (9%), Spain (7%), Netherlands (5%), and Belgium (3%), are interacted with bordercross in Table 6.8. In Table 6.8 a conditional logit model is estimated to reveal average effects of country-specific borders. For instance, *Italy border* equals one if an Italian shipper must cross the Italian border to reach a port, zero otherwise. The main bordercross variable is removed; finally the remaining countries, such as Portugal, Sweden, and Greece comprise the “omitted category”.

Table 6.8 indicates that depending on which country a shipment originates from, the border effect differs: Spanish shippers are the least reluctant to cross their own border and ship via a forwarding port in a third country towards the United States, while the Italians are the most reluctant. This is confirmed by the LCL results reported in Figure 6.7, as it is shown that if the strong barrier reflected by the Italian and French borders (language, cultural, etc barriers) were removed, French, Italian and British ports would become exposed to competitive forces from wider regions; as a result, many French and Italian shippers would choose to ship through ports like Barcelona, Antwerp, and Rotterdam.

Table 6.8: Average border effects over countries of origins. The value in parentheses represents standard errors. All values are statistically significant at 1% (except Spain border, which is significant at 10%)

Control variables		Border variables	
Distance	-0.008 (0.0001)	Italy border	-3.097 (0.676)
Homeport	0.452 (0.151)	France border	-2.617 (0.043)
Distance to U.S.	-0.001 (0.0005)	UK border	-2.322 (0.054)
Distance * efficiency	0.003 (0.0002)	Belgium border	-1.318 (0.065)
Efficiency	0.677 (0.524)	Germany border	-1.156 (0.025)
Land access	0.152 (0.026)	Netherlands border	-0.680 (0.054)
Oceanic connectivity	0.029 (0.0002)	Spain border	0.127* (0.067)

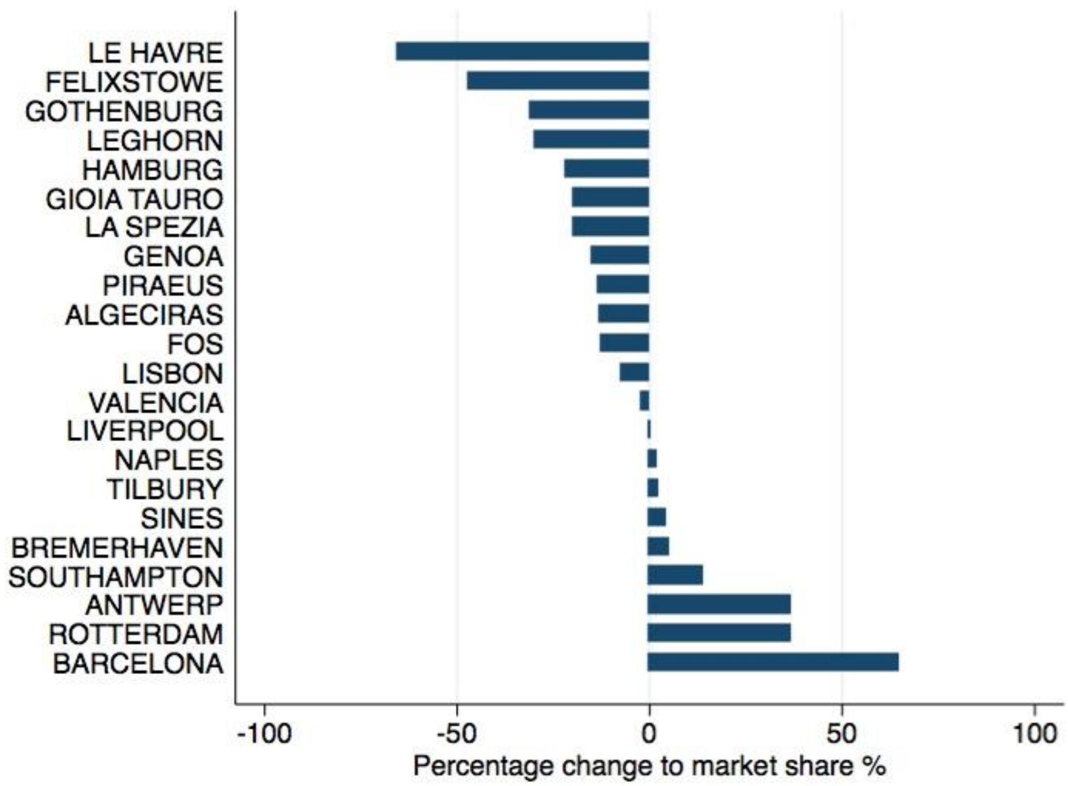
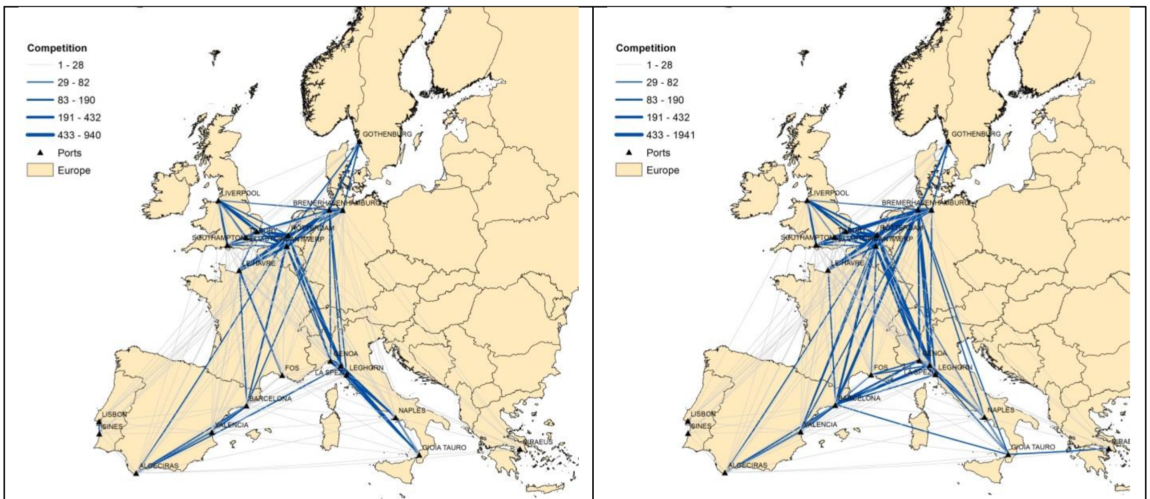


Figure 6.7: Percentage changes in port market shares after border dissolution

Figure 6.8 represents the underlying interactions that lead to aggregated market share in Figure 6.7. A first glance reveals that removing borders leads to a denser and more connected network. Barcelona gains a high degree of connectivity in this network by extending its hinterland to Italian ports. Valencia also becomes connected to the North

European ports by serving large share of customers from land-locked countries. While the thickness of edges decreases between Italian ports, it increases between Italian ports and Non-Italian ports (Spanish and Northern ports). The port of Le Havre that loses the largest share of customers loses its competitive connection with the ports of Fos, Tilbury, and Antwerp, as many French consumers that previously switched between Le Havre and one of these ports, now chooses not to ship via Le Havre at all once the border is removed. However, Felixstowe, which is the second port most negatively affected by border dissolution, experiences a different pattern. Felixstowe becomes part of North-western community as the British shippers, who used to ship only via Felixstowe, start to change to non-domestic ports.



Original competition network	Competition network after border dissolution

Figure 6.8: Competitive interactions with and without borders

6.4.2. Improving Efficiency

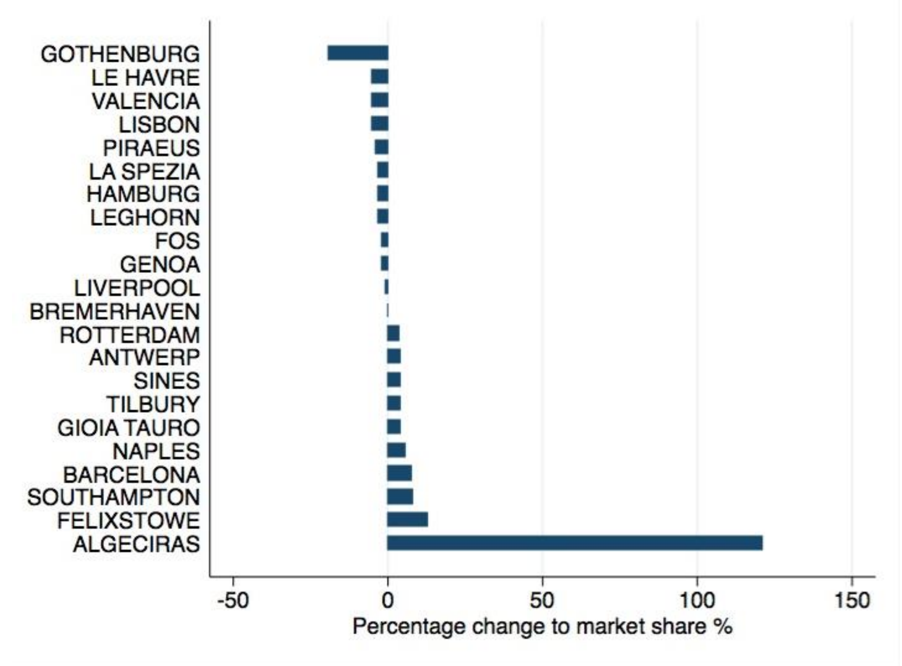
As discussed in the literature review, port efficiency is the key determinant of port choice and port competition. Relying on an estimated model, we investigate two different strategies that could affect ports attributes. This scenario contains 2 sections:

a) changing efficiency of all ports by 0.1

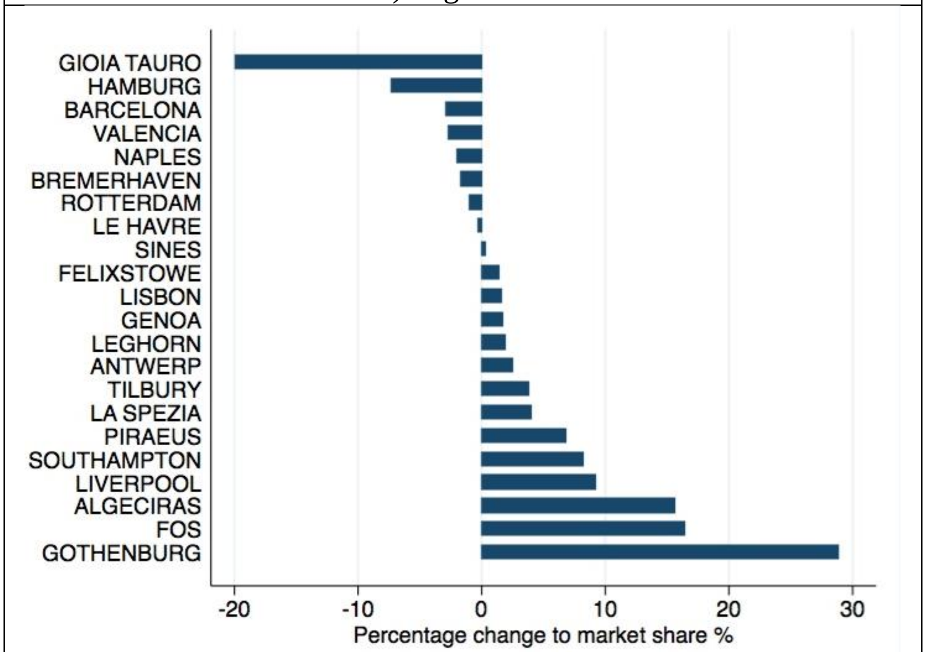
b) changing efficiency of ports by 0.1, one in a time

Scenario 2a- changing efficiency of all ports by 0.1

We improve efficiency of all ports by 0.1 (efficiency + 0.1), simulate the shippers' choices and aggregate them in each segment. This leads to a fundamental rise in Algeciras's throughput (which is the most efficient port) in segments 1, 5, and especially in segment 4. Shippers in these segments are willing to travel longer to reach more efficient ports (larger coefficients for interaction of distance and efficiency). Hence, with improvement in port efficiency, German, and Italian customers in segment 4 more readily cross the borders of their home country to ship via Algeciras that is located very far from them. Figure 6.9 depicts the changes on ports market share as a function of 0.1 changes in its own efficiency, on the segment basis and Figure 6.9f presents the aggregation across segments. Figure 6.9f shows that except Algeciras that remarkably draw customers from other ports, gains in other ports is negligible.

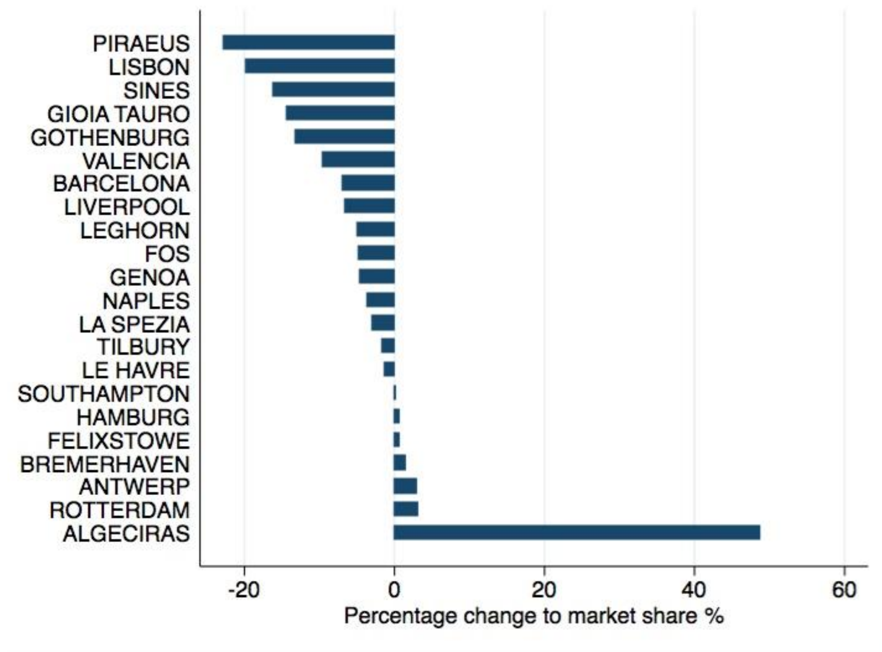


a) Segment 1

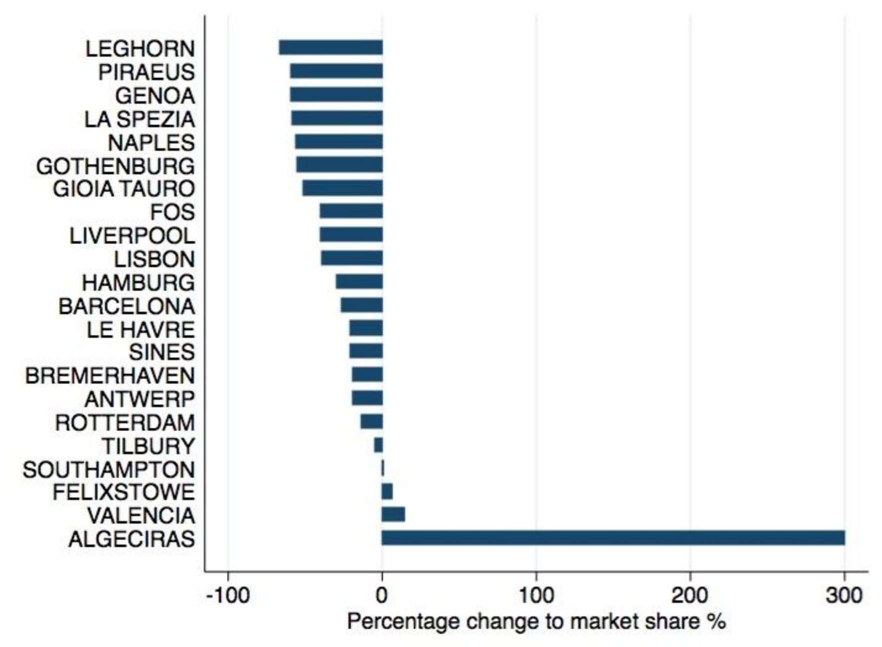


b) Segment 2

Figure 6.9: Percentage change in ports share after increasing efficiency of all ports by 0.1

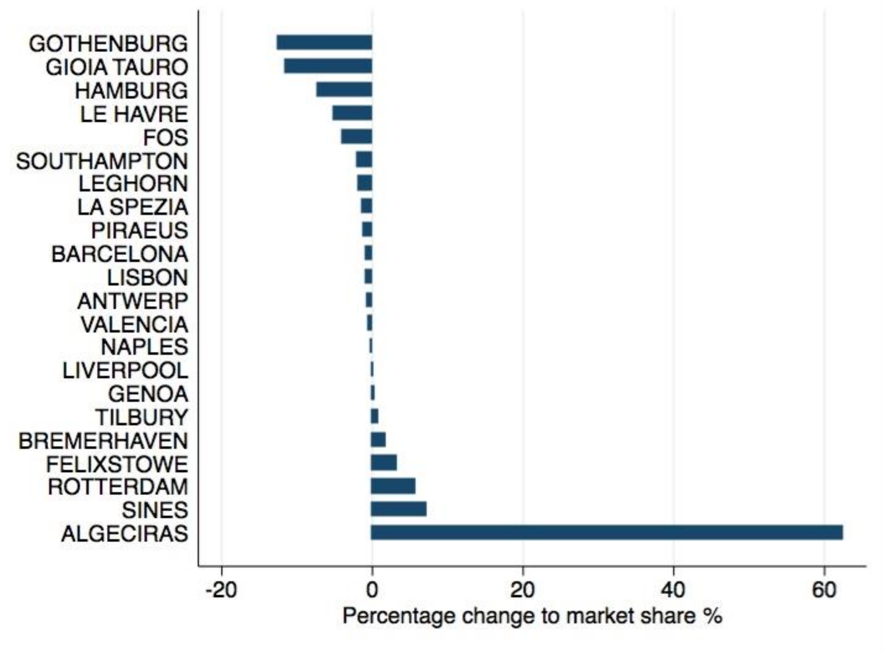


c) Segment 3

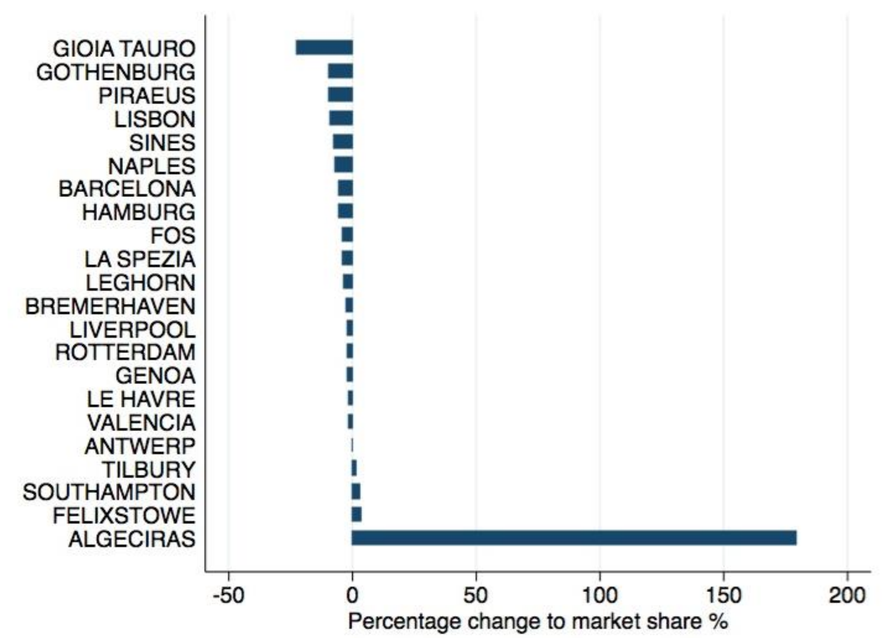


d) Segment 4

Figure 6.9: (continued)



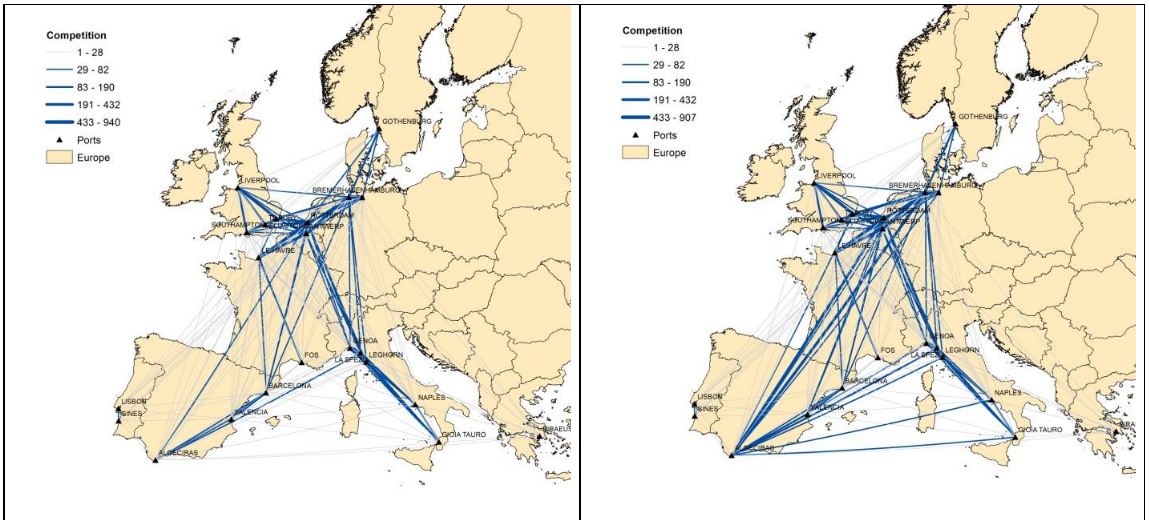
e) Segment 5



f) Overall

Figure 6.9: (continued)

Figure 6.10 compares the competitive interactions between ports with original efficiency (left) and after increasing efficiency by 0.1. We see that long edges are added between Algeciras and Italian/Northern ports implying that many shippers who used to ship only via Italian or North-European ports and did not consider Algeciras because of its location (with choice probability of 0), now have the tendency to consider Algeciras because its enhanced level of efficiency is worth travelling further (choice probabilities increase from 0 to 0.1 or 0.2).



Original Competition network	Increasing ports efficiency by 0.1

Figure 6.10: competitive interactions before and after overall increase to efficiency

Scenario 2b- changing efficiency of a port by 0.1

In this scenario, we increase efficiency of a port by 0.1 (efficiency + 0.1), and keep it constant for other ports, simulate choices and calculate percentage changes of port market shares. We repeat this for each port. Figure 6.11 plots the percentage changes in market shares against port efficiency levels. The upward sloping relationships imply that the demand response is large for ports that are already efficient.

This result reflects the parameter estimates in the LCL model. An increase in efficiency may induce shippers that attach high preference to efficient ports and then put relatively less value on distance for reaching efficient ports to change to the port with improved efficiency. These shippers are both likely to only consider efficient ports and willing to change ports in response to an increase in efficiency at another efficient port, even one that is located further away. Therefore, when an efficient port increases its efficiency, shippers that previously used this port are unlikely to change their decision in response to change in efficiency. Furthermore, many other shippers that are located over a broad geography and used to ship via other ports may well switch to this more efficient port now. On the other hand, diminishing returns to efficiency may observe for an inefficient port that reaches some level of local monopoly over neighboring shippers. Hence, efficiency improvement does not affect their market share much because they are often attracted to customers that have no preference for efficiency, may be due to cost considerations, and therefore, they are not elastic to efficiency improvement.

These results imply that the incentives to focus on efficiency are larger for more efficient ports, since ports above a threshold of efficiency compete intensely for the segments of the customers that are elastic to efficiency, while inefficient ports that are in a monopolistic position over their neighboring shippers do not have that incentive to improve efficiency.

Similar to what was experienced in scenario 2a, the throughput of Algeciras changes by a factor of 3, while the other ports are in the range of 0 to 0.6. Therefore, Algeciras is not shown in Figure 6.11 since it masks variations in the vertical axis. Also, ports located on the bottom left of the Figure 6.11 would lose some customers when they improve efficiency. This is due to the negative value that some shippers place on efficiency and on the interaction of efficiency and distance. It can be explained by the possible higher costs (price) associated with more efficient ports.

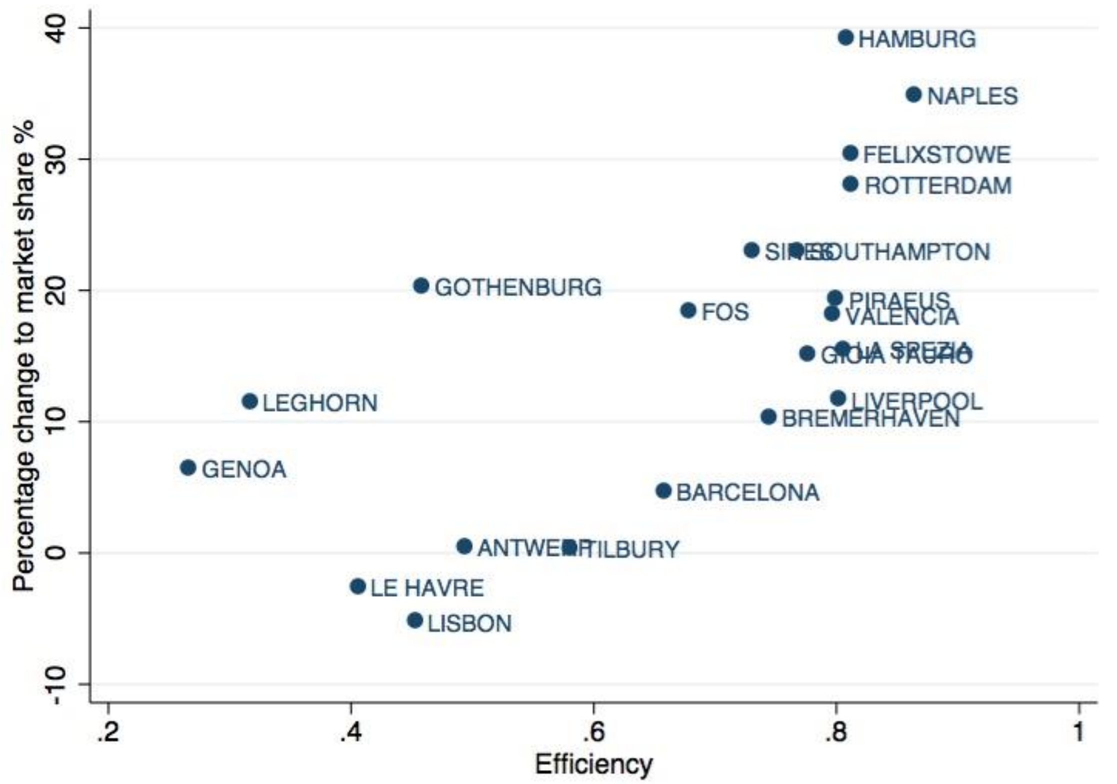
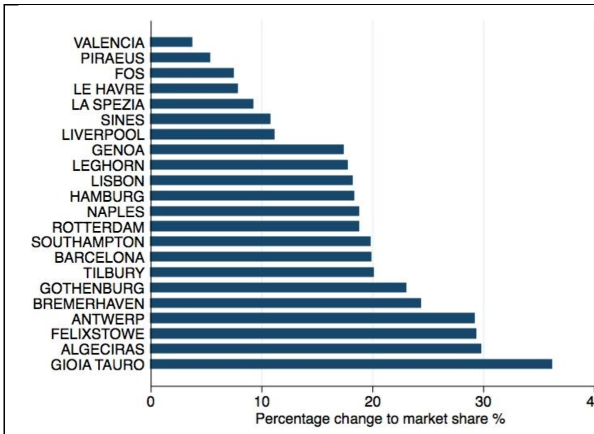
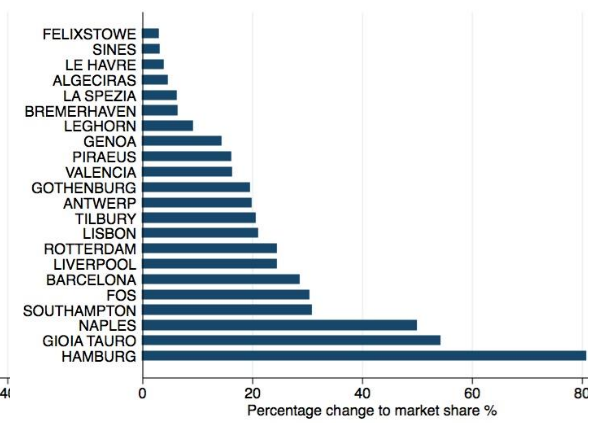


Figure 6.11- Percentage change in market share in response to a 0.1 rise in efficiency against the efficiency level of the port

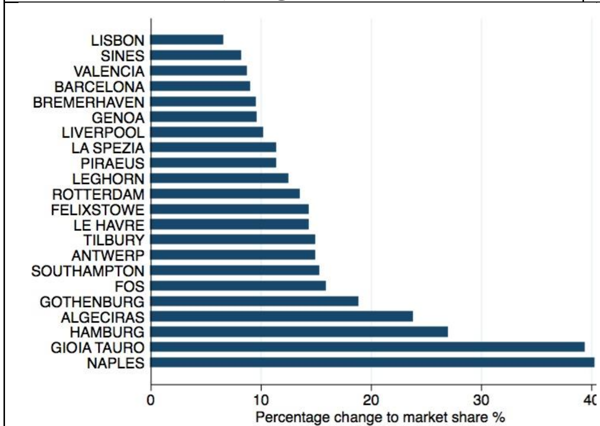
6.4.3. Reducing Distance to Ports

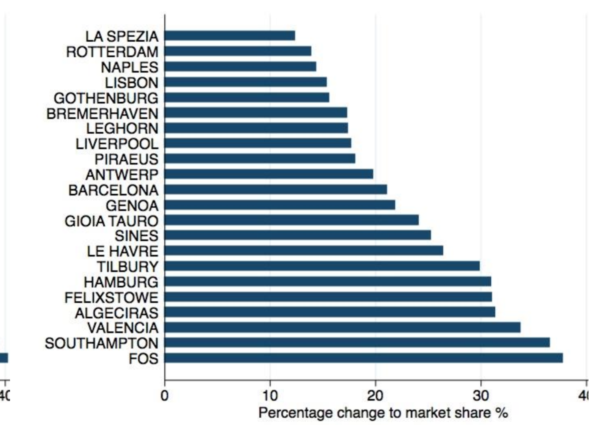
Recently, rail and barge services have brought fundamental changes to hinterland accessibility and inter-port competition, which used to largely rely on trucks and road haulage (Notteboom 2009). Investments in multimodality draw significant attention in economic and transportation research as a driving factor of port competitiveness. Under this scenario, we are interested to examine changes to a port's aggregated share if it is



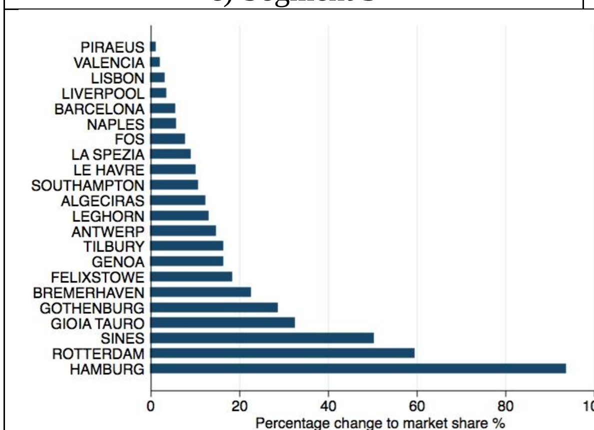


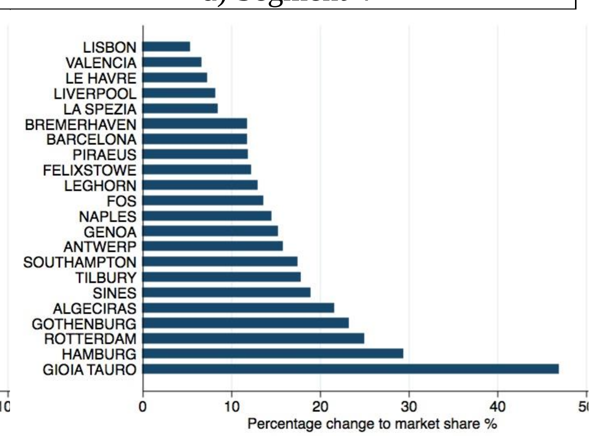
a) Segment 1 | b) Segment 2





c) Segment 3 | d) Segment 4





e) Segment 5 | f) Overall

Figure 6.12: Percentage change in each port market share in response to decreasing distance to it by 10% and keeping distance constant for all the other ports

more accessible because of dedicated train travelling to the port. We reduce distance to a port by 10% and keep it constant for the other port, simulate shippers' choices and calculate changes of aggregated choices.

Figure 6.12 depicts that this strategy affects ports share differently across segments; some ports might gain more market share in one segment but not the other, as positions of ports vary across segment in bar charts, for instance the share of Rotterdam changes approximately 0.3 in segment 5 and only 0.13 in segment 3. In other words, Rotterdam would attract around 260 more customers from segment 5 and 130 from segment 3.

In addition to what is presented in aggregated responses in Figure 6.12, this modeling framework that accommodates heterogeneous preferences allows tracking changes to individual responses and identifying contestable customers that are more responsive to different strategies. Therefore, ports can discriminate among customers depending on their degree of responsiveness (contestability) when they offer better transportation services. While switching (contestable) customers will be induce by these transportation service offers, captive customers will remain loyal no matter what strategies the port adopts.

Figure 6.12f reports changes to port share for all segments combined. While the ports of Le Havre, Lisbon, Liverpool, and Bremerhaven take advantage of improved transport services the least, ports like Hamburg, Gioia Tauro, Algeciras, and Southampton will benefit a lot. It could be related to the monopolistic position of each of these ports: ports that already take advantage of their monopoly over hinterlands are not affected by this strategy. Conversely, ports that are in competitive position with large ports will benefit largely from such enhanced transportation service strategies.

To test if there is any relation, we mapped the changes of Figure 6.12f against the contestability index presented in Table 6.4. This index is the proportion of switching customers to total customers. Figure 6.13 shows that ports whose market is predominantly captive do not need desperately to adopt transportation strategies as they have their loyal customers. On the other hand, such strategies will affect largely ports that are in competitive positions with other ports.

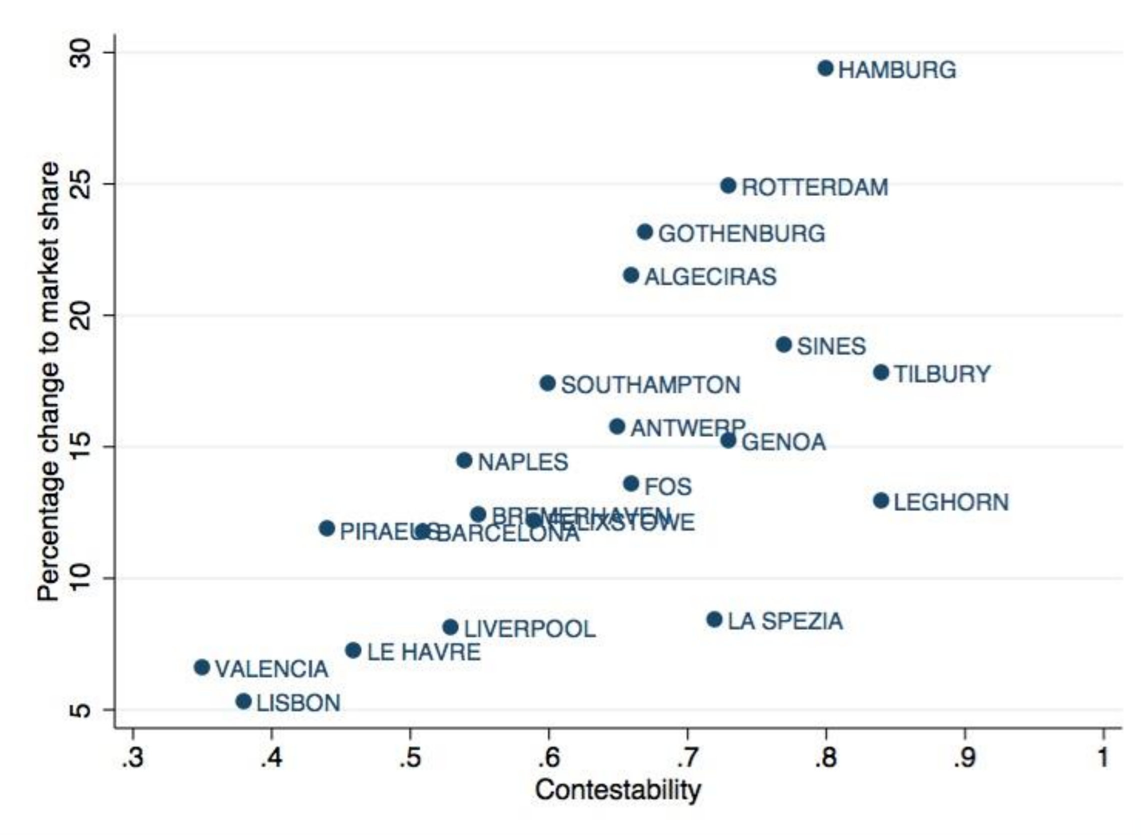


Figure 6.13: simulated change to ports customers against ports contestability index presented in Table 6.4

Following Kamakura and Russel (1989) we call what is presented in Figure10f a measure of ‘competitiveness’ of ports with respect to changes to distance to them. In fact the competitiveness of port i aggregate changes in choice probabilities for port i in response to strategies adopted by port i. Similarly, the "vulnerability" of port i aggregate changes to the choice probabilities for port i in response to strategies adopted by rest of the ports. Figure 6.14 presents ports vulnerability as a result of reduction of distance to other ports. It is worth reminding the reader here that the analysis does not bear on the portion of a port's business generated by the shippers that ship less than three shipments Therefore the measure of vulnerability concerns out the business of a port that is contestable.

Figure 6.14 implies that the ports of Le Havre, Bremerhaven, and Felixstowe, which have the lowest competitiveness indices, are least vulnerable as well. The Pearson correlation between port competitiveness (presented in Figure 6.12f) and port vulnerability (presented in Figure 6.14) is -41.62% and Spearman correlation is -57.76%. This negative association shows the strong reliance of the least competitive ports on their loyal customers that hardly switch to the other ports. It also confirms elastic (contestable) customers of the most competitive ports that readily switch to the ports' competitors.

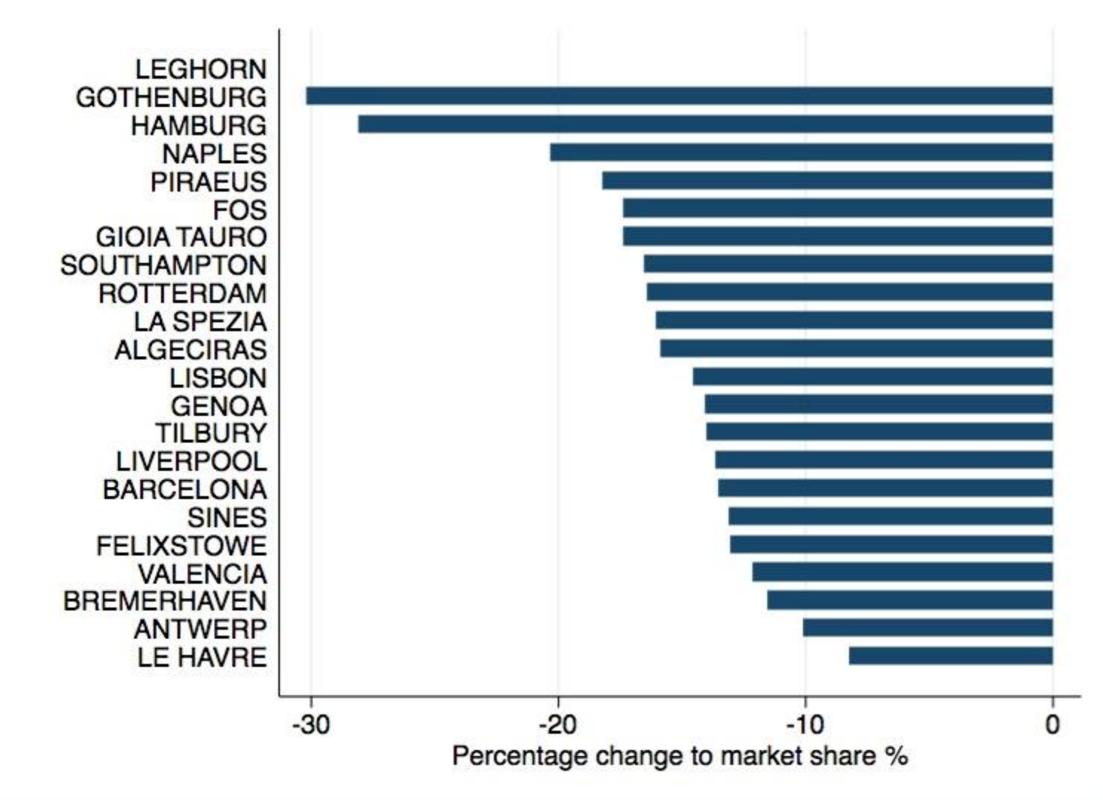


Figure 6.14: vulnerability of ports in response to changes of transportation costs to the rest of ports

The above discussion aggregates change of port probabilities for each segment in response to changes of distance to ports. Figure 6.15 presents disaggregated changes of choice probability for each customer whose original choice probability is more than 20% in relation to distance traveled. We exclude customers with choice probability less than 20% because they are often located very far from the ports and are not affected by reduction in distance. If they were included the plots in Figure 6.15 would be cluttered with many dots located on the horizontal axis.

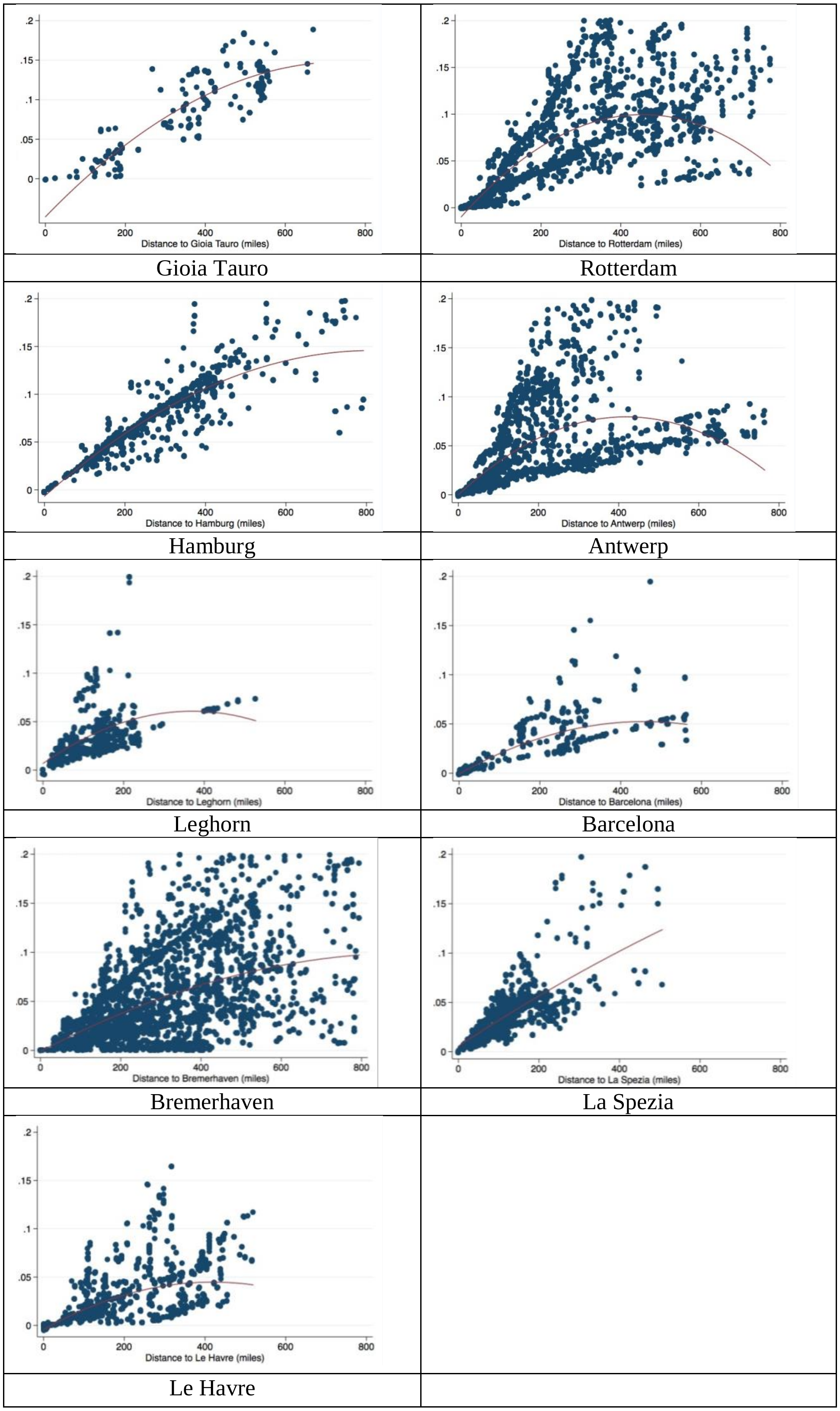


Figure 6.15: Changes in customer choice probabilities in response to 10% decrease in distance to the port (vertical axis) against distance of customer to the port (horizontal axis, miles)

Figure 6.15 depicts how the degree of hinterland contestability varies by distance from a port. Figure 6.15 does not contain all 22 ports, but it covers ports with noteworthy patterns of competitiveness/vulnerability. The general pattern is that the immediate hinterlands are not very elastic to distance changes; as distance becomes larger, demand elasticities increase up to some points. Demands for Gioia Tauro, Rotterdam, Hamburg, La Spezia are the most responsive to distance changes, conveyed by the steep slope of the trend (the fitted red line). Also, changes to demand vary by ports; while these changes range from 0 to 0.2 for ports of North European ports of Bremerhaven, Rotterdam, Hamburg, Antwerp, as well as Gioia Tauro, they are limited to 0 to 0.1 for Mediterranean ports of La Spezia, Leghorn, Barcelona as well as Le Havre, reflected in scale of the vertical axis. This is also reflected on the high value of port competitiveness depicted in Figure 6.12f.

Scales of the horizontal axis shows that demand for the ports of Rotterdam, Hamburg, Gioia Tauro, Antwerp, and Bremerhaven extends over larger areas (up to 800 miles) compared to demand for the Italian ports of La Spezia and Leghorn (200 miles) that are shaped by the constricted geography of Italy. However, the Italian port of Gioia Tauro, which is a transshipment hub, attracts customers from a broad geography area.

Variations in responses are another interesting finding revealed by Figure 6.15, presented by large fluctuation of demand for Bremerhaven and Le Havre and sharp increasing trend for La Spezia and Hamburg. It is worth reminding the reader that the presented customers are those that will choose the ports by probability higher than 20%. Therefore, It is interesting to know why a large number of potential customers of Bremerhaven and Le Havre do not respond to distance reduction. For instance, among customers located 200 miles away from Bremerhaven or Le Havre, some are totally inelastic to distance changes (changes equal zero). In fact, they are customers of segment 2 who overwhelmingly patronize the ports of Le Havre and Bremerhaven, and are inelastic to such transportation cost effects, as they have no choice other than choosing these ports.

6.4.4. Structural Change

There is a perception of “death of distance” in the current wave of globalization (Brun et al., 2005) that claims that a significant negative effect of distance on trade should decrease over time. However, scholars have obtained paradoxical results as in many gravity-based international trade studies negative effect of distance on trade is observed to strictly increase rather than decrease. A survey article by Disdier and Head (2008) based on the systematic analysis of 1467 distance coefficient estimates in 103 papers concludes that technological change that has revolutionized the world economy does not cause the impact of spatial separation to decline or disappear (Disdier & Head, 2008).

Under this scenario, we decrease distance to all ports by 10%, simulate the shippers ‘choices and then aggregate them at the port level. We can see that changes to port market shares are very small (around zero) compared to what they gain under scenario 3. Figure 6.17 depicts that there is not much difference in the intensity of competitive interactions before and after reducing distances by 10%. In other words, port market shares and competitive interaction are relatively stable to structural changes to distance. This may explain why the usual gravity model estimation has not found trade becoming less sensitive to distance as a results of transportation technological changes over time.

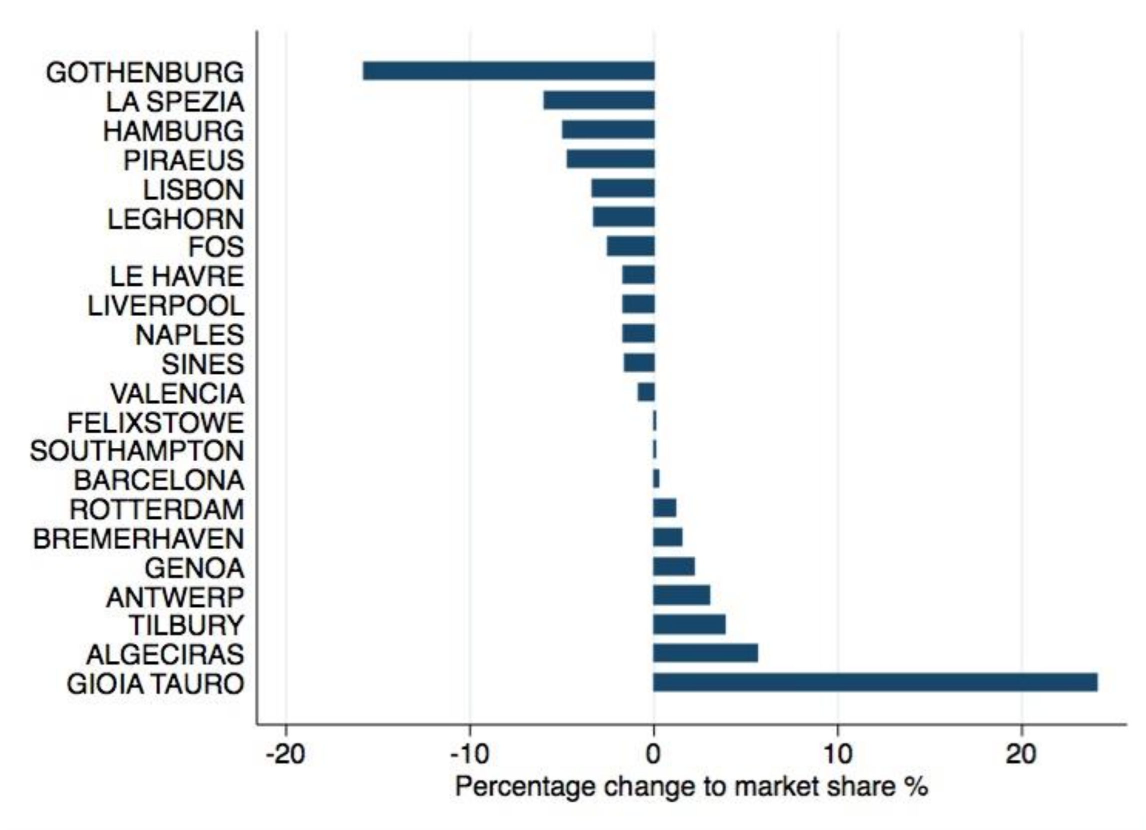
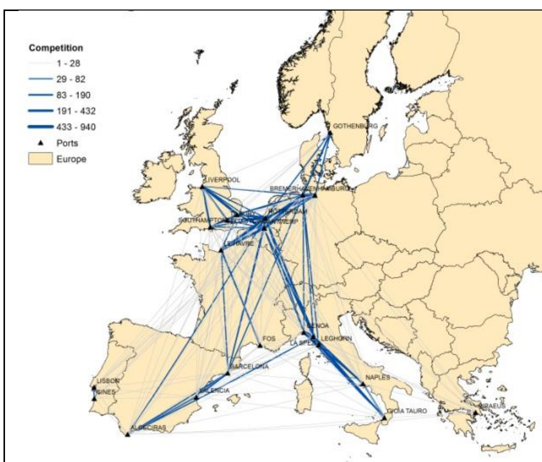
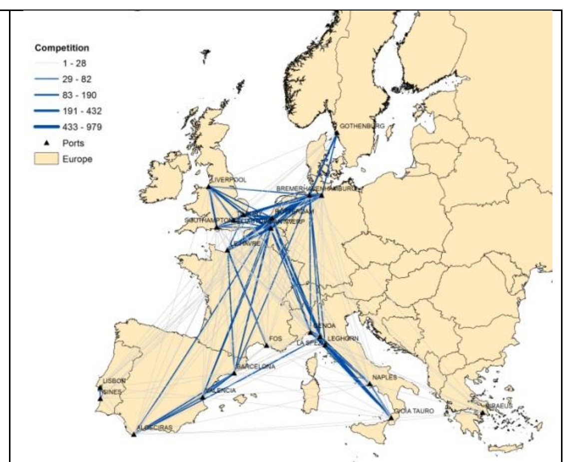


Figure 6.16: changes to port market shares in response to reduction in all distances (structural change) by 10%



Original Competition network



Decreasing Distance by 10%

Figure 6.17: competitive interactions before and after reducing distances by 10%

CHAPTER 7: CONCLUSIONS

This research was built on the recognition that people are different and possess different systems of valuation. It contributed to the modeling of competition by acknowledging the fact that the degree of competition between differentiated firms will depend on the degree of heterogeneity in people's preferences. Also, this research was distinguished by emphasizing the spatial aspects of competition as critical drivers of differentiation. Despite a well-established theoretical literature on concepts of competition, empirical research has paid little attention to the geographical dimension, mainly due to the unavailability of geographical distribution of demand (Miller and Osborne, 2011).

This research proposed an empirical framework to modeling competition by analyzing consumers' choice history in a spatially differentiated market. The observation of consumers' choice history is central to this research, which allows to distinguish between consumer preferences for certain attributes of firms from unobservable factors. This framework employed a Latent Class Logit that accommodates heterogeneous preferences by assuming a discrete mixing distribution for customers' preferences. It allows market segmentation and preference estimation simultaneously. This framework can be used in applications where alternatives are differentiated by their inherent characteristics including geography and other product characteristics and where customers are characterized by significant taste heterogeneity for those characteristics.

The proposed approach was able to delineate captive and contestable demands induced by their preference structure. Segments of consumers whose preferences restrict choice to a single supplier constitute captive hinterlands. Conversely, the preference structure of other consumers may be conducive to multiple firms being selected at different occasions. Also it allowed the estimation of demand in response to different strategies that could be adapted by firms to improve certain attributes of firms. Depending on the degrees of heterogeneity in preferences for these attributes, demand responds differently to the same strategies. This allows firms to identify responsive (contestable) markets and effective competitors and accordingly adopt strategies for attracting contestable markets by offering price discrimination and differential transportation services depending on the degree of responsiveness (contestability).

The applicability of the proposed framework was tested using a rich and novel dataset on shipping route choice. The shipping industry has experienced tremendous expansion under the combined effect of the liberalization of international trade and the continuing decline in transportation costs. Regarding the importance of ports as gateway to global market and the reliance of the national economies on port efficiency and competitive power, inter-port competition has been the focus of a large literature in transportation economics. This research empirically modeled competition between ports as spatially differentiated firms that serve heterogeneous groups of customers including manufacturers, freight forwarders, shipping lines with different market size. As soon as transportation costs decrease as a result of containerization and of the emergence of intermodal rail and barge corridors, customers distribute their shipments between ports depending on how they evaluate transport costs and ports characteristics.

We modeled the choice of European shippers on which port to use when shipping to the United States. Using a Latent Class Logit, we assumed that shippers can be partitioned into segments with different unobserved taste heterogeneity. Our data describe shipper's location, shipper's size, the port's location, and port characteristics. We first investigated how distance, crossing a national border and port characteristics including port efficiency, land accessibility and oceanic connectivity influence port choice across shippers of different size. Variability of preferences across segments showed that consumers are heterogeneous in their valuation for proximity, border crossing, and ports' characteristics. Customers value distance negatively, avoid border strongly and put positive weight on port efficiency. While the smallest shippers have a strong resistance to crossing borders and are not willing to ship further to reach more efficient ports, the largest shippers cross border readily and put less weight on distance to efficient ports.

Competition is rather uneven between ports and its intensity varies based on characteristics of each segment. Competitive interactions between two ports are quantified by the summation of joint probabilities that same customers ship via both ports. Customers that are willing to cross border and travel broad geographical distances enforce international competition while customers that avoid borders and put large negative value on distance forms domestic competition or even leads to the local monopoly of their closest ports. Below are some interesting findings that present demand contestability and changes to competitive forces in response to different scenarios.

Distance and national borders as critical factors of international trade patterns play important role in ports competition. We find that border effects are non-trivial even within the European free-trade zone. We do not have data to explain underlying determinants of border effects but this could be a direction of possible future research. Our results show that effects of borders vary statistically across countries. Italian, French and British ports are advantaged by the barrier imposed by national borders, while Dutch, Belgian, Spanish and German ports would greatly benefit if the borders were removed.

Furthermore, we investigate changes to shipment patterns in response to possible reduction to geographical distance. Our results imply that ports with large share of switching (contestable) markets will benefit more from offering transportation service (e.g. developing barge and rail corridors) than ports that strongly rely on loyal (captive) customers. While demand elasticities are small in areas close to the ports, it increases with distancing from the ports. We also find that port market shares and competitive interaction are relatively stable to structural changes to distance. This may explain why the usual gravity model estimation has not found trade becoming less sensitive to distance as a result of transportation technological changes over time.

This research also contributes to the large body of literature that studies port efficiency and competitiveness as it uniquely combines ports' technical efficiency, estimated using Stochastic Frontier Analysis, to heterogeneous preferences in demand for efficiency. Our results suggest that demand for efficient ports is much more elastic to efficiency improvement policies than demand for inefficient ports. While efficiency-related strategies put efficient ports in more competitive positions, inefficient ports have a local monopoly over neighboring customers that do not have high preferences for efficiency.

Although the dataset employed as an application of the proposed framework is unique and unrivaled among the existing research in the port industry, it still presents a number of limitations:

1. This research used a one-month data consisting of a large number of shippers that ship very few shipments, while they would ship more if we observed them in a longer time period. In fact, focusing on a longer period of data would extend customers choices and enhance the quality of estimation of preferences.
2. The dataset does not contain information on shipping routes and mode of transportation. Instead of assuming one single border between any pair of countries, no matter whether they share a border or they are located some borders apart, we could include effects of different borders that needs to be crossed between origin and ports. Also, being aware of mode of transportation would allow us to separately estimate the effect of distance depending on mode of transportation.
3. The U.S. final destination, as one of the determining factors of destination (U.S.) choice of ports, which indirectly could affect the choice of the port in the European side, is not reliable. While the consignee address is included, this location is often a corporate address.
4. Information on ports and transportation prices, as one of the key determinant of choice, is not observed.

REFERENCES

Adamowicz, W. L., & Swait, J. D. (2013). Are Food Choices Really Habitual? Integrating Habits, Variety-seeking, and Compensatory Choice in a Utility-maximizing Framework. *American Journal of Agricultural Economics*, *95*(1), 17–41.

Ahlers, D., & Boll, S. (2008). Retrieving address-based locations from the web. *Proceeding of the 2nd international workshop on Geographic information retrieval - GIR ’08*, 27.

Aigner, D., Lovell, C. A. K., & Schmidt, P. (1977). Formulation and estimation of stochastic frontier production function models. *Journal of Econometrics*, *6*(1), 21–37.

Amerini, G. (2010). *Statistics in Focus European port activity in 2009 hit by the general economic crisis*.

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

Anderson, S., & Palma, A. De. (1992). The Logit as a Model of Product Differentiation. *Oxford Economic Papers*, *44*(1), 51–67.

Anderson, S., Palma, A. De, & Thisse, J.-F. (1988). A REPRESENTATIVE CONSUMER THEORY OF THE LOGIT MODEL. *INTERNATIONAL ECONOMIC REVIEW*, *29*(3), 461–466.

Anderson, S., Palma, A. De, & Thisse, J.-F. (1989). Demand for Differentiated Products, Discrete Choice Models, and the Characteristics Approach. *The Review of Economic Studies*, *56*(1), 21–35.

Barros, P. (2003). The Measurement of Efficiency of Portuguese Sea Port Authorites Sea Port Authorities with DEA. *International Journal of Transport Economics*, *30*(3), 335–354.

Battese, G. E., & Coelli, T. J. (1993). *A stochastic frontier production function incorporating a model for technical inefficiency effects*. University of New England.

Baumol, W. (1982). Contestable Markets: An Uprising in the Theory of Industry Structure. *American economic review*, *72*(1), 1–15.

Behar, A., & Venables, A. J. (2011). Transport Costs and Costs and International Trade. In A. de Palma, R. Lindsey, E. Quinet, & Roger Vickerman (Eds.), *A Handbook Of Transport Economics* (pp. 97–115). Cheltenham: Edward Elgar.

Behrens, C., & Pels, E. (2012). Intermodal competition in the London–Paris passenger market: High-Speed Rail and air transport. *Journal of Urban Economics*, *71*(3), 278–288.

Belotti, F., Daidone, S., Ilardi, G., & Atella, V. (2012). Stochastic Frontier Analysis Using Stata. *SSRN Electronic Journal*, (ii), 1–39.

Ben-Akiva, M, & Lerman, S. (1985). *Discrete Choice Analysis: Theory and Application to Travel Demand*. Cambridge, MA: MIT Press.

Ben-Akiva, Moshe, De Palma, A., & Thisse, J.-F. (1989). Spatial Competition with Differentiated Products. *Regional Science Urban Economics*, *19*, 5–19.

Biscaia, R., & Mota, I. (2013). Models of Spatial Competition: A critical Review. *Papers in Regional Science*, *92*(4), 851–871.

Blonigen, B. A., & Wilson, W. W. (2007). Port Efficiency and Trade Flows*. *Review of International Economics*, *16*(1), 21–36.

Blonigen, B., & Wilson, W. (2006). International Trade , Transportation Networks and Port Choice. *Water Resources*.

Boone, J. (2008). A NEW WAY TO MEASURE COMPETITION *. *The Economic Journal*, *118*(2002), 1245–1261.

Boone, J., Griffih, R., & Harrison, R. (2004). A new way to measure competition. *Journal of Institutional and Theoretical Economics*.

Borges, K. A. V, Laender, A. H. F., Medeiros, C. B., & Davis, C. A. (2007). Discovering geographic locations in web pages using urban addresses. *Proceedings of the 4th ACM workshop on Geographical information retrieval - GIR ’07* (p. 31). New York, New York, USA: ACM Press.

Brooks, M. A. (1990). Ocean Carrier Selection Criteria in a New Environment. *Logistics and Transportation Review*, *26*(4), 339–355.

Brooks, M. R. (1995). Understanding the ocean container market—a seven country study [1]. *Maritime Policy & Management*, *22*(1), 39–49.

Brun, J.-F., Celine, C., Guillaumont, P., & De Melo, J. (2005). Has Distance Died? Evidence from a Panel Gravity Model. *The World Bank Economic Review*, *19*(1), 99–120.

Cahoon, S., & Notteboom, T. (2008). Port Marketing Tools in a Logistics Restructured Market Environment: the Quest for Port Royalty market environment: the quest for port loyalty.

Cairncross, F. (2001). *The death of distance: How the communications revolution is changing our lives* (Harvard Bu.).

Cao, H., Mamoulis, N., & Cheung, D. W. (2005). Title Mining frequent spatio-temporal sequential patterns Author (s) The 5th IEEE International Conference on Data Mining , © 2005 IEEE .

Carluer, B. F. (2008). *GLOBAL LOGISTIC CHAIN SECURITY : Economic Impacts of the US 100 % Container Scanning Law*. Brussels, Belgium.

Chang, S. (1978). Production function, productivities, and capacity utilization of the Port of Mobile. *Maritime Policy & Management*, *5*(4), 297–305.

Chang, Y., & Lee, P. T. W. (2007). Overview of interport competition : Issues and methods. *JOURNAL OF INTERNATIONAL LOGISTICS AND TRADE*, *5*(1), 99–121.

Chaudhuri, S., Ganjam, K., Ganti, V., & Motwani, R. (2003). Robust and efficient fuzzy match for online data cleaning. *Proceedings of the 2003 ACM SIGMOD international conference on on Management of data - SIGMOD '03*, 313.

Chen, N. (2004). Intra-national versus international trade in the European Union: why do national borders matter? *Journal of International Economics*, *63*(1), 93–118.

Clark, X., Dollar, D., & Micco, a. (2004). Port efficiency, maritime transport costs, and bilateral trade. *Journal of Development Economics*, *75*(2), 417–450.

Combes, P.-P., Mayer, T., & Thisse, J.-F. (2008). *Economic geography: The integration of regions and nations* (Princeton.).

Coto-Millan, P., Banos-Pino, J., & Rodriguez-Alvarez, A. (2000). Economic efficiency in Spanish ports: some empirical evidence. *Maritime Policy & Management*, *27*(2), 169–174.

Cullinane, K., & Song, D.-W. (2006). Estimating the Relative Efficiency of European Container Ports: A Stochastic Frontier Analysis. *Research in Transportation Economics*, *16*(06), 85–115.

Cullinane, K., Wang, T.-F., Song, D.-W., & Ji, P. (2006). The technical efficiency of container ports: Comparing data envelopment analysis and stochastic frontier analysis. *Transportation Research Part A: Policy and Practice*, *40*(4), 354–374.

De Langen, P. W. (2007). Port competition and selection in contestable hinterlands ; the case of Austria.

De Palma, A., Ginsburgh, V., Papageorgiou, Y., & Thisse, J.-F. (1985). The Principle of Minimum Differentiation Holds under Sufficient Heterogeneity. *Econometrica*, *53*(4), 767–781.

Disdier, A.-C., & Head, K. (2008). The Puzzling Persistence of the Distance Effect on Bilateral Trade. *Review of Economics and Statistics*, *90*(1), 37–48.

Dixit, A., & Stiglitz, J. (1977). Monopolistic Competition and Optimum Product Diversity. *The American Economic Review*, *67*(3), 297–308.

Djankov, S., Freund, C., & Pham, C. S. (2010). Trading on Time. *Review of Economics and Statistics*, *92*(1), 166–173.

D'Aspremont, C., Gabszewicz, J. . J., & Thisse, J. .-F. . (1997). On Hotelling's "Stability in Competition". *Econometrica*, *47*(5), 1145–1150.

Ferrari, C., Parola, F., & Gattorna, E. (2011). Measuring the quality of port hinterland accessibility: The Ligurian case. *Transport Policy*, *18*(2), 382–391.

Gan, Q., Attenberg, J., Markowetz, A., & Suel, T. (2008). Analysis of geographic queries in a search engine log. *Proceedings of the first international workshop on Location and the web - LOCWEB '08*, 49–56.

Glaeser, E. L., & Kohlhase, J. E. (2004). Cities, regions and the decline of transport costs. *Papers in Regional Science*, *83*(1), 197–228.

Goettler, R., & Ronald, L. (2001). Spatial competition in the network television industry. *RAND Journal of Economics*, *32*(4), 624–656.

Gonzalez, M. M., & Trujillo, L. (2009). Efficiency Measurement in the Port Industry: A Survey of the Empirical Evidence. *Journal of Transport Economics and Policy*, *43*(2), 157–192.

Greene, W. H., & Hensher, D. a. (2003). A latent class model for discrete choice analysis: contrasts with mixed logit. *Transportation Research Part B: Methodological*, *37*(8), 681–698.

Grossman, G. M. (1998). "comments". In J. A. Frankel (Ed.), *The Regionalization of the World Economy* (pp. 29–31). Chicago, IL: The University of Chicago Press.

Grover, R., & Srinivasan, V. (1987). A Simultaneous Approach to Market Segmentation and Market Structuring. *Journal of Marketing Research*, *24*(2), 139–153.

Hamilton, J., & Nayak, T. (2001). IEEE Data Engineering Bulletin, *24*(4).

Hastings, J. S., Kane, T. J., Staiger, D. O., Berry, S., Haile, P., Lange, F., Oster, S., et al. (2005). Parental Preferences and School Competition: Evidence From A Public School Choice Program.

Heaver, T., Meersman, H., & Van De Voorde, E. (2001). Co-operation and competition in international container transport: strategies for ports. *Maritime Policy & Management*, *28*(3), 293–305.

Helliwell, J. F. (1996). Do National Borders Matter for Quebec's Trade?

Helliwell, J. F. (1998). *How much do national borders matter?* Brookings Institution Press.

Hess, S., Ben-akiva, M., & Walker, J. (2011). Advantages of latent class over continuous mixture of Logit models.

Hess, S., & Polak, J. W. (2005). Mixed logit modelling of airport choice in multi-airport regions. *Journal of Air Transport Management*, *11*(2), 59–68.

Hill, L. L. (2000). Core Elements of Digital Gazetteers : Placenames , Categories , and Footprints. *Research and Advanced Technology for Digital Libraries* (pp. 280–290).

Hillberry, R., & Hummels, D. (2005). Trade Responses to Geographic Frictions : A Decomposition Using Micro-Data. *Russell The Journal Of The Bertrand Russell Archives*.

Hotteling, H. (1929). Stability in Competition. *Economic Journal*, *39*, 41–57.

Hummels, D. (1999). Have international transportation costs declined?

Hummels, D. (2001). Time as a Trade Barrier. *Time*.

Hummels, D. (2007). Transportation Costs and International Trade in the Second Era of Globalization. *Journal of Economic Perspectives*, *21*(3), 131–154.

Irmen, A., & Thisse, J.-F. (1998a). Competition in Multi-characteristics Spaces: Hotelling Was Almost Right. *Journal of Economic Theory*, *78*(1), 76–102.

Ishii, J., Jun, S., & Van Dender, K. (2009). Air travel choices in multi-airport markets. *Journal of Urban Economics*, *65*(2), 216–227.

Jessee, S. a. (2009). Spatial Voting in the 2004 Presidential Election. *American Political Science Review*, *103*(01), 59.

Joint Transport Research Centre. (2009). *Port Competition and Hinterland Connections*. OECD Publishing.

Jones, C. B., & Purves, R. S. (2008). Geographical information retrieval. *International Journal of Geographical Information Science*, *22*(3), 219–228.

Kamakura, W. A., & Russell, G. J. (1989). A probabilistic choice model for market segmentation and elasticity structure. *Journal of Marketing Research*, *26*(4), 379–390.

Kumbhakar, S. C. (2003). *Stochastic Frontier Analysis*. Cambridge University Press.

Larson, R. (1996). *Geographic Information Retrieval and Spatial Browsing* (Vol. 32, pp. 81–124). Graduate School of Library and Information Science, University of Illinois at Urbana-Champaign.

Leamer, E. E., & Levinsohn, J. (1995). International trade theory: the evidence. In G. Grossman & K. Rogoff (Eds.), *Handbook of international economics*. New York: Elsevier.

Liu, F., Yu, C., Meng, W., & Chowdhury, A. (2006). Effective keyword search in relational databases. *Proceedings of the 2006 ACM SIGMOD international conference on Management of data - SIGMOD '06*, 563.

Luo, M., & Grigalunas, T. A. (2003). A Spatial-Economic Multimodal Transportation Simulation Model For US Coastal Container Ports. *Maritime Economics & Logistics*, *5*(2), 158–178.

Magala, M., & Sammons, A. (2008). A New Approach to Port Choice Modelling. *Maritime Economics & Logistics*, *10*(1-2), 9–34.

Malchow, M. B., & Kanafani, A. (2004). A disaggregate analysis of port selection. *Transportation Research Part E: Logistics and Transportation Review*, *40*(4), 317–337.

Malchow, M., & Kanafani, A. (2001). A disaggregate analysis of factors influencing port selection. *Maritime Policy & Management*, *28*(3), 265–277.

Markowetz, A., Brinkhoff, T., & Seeger, B. (2008). *Encyclopedia of GIS* (pp. 1–10). Boston, MA: Springer US.

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

McCurley, K. S. (2001). Geospatial mapping and navigation of the web. *Proceedings of the tenth international conference on World Wide Web - WWW '01* (pp. 221–229). New York, New York, USA: ACM Press.

McFadden, D. (1981). CONSUMER BEHAVIOR Econometric Models for Probabilistic Choice among Products *. In C. Manski, D. McFadden, & Eds (Eds.), *Structural Analysis of Discrete Data with Econometric Applications* (Vol. 53). MIT Press.

Meersman, H., Van De Voorde, E., & Vanelslander, T. (2010). Port Competition Revisited. *Review of business and economics*, *55*(2), 210–232.

Murphy, P. R., & Daley, J. M. (1994). A Comparative Analysis of Port Selection Factors. *Transportation*, *34*(1), 15–21.

Nir, A.-S., Lin, K., & Liang, G.-S. (2003). Port choice behaviour--from the perspective of the shipper. *Maritime Policy & Management*, *30*(2), 165–173.

Nitsch, V. (2000). National borders and international trade: evidence from the European Union. *Canadian Journal of Economics/Revue Canadienne d`Economique*, *33*(4), 1091–1105.

Notteboom, T. (2009). *Port Competition and Hinterland Connections*. *Transport*. OECD Publishing.

Notteboom, T. E., & Rodrigue, J. (2005). Port regionalization: towards a new phase in port development. *Maritime Policy & Management*, *32*(3), 297–313.

Notteboom, Theo, Coeck, C., & Broeck, J. Van Den. (2000). Measuring and Explaining the Relative Efficiency of Container Terminals by means of Bayesian Stochastic Frontier Models. *Maritime Economics & Logistics*, *2*(2), 83–106.

Novshek, W., & Sonnenschein, H. (1979). Marginal Consumers and Neoclassical Demand Theory Hugo Sonnenschein. *The Journal of Political Economy*, *87*(6), 1368–1376.

Ortúzar, J. de D., & Simonetti, C. (2008). Modelling the demand for medium distance air travel with the mixed data estimation method. *Journal of Air Transport Management*, *14*(6), 297–303.

Ottaviano, G. I. P., & Thisse, J.-F. (2011). Monopolistic Competition, Multiproduct Firms and Product Diversity*. *The Manchester School*, *79*(5), 938–951.

Poole, K., & Rosenthal, H. (1985). A Spatial Model for Legislative Roll Call Analysis. *American Journal of Political Science*, *29*(2), 357–384.

Robinson, R. (2002). Ports as elements in value-driven chain systems: the new paradigm. *Maritime Policy & Management*, *29*(3), 241–255.

Roy, J. R., & Thill, J.-C. (2003). Spatial interaction modelling. *Papers in Regional Science*, *83*(1), 339–361.

Slack, B. (1985). Containerization, inter-port competition, and port selection. *Maritime Policy & Management*, *12*(4), 293–303.

Sloev, I., Ushchev, P., & Thisse, J. (2013). Do we go shopping downtown or in the bubs ? Well , why not to both?

Song, D.-W., & Yeo, K.-T. (2004). A Competitive Analysis of Chinese Container Ports Using the Analytic Hierarchy Process. *Maritime Economics & Logistics*, *6*(1), 34–52.

Stokes, N., Li, Y., Moffat, A., Rong, J., & Engineering, S. (2007). An empirical study of the effects of NLP components on Geographic IR performance, *00*(00), 1–14.

Swait, J., & Ben-Akiva, M. (1987). Incorporating random constraints in discrete models of choice set generation. *Transportation Research Part B: Methodological*, *21*(2), 91–102.

Sánchez, R. J., Hoffmann, J., Micco, A., Pizzolitto, G. V, Sgut, M., & Wilmsmeier, G. (2003). Port Efficiency and International Trade: Port Efficiency as a Determinant of Maritime Transport Costs. *Maritime Economics & Logistics*, *5*(2), 199–218.

Tang, L. C., Low, J. M. W., & Lam, S. W. (2008). Understanding Port Choice Behavior—A Network Perspective. *Networks and Spatial Economics*, *11*(1), 65–82.

Thill, J-C. (1992). Spatial competition and market interdependence. *Papers in Regional Science*, *73*(3), 259–275.

Thill, J., & Rushton, G. (1992). Demand Sensitivity To Space-price Competition with Manhatan and Euclidean Representation of Distance. *Annals of Operations Research*, *40*, 381–401.

Thill, J.-C. (1992). Choice set formation for destination choice modelling. *Progress in Human Geography*, *16*(3), 361–382.

Thill, Jean-Claude, & Lim, H. (2010). Intermodal containerized shipping in foreign trade and regional accessibility advantages. *Journal of Transport Geography*, *18*(4), 530–547.

Tiwari, P., Itoh, H., & Doi, M. (2003). Shippers' Port and Carrier Selection Behaviour in China: A Discrete Choice Analysis. *Maritime Economics & Logistics*, *5*(1), 23–39.

Tongzon, J. (2001). Efficiency measurement of selected Australian and other international ports using data envelopment analysis. *Transportation Research Part A: Policy and Practice*, *35*(2), 107–122.

Tongzon, J., & Heng, W. (2005). Port privatization, efficiency and competitiveness: Some empirical evidence from container ports (terminals). *Transportation Research Part A: Policy and Practice*, *39*(5), 405–424.

Tongzon, J. L. (1995). Systematizing international benchmarking for ports. *Maritime Policy & Management*, *22*(2), 171–177.

Tongzon, J. L. (2009). Port choice and freight forwarders. *Transportation Research Part E: Logistics and Transportation Review*, *45*(1), 186–195.

Tongzon, J. L., & Sawant, L. (2007). Port choice in a competitive environment: from the shipping lines' perspective. *Applied Economics*, *39*(4), 477–492.

Train, K. E. (2003). *Discrete Choice Methods with Simulation*. Cambridge: Cambridge University Press.

Turner, H., Windle, R., & Dresner, M. (2004). North American containerport productivity: 1984–1997. *Transportation Research Part E: Logistics and Transportation Review*, *40*(4), 339–356.

Turrini, A., & Van Ypersele, T. (2010). Traders, courts, and the border effect puzzle. *Regional Science and Urban Economics*, *40*(2-3), 81–91.

Veldman, S., Garcia-alonso, L., & Vallejo-Pinto, J. Á. (2011). A PORT CHOICE MODEL WITH LOGIT MODELS: A CASE STUDY FOR THE SPANISH CONTAINER TRADES, (October).

Veldman, S. J., & Bückmann, E. H. (2003). A Model on Container Port Competition: An Application for the West European Container Hub-Ports. *Maritime Economics & Logistics*, *5*(1), 3–22.

Wang, T.-F., & Cullinane, K. (2006). The Efficiency of European Container Terminals and Implications for Supply Chain Management. *Maritime Economics & Logistics*, *8*(1), 82–99.

Wanke, P. F., Barbastefano, R. G., & Hijjar, M. F. (2011). Determinants of Efficiency at Major Brazilian Port Terminals. *Transport Reviews*, *31*(5), 653–677.

Wen, C.-H., & Lai, S.-C. (2010). Latent class models of international air carrier choice. *Transportation Research Part E: Logistics and Transportation Review*, *46*(2), 211–221.

Wilson, A. G. (1970). *Entropy in Urban and Regional Modelling*. Pion, London.

Witten, I. H., & Frank, E. (2000). *Data Mining : Practical Machine Learning Tools and Techniques with Java Implementations*.

Wolf, H. C. (1997). Patterns of Intra- and Inter-state trade. *National Bureau of Economic Research*.

Woo, S.-H., Pettit, S. J., Kwak, D.-W., & Beresford, A. K. C. (2011). Seaport research: A structured literature review on methodological issues since the 1980s. *Transportation Research Part A: Policy and Practice*, *45*(7), 667–685.

Zhang, A. (2009). *The Impact of Hinterland Access Conditions on Rivalry between Ports*. *Transport*. OECD Publishing.