ABSTRACT

ANALYSIS OF CONTAINER THROUGHPUT: DEMAND FORECAST AND SEAPORT COMPETITIVENESS ASSESSMENT

by
Hussain Talat Sulaimani

Seaports play a crucial role in the container industry, where they act as important nodes in the transport chain to facilitate international trade. In a competitive market, port capacity plays a significant role in defining its competitive position to attract demand and avoid congestion. Failing to provide suitable capacity results in the loss of market share. Therefore, port decision-makers face the challenge of maintaining and developing suitable port facilities to provide efficient services to port users. One of the aspects that decision-makers consider in the planning and development process is analyzing container demand. The analysis of container demand can be challenging due to the dynamic changes in international trade, port location and accessibility, competition from other ports in the same geographic region, and port selection behavior of shippers and liner companies.

This dissertation focuses on analyzing container demand; specifically, it has two main objectives: Forecasting short-term container demand and assessing the competitiveness position of the port. To forecast demand, the univariate time series stochastic approach is applied based on the methodology of Box-Jenkin, and because it only requires the historical container throughput. The developed model is used to forecast container demand of Jeddah port. The proposed model provides accurate forecasts with a

confidence interval of 93 Percent. The systematic forecasting approach provides the ability to update and apply the methodology continuously in the future.

To assess port competitiveness, spatial interaction models (SIM) are applied to estimate the impact of port performance, hinterland accessibility, and geographic location on the container flow. Both temporal and spatial data are collected for the four major ports in Saudi Arabia, which are analyzed in the case studies and SIM calibrations. The analyses performed in this study revealed that port users, as the results of modernization and privatization of the transport sector of the country, are provided with feasible port alternatives to efficiently transport freight, leading to fierce inter-port competition. The analysis also reveals that maritime connectivity of ports located in the Red Sea have a competitive advantage that allow them to attract more container flow and reach further hinterland regions when freight rates increase. This is due to their strategic location in the major maritime shipping routes. However, the availability of railway connectivity provides cheaper inland alternative that restricts the importance of maritime accessibility.

This dissertation should be of interest to policy and port-decision makers. The applied forecast model is important in the planning phase of resource allocation and facility improvements because it provides a reliable instrument to obtain insight into the future demand. The assessment of port competition helps decision-makers in evaluating the impact of port strategies by understanding the competitive position of the ports. Recognizing the scarcity of systematic research on Saudi Arabian seaports suggests that these forms of forecast analysis and competitive assessment will benefit the port sector in the country.

ANALYSIS OF CONTAINER THROUGHPUT: DEMAND FORECAST AND SEAPORT COMPETITIVENESS ASSESSMENT

by

Hussain Talat Sulaimani

A Dissertation
Submitted to the Faculty of
New Jersey Institute of Technology
in Partial Fulfillment of the Requirements for the Degree of
Doctor of Philosophy in Transportation

Interdisciplinary Program in Transportation

August 2021

APPROVAL PAGE

ANALYSIS OF CONTAINER THROUGHPUT: DEMAND FORECAST AND SEAPORT COMPETITIVENESS ASSESSMENT

Hussain Talat Sulaimani

Dr. Rongfang Liu, Dissertation Advisor — Date
Professor of Civil and Environmental Engineering, NJIT

Dr. Athanassios Bladikas, Committee Member — Date
Associate professor of Mechanical and Industrial Engineering, NJIT

Dr. Steven I-Jy Chien, Committee Member — Date
Professor of Civil and Environmental Engineering, NJIT

Dr. Janice R. Daniel, Committee Member — Date
Professor of Civil and Environmental Engineering, NJIT

Dr. Shmuel Yahalom, Committee Member — Date
Professor of Economics, SUNY Maritime College, Bronx, NY

BIOGRAPHICAL SKETCH

Author: Hussain Talat Sulaimani

Degree: Doctor of Philosophy

Date: August 2021

Undergraduate and Graduate Education:

- Doctor of Philosophy in Transportation
  New Jersey Institute of Technology, Newark, NJ, 2021

- Master of Science in International Transportation Management
  SUNY Maritime College, Bronx, NY, 2013

- Bachelor of Science in Port and Maritime Transportation
  Faculty of Maritime Studies, King Abdulaziz University,
  Jeddah, Saudi Arabia, 2008

Major: Transportation

I dedicate this dissertation to:

my parents: baba Talat and mama Zizi,

my son Talat,

my fiancé Samaher,

my brothers and sisters,

and to all my former and current teachers and professors, specially, Dr. Aly Bassiouny who inspired my pursuit of maritime shipping and economics.

ACKNOWLEDGMENT

Along the journey of this Ph.D. dissertation, I have received so much help, which I am genuinely grateful. I would like to express my sincere gratitude to Prof. Rongfang Liu as my advisor for her devoted advice, guidance, and countless contributions throughout my dissertation. and most importantly her support, understanding, and encouragement that went beyond the formal responsibility as an advisor. It has helped me to survive difficult times and unexpected circumstances throughout my Ph.D. journey.

I would like to express my appreciation to my committee members: Drs. Steven Chien, Janice Daniel, Athanassios Bladikas, and Shmuel Yahalom who has been very supportive and patient in supervising my research. Without their guidance, teaching, and assistance, this dissertation would not have been possible, particularly, Prof. Yahalom who have helped me to improve the quality and clarity of my dissertation further. I would also like to extend my gratitude to every professor, administrator in the transportation program and the department of civil engineering that assisted me during my Ph.D. journey. I am also grateful to Ms. Clarissa Gonzalez in the office of Graduate Studies and Dr. Taha Marhaba for their support and guidance.

I also would like to express my deepest appreciation to the government of the Kingdom of Saudi Arabia for sponsoring me and funding my Ph.D. study. I would like to also thank the faculty of Maritime Studies for the opportunity to pursue my studies.

This dissertation is dedicated to my family. A special feeling of gratitude to my parents, Talat and Zizi for their encouragement, patience, unlimited prayers, and devoted support throughout my time away from home. Huge thanks to my son and friend, Talat for his unconditional love, inspiration, and patience with my absence. My parents and son, you are the reason that enabled me to go through the difficulties to finish my studies. My fiancé, Samaher who never left my side and has been my cheerleader since we engaged, thank you for helping me get past the hardest point during my studies.

My brothers and sisters, Manal, Shokran, Abdulrahman, Mohamed, Hala, and Fahad, thanks for believing in me, countless support, phone calls, and visits. I love you all.

Also, thanks to Dr. Mona, my sister-in-law and friend, for your guidance and assistance. Throughout the difficult times, many friends have stood by me. Sincere thanks and gratitude to Amin Attar, Abdulellah Shahrani, and Yasine Mokhtar for all the moments that we shared, support, and the stimulating and stress-relieving discussions. Special thanks to Yasser Farha for his support and advice. I would also like to extend my gratitude to everyone who helped me toward the completion of my dissertation.

TABLE OF CONTENTS

Chapter **Page**

TABLE OF CONTENTS
(Continued)

TABLE OF CONTENTS
(Continued)

Chapter **Page**

TABLE OF CONTENTS
(Continued)

Chapter **Page**

LIST OF TABLES

LIST OF FIGURES

LIST OF FIGURES

(Continued)

Figure **Page**

LIST OF ACRONYMS

ACF	Autocorrelation function
ADF	Augmented Dickey-Fuller
AIC	Akaike information criterion
ARIMA	Autoregressive integrated moving average
GDP	Gross Domestic Product
KAP	King Abdullah Port
KPSS	Kwiatkowski-Phillips-Schmidt-Shin
LSCI	Liner Shipping Connectivity Index
MAE	Mean absolute error
MAPE	Mean absolute percentage error
MSE	Mean square error
OLS	Ordinary Least Square
PACF	Partial autocorrelation function
PPML	Poisson Pseudo-Maximum Likelihood
SARIMA	Seasonal Autoregressive integrated moving average
SIM	Spatial Interaction Model
TEU	Twenty-foot equivalent unit
UNCTAD	United Nation Conference on Trade and Development

CHAPTER 1

INTRODUCTION

The globalization of the world economy and the technological evolution have contributed to the rapid growth of the international trade (Dicken, 2015). These developments have triggered the growth of the transportation industry, which went through a paradigm shift that was fueled by containerization and the expansion of intermodalism (Vanelslander, 2008). Maritime shipping has played an important role in these transformations and developments because it is not only the most economical transport mode, but also scales well in the production processes (Stopford, 2009). According to the UNCTAD (2017), approximately 80% of overall world trade volume is transported by sea, from which containerized trade represents 24%.

The port industry plays a crucial role in supporting maritime shipping. Container shipping and ports are highly interrelated, where changes in container shipping are reflected in the port sector. This relationship is seen in the modernization of ports, where the strategic behavior of shipping liners makes a major impact on the port operations and the development of port infrastructure. In a similar vein, technological advancements have positively influenced the operation of containerships through modern handling equipment at ports as well as a wider range of operational concepts for transporting containers globally by integrating shipping with ports and hinterland transportation (Grant, 2016). These advancements are reflected in the growth of container shipping. According to the UNCTAD (2011), world container throughput increased by an estimated 13.3% rate to 531.4 million TEUs in 2010.

Container terminals have a major impact on economic growth as they act as important hubs and gateways for the facilitation of international trade and associated logistics and supply chains, since their development is reflected in the efficiency of services provided to containerships. Therefore, it is imperative to ensure efficient operation, high quality of logistic services, security, and effective supply chain management to keep a specific port globally competitive (Notteboom, 2004). However, ports are considered as nodes in value-driven logistic chains, and the port selection depends on the entire transport chain (Robinson, 2002). The determinant factors that influence the port selection, thus, port competitiveness, can be categorized into three interrelated groups: Efficiency of port operation, hinterland-related, and maritime-related (Grossmann et al., 2007; T. Notteboom, 2009) (More information in Subsection 2.3.2).

Traditionally, port hinterlands, inland markets served by a port, were relatively fixed in nature with individual captive markets due to the existence of undeveloped land transport infrastructure and the long distances (Alderton, 2008), where ships had to call each individual port where freights shipped to and from them (Chlomoudis et al., 2003). However, the improvements of land transport networks allow ports to gain access to inland markets in further geographic regions, extending their hinterland region (Rodrigue et al., 2017). Such practice has contributed to the phenomenon of "Inter-port competition" in the maritime industry. This competition takes place between ports located in the same service region, where advantageous location near markets and better foreland and hinterland accessibility of a particular port provides it a competitive advantage over rival ports.

Additionally, port capacity plays a significant role in defining its competitive position to meet demand and avoid congestion. The lower the time a ship spent in a port, the more attractive the port is to port users. The competitive position of a particular port is one of the factors that influence its potential demand; thus, port competitiveness is a crucial aspect for decision-makers in terms of investment decisions in increasing the capacity of the port (Janssens et al., 2003). Consequently, the main objective of port decision-makers is to attract port users by increasing the capacity for the potential demand growth to be able to retain and expand the market share of their port.

To support decision-making in port development, an accurate forecast of container demand is crucial to determine the scale of developments. In addition, port shareholders and planners must perform their due diligence by analyzing the strategic and competitive position of the port in the context of ever-changing competition (Haezendonck et al., 2006). Hence, when planning strategies for retaining or increasing the market share, the competitive position of the port must be assessed by the shareholders to understand the factors that influence the market and to identify the port position within the competition.

The purpose of this dissertation is to provide an instrument to support the port stakeholders in the planning process by developing suitable tools to forecast container demand, analyze the competitiveness position of the port, and investigate the impact of the port, foreland, and hinterland determinants on the container flow.

1.1 Background

The potential demand growth for container ports and the dynamic changes in port competition are two major issues that decision-makers face and need to deal with in the process of port planning and development (Benacchio et al., 2001; Rodrigue & Notteboom, 2017). Analyzing container demand can be challenging due to the dynamic changes in international trade, port location, port accessibility to markets and multimodal facilities, competition from other ports in the same geographic region, and port selection behavior of shippers and liner companies (Arvis et al., 2018; Fagerholt, 2000; Gerrits, 2007; Tsamboulas & Kapros, 2000). Moreover, the complexity of the problem, its rigorous computation, and data availability are other factors that add to the difficulty of analyzing demand (Luo & Grigalunas, 2003).

The demand for competing ports and their services are interrelated. For example, when services are improved in a particular port, it influences the demand for its competing ports. This is due to the increasing requirements and port selection strategic behavior of port users. Therefore, investments in developing port capacity and intermodal connections are also determined by inter-port competition (Meersman & Van de Voorde, 2013). Given that ports nowadays are considered nodes in the transport chain, determinants affecting port competitiveness can be grouped into three main interrelated categories: Efficiency of port operation, hinterland-related, and maritime-related (Grossmann et al., 2007; Notteboom, 2009). The port selection decision of port users is influenced by the quality, efficiency, and variety of services provided in the determinants mentioned above. Hence, the developments of the container industry suggest that well-connected container terminals in terms of the hinterland and foreland are more attractive for port users.

1.2 Problem Statement

Container demand analysis is a crucial requirement that decision-makers need to consider in the planning and development of a port. The potential growth in demand for container ports and the dynamic changes in port competition are two major issues that decision-makers face, and need to deal with in the port planning and development process (Benacchio et al., 2001; Rodrigue & Notteboom, 2017). Container ports encounter significant challenges to increase competitiveness and handle the growth in demand, where competitive strength can be achieved by providing efficient capacity through increasing productivity, reducing port congestion, and, in some markets, improving hinterland accessibility (Gerrits, 2007). Providing efficient port capacity is essential to all the stakeholders (e.g., port authority, shipping companies, shippers, and terminal operators). To increase capacity, container demand analysis is one of the requirements that port decision-makers need to consider in the planning to develop the port. However, port stakeholders have to deal with an increasing number of uncertainties in the container industry.

In a competitive market, container demand is not only associated with the economic activities but also the competitive position of the port and its competitors. The relative port competitiveness position is one of the factors that influence demand, which shareholders take into account to determine the investment decisions regarding increasing port capacity (Meersman et al. 2003). Moreover, the importance of the competitive position of a particular port increases when competing ports provide an advantage, compared to its competitors, to port users (e.g., shipping companies, shippers, and freight forwarders).

Cullinane and Wang (2009) argued in their research that "inter-port competition occurs when the user of port infrastructure or a particular port service has an economically feasible substitute for those facilities in another location." A port may be threatened by substitution by other ports due to the selection behavior of port users. Such a selection process may be affected by the introduction of mega-containerships, changes in fuel prices, and changes in the transportation network. For example, the emergence of alternative routes via a competing port may decrease the market share for a particular port.

Furthermore, port selection is a part of the supply chain selection as ports are considered nodes in the supply chain. The selection made by port users for a particular port of call is influenced by a set of feasible substitutes where port, routes, transport modes, or various combinations of them are considered as alternatives (Merkel, 2017). Thus, the competitive position of a port is influenced by factors related not only to the port but to the whole supply chain, where competition between alternative routes is a determinant of market share.

These developments in the liner shipping network revolutionized the way ports are operated, financed, managed, regulated, and how ports compete. In essence, due to the continuous transformations in the maritime shipping industry, and subsequently in the port sectors, the competition between ports intensified in some regions; thus, increased the importance of port competitiveness. Therefore, port shareholders need to understand their competitive position, the forces shaping the market, and other ports' competitive standing to strengthen their competitiveness.

With the complex relationship between port of call and various factors, such as port capacity/efficiency, hinterland/foreland access, and other political/economic impact, it is simply not enough to forecast the short-term demand for a particular port based on historical throughput data, which has been the main approach in the maritime analysis for several decades. Therefore, this dissertation is set out to identify the causal factors that affect seaport demand by examining the port efficiency, hinterland access including multimodal land transport networks, and the impact of unique geographical locations.

1.3 Research Objectives

The main purpose of this dissertation is to analyze container demand on ports by developing quantitative models to assist in the decision-making process of future investments. To achieve this, the dissertation has three objectives.

The first objective is to develop a model to forecast short-term container throughput. As port stakeholders aim to reduce operation deficiency by reducing congestion and improving container handling efficiency; their short-term planning regarding operations and resource allocation decisions is critical. The short-term forecast model contributes as a tool to support decision-makers in operational decisions, developments, and modifications to deal with the rapid changes and variability of future demand.

The second objective is to investigate the impact of inland distance on the distribution of container flow and to assess its role in the competitive position of competing ports. The evolution of inland distance role in defining the hinterland region is analyzed at the port-provincial level by determining and comparing the impact in two time periods. Additionally, this analysis will reveal the influence of other barriers that might impact the explanatory power of distance on the competing ports, individually. In other words, it provides an understanding of the geographic competitive advantage of each port with respect to the other ports.

However, factors other than inland distance impact the distribution of container flow and port competitiveness are considered in the third objective. The third objective is to analyze the impact of the geographic characteristics and intermodal connectivity in port competitiveness for inland distribution of maritime traffic. This analysis reveals the unexploited potentialities of the ports based on the factors considered in modeling the spatial interaction.

1.4 Approach and Unique Perspectives

To achieve the dissertation objectives, container demand analysis is conducted. The analysis is twofold: (1) forecasting short-term container demand of the port and (2) assessing its competitive position. Dissertation part one is on forecasting short-term demand at the port level. Dissertation part two is on assessing the port competitiveness at the country level. The case study of Saudi ports is used in this dissertation where massive port and inland infrastructure development, and recent changes in container terminal concessions in the port sector of Saudi Arabia led to the increase of port competition.

To forecast container demand, the univariate time series stochastic model is developed based on the Box and Jenkin methodology. The advantage of this model is its independence from other variables in forecasting container throughput and its ability to consider seasonal variation. In addition, the conducted methodology provides a systematic approach in identifying time series patterns, estimating model parameters, and generating forecasts, which provides the ability of continuously using the methodology in the future to forecast demand. The forecasts are obtained by regressing historical observations of container throughput and the current value with the error terms of the past values at different lags. As the model forecasts demand by only considering container throughput, and the volatility and uncertainty of container demand, the model is only applicable to forecast short-term container throughput. Therefore, monthly container throughput of Jeddah port for the period 2003-2018 is used to forecast short-term demand.

To assess port competitiveness, two steps are conducted to investigate the impact of the determinants of port hinterland and foreland in the distribution of container flow in the hinterland. First, the impact of inland distance on the distribution of container flow is investigated, where the distance decay parameter is estimated by using Spatial Interaction Model (SIM) for two different years to investigate the evolution of distance role in the distribution of container flow. Secondly, the impact of port location and intermodal connectivity is analyzed along with inland distance. This is done by estimating the variables parameter using SIM and the model outcomes to investigate their impact on port competitiveness. The unexploited potentialities of the competing ports are assessed by analyzing the gap between actual and estimated container flows in both steps mentioned above. The gap analysis will reveal the impact of distance and other barriers on each port-province pair individually. The analysis outcomes also provide an insight into the competitive position of the ports based on each hinterland region.

1.5 Dissertation Structure

This dissertation is presented in six chapters. The first chapter briefly introduces the research area, defines the problem statement, and outlines the dissertation objectives.

The second chapter provides an overview of the literature review. The chapter consists of three Sections. Section 2.1 includes a brief background of the shipping industry and the recent evolution of container ports. Section 2.2 presents previous studies on forecasting container demand and various forecasting models. Section 2.3 provides an overview of port competition and previous studies on the assessment of port competition. Finally, Section 2.4 outlines the research questions.

Chapter three presents the proposed methodology to answer the research questions. The methodology involves three quantitative modeling techniques to analyze container demand. That is, (1) Forecasting short-term container demand, (2) investigating the impact of inland distance on container flow, and (3) Analyzing the factors impacting port competitiveness. Thus, The chapter is divided into three Sections. In Section 3.1, Time series stochastic forecasting models are presented. Sections 3.2 and 3.3 offer the models used to assess port competitiveness. Section 3.2 explains the maximum entropy version of SIM, which investigates the impact of inland distance on port competitiveness. In Section 3.3, the statistical approaches to estimate SIM are exhibited, focusing on Poisson-based SIM, which investigates the impact of other factors along with the inland distance on port competitiveness.

Chapter four presents the case study used to forecast container demand and analyze port competitiveness. In Section 4.1, the background of Jeddah port is presented. Then, section 4.2 gives an overview of the development of the port sector in Saudi Arabia.

Chapter five presents the model fitting and research outcomes. The chapter also includes a discussion about the dissertation findings. In Section 5.1, the univariate Time Series stochastic model is applied to forecast short-term container throughput at Jeddah port. In Section 5.2, the explanatory power of distance on inland distribution of maritime traffic is investigated by applying Spatial Interaction Model (SIM) and gap analysis. After that, Poisson-based SIM is used in Section 5.3 to investigate competing ports' competitive position by analyzing the impact of inland distance, railway availability, and port location within major shipping routes on the inland distribution of maritime traffic.

Lastly, chapter six provides a summary of the conducted methodology, findings, and contribution in Sections 6.1 and 6.2. In Sections 6.3 and 6.4 outlines limitations and recommendations for future research, respectively.

CHAPTER 2

REVIEW OF THE LITERATURE

This chapter is organized into four Sections, each presenting research gaps at the end. In Section 2.1 presents the recent development of the global container shipping industry, its implications on container ports, and the evolution of hinterland. Section 2.2 reviews various forecasting models applied in previous studies on forecasting container demand, with a focus on the time series stochastic approaches. In Section 2.3 presents the evolution of port competition, its influence on hinterland expansion, and the impact of port characteristics and geographical location in its competitive position. Previous studies applying spatial interaction models to assess hinterland accessibility and port competitiveness are reviewed as well. Research gaps are presented at the end of each Section. Lastly, Section 2.5 introduces the research questions.

2.1. The Evolution of Shipping Industry

This section presents recent developments of the global container shipping industry and the role of port planning in the development of port capacity.

2.1.1 Recent developments in container industry

Containerization plays a significant role in facilitating world international trade. Ever since launching the first container ship in 1956, the container shipping industry has grown remarkably, reflecting the importance of containerization in a global economy. Because of the standardization aspect, containerization provides the cost and time needed of freight shipping dropped as economies of scale rose dramatically. In addition, during the last three decades, as globalization reshaped the global market economy, global economic development boosted the growth of the container industry, bringing enormous international trade growth. The expansion resulted from increased off-shore manufacturing activities within the production chain (De Langen, 2003). These changes in international trade paved the way for the container shipping sector to grow significantly.

As container seaborne trade has proved its worth globally, it became the backbone of international trade. Compared to other shipping sectors, container shipping has been the fastest growing cargo segment at an annual average growth rate of 8.2% over 1990-2010 (UNCTAD, 2011). According to the annual report of "Review of Maritime Transport 2011" published by the World Economic Forum, global container seaborne trade accounted for approximately 16% of the total volume of international seaborne trade in 2010 compared to just 6% in 1990. However, container demand growth is expected to continue in the upcoming years (UNCTAD, 2018).

The rapid growth in container seaborne trade led to the evolution of mega containerships. To cope with the expansion of global container seaborne trade, liner companies seek economies of scale, container ship size has been increasing at a high pace over the last twenty years. The new generation of mega containerships has a capacity of more than 24,000 TEU (UNCTAD, 2019) compared to a little over 4,800 TEUs capacity of containerships in 1990 (Merk et al., 2015). This rapidly increasing pace of container ship size had significant consequences on the whole container shipping industry. They brought cost savings per container for carriers and reduced freight cost per container for shippers.

Furthermore, the continuous increase of containership sizes largely impacts the competitiveness of ports. As mega containerships have large dimensions, ports face the challenge of providing the infrastructure needed to meet mega ships requirements and carriers' expectations (Merk, et al., 2015). Carriers argued that container ports that are not capable of providing the required port facilities will not get called by liners (Haralambides, 2017). As a result, ports that are incapable of getting called by mega containerships face the risk of losing market share.

2.1.2 Port Capacity and Planning

The main objective of port decision-makers and planners is to attract port users by increasing the capacity for the potential demand growth to be able to retain and expand the market share of their port. Port planning is a complex process that involves various options and scenarios (Sanders et al., 2007). Having enough capacity is not a goal by itself, as port decision-makers must increase port utilization. Excessive port capacity compared to limited container demand results in low utilization of port facilities (Figueiredo et al., 2015). On the contrary, exceeded demand leads to congestion in port facilities, an increase in a shipper's cost while using ports, and extra time needed due to waiting for service and container handling, leading to the loss of market share.

To improve port capacity, port planners must decide at different planning levels. Based on the time horizon, three planning levels are considered by decision-makers: (1) the strategic level concentrates on large investment for long-term plans to provide strategic competitive position to the port.; (2) the tactical level concerns utilizing available resources and increasing facilities efficiency; and (3) the operation level increases operational efficiency (Figure 2.1).

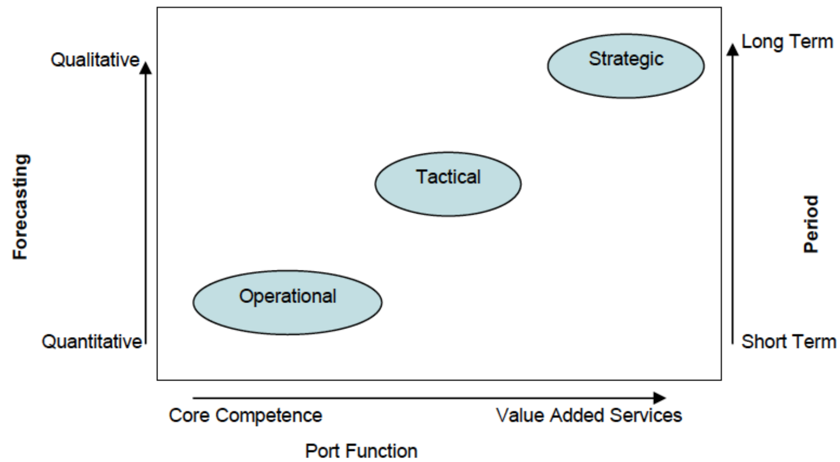


Figure 2.1 Demand analysis as strategic tool.
Source: Gaur (2005) P:40.

The analysis of container demand is a crucial requirement that the decision-makers need to consider in the planning process (see Figure 2.2). The potential demand growth for container ports and the dynamic changes in port competition are two major issues that decision-makers face and need to face in port planning and development.

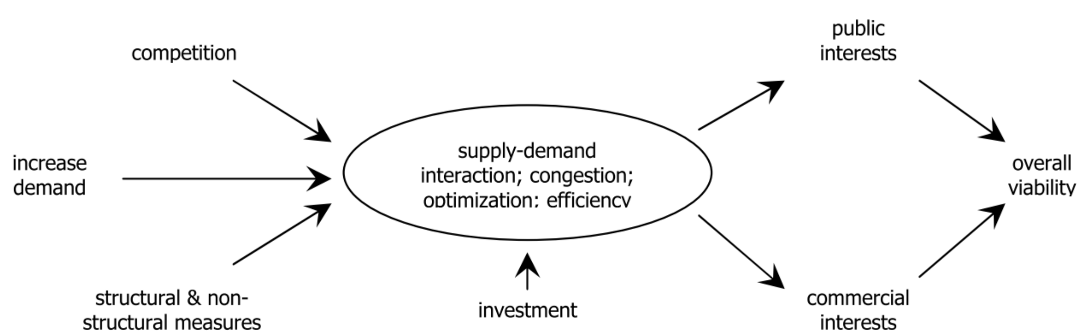


Figure 2.2 Port planning elements.
Source: Dekker (2005) P: 71.

2.2 Forecasting Container Demand

This Section is organized in the following way: 2.2.1 reviews various forecast techniques applied to forecast freight demand. Subsection 2.2.2 reviews studies applying forecast techniques to predict freight demand. Subsection 2.2.3 focuses on comparing short-term freight forecasts with various parametric and non-parametric models. Subsection 2.2.4 focuses on the application and differences of parametric time series models. Also, studies comparing various time series models are presented to predict container throughput. Subsection 2.2.5 presents the characteristics of time series stochastic methodology. Lastly, Subsection 2.2.6 illustrates research gaps.

2.2.1 Forecasting as a Strategic Tool

Demand forecast has been a recurrent topic in the shipping industry. The planning for developing an existing transport infrastructure or investing in a new project raises the question of how the demand growth is expected to evolve and how the service level of the under-planned projects is going to be. Therefore, port planners and decision-makers rely heavily on forecasting container demand which contributes as a tool in making suitable decisions and reducing uncertainty (Stopford, 2009). For example, port authorities need container forecasts to assess future infrastructure projects, while terminal operators use the forecasts to plan for operational decisions.

Before proceeding to the literature, however, it is important to define the major terms from this section. Traffic forecast is defined by Flyvbjerg (2007) and Nicolaisen (2012) as being the estimation of demand, ridership, or traffic for a transport project, either for the short-term, medium-term, or long-term. In addition, Jansson and Shneerson (1982) defined demand forecast as a tool to assess operational and planning decisions by forecasting future demand. Container throughput is a measure that reflects the container volume, expressed in Twenty feet Equivalent Units (TEUs), where TEU is a standard measure of container demand and capacity of ports and containerships.

In the literature of freight forecast, many studies forecasted freight demand at seaports. The diversity across these forecast studies is large in terms of forecast objective, techniques, and study span. Therefore, these studies can be classified into different perspectives:

1. **From the methodology perspective:** Forecast methods can be quantitative, some are qualitative, and others are a combination of both (De Langen et al. 2012).

2. **Form the analysis level perspective**: Forecasts can be applied at the level of analysis such as total port throughput, specific cargo category, or disaggregate commodity level.
    a. Studies may aim to forecast freight at the port or country level, such as the study of Dragan (2017), who predicted seaborne flow of Sweden.
    b. Specific cargo category (container, dry bulk) as in the case of Veenstra et al. (2001) who forecasted seaborne trade flows of crude oil, iron ore, grain, and coal in various global routes.
    c. At a disaggregate commodity level such as the study conducted by Milad (2012) forecasted imports of machinery equipment. De Langen et al.(2012) emphasized the importance to apply forecast models on one type of cargo or category instead of the overall freight volume of the port since cargo characteristics influence model selection.

3. **From the application perspective**: Some studies forecasted freight traffic at the port level, such as Antwerp port in Rashed (2014). Other studies were conducted at a country level, such as Spain in Coto-Millán et al. (2011), Sweden in Henesey (2014), and Indonesia in Milad (2014). Also, some studies were conducted at a regional level, such as van Dorsser et al. (2012), who forecasted port demand at the range of Hamburg – La Herve ports in Northern Europe, and the study of Xie et al. (2013) who forecast container demand of ports at Yangtze River delta in South China.

4. **The objective and forecast horizon perspective:** Forecast studies can also be categorized according to the length of time series data used, which also determines the forecast horizon. Some studies forecast short-term by using monthly or quarterly data; others are aimed at long-term forecasts by using annual data. The forecast horizon is determined based on the research objective where short-term

forecasts are used to assess operational planning, and Long-term forecasts assist in capital investments.

For this dissertation the aim is to forecast short-term container throughput at the port level measured in twenty-foot equivalent units (TEUs), that represents the demand side for the port. Various types of models have been extensively applied to modeling and forecasting freight demand in the literature. Therefore, an overview of the literature is provided in this section by focusing on the empirical studies to forecast short-term demand. The classification adopted in the literature review is according to the conducted forecast method. An overview of overall studies in freight forecast is provided first, and then, the literature review is narrowed down to concentrate on short-term container throughput.

2.2.2 Forecasting Freight Demand

Various forecasting models can be applied to deal forecast freight demand. Choosing the most appropriate model depends on a number of factors: i.e., the forecast context, data availability, the desired degree of accuracy, and forecast horizon (McCarthy et al. 2006). Based upon the extent to which mathematical and statistical methods are used, forecast techniques can be broadly classified into two major categories: qualitative and quantitative techniques. (Chambers et al., 1971). Quantitative techniques focus on analyzing historical demand to generate forecasts based on the assumption that the underlying historical trend will continue in the future. On the other side, qualitative techniques do not rely on analysis methods that heavily require existing historical data. Instead, it mainly employs experts' judgment and accumulated knowledge by using expert surveys as the main source of information to predict future demand. A wide range of forecast approaches, both qualitative and quantitative, were used in the literature. However, applied quantitative techniques are more commonly used to forecast freight demand because of the systematic apparatus it provides and the availability of time series data (Zhao et al., 2013). This section focuses on applied quantitative techniques to forecast freight demand.

Various quantitative methods have been applied in the literature to forecast freight demand. These methods can be categorized into two major types, namely, time series models and cause and effect models. Time series models rely on past observations to forecast future demand. Examples of time series models include classical exponential smoothing, Holt-Winters' exponential smoothing, growth curve, and Stochastic models (i.e., ARMA and ARIMA). On the other hand, cause and effect models identify and measure the explanatory relationship between the forecasted variable and one or more predictor independent variable(s). Examples include linear regression, system dynamics, nonlinear regression, traditional neural networks, and artificial neural networks.

Various forecast models are applied in the literature to predict freight demand with causal models being widely applied in previous studies e.g., (C. C. Chou et al., 2008; Patil & Sahu, 2016; Seabrooke et al., 2003). Since causal models rely on the causal relationship between variables, the determinants of freight demand behavior have to be analyzed as with Coto-Millán et al. (2005). However, estimating demand by using traditional regression approaches is based on time series data (a data that is non-stationary in real-life time series) where models such as Ordinary Least Square (OLS) assumes that the time series is stationary. Broadly, a time series is defined as nonstationary when its mean, variance and covariance values between equal lag length are constant over time (Schmukler et al., 1997). Ignoring non-stationary and not modifying the errors leads to spurious regression (Rashed et al., 2015; Syafi'i et al., 2005).

To improve forecast accuracy, other studies combined regression models with various forecasting models such as Linear regression and Neural Network (Gosasang et al., 2011; Lam et al., 2004), Linear with gray model Chen and Gu (2010), Grey model with logistic growth curve models (Zhao et al., 2013), and traditional fuzzy set theory and regression analysis (Chou et al., 2003). Combined regression, compared to singly models provides better forecasts.

It is shown in the literature that studies applied to forecast freight demand were mostly conducted to predict long-term demand where traditional regression models were widely adopted. On the contrary, limited number of studies were conducted on short-term freight demand at seaports. In the next subsection, studies conducted to forecast short-term freight demand are reviewed.

2.2.3 Short-term Freight Demand Forecast

To forecast short-term demand, both parametric and non-parametric time series models are applied in the literature. In recent years, the popularity of non-parametric in forecasting short-term freight demand increased due to their ability to capture time-series patterns and to improve forecast results by trial-and-error methods (Gautam & Singh, 2020). Examples of non-parametric models includes Neural Networks (NNs), Genetic Programming (GP) and Error Correction Models (ECM). These models, among others, were applied in the literature to forecast freight demand.

NN models are soft computing methods that are able to model seasonality and changes in the time series trend (Lam et al., 2004). Lam et al., (2004) used the Neural Network model to forecast freight demand at Hong Kong Port. Zhang and Kline (2007) applied 44 NN models into a large dataset of 756 quarterly time series. NN models are able to deal with changes in the time series pattern to overcome the linearity assumption (Lam et al., 2004). Other non-parametric models applied to forecast freight demand includes Error Correction Models (ECM) as in the studies of (Fung, 2002; Lam et al., 2004). Genetic Programming as in (Chen & Chen, 2010; C. C. Chou et al., 2008), and Least Squared Support Vector Regression (LSSVR) (Xie et al., 2013). However, non-parametric models have some deficiencies.

Despite the fact that non-parametric models provide accurate forecasts, neither provide a structural mathematical formation or functional form (Farhan & Ong, 2018); thus, making it difficult to understand how the forecasts are generated. Also, non-parametric models require a large number of time series observations, proper time series preprocessing and transformation where there is no literature that provides suitable transformation to be utilized to improve forecasts (Gautam & Singh, 2020).

On the contrary, parametric models are known for their transparency where they assume that the stochastic stationary process has a specific structure that can be described by mathematical expressions that contain a specific number of parameters (Farhan & Ong, 2018). Furthermore, parametric models allow the historical pattern to be differentiated from random components in the time series and avoids overfitting by applying various parameter selection criteria. The above-mentioned limitations of non-parametric models led to the selection of parametric over non-parametric models to forecast short-term container demand.

2.2.4 Univariate Time Series Forecasting Model

Various types of time series forecast methods are used to predict short-term demand based on the historical observations of demand. Examples of time series models are exponential smoothing, gray model, classical decomposition, and Box-Jenkin (ARIMA) models. Various models were applied in previous studies to forecast freight demand.

Choosing the right method depends on the port characteristics and behavior of historical freight throughput. Exponential smoothing (ES) techniques are simple tools used for smoothing and forecasting the time series by assuming that most recent observations are more important for a forecast; thus, applying weights on observations that decrease exponentially with time (Lemke, 2010). Ee et al. (2014) forecasted container throughput by using Halts and Winters Exponential Smoothing and SARIMA model and stated that both models provide reliable forecasts; but they indicated that the flexibility of SARIMA model in dealing with autocorrelation makes it more suitable in providing reliable forecasts.

Another time series model is Grey Model (GM). This model is suitable to forecast demand when issues such as insufficient information and uncertain behavior are encountered (Deng 1982). It also does not require a large dataset nor prior knowledge (less computational work). GM is applied in the studies of Jiang and Lei (2009); Huang et al. (2003) to forecast cargo throughput. The selection of a proper model doesn't only depend on the forecast purpose and horizon (short/medium/long term) but also on port characteristics and historical time series structure and pattern (seasonality and trend).

However, capturing the seasonality variation on the time is crucial for the operational planning of ports. Liang and Chou (2003) and Chen (2010) analyzed container throughput at three Taiwanese ports and argued that Chinese New Year has a significant impact on container throughput at all ports in Taiwan. Furthermore, capturing the seasonality effect in the historical demand increases forecasting reliability (Xie, 2013). This is due to the impact of seasonal fluctuations on the variations of the forecasted dataset. Gray and exponential smoothing models lack the ability to deal with seasonal variations. On the other hand, the classic decomposition and Box-Jenkins time series methods allow to model seasonal time series; thus, they provide the ability to remove seasonal factors before forecasting demand. This is done by decomposing the time series into four separate components: trend, cyclical, seasonal, and random components.

Various studies were conducted to compare the reliability of time series models where the impact of ignoring seasonal variation in some models can been seen in the forecast results of some models. The model reliability, in these studies, is based on measuring the forecast error in the predicted results by using various error measures such as the mean absolute error, mean absolute percent error, and the root mean square error. A comparison of six univariate and causal models is conducted by Peng & Chu (2009) in forecasting container demand of three Taiwanese ports of the period 2003-2006. The applied models are a combination of causal and time series models, namely, the classical decomposition model, trigonometric regression model, regression model with seasonal dummy variables, grey model, hybrid grey model, and SARIMA model. The comparison showed that classical decomposition model, and to a lesser extent SARIMA model, provides the most accurate forecasting results.

In addition, the Grey Model provided the poorest forecast accuracy due to its inability to deal with seasonality properly. The authors indicated that the monthly observations of the three ports exhibit a seasonal pattern and sharp decrease in the container throughput in February of each year due to the Chinese New Year holidays. A study is conducted by Huang et al. (2020) based on the same dataset, forecast models, and time horizon as the study of Peng & Chu (2009). However, the authors stated that using the formal statistical theory to estimate SARIMA model and forecast container demand revealed its superiority compared to Classical decomposition.

Furthermore, Dragan et al. (2014) compared the results of three univariate time series models, namely, classical decomposition, Holt-Winters, and ARIMA models. By applying the models on the quarterly container throughput for the period 2002-2012 of four Northern Adriatic ports, the outcomes showed that the ARIMA model resulted in better forecasting results compared to less complicated models. They indicated that this is due to the capability of SARIMA in delineating the seasonal variation. Despite the fact that many forecasting models have been applied in previous studies, ARIMA models tend to outperform the other time series models due to their ability to deal with seasonality, stationary, and autocorrelation.

2.2.5 The Characteristics of Time Series Stochastic Models

Unlike its counterparts, the formal and structured approach of ARIMA, developed by (Box and Jenkins 1976), flexibly incorporates the dynamic structure of the time series, which provides ARIMA an advantage in forecasting short-term demand (Farhan, 2018). Therefore, ARIMA approach is drawn attention in the literature due to its capability of identifying and analyzing: a) the dynamic and systematic variation in the series of data (seasonality); b) the cyclical patterns; c) the presence of trend patterns; and d) the historical growth rates.

Accordingly, the process of the Box-Jenkins forecast approach provides a systematic methodology in identifying and dealing with seasonal variation and non-stationary nature of real-life time series observations (Siami-Namini et al., 2018). To identify model parameters and forecast demand, Box and Jenkins (1976) proposed a systematic methodology, known as Box-Jenkins methodology. This methodology consists of three phases: 1) model identification, 2) model estimation and diagnostic testing, and 3) forecast application.

The literature in freight demand forecast shows that applying Box-Jenkins time series approach provides high forecast performance in the short-term horizon (Kim, 2008; Shin, 2011; Pak & Yeo, 2011; Rashed, 2016; Farhan (2018)). To account for the seasonal component, other studies used the Seasonal Autoregressive Integrated Moving Average (SARIMA), an extended version of ARIMA, to forecast freight throughput (Chou, 2003; Kim, 2007). These studies reveal the significance of ARIMA and SARIMA models in forecasting freight demand. However, as the forecast horizon increases, the ability of ARIMA models to generate accurate forecasts deteriorates, because these models are ineffective at capturing long-term variation in the time series. Nevertheless, ARIMA models are proven to provide accurate short-term forecasts.

2.2.6 Conclusion

To sum up, the review performed on extant forecast literature provides some insights and recommendations. From the literature review, it is concluded that a reliable forecasting model from a specific port does not mean it suits other ports' forecasts. Therefore, there is no inclusive best model that outputs reliable results for all ports; so, choosing a forecasting technique does not only depend on the forecast purpose and horizon (short/medium/long term) but also on port characteristics and historical time series structure and pattern (seasonality and trend).

Both short-term and long-term perspectives were considered in forecasting container demand. However, long-term demand was widely forecasted in previous studies, where as a limited number of studies forecasted short-term demand. Most studies applying time series models to forecast short-term demand lack consideration of seasonal variations which appears to be crucial for short-term port planning and operational decisions. Hence, issues related to time series behavior such as periodicity and seasonality may not be appropriately addressed. Furthermore, to deal with these issues, ARIMA models appear to provide a systematic approach to understand time series behavior and deal with issues related to historical patterns of container demand such as non-stationary and seasonal variations.

A limited number of studies applied time series stochastic models to forecast short-term container demand. In most cases, studies applying ARIMA to forecast container demand were either using tones, a naive conversion method from tones to TEUs or lack explaining the conversion method. From a practical perspective, using TEUs to forecast the container throughput is crucial and more beneficial to terminal operators and port authorities since the decision making depends on the number of handled TEU. Therefore, there is a need for developing forecasting models for short-term demand by using monthly TEU and incorporating seasonality and considering time series stationarity.

Additionally, the literature shows that global financial and economic crises, such as the 2008 financial crisis, play a leading role in container demand. Most studies were conducted on time series observations for the periods before the 2008 financial crisis. Also, empirical studies that were conducted to forecast container demand of forecast horizons that covers the crisis period impacts the performance of certain models. To deal with the complexity and dynamic behavior of container demand, and to ensure a rigorous support instrument for the decision-making process, ARIMA models are applied to understand the structure of the series by analyzing the historical demand pattern; thus, avoiding the negative impacts of such major changes in the time series pattern. The literature also shows that the majority of studies applying ARIMA models were concentrated in ports located in developed countries. In this perspective, ARIMA model is applied in this dissertation to forecast demand based on the historical time series from 2003 to 2018 for the port of Jeddah in Saudi Arabia.

2.3 Port Competition

In this section, the content is arranged in the following way: Subsection 2.3.1 presents the evolution of port competition and its impact hinterland expansion. Subsection 2.3.2 investigates the impact of various factors on port choice process, along with the relationship between port competitiveness and port choice. Subsection 2.3.3 investigates previous studies applying Spatial Interaction Models (SIM) to analyze the distribution of container flow. Also, a brief explanation of SIM. Lastly, Subsection 2.3.4 reviews previous studies applying gravity models and introduces research gaps.

2.3.1 The Evolution of Port Competition and Hinterland

During the noticeable transformation of the maritime industry, the traditional concept of port hinterland developed throughout the years. For instance, Intermodalism expanded port hinterland by transforming the port market from being monopolistic to becoming more competitive. Container ports, as a result, are able to expand market and reach further inland markets. Based on these developments, the competition among ports, in some regions, serving the same hinterland intensified (Ferrari et al., 2011). As a result, the competition transformed from port competition to competition between supply chain. Therefore, a port's ability to attract containers depends on the overall supply chain where the port is involved (Figure 2.3). The relationship between ports is considered as inter-port competition when port users have various economic port alternatives (Cullinane & Wang, 2009). More information on supply chain is provided in Subsection 2.3.2.

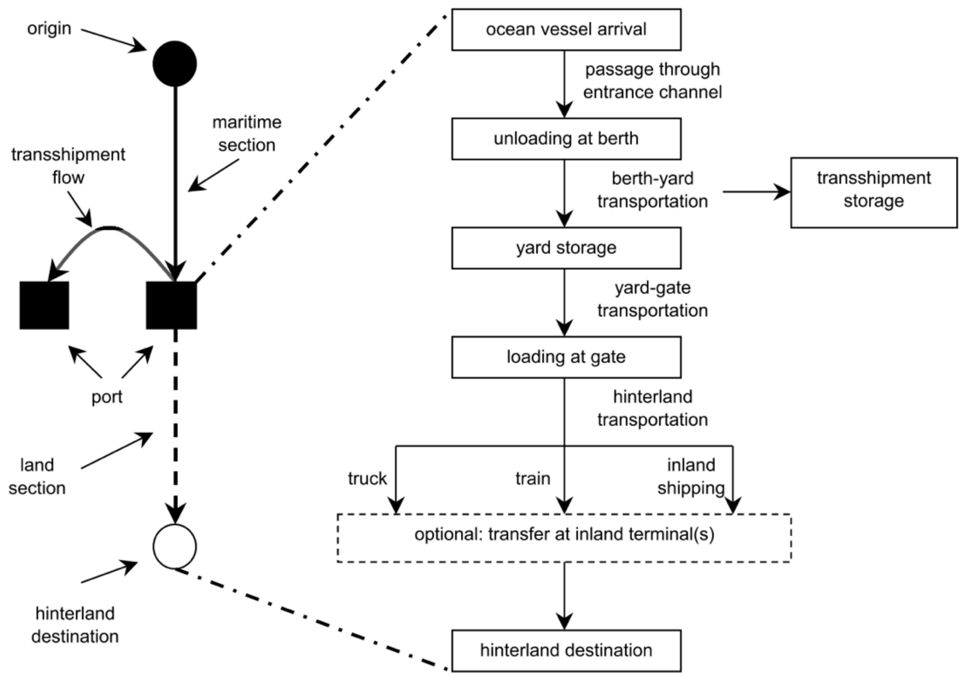


Figure 2.3 Container transfer process at a seaport and within the supply chain.
Source: Dekker (2005) p:38.

Winkelmans et al. (2002) classified port competition into three levels: (1) inter-port competition at port authority level between different ports; (2) inter-port competition at operator level between operators from different ports located in the same range; and (3) intra-port competition at operator level between operator in the same port (Figure 2.4).

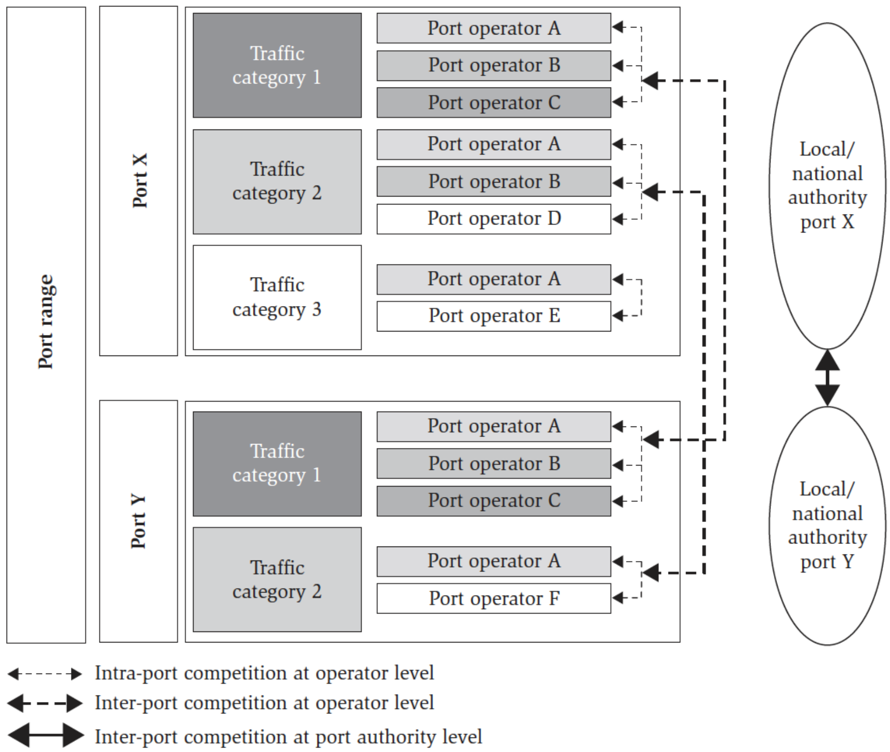


Figure 2.4 Port competition levels.
Source: Winkelmans et al. (2002) p. 12.

2.3.2 Determinants of Port competition

The evolution of the container industry has changed the strategic function of the port where ports are increasingly competing as essential nodes in the supply chain rather than individual loading/discharging nodes (Robinson, 2002; Carbone and Gouvernal, 2007). Supply chain is defined by Christopher (1992) as "*the network of organizations that are involved, through upstream and downstream linkages, in the different processes and activities that produce value in the form of products and services delivered to the ultimate consumer*.".

Such changes in the industry necessitate a supply-chain-oriented approach in the port choice decision of liner companies and shippers; thus, this viewpoint has become dominant has become in the assessment of port competitiveness (Guy and Urli, 2006).

Consequently, the supply-chain-oriented approach indicates that the competitive position relies more on external factors (Ferrari, 2011). Port competitiveness, as a result, is determined not only by port efficiency but also by the efficiency of factors outside the port domain, particularly connectivity to the hinterland and accessibility to global shipping networks

Van der Berg and van Den Berg (1998) defined port hinterland as *"the continental area of origin and destination of traffic flows through a port, in other words, it is the interior region served by the port."*. On the other side, Parola et al. (2017) state that Maritime connectivity reflects the adequacy of transport networks (e.g., number of ships calling at a port, service frequency, distance to/from the global shipping network). Figure 2.5 depicts the transport network where ports act as nodes, connecting the hinterland and maritime sections.

In a competitive environment where ports provide feasible substitutes, the importance of port competitiveness increases. Consequently, the demand for competing ports and capacity becomes interrelated. For example, when services are improved in a particular port, it influences the demand of the particular port and its competitors. This is due to the increasing requirements caused by the selection behavior of port users where they choose the most efficient port. Given that port selection is influenced by port efficiency, port selection determines the port that is thought to be the most competitive one (Moura et al., 2018). Therefore, the determinants impacting port competitiveness impacts port choice as well.

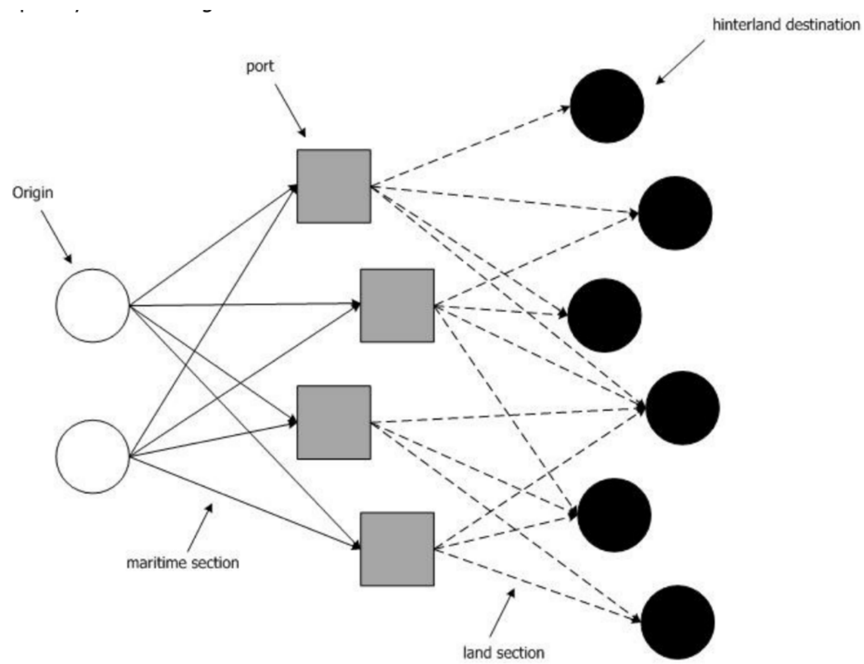


Figure 2.5 Ports as nodes in a transportation network.
Source: Dekker (2005) P:9.

Numerous studies analyzed the factors that impact port selection, which are shipping cost, the efficiency of port infrastructure, port location, transport time and cost, frequency of liner services, foreland and hinterland accessibility, and inland transport cost and time (Talley & Ng, 2018; Tiwari et al., 2003). The importance of factors impacting port selection is difference for liner companies compared to shippers, as each one has different preferences. Therefore, factors impact the selection behavior of shipping companies and shippers varies for different port users. While shipping companies tend to consider costs at port as the major factors that influence the choice of calling at a particular port, the decision of shippers to use a particular port is influenced by port location and hinterland connections (Acosta, 2007; Parola et al., 2017b).

These studies, among others, emphasize the impact of port location on ports' competitive position. The importance of port location raises from its role on connecting the port with inland markets and maritime network where well-geographically located ports contribute positively by connecting shippers and consignees in the port's hinterland to international markets. Consequently, Tongzon (2001) emphasized the importance of port connectivity on improving the performance of the port. Port connectivity can be broadly defined to encompass the whole transportation network by including connections to both hinterland and maritime (de Langen et al., 2016). Therefore, The higher a port's connectivity, the more value it provides to its users.

Factors influencing the port selection process, whether they are in the port domain or not, are of importance to port authorities. They can be categorized as internal and external factors. Where Internal factors, such as port efficiency and cost, are under direct control of port authorities where they have full capability to improve them. On the other side, external factors such as those related to port location influence port selection. Although port choice depends on factors beyond the port domain, port authorities can indirectly influence port choice. A particular port, for instance, can adopt more proactive measures to improve its competitive position by identifying the preferences of port users for other ports (Martnez Moya & Feo Valero 2017). Malchow and Kanafani (2004) emphasized that factors **out of port authorities' control** have the most significant impact and stated that port location has the most significant impact on shipment volumes. This is supported by other studies including (Moura et al., 2018; Parola et al., 2017a; Veldman et al., 2011).

Considering the hinterland connectivity, one of the factors that heavily determines the port competitiveness over market share is the inland connectivity. Inland connectivity refers to the efficiency of inland transport networks (Parola, et al., 2017). Halim et al.

(2016) highlighted that the key factor that impact port selection is the hinterland connectivity. Therefore, inland distance between ports and the hinterland of region impacts the container flow of ports. The distance between port and users in the hinterland has a key impact on inland transport costs (Tiwari et al., 2003). Notteboom and Winkelmans (2001) emphasized that inland transport cost has an average of approximately 40% of total container transport cost. port choice is influenced by inland transport cost (Blonigen & Wilson, 2006). The shippers' choice of a specific route depends on the lowest route cost among alternatives (Luo & Grigalunas, 2003). In this sense, cargo volume decreases when the inland distance between the port and the inland market increases (Guerrero, 2018; Levine et al., 2009; Moura et al., 2018; L. Wang et al., 2018). From a statistics perspective, (Malchow and Kanafani (2004) compared the impact of inland distance and maritime distance on port choice, from the perspective of liner companies. They stated that inland distance is more significant than maritime distance in case of the availability of feasible alternatives.

The impact of inland distance on the freight flow from/to the port hinterland varies among different regions and depends on the geographic configuration where the distance impact is weaker in geographically wider countries. This can be seen in the case in the United States (Levine et al., 2009) where the friction of distance is relatively weak compared to other countries that are geographically smaller such as Italy (C. Ferrari et al., 2011), Spain (Moura et al., 2017) and France (Guerrero, 2018). In addition to geographic characteristics, the availability of alternative intermodal connectivity, transport infrastructure development, technology and transport cost influence the distance on container flow (Guerrero, 2014; J. Rodrigue, 2012).

Another determinant that heavily determines the port competitiveness is the maritime connectivity. Maritime connectivity is vital for ports since the length of maritime haulage impacts shipping cost. Maritime connectivity is referred to as the accessibility to maritime routes and liner services in a particular geographical location (Pitoski et al. 2015). In this dissertation, foreland accessibility and maritime connectivity are used interchangeably. Various studies have emphasized the impact of maritime connectivity to shipping networks on trade costs (Arvis et al., 2016; Wilmsmeier & Hoffmann, 2008). Furthermore, (Hoffmann, 2012) demonstrated the importance of maritime connectivity to container shipping routes in determining port competitiveness. The better maritime connectivity, the more options port users have for transporting containers to and from international destinations and origins; thus, maritime connectivity is of relevance to port users.

To assess maritime connectivity, various centrality measures have been employed where closeness centrality degree centrality, and betweenness centrality are the most commonly used (Dinu et al., 2016; Ducruet et al., 2010; Li et al., 2015; Wang & Cullinane, 2016; Zheng et al., 2017). These connectivity measures are calculated based on the number of links in the network and/or the number of connections available at a particular port in the network. To account for the quality of port connections, other connectivity measures have been developed to assess the quality of port connectivity based on factors such as connection capacity and competition level in the connection (Burghouwt & Redondi, 2013). Lam and Yap (2011), for example, simply used both the number of ships calling at a particular port with ship capacity in TEUs to take into account the dynamics of liner shipping network.

However, in global container shipping, maritime connectivity, as well as the frequency and intensity of liner shipping services, are essential elements to determine port competitiveness (Hoffmann, 2012). Several indicators of connectivity have been developed specifically for liner shipping networks to reflect the dynamic structure of container shipping networks such as the Lloyd's Shipping Index and Liner Shipping Connectivity Index (LSCI). The following components were considered in developing maritime connectivity indicators in previously developed maritime connectivity indicators, according to a review study by Pitoski et al. (2015): Ship capacity, number of ships call in the port, frequency of port calls, number of liner services to/from ports, number of liner shipping providers and transit time. These components are believed to have an impact on the generalized transport costs of port users. Certain components such as ship size and the number of liner services provided to/from the port) are regarded as proxies for transport costs, whereas others are related to the quantity and level of connections, such as the number of direct port calls (de Langen et al., 2016).

United Nations Conference on Trade and Development (UNCTAD) developed the Liner Shipping Connectivity Index (LSCI) to measure the shipping Connectivity and performance of individual ports (UNCTAD, 2021). The index, published in 2019, covers more than 900 container ports globally and generates the degree of port connectivity. This index is the normalized average of six components: (a) The number of ship calls per week in the port; (b) The total deployed annual capacity in TEU offered at the port; (c) The number of liner shipping services from/to the port; (d) The number of liner services providing services to/from the port; (e) The average carrying capacity in TEU of deployed ships by scheduled services with the largest average ship size; and (f) The number of other ports that are directly connected to the port via liner services. A direct service is regarded as a regular liner service between two ports which may include other port calls in the route, but does not require the transshipment of the transported containers. (UNCTAD, 2021). The port's value of each component is divided by the maximum value of that component in the base year of 2004, then, the total average of all six components for the particular port is calculated and divided again by the maximum average in 2004 and multiplied by 100. This is done for all the ports to generate the LSCI value where the higher the index value, the better the ports connectivity in the global liner shipping network (UNCTAD, 2021).

A distinctive feature of LSCI is that it measures port connectivity based on the strength position of the port within the dynamic structure of container shipping network, rather than simply relying on local data such as ship numbers, container volumes handled or direct connections to other ports (Bartholdi et al., 2016). It is also noted that no study in the existing literature was found to apply port LSCI in the SIM. However, a number of studies used the country version of the andix in various analysis. Given that LSCI reflects connectivity strength and the port, the study of De Oliveira and Cariou (2015) demonstrated that LSCI is a key factor in determining freight rates and competitive position of ports. This is in line with Wilmsmeier and Hoffmann (2008) who emphasized the importance of port connectivity on transport cost. Consequently, the index can be employed as a proxy of the port competitiveness, with the view that it is positively correlated to port efficiency (De Oliveira and Cariou, 2015). Therefore, LSCI is used in this dissertation to investigate the impact of port efficiency and assess port competitiveness by analyzing connectivity position of different ports.

2.3.3 Assessing Port Competitiveness: Spatial Interaction Model

Port competitiveness and port selection have been widely analyzed in previous studies. The theoretical foundation of applied models in the majority of these studies is the Discrete Choice Theory where Multinomial Logit Models (MNL) were mostly applied to analyze port selection process (Moya and Valero, 2016). MNL models were also applied to analyze freight flow in the maritime industry. Other models, such as Spatial Interaction Models (SIM), were also used to analyze the spatial interaction between origin/destination nodes. However. assessing port competitiveness by considering the spatial implications is limited in previous studies. Spatial Interaction Models (SIM) analyze flow between bilateral points (origin and destination) as a function of attractiveness and emissiveness of both origin and destination, respectively, and a vector of distance variable (Wilson, 1976). Even though both SIM and MNL approaches use an explicative-stochastic perspective and revealed preferences, their theoretical approaches are different (Anas ,1983).

Compared to Multinomial Logit Models (MNLs), SIM has several advantages. First, despite the structural similarity of SIM and Discrete choice models, SIM is more appropriate in dealing with aggregate data compared to discrete choice models due to the minimal initial assumptions of the former (Anas, 1983; Roy, 2004). Secondly, SIM have the ability to take into consideration the impact of spatial characteristics as well as transport network characteristics (Kerkman et al., 2017). Thirdly, Merkel (2017) argues that as ports have fixed geographical locations, the distance separating them impacts the intensity of their competition; thus, the inland distance separating ports and hinterland regions can be used as a proxy to estimate the unobservable impact of interdependence degree among ports. Considering the difficulty of calculating cost between origin and destination points due to the lack of disaggregate transport cost data, distance can be used as a proxy for trade cost Anderson and van Wincoop (2004).

Spatial interaction model has been a limitedly applied in the transport economics, and its scarcer in analyzing container flow between ports and hinterland regions. Previous studies concerned in analyzing the relationship between distance and traffic flow have demonstrated that SIM can be applied (Ferrari et al. 2011; Guerrero 2014). Debrie and Guerrero (2008) used SIM to investigate the impact of distance on freight flow between French ports and regions. They observed that distance is an important variable in explaining freight distribution in France. Guerrero (2014) investigated the impact of inland distance on different types of cargo and concluded that the impact varies according to the cargo category, where container flows are the least sensitive to its influence. Guerrero (2016) applied SIM to investigate the impact of inland distance and shipper services, finding out that considering both variables improve the model results. Tiller and Thill (2017) analyzed the degree of trade impedance in South American exports by using a reverse doubly constrained SIM.

Additionally, the expansion of port hinterland can be analyzed by the use of Spatial interaction model (SIM) (Garcia-Alonso et al., 2016). During the last decade, only four relevant articles have been published focusing on the distribution of the container flow from the perspective of the spatial interaction models. Ferrari et al (2011) investigated the explanatory power of distance in explaining freight distribution in North Italian ports, concluding that distance is has high influence but other barriers define the hinterland. To delimit the hinterland shape, Moura et al (2017) applied origin constrained SIM by using the variables of port throughput and travel time to inland regions. Moura et al (2018) applied origin constrained SIM to investigate if the impact of port size and inland distance on inland flow distribution vary according to the foreign trade pattern of Spain. Then, the authors used the predicted results to assess the discrepancy of each port by considering the gap between predicted and actual flows. Lately, Gruerrero (2019) assessed the importance of transport connections and geographical proximity in the Freight distribution of France. The author also investigated if these factors have changed between the years 2008-2012.

Analyzing the difference between the outcomes of SIM and actual flow may reveals the ability of each port to compete over the hinterland regions; thus, provides an insight into the impact of other barriers in preventing each port from expanding hinterland region; thus, the competitive advantage. Among the above-mentioned studies, Ferrari et al. (2001), Moura et al. (2017) and Moura et al. (2018) attempted to assess port competitiveness and delimit hinterland scope by investigating the unexploited potentialities of ports. However, the former study merely considered the impact of inland distance, while the other two studies did not consider the impact of foreland determinants.

The literature also shows a lack of studies for port competitiveness in developing countries. Guerrero (2018) pointed out that the issues of port selection are not of relevance in the case of developing countries where high inland costs might cause stronger effects of distance that leads to limiting the contestable hinterland. However, no studies were conducted in the literature to analyze the impact of distance on the hinterland expansion on developing countries.

2.3.4 Conclusion

To sum up, the impact of various factors based on port selection process where, in addition to port efficiency, foreland and hinterland connectivity are two important determinants of port connectivity. It is worth mentioning that the factors impact port competitiveness have direct impact on container flow between ports and hinterland regions, where the characteristics of these factors impact cost, capacity and reliability of transport services. The interaction between these factors raises the need to investigate their impact on port competitiveness. In the literature, numerous studies are concentrated on identifying the factors that determine port competitiveness. However, limited number of studies assessed port competitiveness by considering the impact of spatial interaction on container flow. In the next subsection, literature review on port competitiveness is presented and the theoretical foundation of various models are briefly presented.

Spatial Interaction Models are gaining more attention and being employed to investigate the pattern of maritime throughput on the hinterland. However, a limited number of studies applied SIM to assess port competitiveness by analyzing their unexploited potentialities (C. Ferrari et al., 2011; Moura et al., 2018). These two studies did not consider the influence of maritime; that is, factors related to the determinants of the overall transport chain must be considered. Additionally, it is also noted that no study in the existing literature was found to apply port LSCI in the SIM, therefore, contributions of this dissertation. this study aims to use Port Liner Shipping Connectivity Index (LSCI) as an attractiveness factor that explains port connectivity.

This dissertation aims to investigate the impact of geographic characteristics of competing ports and intermodal availability in the distribution of container flow in the country of Saudi Arabia and to assess port competitiveness in Saudi Arabia or not. Furthermore, Gap analysis is applied to analyze unexploited potentialities of competing port to assess their competitive position. Because ports competitiveness relies on the capability of attracting traffic from global maritime networks (foreland) and their inclusion in the hinterland, the impact of factors related to both of hinterland and foreland are considered.

2.4 Research Questions

RQ1: How to develop a quantified model to forecast short-term container demand by considering the time series pattern?

The aim of this question to provide an instrument to analyze historical behavior of container throughput and forecast short-term demand by using monthly container throughput. This is done by developing a forecasting model using the univariate time series stochastic approach.

RQ2: How does the explanatory power of inland distance evolve over time in the distribution of containers over the hinterland?

The aim of this question is to test the relevancy of inland distance in shaping the hinterland of ports. To do so, a spatial interaction model (SIM) is employed to investigate the role of inland distance in drawing market share of competing ports over hinterland. Thereafter, gap analysis is applied to investigate the unexploited potentialities of competing ports by comparing the actual and predicted container flows. The gap implies the presence of barriers has not included other than distance.

RQ3: How does the geographic characteristics and intermodal connectivity impact the distribution of container flow and their role in the competitive position of the competing ports?

The aim of this question is to investigate the impact of port characteristics and geographical location on the inland distribution of maritime traffic and to assess port competitiveness. The impact of maritime connectivity, railway availability, inland distance, and port location within maritime shipping routes on the container flow is investigated. Attraction-constrained Spatial Interaction model is developed where the influence of these factors is investigated. Thereafter, port competitiveness is assessed by analyzing the gap between predicted and actual container flow to reveal the impact of the various factors on the hinterland expansion of the competing ports.

CHAPTER 3

METHODOLOGY

In this section, the methodology conducted to achieve the dissertation objective is presented consists of two major sections. Section 3.1 presents the time series stochastic models that is used to forecast short-term container demand, and the methodology followed to analyze time series historical behavior, identify and estimate the tentative models, and generate forecasts.

Section 3.2 introduces two different version of Spatial Interaction Models (SIM) to analyze the impact of various factors on container flow and assess port competitiveness, where Subsection 3.2.1 presents the development and types of Wilson maximum entropy SI models. Followed by explaining the model calibration. Attraction and doubly constrained maximum entropy models are used to investigate the explanatory power of distance in defining container flow.

Subsection 3.2.2 presents the development of the Poisson-based SI models. This Subsection also presents and compare the assumptions of various statistical approaches that are used to calibrate SIM. Poisson-based SIM models are known for being flexible to fit additional variable. Poisson-based attraction constrained model is used to investigate the impact of maritime connectivity, railway availability, inland distance, and port location within maritime shipping routes on the container flow between ports and hinterland region.

3.1 Forecasting Container Demand

The Autoregressive Integrated Moving Average (ARIMA) is a model in which forecasts are obtained by regressing historical observations of the variable itself and the current value with the error terms of the past values at different lags. This model has the advantage of being independent in finding and forecasting explanatory variables by analyzing the impact of time on the relationships among time-ordered variables. Considering the rapid, volatile, and unexpected changes of container demand that are due to factors other than its own trend, the forecasts of this model are for a short period. The purpose of applying this methodology is to investigate the historical attitude of container throughput and understanding its future direction and development, evaluate the time series pattern, and produce short-term forecasts.

3.1.1 Univariate Time Series Model

The idea of a univariate time series model is using the past values in the series to forecast future values, where Autoregressive AR is a combination of the past value of the series $(Y_{(t-1)}, Y_{(t-2)}, \dots, Y_{(t-p\,)})$ and Moving Average MA is the combination of the historical random errors $(a_{(t-1)}, a_{(t-2)}, \dots, a_{(t-q)})$ as explanatory variables. The AR and MA models are, respectively, represented as:

$$Y_t = c + \emptyset_1 Y_{(t-1)} + \emptyset_2 Y_{(t-2)} + \cdots + \emptyset_p\, Y_{(t-p\,)} + \varepsilon_t \tag{3.1}$$

$$Y_t = \mu + \theta_1\, a_{(t-1)} + \theta_2\, a_{(t-2)} + \cdots + \theta_q\, a_{(t-q)} + \varepsilon_t \tag{3.2}$$

where Y_t is the time series value at period; t, ε_t is the random error at time period t; $\emptyset_p$ is the AR parameter,;θ_q is the MA parameter; and μ is the time series mean.

Autoregressive and moving average models can be effectively combined to form an extended time series model. The extension of AR and MA is called Autoregressive Moving average ARMA(p,q), where p and q are the AR and MA are the parameter orders. The ARMA model is represented as:

$$Y_t = \emptyset_0 + \emptyset_1 Y_{(t-1)} + \ldots + \emptyset_p Y_{(t-p)} + a_t + \theta_1 a_{(t-1)} + \ldots + \theta_q a_{(t-q)} \quad (3.3)$$

where $\emptyset$s and θs are the autoregressive and moving average parameters, respectively, to be estimated. The random error term a is assumed to be to be a white noise process and follow the normal probability distribution $a_t \sim N(0, \sigma^2)$.

Most of the economic time series exhibit nonstationary behavior. Non-stationarity indicates that the distribution of the series depends upon time. Usually, the nonstationary behavior of time series is due to a trend, a change in the mean, or seasonal variation. Since stationarity is one of the fundamentals for Box-Jenkins methodology, the nonstationary series must be made stationary around mean and variance. Therefore, if non-stationarity is found, the data must be transformed by linear detrending or differencing to remove trends and achieve stationarity. Making this modification to the series is called an integration process. By doing so, the ARMA model becomes an Autoregressive Integrated Moving Average ARIMA(p,d,q) model, where the integration order (I) refers to the integration process and (d) is the order of differencing (the number of time differencing applied into the series to get it stationary). It is referred to as Δ^d. The differencing also represents the number of unit roots.

$$\Delta^d Y_t = \emptyset_0 + \Delta^d \emptyset_1 Y_{(t-1)} + \ldots + \Delta^d \emptyset_p Y_{(t-p)} + a_t + \theta_1 a_{(t-1)} + \ldots + \theta_q a_{(t-p)} \tag{3.4}$$

Some economic series exhibit seasonality also. If that occurs, another modification is done to the ARIMA equation to deal with seasonality, meaning that the ARIMA model is extended to SARIMA(p,d,q)(P,D,Q) namely Seasonal Autoregressive Integrated Moving Average. In the SARIMA model, seasonal differencing of appropriate order is used to remove non-stationarity from the series, where d is the differencing order to make the series stationary, D is the order of seasonal differencing, and s is the seasonal period. The model is expressed as:

$$\emptyset_p(B) \Phi_p(B^s) \Delta^d \Delta_s^D Y_t = \theta_q(B) \Theta_Q(B^s) a_t \tag{3.5}$$

Where:

$\emptyset_p(B)$ = nonseasonal autoregressive operator of order p. = $1 - \emptyset_1 B - \emptyset_2 B^2 \ldots - \emptyset_p B^p$

$\theta_q(B)$ = nonseasonal moving average operator of order q. = $1 - \theta_1 B - \theta_2 B^2 - \cdots - \theta_q B^q$

$\Phi_P(B^s)$ = the seasonal AR operators of finite orders P= $1 - \Phi_1 B^s - \Phi_2 B^s - \cdots - \Phi_p B^{sp}$

$\Theta_Q(B^s)$ = the seasonal MA operators of finite orders Q= $1 - \Theta_1 B^s - \Theta_2 B^s - \cdots - \Theta_Q B^{sQ}$

Δ^d = the nonseasonal operator = $(1 - B)^d\, Y_t = Y_t - Y_{t-1}$

Δ_s^D = the seasonal differencing operator = $(1 - B^s)\, Y_t = Y_t - Y_{t-s}$

B = the lag operator

In order to determine the suitable model, the following steps are required:

1) Determine the difference order of d by using Autocorrelation Function (ACF), Partial Autocorrelation Function (PACF), and Augmented Dickey-Fuller (ADF) unit root tests.

2) Use correlation coefficients to determine the AR and MA lags p and q, respectively.

3) Estimate the value of $\emptyset_p$ and θ_q parameters.

4) Envision and determine the effect of seasonality with ARIMA.

To perform these steps, the methodology of Box-Jenkins is used as explained in the following subsection.

3.1.2 Box-Jenkins Methodology

The process of Box-Jenkins methodology consists of three phases: (1) model identification, (2) model estimation and testing and (3) forecast application. Each phase consists of various steps as illustrated in the following flowchart (Figure 3.1):

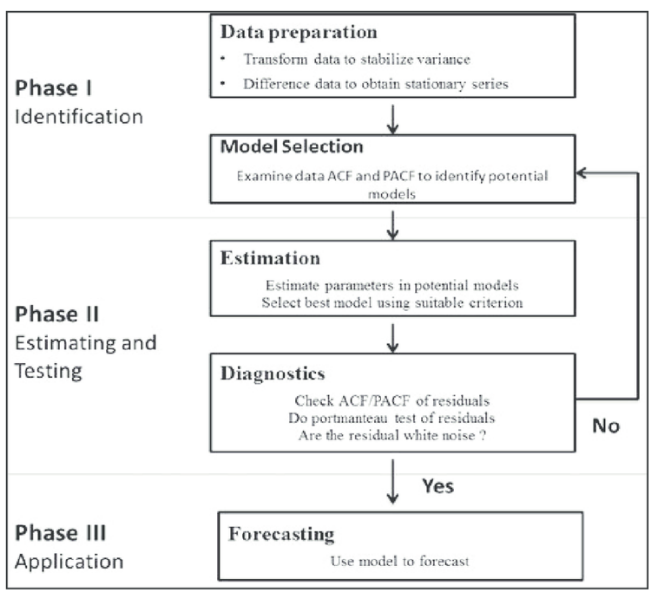


Figure 3.1 Box-Jenkins methodology for time series forecasting

Source: Box et al. (2013)

These phases are implemented in an iterative process by suggesting various tentative models. Thereafter, tentative models are checked by using diagnostic tests and visual inspection to ensure: data fitting and model adequacy, parameter significance, residuals are random (white noise) and the model's ability to generate reliable forecast results. The potential models passing through these tests and inspections are used to produce forecasts. Then, forecast errors are compared to ensure model reliability. A detailed description of these phases is represented in the model identification, model estimation and diagnosis and forecast application.

<u>Phase I: Model identification</u>

The purpose of this phase is to investigate the time series pattern and to determine the degree of integration (d), the lag terms of Autoregressive (AR), and Moving average (MA). To do so, this phase consists of data preparation and model selection. In the former, or data preparation, the following steps are conducted to identify the changes that occurred on container throughput and ensure series stationarity:

1) Time series data and various transformations are plotted. Trend, seasonality, and random components of data series are plotted as well. Time series data is decomposed to extract the three components mentioned above. By composing seasonality, the trend component corresponds to the long run behavior of the series. The random (stochastic) component is used to visualize if the series encounters structural breaks. If structural breaks are found, they have to be considered in constructing the model.

2) In order to use the ARIMA model, time series data must be stationary around mean and variance, if not, stationarity must be obtained by applying differencing or linear detrending filters. This step will lead us toward determining the degree of integration (d).

In the latter, or model selection, tentative models are determined. Stationary is checked as well. To do so, the followings tests and visualizations are conducted:

1) Diagnostic tests: Augmented Dickey Fuller (ADF) and (KPSS) tests are applied to the original and the transformed time series to verify the stationarity of time series. This diagnosis avoids spurious regression.

2) Visual inspection: Box-Jenkins methodology suggests the use of Autocorrelation Function (ACF) and the Partial Autocorrelation Function (PACF) to identify the order of lag structures of Autoregressive and Moving Average (p and q). ACF and PACF measure the correlation coefficient to reflect how the time series observations relate to each other. The plots also help identify seasonal patterns. If seasonality is identified, the SARIMA model is used.

In case of seasonality, SARIMA models or transforms the dataset by taking the seasonal difference. The seasonal differencing (year-over-year monthly growth rate) is defined:

$$\Delta\Delta_{12}Y_t = (1-B)(1-B^{12})Y_t = \Delta Y_t - \Delta Y_{t-12} = Y_t - Y_{t-12} - Y_{t-1} + Y_{t-13} \qquad (3.6)$$

<u>Phase II: Estimating and testing</u>

The accuracy of parameters and forecasts depends on choosing the right lag length. On the one hand, choosing a higher lag order than needed results in parameter overestimation which leads to forecasting errors. On the other hand, selecting a lower lag length underestimates the coefficients and autocorrelation errors. Therefore, tentative models from the previous phase are checked by using diagnostic tests and visual inspection to ensure model appropriateness. Additionally, other model parameters are considered to generate appropriate lag length by using the selection criteria of Akaike's information criterion (AIC) and Schwarz information criterion (SIC) to estimate model parameters. To avoid over-differencing, parameter significance and invertibility conditions must be met. Therefore, the following tests are conducted to ensure model appropriateness:

1) Model-fit adequacy

2) Residual randomness and normality to ensure white noise

3) Parameters' significance and their relationship

Phase III: Forecasting

In this final phase, the potential models are used to forecast container throughput. To test accuracy, the forecasts of tentative models are compared to the actual values with error measures: mean square error (MSE), the root mean square error (RMSE), and the mean absolute percentage error (MAPE). To evaluate forecast accuracy, the following expression for the RMSE is used:

$$RMSE = \sqrt{MSE} \tag{3.7}$$

The developed model with the lowest forecast errors is the most reliable model. Considering the difference between the actual and the forecasted time series, forecast error is represented as $e_t = y_t - \hat{y}_t$. After finding the most suitable model, it is used to forecast the out-of-sample set. The forecast methodology is applied to in Section 5.1.

3.2 Assessing Port Competitiveness

Another purpose of this dissertation is to conduct a spatial analysis to assess port competitiveness by analyzing the impact of port accessibility and geographic characteristics on inland distribution of container traffic. An in-depth discussion on various gravity model approaches, therefore, is presented in this section, where Spatial Interaction Models (SIM) is applied to assess port competitiveness.

The conducted port competitiveness analysis has two elements. The first one is analyzing the spatial relationship between container flow and distance; Wilson maximum entropy SI models is used to analyze the impact of inland distance on port competitiveness by investigating the role of distance as an explanatory power in determining the market share among competing ports. The second one is analyzing the importance of port characteristic and accessibility in hinterland region. This is done by applying Poisson-based SI model to investigate the impact of port accessibility, inland distance, railway availability and port location within shipping routes on the inland distribution of container flow within hinterland. The model seeks to investigate whether the preference of shippers favor inland accessibility or not.

Lastly, this section, is organized as follows: Subsection 3.2.1 presents the spatial interaction model used to analyze the inland distance in container flow. The Spatial Interaction Model based on Poisson-based SI model is presented in Subsection 3.2.2 to include factors other than distance in the spatial analysis,

3.2.1 Spatial Interaction Analysis: Wilson Entropy-maximizing Models

Spatial Interaction Model (SIM) is one of the well-known gravity model approaches that analyzes the geographic distribution of flow between origins and destinations, or the spatial interaction, it consists of three factors typically relied on upon the gravity model.

Factor 1	Origin-specific describes the ability of origins in generating flows.
Factor 2	Destination-specific tend to reflect destination attributes in attracting flow.
Factor 3	Impediment function demonstrates the spatial separation between origin and destination between constraining flows.

On a broad geographical scale, the spatial separation between origin and destination is measured in terms of the spatial impediment between origin and destination. The traditional gravity model is formulated as follows (Wilson, 1967):

$$F_{ij} = \frac{O_i D_j}{d_{ij}^2} \tag{3.8}$$

Where the attributes of emissiveness O_j and attractiveness D_j are directly proportional to the spatial interaction between origin i and destination j, and the impedance attribute is the road distance between i and j d_{ij}, where the squared distance is inversely proportional to it. However, the model in equation X shows a clear flaw. When the mass of an origin/destination doubles, the volume of movements between them quadruples, and not just doubles, as could be expected.

To overcome this deficiency, two constraints have to be considered in relation to origins and destinations (Wilson, 1967). The sum of O_i over all origins must be equal to the total productions and the sum of D_j over all destinations must be equal to the total attraction. Equations (3.9) and (3.10) define the two constraints of production and attraction.

$$O_i = \sum_j F_{ij} \tag{3.9}$$

$$D_j = \sum_i F_{ij} \tag{3.10}$$

Considering these constraints is ensured by using two balancing factors namely A_j and B_j. These balancing factors are associated with production and destination, respectively. By introducing the balancing factors into Equation 3.11, the gravitational is formulated as follows:

$$F_{ij} = A_i B_j O_i D_j f(d_{ij}) \tag{3.11}$$

Where:

$$A_i = \left[\Sigma_j B_j D_j f\left(d_{ij}\right)\right]^{-1} \tag{3.12}$$

$$B_j = \left[\Sigma_i A_i O_i f(d_{ij})\right]^{-1} \tag{3.13}$$

Alternative types of the gravity model can be used by adding exogenous constraints on the main spatial interaction model. These model variants consist of the balancing factor of origin-specific and/or destination-specific. These balancing factors work as constraints to ensure the totals of origin and destination constraints, specified above, are exactly predicted (Ferrari et al., 2011). In case that either the origin or the destination constraints are used, the mode is singly constrained; otherwise, it is not constrained. The model is doubly constricted when both balancing factors of origins and destination hold for every location. The new matrix F'_{ij} is generated by setting the following alternative types of constraints:

1) **Production constrained:** where the container flow generated by each origin remains constant both in the observed and predicted matrix, while the traffic attracted by each destination is redistributed among them.

$$F'_{ij} = A_i O_i D_j f(d_{ij}) \tag{3.14}$$

2) **Destination constrained:** where the overall container flow is redistributed among the different origins, while the traffic flows assigned to each destination in both matrices are constant.

$$F'_{ij} = B_j O_i D_j f(d_{ij}) \tag{3.15}$$

3) **Doubly constrained:** where both the container flow of the ith origin and jth destination are kept constant. The difference between this constraint and previous ones is that it has two balancing factors A_i for origin and B_j for destination, whereas each of the production and attraction constraints has only a single balancing factor A_i and B_i, respectively. The formula of the doubly constraint is presented in equation (3.16), where the constants of the origin A_i and destination B_j are dependent on each other. Depending on the calibration technique, they may need to be computed iteratively.

$$F'_{ij} = A_i B_j O_i D_j f(d_{ij}) \tag{3.16}$$

There are two container flow matrices that need to be distinguished. The actual container flow CF_{ij} matrix and the predicted container flow CF'_{ij} matrix. In addition, the geographical inland distance matrix. Input data are the observed origin-destination O-D matrix, where CF_{ij} represents the flows produced by port i and attracted by province j, and the distance matrix where element d_{ij} is the geographical distance between the ith region and the jth destination. The model produces as output a new O-D matrix (predicted OD matrix), which represents the redistribution of the observed traffic flows.

3.2.1.1 Model calibration: Starting from the general Equation 3.11, distance function is introduced into the gravity model. Various decay functions presume different responses to the increasing costs associated with long distance. To determine the distance function $f(d_{ij})$, there are various options that have been considered in the literature. The two major functions are: the exponential function ($e^{\beta d_{ij}}$) and the power function (d_{ij}^{β}). For locations separated by short spatial distances, the negative exponential function ($e^{\beta d_{ij}}$) can work properly (C. Ferrari et al., 2011). In case of long distance between origins and destinations, the alternative of power function (d_{ij}^{β}) is preferable. The mainstream of the literature suggests the use of the latter for long distances; Thus, power function is applied in the case study of this dissertation. By applying power distance, the predicted O-D matrix CF'_{ij} is calculated as follows:

$$F'_{ij} = A_i B_j O_i D_j d_{ij}^{\beta} \tag{3.17}$$

Where all the parameters and variables have the same meaning specified earlier. (see for e.g. Equation 3.8). The Interpretation of distance parameters will be clarified in the calibration phase.

3.2.1.2 Goodness-of-fit: An essential step of modeling the spatial interaction is evaluating its ability to predict a set of flow distribution. Accurate prediction supports the theoretical propositions on which the design is based: that is, it supports a specific model type over the others. It likewise provides confidence in the accuracy of parameter estimates and in the capability of a model to predict system flows. Various goodness-of-fit statistics have been used in the literature to assess the ability of the SI models in predicting flow. These goodness-of-fit statistics provide quantitative description by comparing the actual and predicted flow matrices. SRMSE and information gain are two measures of model goodness-of-fit where SEMSE is used to measure the error in the model outcomes. However, there has been little consistency in using a specific accuracy measure which prevents the comparison of model ability throughout the studies (Williams & Fotheringham, 1984).

The aim of this analysis is not the model's goodness-of-fit, however. Ferrari et al., (2011) applied Absolute Entropy Difference (AED) and statistics deviation (d) to evaluate the degree of flow redistribution in calibrating the constrained models. In this dissertation, these two parameters are applied in the outcomes of the analysis in Section 5.2. The two parameters are calculated for both models. AED can be conceptualized as a statistical index defining systems' entropy and d is a fitting parameter. AED is defined as the difference in the variance of the actual and predicted probability distribution in absolute value. AED is not used to measure the model fit but to evaluate the degree of influence that distance has on container flow. It can be expressed:

$$AED = |H_p - H_q| \tag{3.18}$$

Where H_p and H_q denotes Shannon Entropy measures and they refer to the variances of observed and predicted probability distribution, respectively. Therefore:

$$H_p = \Sigma_i \Sigma_j \, p_{ij} ln \, (p_{ij}) \tag{3.19}$$

$$H_q = \Sigma_i \Sigma_j \, q_{ij} ln\,(q_{ij}) \tag{3.20}$$

Where:

$$p_{ij} = \frac{f_{ij}}{\Sigma_i \Sigma_j \, f_{ij}} \tag{3.21}$$

$$q_{ij} = \frac{f'_{ij}}{\Sigma_i \Sigma_j \, f'_{ij}} \tag{3.22}$$

AED shows how much the real system is predictable where the lower limit of AED can be zero when the system is fully predictable ($H_p = H_q$). In contrast, the upper limit of AED represents the maximum entropy with maximum level of uncertainty, when $H_p = 0$ and $H_q = ln\,(n)$. n is determined by the size of the system, where n is the number of destinations divided by the number of origins. The higher bound of AED equals ln(n), where n is the maximum size of the system (the number of provinces divided by the number of ports). AED provides evidence of how much the variance of predicted flows compared to the variance of the actual flow, where a close value of AED to zero indicates a minimal difference between the two variances, whereas a higher AED indicates the model does not reliably fit the real system (Knudsen & Fotheringham, 1986).

Parameter d also measures the deviation in the actual container flow F_{ij} compared to the predicted container F'_{ij}. It is measured in percentage where d=100% indicates that predicted flow distribution has no deviation from the actual flow. The deviation statistics is characterized by the functions of f'_{ij} and f'_{ij}, where f_{ij} is an element of the actual container flow matrix, and f'_{ij} is an element of the predicted container flow matrix. The deviation is defined as:

$$d = \frac{\Sigma_i \Sigma_j \, f_{ij} - f'_{ij}}{\Sigma_i \Sigma_j \, f_{ij}} \times 100 \tag{3.23}$$

3.2.2 Spatial Interaction Models: Poisson-based models

The Spatial Interaction Models presented in the previous Subsection focused solely on inland distance as a factor that impacts container flow. However, other determinants also have an impact on the distribution of container flow within the considered hinterland. In this subsection, the Poisson-based spatial analysis model is proposed where factors related to geographic and port accessibility fitted in the analysis of container flow.

As shown in Subsection 3.2.1, there are three variants of the spatial interaction gravity model: doubly constrained, production constrained, and attraction constrained exponential gravity models. These models include either origin or destination-specific or both as balancing factors to ensure that the total volume of the predicted equals that of the actual data. These can also be estimated statistically. If data typically counts and follows Poisson distribution, for instance, the Poisson model provides better statistical distribution, so that unbiased results can thus be compared to OLS regression. The probabilistic Poisson allows the use of origin and/or destination fixed effects for model estimation. The development of Poisson-based SIM is presented in this Subsection.

The classical concept of the gravity model states that every point mass O_i (emissiveness/ propulsiveness) attracts every other point mass D_j (attractiveness) with a gravity force F (interactions) that is directly proportional to the product of their masses O_i and D_j and inversely proportional to the square of the distance d between them:

$$F = k \frac{O_i \cdot D_j}{d_{ij}^2} \tag{3.24}$$

3.2.2.1 Calibration framework: Linear regression is the typical method to calibrate the Spatial Interaction Models. However, linear programming and nonlinear optimization can also be used. For this study, the regression framework is used as a calibration technique due to its flexibility and extensibility. The natural logarithm is applied on both sides of the basic gravity model in Equation (3.24) to obtain the log-linear version of the gravity model (Equation 3.25). By including the error component, the log-normal gravity model is obtained (Equation 3.26).

$$lnCF_{ijt} = k + \mu InV_i + \alpha InW_j - \beta lnd_{ij} \tag{3.25}$$

$$lnCF_{ijt} = k + \mu InV_i + \alpha InW_j - \beta lnd_{ij} + \varepsilon \tag{3.26}$$

here ε is the error component that is normally distributed with a mean of 0. β represents the distance parameter. To validate the underlying hypothesis, A negative value of β indicates that spatial interactions decrease when distance is higher.

The constrained spatial interaction may have the following model components: production constrained or the fixed effects for the origins; attraction-constrained or the fixed effects for the destinations; or fixed effects for both attraction and production constrained (doubly constrained). However, it is important to note the limitation of the log-normal gravity models. First, the flows are considered discrete entities representing counts of flow units. Second, the flows are abnormally distributed. Third, it produces estimates of logarithmic flows instead of actual flows resulting in biased predictions. Fourth, it cannot deal with zero flows as the logarithm of zero is undefined, consequently applying log-normal results in the exclusion of observations with zero value.

Considering these limitations, Flowerdew and Aitkin (1982) proposed Poisson log-linear regression to estimate spatial interaction models. The inherent assumption of this specification is that the flows observations between i and j is drawn from a Poisson distribution with mean $\lambda_{ij} = CF_{ij}$, where λ_{ij} is assumed to be logarithmically linked to the linear combination of variables. The equation of log-linear regression is as follows:

$$ln\lambda_{ij} = k + \mu InV_i + \alpha InW_j - \beta lnd_{ij} \quad (3.27)$$

Removing the log-operator provided from both sides yields the following equation that represents unconstrained Poisson log-linear gravity model with a distance-decay power function.

$$CF_{ij} = exp\ (k + \mu lnV_i + \alpha lnW_j - \beta lnd_{ij}) \quad (3.28)$$

Therefore, by using the balancing factors in the Equations from (3.11) to (3.13), the constrained variants of production, attraction, and doubly can be respectively defined:

$$CF_{ij} = exp\ (k + \mu_i + \alpha_i lnW_j - \beta lnd_{ij}) \quad (3.29)$$

$$CF_{ij} = exp\ (k + \mu_i lnV_i + \alpha_i - \beta lnd_{ij}) \quad (3.30)$$

$$CF_{ij} = exp\ (k + \mu_i + \alpha_i - \beta lnd_{ij}) \quad (3.31)$$

Where μ_i and α_i are origin and destination balancing factors, respectively (Tiefelsdorf, Boots 1995). To ensure the flow observation totals are conserved, the value of k is included, and it represents the estimated intercept in equations from 3.25 to 3.31. Using Poisson regression addresses the above-mentioned limitation 1-4 of the linear model as the flow is no longer logarithmic. Poisson regression can be calibrated by using iteratively weighted least squares (IWSL) within a generalized linear modeling (GLM) framework (Nelder & Wedderburn, 1972).

CHAPTER 4

CASE STUDY

In this chapter, the case study is presented where Section 4.1. covers the background of Jeddah port, the port in which the forecast methodology is implemented to estimate potential demand. In Section 4.2, the characteristics of the inter-port competition within country-level are discussed.

4.1 Background of Jeddah Port

The port of Jeddah is located in the red sea on the western coast of Saudi Arabia. Its strategic location in the major East Asian-European liner shipping route increases its importance as a major port in Saudi Arabia and as a transshipment hub in the Middle East region (Figures 4.1 and 4.2). According to General Authority for Statistics of Saudi Arabia, in 2018, Jeddah Port was ranked number one among the Saudi ports in terms of container volumes where over 4.5 million TEUs were handled by the three container terminals of the port in 2013, making the port one of the major ports in the Middle East region and among the 30 busiest container ports in the world.

The container terminals are managed, operated and maintained by private companies under the umbrella of the Port Authority of Saudi Arabia (PASA) which acts as a regulatory body for all ports located in both the West and East coasts except the newly-emerged port of King Abdullah which is privately owned and not subject to the regulatory body of the PASA.

Figure 4.1 Major container shipping routes and container volumes (million TEUs) in 2007.
Source: UNCTAD (2008).

Figure 4.2 Geographic location of Jeddah port.

4.2 Competition Among Saudi Ports

Saudi Arabia has seven commercial seaports distributed between the West and East coasts of the country. Four of them are considered as the critical ports of the country. Ports of Jeddah and King Abdullah are located in the Red Sea. Ports of Dammam and Jubail are located in the Arabian Gulf. Over the last decade, nonetheless, the port sector of Saudi Arabia developed and expanded rapidly to provide efficient performance and reliable logistics.

In addition to the development of existing ports, colossal investments permitted the port of King Abdullah (KAP) to be built. It is the first port in the country to be privately funded, owned and run outside the umbrella of Saudi Port Authority. The port was opened in late 2013. The port is strategically located in the newly-built King Abdullah Economic City on the Red Sea coast. It is located 130 kilometers north of Jeddah port. During the first phase of operation, the port had a total initial capacity of 3 million TEUs, with a depth of 18 meters that allows the biggest containerships to access the port. It had a total quay length of 1,470 meters and fit ship-to-shore gantry cranes with 25 containers outreach.

The port has a long-term strategic expansion plan consisting of seven phases to extend its quay length to 11,707 meters and capacity to reach 20 million TEUs in the future. Phase 2 later on was brought into operation in 2017, where two more berths were commissioned which increased the total capacity by 25% to reach 4 million TEUs. As a result, Jeddah Port has been facing increasing competition from KAP as they are relatively located close to each other.

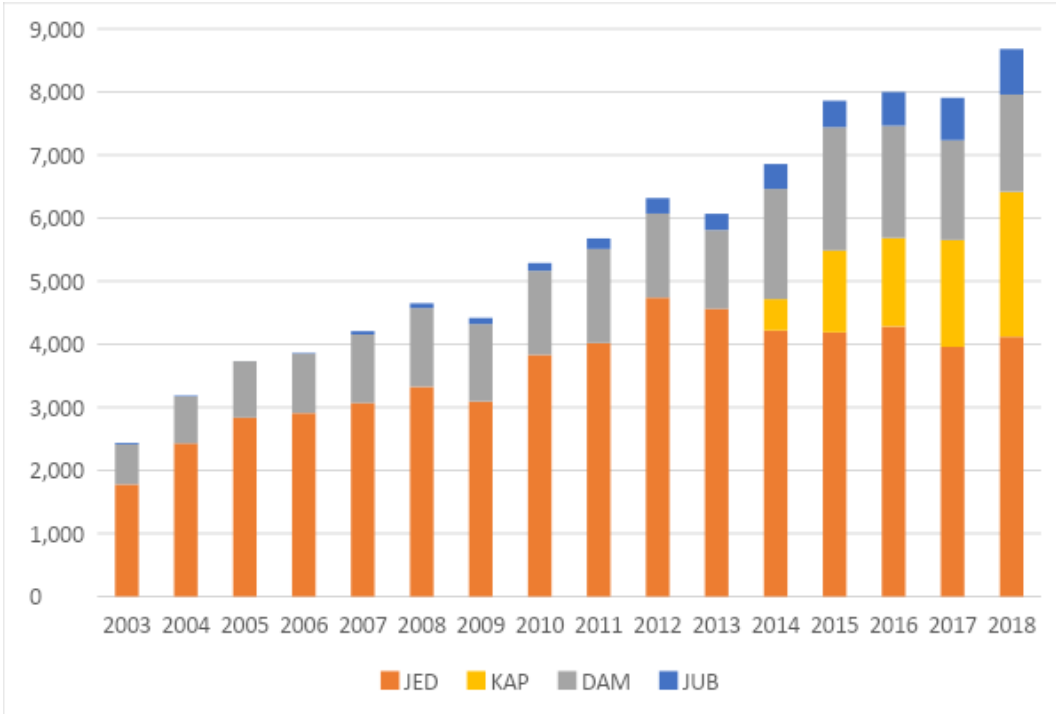


Figure 4.3 Market Concentration of Saudi Ports Measured by TEUs (2003-2018).

Figure 4.3 presents the container volume of the four ports for the period 2003-2018. Jeddah port experienced a strong growth during the period 2003-2008. In 2018, the port handled 4.5 million TEUs compared to 1.7 million TEUs. Following the global economic crisis of 2008, The port faced a decline by 7% in 2009. Jeddah container throughput faced approximately the same decline rate as the average global rate reaching 3.3 million TEUs. In the following years, the container throughput resumed its growth to reach 4.5 million TEUs in 2013.

Container throughput of Jeddah port has been fluctuating during the period 2014-2018. It reflected the loss of market share to its rival KAP. Since the opening of KAP, it has gained a remarkable share of Jeddah Port's demand. Despite the slowdown in the global economy, the number of containers handled in KAP quintupled from almost 500 thousand TEUs in 2014 to 2.3 million TEUs in 2018, making it the fastest growing port in the world.

At the same time? container throughput in Jeddah Port decreased by 2.4% reaching 4.117 million TEUs in 2018 compared to 4.218 million TEUs in 2014 (as shown in Table 4.1).

Table 4.1 Container Throughput and Growth Rates of Jeddah and KAP

Port	**2014**	**2015**	**2016**	**2017**	**2018**	**Avg. annual growth rate (2014-2018)**	**Cumulative growth rate (2014-2018)**
Jeddah Port	4,218	4,188	4,283	4,154	4,117	-0.6%	-2.4%
Ann. growth		-0.7%	2.2%	-3%	-0.9%		
KAP	497	1,270	1,402	1,695	2,302	46.7%	363.2%
Ann. growth		161.6%	7.8%	20.9%	35.8%		

From 2003 to2014, Jeddah port dominated the container throughput in Saudi Arabia by an average annual market share of 73%, while the majority of the remaining market share was for Dammam port in the east coast of the country as shown in Table 4.2. Ever since the opening of the first privately-owned KAP in 2014, Jeddah port has been facing an increase in competition which resulted in a decrease of market share of Jeddah port by 14% the same year. The market share decline continued until 2018, where Jeddah port had a market share of 46.4% in comparison to 26% for KAP. This decrease was due to the close location of KAP port to the local hinterland of Jeddah port.

Development of the dry terminal in the capital of Riyadh in 2014 caused another decrease. The improved efficiency of the railway link between the capital and port of Dammam increased the container handled in the dry terminal by 29% to reach 350,000 TEUs in 2015. The railway connectivity provides Dammam port an advantage over the strategically located ports of Jeddah and KAP, however, this railway connection is the only one specialized in container transport in the country.

Table 4.2 The Market Share of Container Ports in Saudi Arabia (2003-2018)

Year	Red Sea ports		Arabian Gulf ports	
	Jeddah	**KAP**	**Dammam**	**Jubail**
2003	73.0	-	26.0	1.0
2004	76.2	-	23.3	0.4
2005	76.0	-	24.0	0.0
2006	75.3	-	24.4	0.3
2007	72.9	-	25.8	1.2
2008	71.5	-	26.8	1.6
2009	70.0	-	27.8	2.2
2010	72.4	-	25.2	2.3
2011	70.8	-	26.3	2.9
2012	75.0	-	21.1	3.9
2013	75.2	-	20.6	4.2
2014	61.5	7.2	25.5	5.8
2015	53.3	16.5	24.9	5.3
2016	53.5	17.5	22.3	6.7
2017	50.0	21.4	20.0	8.5
2018	47.4	26.5	17.7	8.3

The overall market share of Dammam port declined by approximately 7% from the year 2015 to 2018 due to the development of the port of Jubail. Given that Jubail port is also located on the East coast, it shares the same market as Dammam port. The massive investments in the development of the port sector have also benefited Jubail port resulting in the increase of its market share to 8% in 2018.

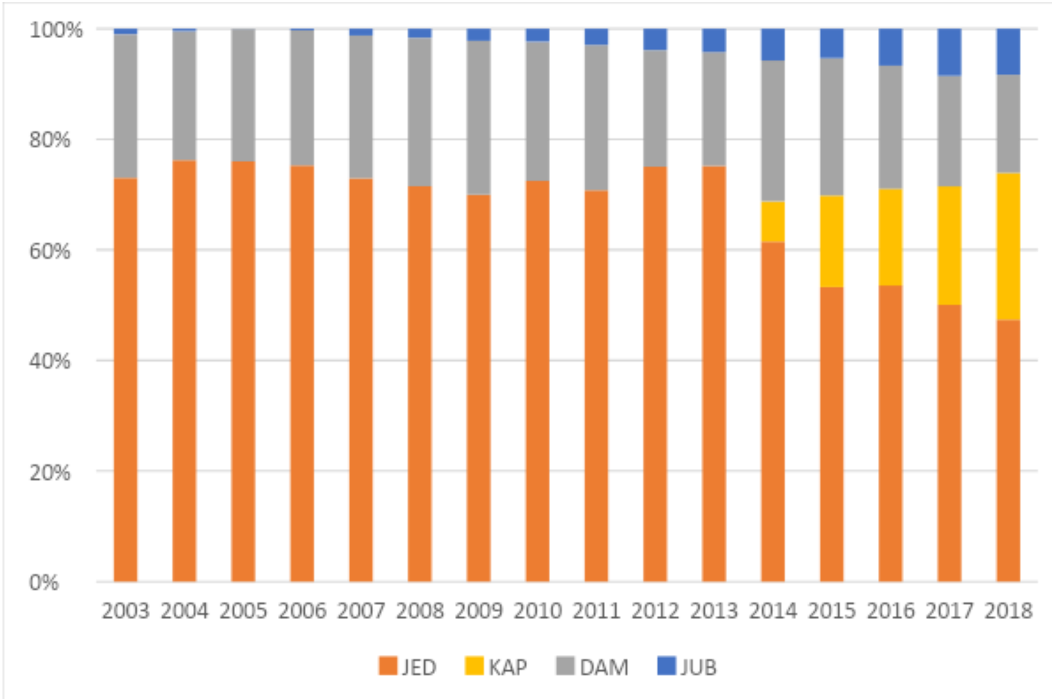


Figure 4.4 The market share of the four Saudi ports (2003-2018).

4.2.1 Market structure and condition

To understand the structure and condition of the container market in Saudi Arabia, Shift-Share Analysis is conducted in the container throughput of the four ports. Table 4.3 presents shift and share effects, supported by Figure 4.5 that illustrates these effects. The figure reveals the following interesting findings.

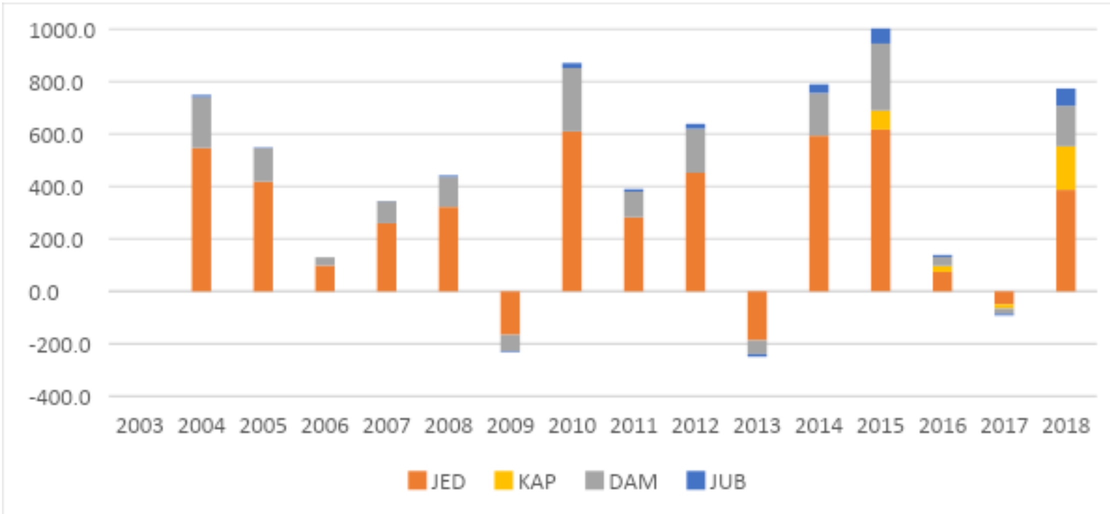
Since the opening of KAP in 2014, Jeddah port has continuously lost significant container share. As the shift effect analysis in Figure 3 shows, the total container volume of over 2,300,000 TEUs shifted from Jeddah to KAP in just five years. This shift might explain the declined volume of containers that Jeddah port went through, while its throughput decreased by 2% in the 2014-2018 five-year- period, compared to an increase of 10% during the prior-five-year-period of 2010-2014.

On the East coast, another clear shift effect occurs between Dammam and its neighbor port of Jubail, where approximately 300,000 TEUs shifted from Dammam port to Jubail port during the same period (2014-2018). Further to this noticeable shift, there is a gradual increase over the previous period 2003-2013. These two periods have led to a remarkable decrease in the gap between Jubail and Dammam ports. These shifts might explain the declined volume of container throughput that Dammam port went through as the port throughput decreased by 11% in the 2014-2018 five-year- period, compared to a remarkable 31% increase during the prior-five-year-period (2010-2014).

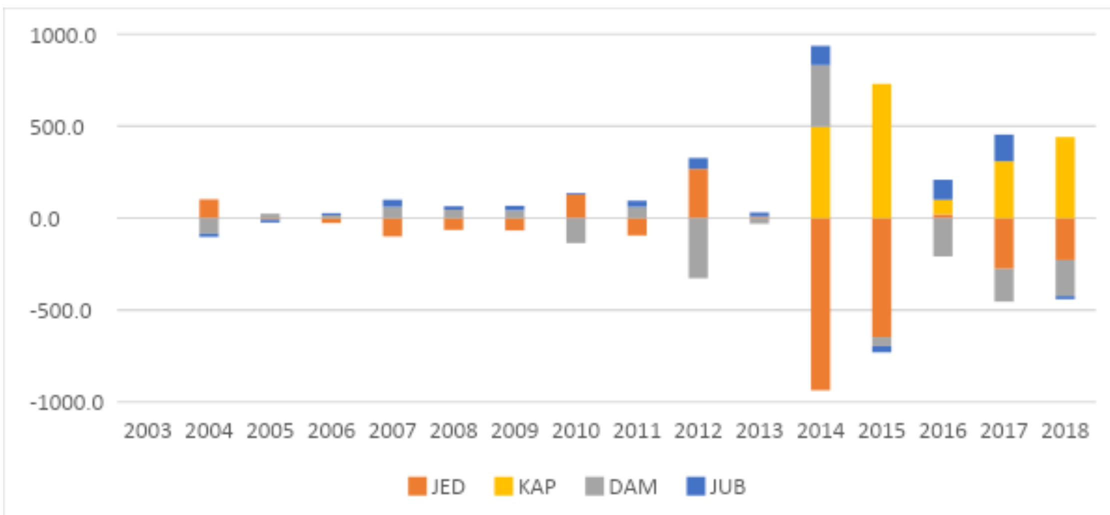
On the other hand, the container throughput increase in Jubail port was 83% in the 2014-2018 period, compared to 220% in the previous period. The increase of container throughput in Jubail port slowed down from 220% to 83% in the periods 2014-2018 and 2010-2014, respectively. This led to a gap reduction of container throughput in the two Eastern ports from 1.2 million TEU in 2010 to 0.7 million TEU in 2018. Shift Share Analysis is a useful and effective tool to measure structural changes of ports as it represents the shift and shares that occur in competing ports. However, this technique does not explain the reasons behind changes on shifts nor shares. In the next subsection, market share of the four ports is presented at the province level.

Table 4.3 Shift-Share Analysis results for container ports in Saudi Arabia (2003-2018)

	Share effect				**Shift effect**			
	Red Sea		**Arabian Gulf**		**Red Sea**		**Arabian Gulf**	
Year	**Jeddah**	**KAP**	**Dammam**	**Jubail**	**Jeddah**	**KAP**	**Dammam**	**Jubail**
2003	546.8	0	194.8	7.4	102.2	0	-84.8	-17.4
2004	418.4	0	128.2	2.4	-8.4	0	23.8	-15.4
2005	98.8	0	31.2	0.0	-26.8	0	15.8	11.0
2006	259.8	0	84.2	1.1	-99.8	0	60.8	38.9
2007	322.3	0	114.2	5.5	-64.3	0	45.8	18.5
2008	-166.0	0	-62.2	-3.8	-67.0	0	42.2	24.8
2009	609.9	0	242.0	19.1	128.1	0	-136.0	7.9
2010	282.5	0	98.3	9.1	-94.5	0	60.7	33.9
2011	452.3	0	167.9	18.8	267.7	0	-326.9	59.2
2012	-186.8	0	-52.5	-9.7	8.8	0	-30.5	21.7
2013	593.8	0	162.7	33.5	-936.8	497.0	335.3	104.5
2014	616.9	72.7	255.6	57.8	-646.9	730.3	-49.6	-33.8
2015	74.1	23.0	34.6	7.4	20.9	79.0	-208.6	108.6
2016	-49.3	-16.1	-20.5	-6.2	-276.7	309.1	-177.5	145.2
2017	387.3	165.9	154.8	66.0	-228.3	441.1	-195.8	-17.0
2018	4260.9	245.5	1533.2	208.4	-1921.9	2056.5	-625.2	490.6



(a) Share analysis



(a) Shift analysis

Figure 4.5 Share and Shift effect of container ports in Saudi Arabia (2003-2018)

4.2.2 Port Competition Over Hinterland

Saudi Arabia is divided into 13 governmental provinces where Riyadh, the capital, is in the province of Riyadh. Table 4.4 presents the demographic and economic characteristics of the 13 provinces. The table also shows the distance from these provinces to the four ports. The Western ports of Jeddah and KAP are in the province of Makkah, and the ports of Dammam and Jubail are located in the Eastern province. Figure 4.6 shows the location of the four ports and the inland connectivity to the major cities.

Table 4.4 The characteristics of the provinces of Saudi Arabia (2012)

			GDP (million	Distance to each province (km)			
Province	**Capital**	**population**	**US$)**	**JED**	**KAP**	**DAM**	**JUB**
Riyadh	Riyadh	8,002,100	152.012	951	1,014	432	476
Makkah	Makkah	8,325,304	89.340	0	139	1,379	1,422
Madinah	Madinah	2,080,436	42.670	422	316	1,245	1,273
Qaseem	Buraidah	1,387,996	41.069	917	814	773	780
E. Province	Dammam	4,780,619	80.811	1,360	1,402	0	92
Aseer	Abha	2,164,172	58.137	632	760	1,379	1,421
Tabouk	Tabouk	890,922	22.001	1,023	905	1,775	1,687
Hail	Hail	684,619	33.336	883	780	1,032	948
N. Borders	Arar	359,235	20.268	1,429	1,328	1,056	978
Jazan	Jazan	1,533,680	23.201	711	840	1,561	1,610
Najran	Najran	569,332	20.641	924	1,053	1,366	1,408
Baha	Baha	466,384	11.734	421	480	1301	1,350
AlJouf	Sakaka	497,509	26.668	1,264	1164	1237	1,159

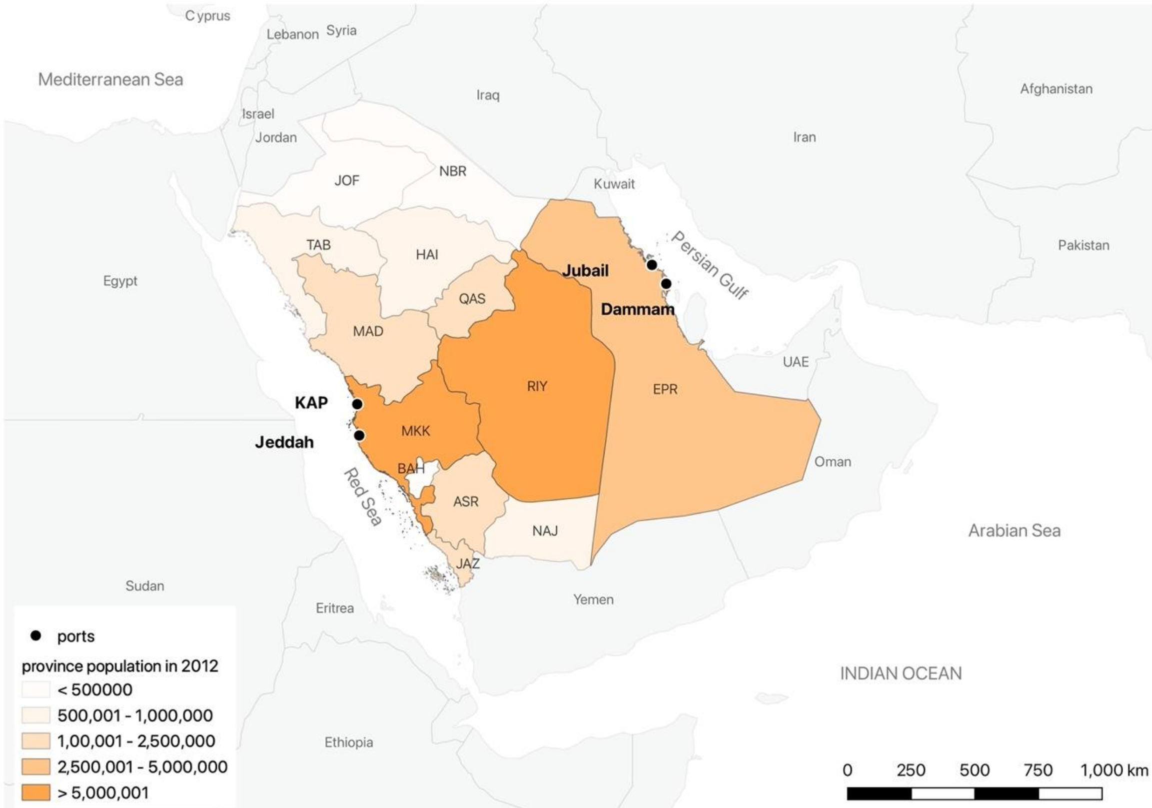


Figure 4.6 Port locations and population size of each province in 2012.

Figures 4.7 and 4.8 depict the container flow market shares of each port in the 13 provinces for the years from 2011 until 2017. The provinces can be geographically categorized into three main regions: west, east, and central. The above two figures, based on each region, reveal the following:

a) **The western region:** Until 2014, the market of seven provinces of Makkah, Aseer, Madinah, Jazan, Tabouk, Najran and Al-Baha was dominated by Jeddah port where it had a market share of more than 95%. This picture has changed after the emergence of KAP. In 2018, Jeddah port continued to be the dominant port for only three South Western provinces of Aseer, Jazan and Najran. These changes show that the emerge of KAP has reduced the domination of Jeddah port over its immediate hinterland (Makkah province) to 82% in 2017

b) **The Eastern region:** The oil-rich Eastern Province, where the two ports of Dammam and Jubail are located, is the third largest province in terms of population size. It has been the captive hinterland of Dammam port. In recent years, a remarkable increase of Jubail's share reduced the domination of Dammam. In the Eastern region, the small province of "North Borders" is also located. The growing port of Jubail has gained 26% market share in the Eastern province from the neighboring port of Dammam in 2017 compared to 2011.

c) **The Central region:** The hinterland of the northern region consists of the provinces of Riyadh, Qaseem, Hail, Jouf. Riyadh province is the second most populous province and has the highest regional GDP. It is also home of the capital city of Riyadh. Moreover, the province is the only one that has a railway connection to Dammam port, the port that is strategically located 432 km away from the capital city of Riyadh, sustained its market share at the same range of 72%. While the Jeddah port, located 950 km away from the capital, lost more than 13% of its market share. Jubail, on the other hand, increased its market share from just 3% to 14%. Qaseem province, located north of Riyadh, has been under fierce competition among the 4 ports. The market share of Jeddah and Dammam ports declined by 22% and 10%, respectively, during the period 2011-2017. In contrast, the port of Jubail and KAP dramatically increased their market share to 14% and 26%.

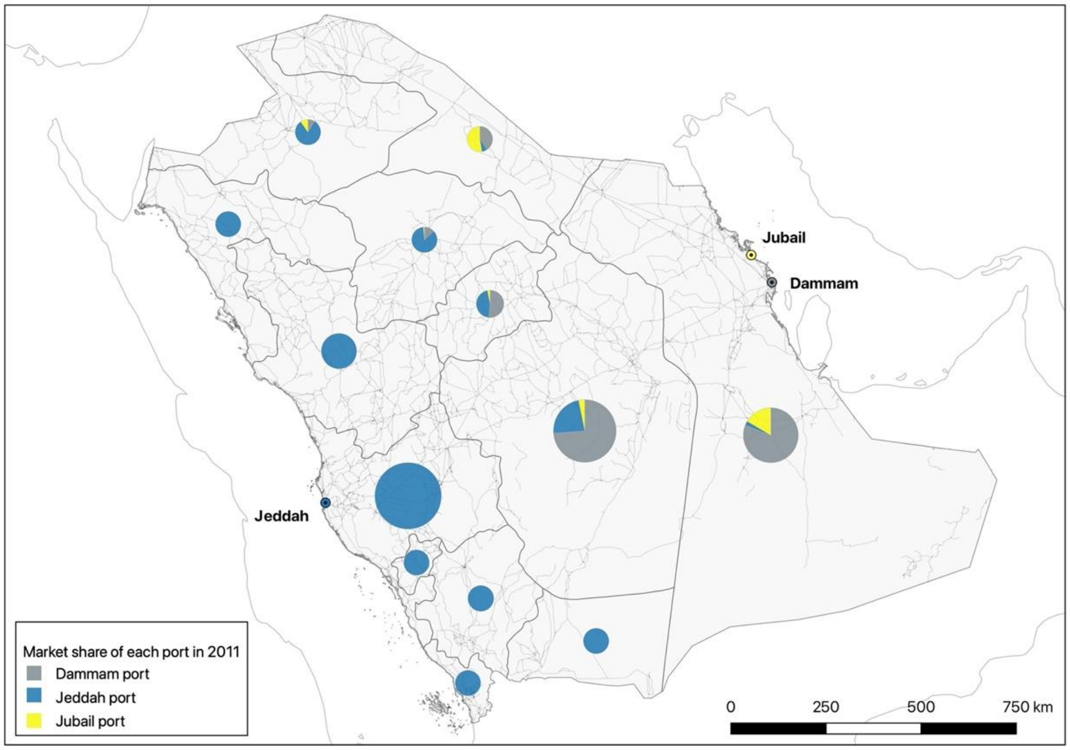


Figure 4.7 Port share of the imports of each province in 2011.

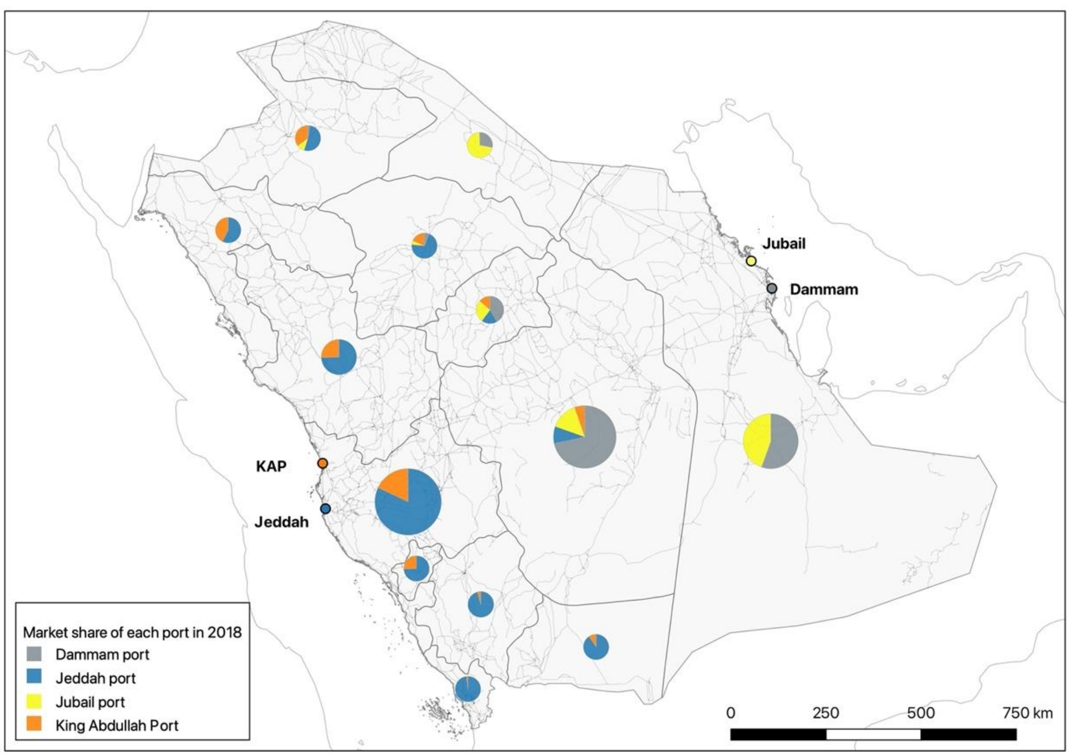


Figure 4.8 Port share of the imports of each province in 2018.

To sum up, changes in container flow occur as a result of various factors. The location of Jeddah and KAP ports gave them an advantage over the Eastern ports of Dammam and Jubail in regard to expanding their market shares specially during the increase of freight rates. Concomitantly, the ability of Dammam port to sustain its share might be caused by the availability of the railway connection to Riyadh province.

These findings suggest the importance of investigating the impact of port size and connectivity to foreland and hinterland, and the availability of railway connectivity in the competitive position of the ports. Therefore, in Section 5.2, the role of inland distance in the distribution of container flow is investigated by applying spatial interaction models for the years 2011 and 2018. What is more, the competitiveness of ports is assessed by considering the impact of the port size their accessibility in terms of hinterland and foreland, specifically, inland distance, the availability of railway connections and ports locations within the maritime shipping routes in Section 5.3.

CHAPTER 5

MODEL FITTING AND RESEARCH FINDINGS

To achieve the dissertation objectives and to present the outcomes and findings, this chapter fits, validates, and calibrates the models shown in chapter 3. This chapter is also divided into three sections. In Section 5.1, container demand is forecasted by applying univariate time series stochastic approach. In Section 5.2, the explanatory power of distance on inland distribution of maritime traffic is investigated by applying Spatial Interaction Model (SIM) and gap analysis. In Section 5.3, the statistical version of Spatial Interaction Model is applied to analyze the impact of port characteristics and geographic location in port-hinterland container flow. Considering these factors, the unexploited potentialities of each port is investigated to assess port competitiveness.

5.1 Container Demand Forecast

In subsection 5.1.1, data description is presented followed by model identification in Subsection 5.1.2. After identifying various tentative models, their parameters are estimated, and a diagnosis test is applied to determine the most suitable models in Subsection 5.1.3. Lastly, container forecasts are generated by the qualified models and the model that provides the lowest accuracy error is used to forecast out-of-sample demand in Subsection 5.1.4.

5.1.1 Data Description

The dataset consists of 204 monthly container throughput of Jeddah port for the period of January 2003 to December 2018. To validate the model, the time series data set is divided into two samples: experimental set (the training set) and the validation set (the test set).

First, the training set (80% of the time series observations) consists of 168 observations, starting January 2003 to December 2016. This set is used for model development. Second, the validation set consists of 36 observations, starting January 2017 to December 2019. It is used to evaluates the forecast accuracy of estimated models. Additionally, data exploration is conducted as a part of applying Box-Jenkins methodology in the next subsection.

5.1.2 Model identification

Model identification is conducted in three steps.

<u>I. Data exploration</u>

Prior to estimating the forecast model, it is important to understand the nature of the dataset and its variation through time. To explore the data, container throughput (CONT) dataset is decomposed to analyze trend and seasonality changes (Figure 5.1). The figure depicts the decompositions of CONT which are trend, seasonal and random components. The trend component shows an increase of container throughput indicating nonstationary. Followed by the fluctuation of container throughput starting in 2014 until 2020. This is due to the decline in oil prices that highly impacted the economic activities in Saudi Arabia, as well as the emergence of the neighboring port of King Abdullah. The seasonal changes are contrasted in the seasonal component which shows accompanied cyclical patterns, where container demand reaches its peak in the second quarter. By detrending and de-seasonalizing the dataset, the random component remains.

The stationarity of the time series is an essential condition in building the time series model. Nonstationary is often observed on the trend component of trade dataset. To achieve stationarity, the time series must be detrended to stabilize the variance by transforming the dataset. Therefore, various power transformations and differentiations are applied. Also, visual inspection assesses changes in the transformed data series.

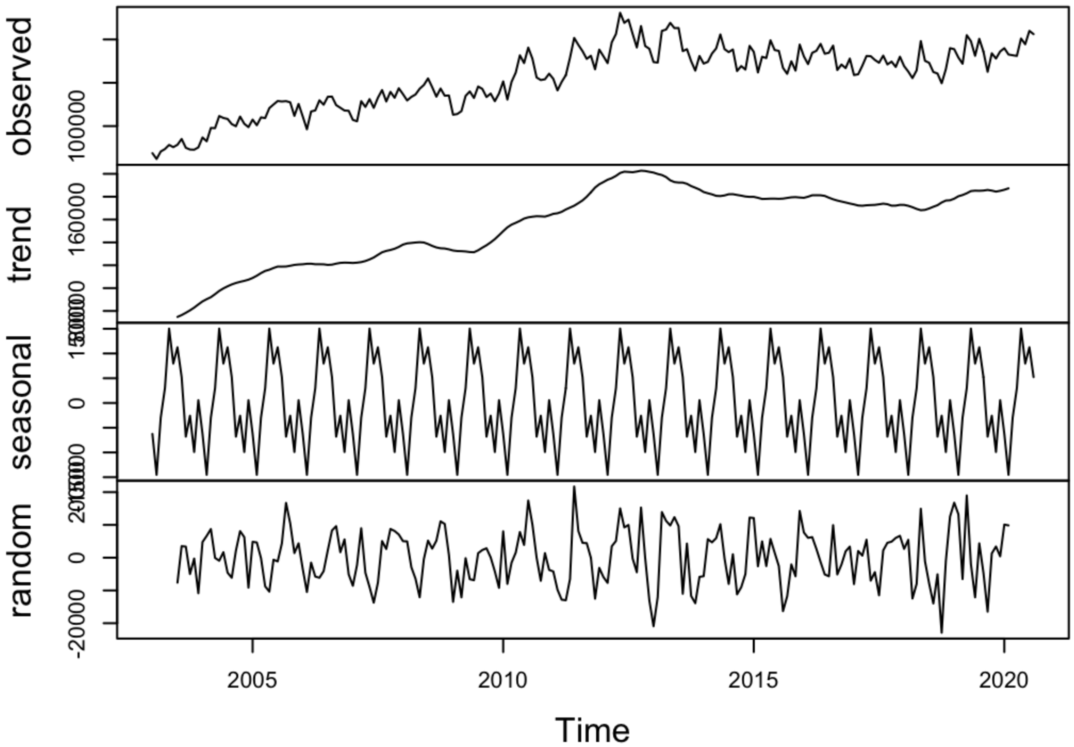


Figure 5.1 Time series decomposition of container throughput.

<u>II. Data transformation and graphical representation</u>

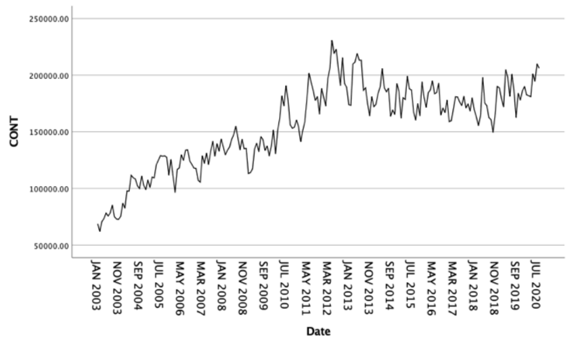
As shown in the previous subsection, the time series of CONT shows trend and seasonality, which are evidence of non-stationarity. To achieve stationarity, four transformations of CONT are conducted. The original and transformed time series datasets are shown in the Figure 5.2, and described as follows:

a) Figure 5.2.a shows the monthly time series of imported container throughput in TEUs (CONTt). The plot shows a gradual growth from January 2003 to May 2012, faced by various downturns. The overall time series shows a positive pattern.

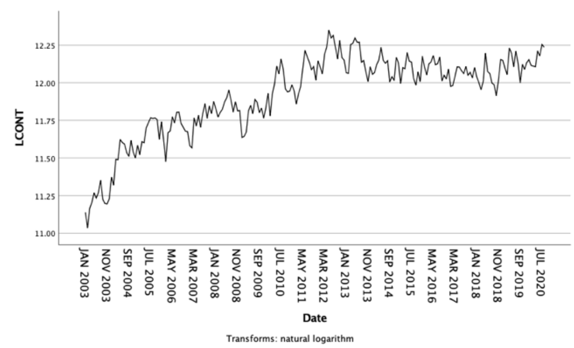
b) Figure 5.2.b shows the natural logarithm of the original series (LCONTt). The logarithm series is generated to reduce variance inconsistency. The plot displays a positive trend until May 2012, followed by a shift in trend level combined with continuous fluctuations. These changes might be caused by seasonal and economic shocks. The economic shocks are associated with the 2008 financial crisis, and the decline of oil prices. Other port-related factors, however, might cause the changes, such as port congestion in 2009 and/or the emerging competition in 2014.

c) Figure 5.2.c depicts unsteady month-to-month growth rate of container throughput. This monthly growth rate is computed as DCONTt= CONTt – CONTt-1.

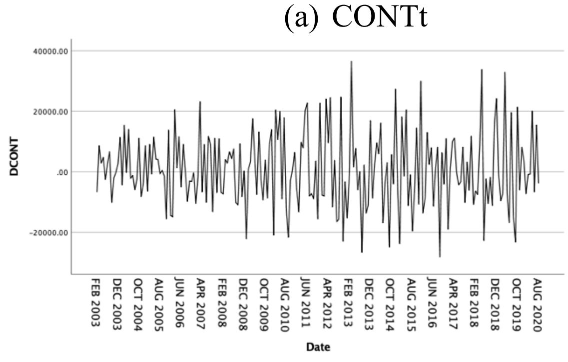
d) Figure 5.2.d shows year-over-year monthly growth rate which is defined as D12CONT= CONTt – CONT t-12. During the 5-year-period to December 2008, the growth rate was around 14% followed by a decline of -6% during the year of 2009. In the following three years (2010-2012), the growth rate returned to 15%. Then, a long period of fluctuation a small amount below the mean at a rate of -2.7% from April 2013 to December 2018 is observed. The problem that encounters the modelling of container demand is the interruption that the dynamic process of time series faces. Most of these dynamic changes are associated with the economic situation and other factors such as port competition to some degree.



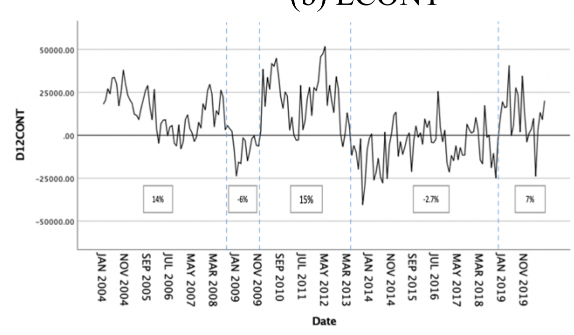
(a) CONTt



(b) LCONT



(c) DCONTt



(d) D12CONT

Figure 5.2 Original and transformed monthly container throughput time series datasets.

III. Identifying lag orders

In this subsection, various tentative models are generated by identifying the various components of them as illustrated in Subsection 3.1.1. First, the degree of integration (I) is identified, and stationarity is tested by using the unit root test of Augmented Duckey Fuller test (ADF) to determine difference order (d). Afterwards, the lag orders of AR and MA are specified with correlation coefficients.

A time series is considered stationary if its mean, variance, and covariance are constant over time. As shown in the previous phase, the original time series (CONT) have shown a clear trend that violates one of the assumptions of a stationary time series. Stationarity is tested for different transformations of CONT. To confirm stationarity, the unit root test of ADF and Kwiatkowski-Phillips-Schmidt-Shin (KPSS) are conducted. Their results are shown in Table 5.1.

From the unit root table results, since the time series (LCONT) has a trend, it is nonstationary at 5% significance. On the other hand, the unit root test of the differenced time series DLCONT, D12LCONT and DD12LCONT are found stationary as their t-statistics (i.e., -20.6, -3.16, and -4.41), are below the critical value of -1.9424 at 5% significance level, meaning that we reject the null hypothesis that the series has a unit root.

To sum up, the LCONT has a unit root, and taking the first difference of LCONT results in the stationary time series of DLCONT. Also, the inclusion of a seasonal difference in both LCONT and DLCONT results in the two stationary time series of D12LCONT and DD12LCONT, respectively.

To generate tentative ARIMA models, the lag orders of AR and MA terms, p and q respectively, are identified. To do so, Autocorrelation Function (ACF) and Partial

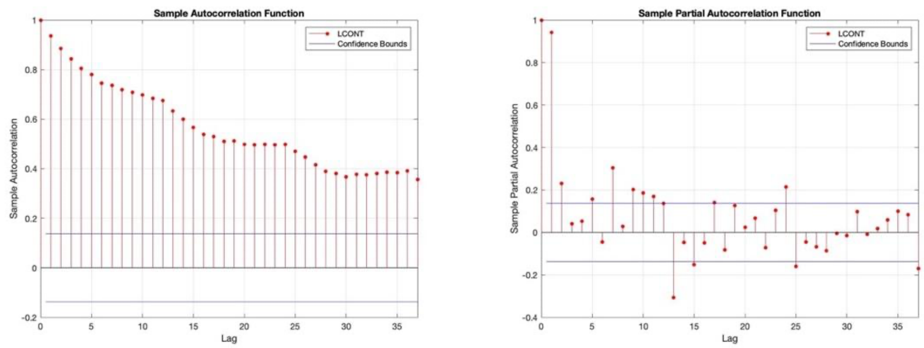
Autocorrelation Function (PACF) are shown in Figure 5.3; the plots are used to confirm the stationary status of the time series and identify model parameters. To identify the ARIMA model, the parameters AR(q) and MA(p) are identified by visualizing the plots. By matching the empirical autocorrelation and partial correlation patterns with the theoretical ones, it is often possible to identify one or several potential models for the given time series.

Figure 5.3a shows ACF and PACF of the natural logarithm container throughput (LCONT). Even though it is transformed, the ACF plot shows that the autocorrelation follows a slow linear decay pattern. Hence, natural logarithm time series (LCONT) is nonstationary. Taking the first non-seasonal deference of (LCONT) to generate (DLCONT). Figure 5.3b represents the ACF and PACF of (DLCONT). ACF plot indicates the presence of significant seasonal lags at 12, 24, and 36 outside the 95% boundaries emphasizing the existence of seasonal effect. PACF plot shows spiked at lag 1 and 6. As the PACF plot also depicts that the 12^{th} lag is positively significant, one might consider adding a seasonal AR. However, these plots do not provide a clear pattern for a definite model. Therefore, many tentative models are considered and tested with reference to the selection criterion of AIC and BIC.

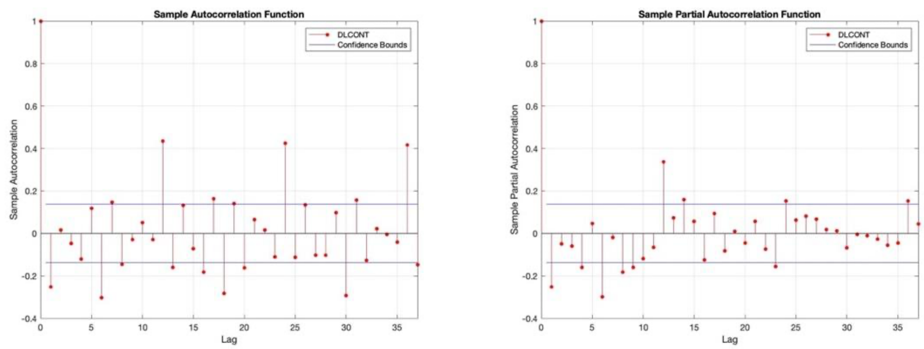
To account for seasonality, the datasets first seasonal differencing (D12LCONT) and second non-seasonal differencing (DD12CONT) are considered. The ACF and PACF for the two different datasets are also shown in Figures 5.3c and 5.3d, respectively. In Figure 5.3c, the ACF plot depicts that the autocorrelation decays slowly. The PACF depicts that partial correlation is significant at lag 1, then dies out fairly quickly afterward, suggesting an AR(1).

Table 5.1 ADF and KPSS Results

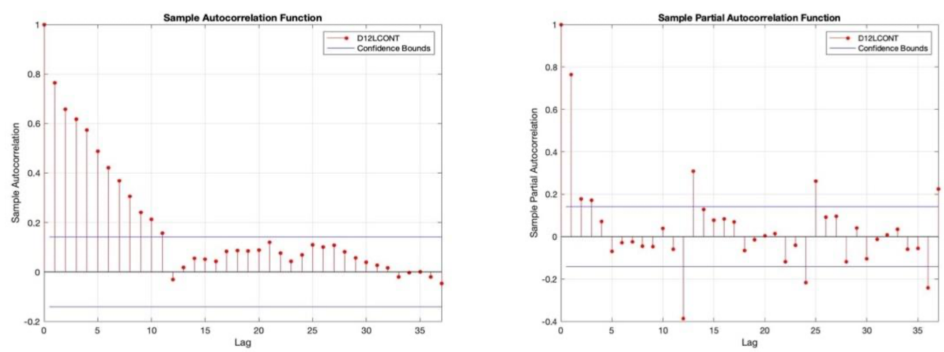
Time series		None	Constant	Constant & Trend	Constant	Constant & Trend
		ADF H0: The series has a unit root			KPSS H0: The series has no unit root	
LCONT	P-value	0.9422	0.0282	0.0229	1.4480	0.3721
	t-statistics	1.2123	-3.0996	-3.7241		
DLCONT	P-value	0.0000	0.0000	0.0008	0.3154	0.1915
	t-statistics	-20.6386	-20.7091	-4.7458		
D12LCONT	P-value	0.0016	0.0139	0.0495	0.5596	0.0968
	t-statistics	-3.1697	-3.3551	-3.4375		
DD12LCONT	P-value	0.0000	0.0004	0.0013	0.1856	0.0523
	t-statistics	-4.4101	-4.4394	-4.6315		
		The CRITICAL VALUES			Asymptotic critical value	
	1%	-2.5765	-3.4629	-4.0044	0.739	0.216
	5%	-1.9424	-2.8758	-3.4323	0.463	0.146
	10%	-1.6156	-2.5744	-3.1399	0.347	0.119



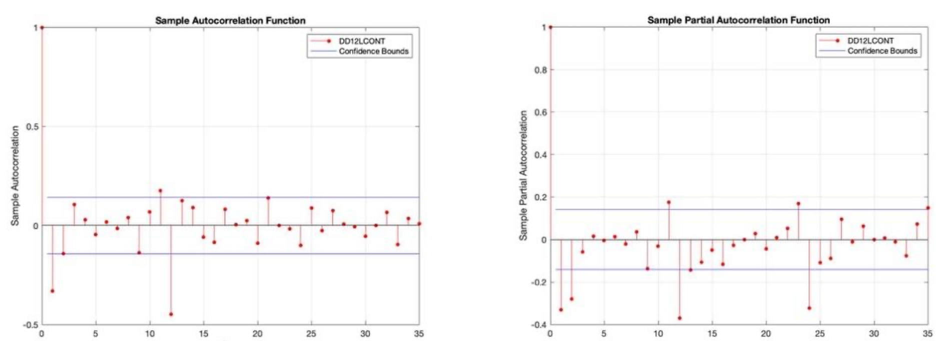
(a) LCONT



(a) DLCONT



(a) D12LCONT



(a) DD12LCONT

Figure 5.3 The ACF and PACF of various transformations.

Also, PACF shows that there is a significant negative correlation that dies out in lags 24, 36, and so on. Thus, seasonal MA might be considered suggesting the multiplicative seasonal model of $\mathrm{SARIMA}(1,0,0)(1,1,0)_{12}$. ACF and PACF plots of (DD12LCONT) series are shown in Figure 5.3d ACF plot shows significant spikes at 1st and 12th lags and dampening effects for the other lags which might imply the need for an MA(1) and seasonal MA(1).

5.1.3 Model estimation and diagnostic checking

Based on the autocorrelation visualization, various tentative models are considered. For time series that do not provide a clear pattern to generate tentative models, alternative models have been considered and estimated, and the best models are chosen based on the selection criterion of Akaike's information criterion (AIC). These information criterion measures are designed explicitly for model selection. Models having the lowest AIC and SIC values are considered the most optimal. To avoid over-differencing, parameter significance and MA invertibility are considered as well. Once the models are estimated, various residual diagnostic tests are applied to ensure variance consistency, homoscedasticity, and stationarity.

Out of the estimated models, the diagnostic checking of 6 models are illustrated in Table 5.2, namely, ARIMA(2,1,1), $SARIMA(2,1,0)(1,0,1)_{12}$, $SARIMA(1,0,1)(0,1,1)_{12}$, $SARIMA(1,0,1)(2,1,0)_{12}$, $SARIMA(2,1,0)(0,1,1)_{12}$, and $SARIMA(1,1,1)(0,1,1)_{12}$. Models not reported in the table may have insufficient parameters, violate the white noise assumption, or have a high selection criteria of AIC. The full diagnostic tests and graphs of these models are presented in Appendix B. In Table 5.2 considering residual normality

Table 5.2 diagnostic test s for the six potential models

		DLCONT		D12LCONT		DD12LCONT	
Test Type	Diagnostic test	(2,1,1)	(2,1,0)(1,0,1)	(1,0,1)(0,1,1)	(1,0,1)(2,1,0)	(2,1,0)(0,1,1)	(1,1,1)(0,1,1)
Model fit	R-squared	0.119731	0.528	0.827	0.708	0.523	0.519
	Adj R-squared	0.108863	0.518	0.825	0.701	0.517	0.513
	Sum squared Residual	0.835186	0.413	0.426	0.432	0.420	0.424
	Std. dev. Of residuals	0.071317	0.052	0.0525	0.0576	0.0525	0.0526
	Inverted MA and/or MA roots	<1	<1	<1	<1	<1	<1
Selection criterion	AIC	-2.411	-3.023	-3.017	-2.813	-2.96	-3.016
Residual normality	Skewness	0.024748	0.044914	0.053315	0.2153	0.030151	0.084628
	Kurtosis	2.751351	2.866969	2.881283	3.639557	2.904097	2.87293
	Jaraque-bera test	0.801757	0.921152	0.921063	0.197434	0.959918	0.866135
Residual serial correlation	Ljung-Box test	0.0000	0.150	0.156	0.243	0.186	0.298
	Breusch-Godfrey test	0.0000	0.3550	0.0929	0.2464	0.3232	0.3161
Residual variance	ARCH test		0.9929	0.4045	0.6619	0.9896	0.6272
Forecast accuracy	RMSE (static forecast)		33176.07	34115.76	35567.59	33002.41	33405.59
	MAPE (static forecast)		6.654	6.678	6.935	6.687	6.685
	RMSE (dynamic forecast)		35217.04	33495.93	32558.96	37906.19	38847.19
	MAPE (dynamic forecast)		7.736	6.327	5.586	8.629	8.950

a) When no AR root lies outside the unit circle, the model is stationary, when no MA root lies outside the unit cycle, the model is invertible.
b) Autoregressive conditional heteroscedasticity test (ARCH) is used to test if residuals have constant variance. The null hypothesis: the residuals are homoscedastic.
c) Ljung-Box test (portmanteau test) is used to test autocorrelation. The null hypothesis is that no serial correlation at residuals.
d) Static forecast uses the actual value for forecasting the subsequent value.

and according to the Jarque-Bera test, the estimated residuals of the proposed models are normally distributed where the probability of their J-B test exceeds 5%.

However, the residuals of ARIMA(2,1,1) are serially correlated which violate the assumption. On the other hand, the issue of serial correlation that is present in ARIMA(2,1,1) is invalid in the other five models, as the probability value of their Breusch-Godfrey test is over 5% meaning that all the other five models satisfy the residual assumptions. Considering variance consistency among residuals, the ARCH test is for testing residual variance. ARCH results prove that the residuals are homoscedastic with a probability exceeding 5% for each of the remaining models. To determine the most suitable model, forecast error is examined in the third phase.

5.1.4 Forecast Application and Findings

The five tentative models that satisfy all the diagnostic tests in the previous phase are used to generate forecasts from Jan 2017 to Dec 2019. Thereafter, the forecast results are compared to the actual observations on the training set specified in Subsection 5.5.1. Table 5.2 depicts the forecast accuracy of the tentative models. Two forecast accuracy measures are used, namely Root Mean Square Error (RMSE) and Mean Absolute Percentage Error (MAPE). The model that provides the lowest RMSE and MAPE forecast accuracy is considered the most reliable one. By comparing the accuracy of the models in Table 5.2, the model $SARIMA(1,0,1)(2,1,0)_{12}$ demonstrates the most reliable forecasts over the training set with a forecast error equal 6.65%. The model estimation output is reported in Figure 5.3, and its estimated equation is as follows:

$$(1-\emptyset_1 B)\cdot(1-\Phi_1 B^{12}-\Phi_2 B^{12})\cdot(1-B^{12})LCONT_t=(1-\theta_1 B)a_t \qquad (5.1)$$

In Table 5.3, it can be seen from the t statistics and p-value of model coefficients that the parameter estimates are significant at 5% with an R squared value of 70.7%. Using the SARIMA(1,0,1)(2,1,0) model, Figure 5.4 represents a comparison between the forecasted and the actual container throughput of Jeddah port for the period of January 2017 to December 2019. Despite the spike in the forecast of July 2019, the model shows accurate performance in explaining the container throughput of the port during the three-year period (2017-2019).

Table 5.3 Model Estimation of SARIMA(1,0,1)(2,1,0) Using D12LCONT Series

Dependent Variable: LOG(CONT)-LOG(CONT(-12))
Method: ARMA Conditional Least Squares (Newton-Raphson / Marquardt steps)
Date: 02/17/21 Time: 11:23
Sample (adjusted): 2006M02 2016M12
Included observations: 131 after adjustments
Convergence achieved after 8 iterations
Coefficient covariance computed using outer product of gradients
MA Backcast: 2006M01

Variable	Coefficient	Std. Error	t-Statistic	Prob.
AR(1)	0.953186	0.025741	37.03020	0.0000
SAR(12)	-0.561460	0.080527	-6.972355	0.0000
SAR(24)	-0.291634	0.077707	-3.752981	0.0003
MA(1)	-0.339851	0.090262	-3.765173	0.0003

R-squared	0.707948	Mean dependent var	0.037137
Adjusted R-squared	0.701050	S.D. dependent var	0.106775
S.E. of regression	0.058381	Akaike info criterion	-2.813604
Sum squared resid	0.432855	Schwarz criterion	-2.725811
Log likelihood	188.2910	Hannan-Quinn criter.	-2.777930
Durbin-Watson stat	1.930046		

Inverted AR Roots	.95	.94+.17i	.94-.17i	.89-.32i
	.89+.32i	.73+.61i	.73-.61i	.61+.73i
	.61-.73i	.32-.89i	.32+.89i	.17-.94i
	.17+.94i	-.17+.94i	-.17-.94i	-.32-.89i
	-.32+.89i	-.61+.73i	-.61-.73i	-.73+.61i
	-.73-.61i	-.89+.32i	-.89-.32i	-.94+.17i
	-.94-.17i			
Inverted MA Roots	.34			

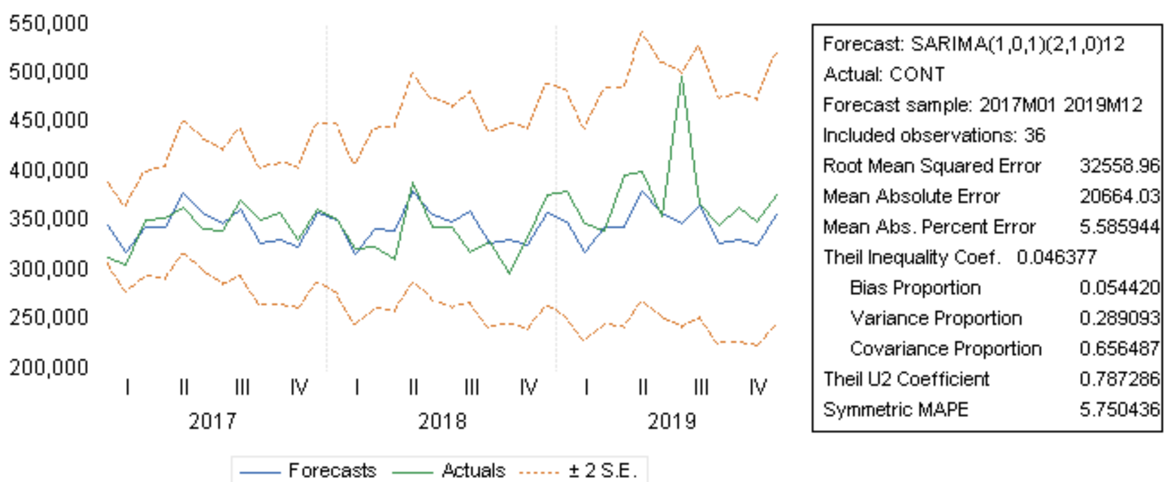


Figure 5.4 Forecasted container throughput of Jeddah port (Jan 2017 – Dec 2019).

5.1.5 Conclusion

The purpose of this subsection is to forecast container demand, understand the behavior and pattern of container throughput, and provide an approach that can be updated and applied in the future. The port of Jeddah is used in the empirical analysis to forecast container demand from January 2003 to August 2020. To validate the model, the time series data set is divided into two samples, that is the experimental set (the training set) and the validation set (the test set) As mentioned in Subsection 5.1.1. The empirical analysis resulted in an important conclusion. First, based on the data generating process, and assuming the historical pattern remains the same, this approach provided a reliable forecast of demand variation for the period of three years. Second, the model of SARIMA(1,0,1)(2,1,0) provides the most accurate forecast results. Based on the error measure of MAPE, the forecast error of the model is 6.6%.

Forecasting short-term container demand is of importance for port decision-makers by supplying an instrument to obtain insight into future demand. Port authorities and terminal operators aim to avoid traffic congestion and improve container handling efficiency. The short-term forecasts help them in the planning process for operation decisions and resource allocation. Therefore, improvements and modifications of operation plans may be applied to improve productivity and can be achieved by considering these forecasts.

5.2 Assessing the Role of Distance in Container Flow

The objective of this section is to measure the diversion of container flow between ports and provinces. It has two folds. First, the explanatory power of inland distance on the distribution of container flow is analyzed. This is done by analyzing the role of distance in the hinterland of competing ports. The analysis will detect the impact of distance in two years to investigate whether the role of distance in the container flow has changed. Second, assessing the potential market share of each port by considering the role of the inland distance between ports and provinces. This assessment will reveal the unexploited potentialities of the ports based on the inland distance. This is done by comparing the actual and predicted container flow based on the spatial analysis model.

In terms of how this section is organized, Subsection 5.2.1 data used to assess hinterland connectivity is described. In Subsection 5.2.2, model calibration and goodness-of-fit statistics are applied; also, model results are presented. In Subsection 5.2.3 findings are presented. Lastly, the conclusion is presented in Subsection 5.2.4.

5.2.1 Data Description

The analysis of this section deals with the study of Saudi ports. Since the container flow in the country is import-driven, only containers transported from ports to provinces are considered in the analysis. The four ports of Jeddah, KAP, Dammam, and Jubail are considered as the origins, and the 13 provinces are the destinations (Attracting container flow). The four ports represent over 95% of the containers transported to the 13 provinces, thus, other ports are not considered in the analysis.

There are two matrices that need to be distinguished. The actual container flow F_{ij} matrix and the inland distance matrix. These two matrices are used as input data, where

F_{ij} represents the actual flows from ith province to jth province. The container flows are measured in TEU. The inland distance matrix where d_{ij} represents the geographical distance between the ith province and the jth port and distance is measured in km. Geographic distance between ports and provinces is used in the analysis. It is calculated between the most populated city in each province. Further matrices are generated as outputs by applying the Spatial Interaction Models in the next section. The SIModel software is used to produce new O-D matrices (predicted OD matrix), which represents the redistribution of the observed traffic flows.

5.2.2 Model calibration

The spatial interaction model (SIM) is performed by applying doubly and constrained models. First, the doubly constrained model is applied to assess the explanatory power of distance in the distribution of container flow between ports and provinces in Saudi Arabia. This is done by estimating the distance decay parameter based on the port-province inland distance and container flows, then investigating the explanatory power of distance in shaping the hinterland regions. Thereafter, the distance-decay parameters β for the two time periods of 2011 and 2017 are compared to assess whether its role has evolved or not. The comparison provides an understanding of the evolution of distance explanatory power which can be beneficial in determining the behavior of spatial interaction.

Second, the attraction constrained model is applied, then, the estimated and the actual container flows are analyzed by using gap analysis to reveal the role of distance in each port-hinterland connection and to investigate unexploited potentialities of the Saudi ports. The attraction model is applied to the container flow by reassigning the container flow within the container ports. The redistribution of container volumes among ports is performed based on the inland distance and the total volume of each port, to iron out the frictions from the system and keep the traffic transported to each province fixed.

In the case of this study, the model proposed by Williams & Fotheringham (1984) is presented in Subsection 3.2.1. This allows the calibration of spatial interaction models through the Maximum Likelihood (ML) estimation. By running the model calibration on the container flow between i-ports and j-provinces, the distance parameter is estimated. The model is calibrated by using Spatial Interaction Model (SIModel) software, which the model calibration provides parameter estimates on the distance decay which provide useful information on the distribution of flow within an investigated system.

For the doubly constrained model, the flows are redistributed based on the distance, inflow totals to each province, and outflow totals from each port. The reason for using inflows to provinces is to ensure that the sum of estimated inflows to a particular province equals to the total actual inflow to it. Similarly, outflow totals are used to ensure the sum of estimated outflows from a particular port equals the total actual outflow from that port. To calibrate attraction constrained model, the model calibration assigns the predicted container flows based on the distance and the inflow totals to provinces, to iron out the frictions from the system and keep the traffic transported to each province fixed.

The calibration of a model involves optimizing the value of the parameter β. This parameter defines how close an estimated value is compared to the observed flows. From a mathematical perspective, the optimization of β ensures the probability that the flows estimated by the model are similar to the actual ones. According to the doubly constraint model, the β measures the role of the distance in defining the relationship between ports and provinces. β is sensitive to the inland distance and represents practical challenges that would cause traffic flow deviations (Clarke et al., 1986). Such challenges include inland transport cost and inefficiency of transport operation. The literature defines the port attractiveness as a function of distance-decay framework. The effect of β may take any value below zero, where:

$0 \leq \beta < -1$ friction is proportionately lower l with the distance.

$\beta < -1$ container flow is proportionately higher with the distance.

$\beta = -1$ container flow is directly proportional to the distance.

When beta equals zero, distance has no role at all in shaping the hinterland which results in a perfectly contestable hinterland. The closer to zero the β value, the lower the impact of distance in shaping the port hinterland which means that the role of distance in container flow is limited. In contrast, the role of distance in the distribution container flow is higher when the beta value is away from zero. In other words, the container flow is highly influenced by distance, thus, the explanatory power of distance is stronger in shaping hinterland resulting in a captive hinterland (C. Ferrari et al., 2011).

Two other parameters are used: Absolute Entropy Difference (AED) and corresponding statistical deviation (d) for Spatial Interaction models. These parameters are calculated for both models. AED can be conceptualized as a statistical index defining systems' entropy and d is a fitting parameter. AED is defined as the difference in the variance of the actual and predicted probability distribution in absolute value. AED is not used to measure the model fit but to evaluate the degree of influence that distance has on container flow.

5.2.3 Findings

To investigate the role of distance, a doubly constrained model is applied from 2011 and 2017. The β parameters of both years are compared to find out whether the impact of distance encountered any changes or not. The outcomes of the model calibrations are reported in Table 5.4.

Table 5.4 The Outcomes of the Doubly and Attraction Constraint Models for 2011 and 2017

year	constraint	β	AED	d(%)
2011	Attraction	-0.989	0.097	27.4
	Doubly	-1.044	1.333	29.3
2017	Attraction	-0.5815	1.216	55.7
	Doubly	-0.6671	2.523	57.2

Table 5.4 depicts the outcomes of the models of doubly and attraction constraints. In order to analyze the explanatory power of distance, the β value of the doubly constraint model is considered for the years 2011 and 2017. The β value of the doubly model in 2017 is -1.044 compared to -0.671 in 2011. This increase in β value (getting closer to zero) indicates that container flow became less sensitive to distance in 2017. This suggests an increased grade of contestability in the inland provinces, where the container flow to further inland markets from ports is less influenced by distance due to the decreased role of distance in 2017. Following the study conducted by Ferrari et al. (2011), two additional parameters are used to interpret the model results, namely, Absolute Entropy Difference

(AED) and corresponding statistical deviation (d). The two parameters are calculated for both models.

AED provides evidence of how much the variance of predicted flows compared to the variance of the actual flow, where a close value of AED to zero, indicates a minimal difference between the two variances whereas a higher AED indicates the model does not reliably fit the actual system. AED shows how much the actual system is predictable where the lower limit of AED can be zero when the system is fully predictable ($H_p = H_q$). In contrast, the upper limit of AED represents the maximum entropy with maximum level of uncertainty, when $H_p = 0$ and $H_q = ln\ (n)$; n is determined by the size of the system. Given that there are 13 provinces and 4 ports, the higher bound of AED equals to ln(n), where n is the maximum size of the system (the number of provinces divided by the number of ports). Therefore,

$$0 \leq AED \leq 3.6665 \qquad \text{for 2011}$$

$$0 \leq AED \leq 3.951 \qquad \text{for 2017}$$

The imbalance in serving the inland provinces between the western ports of Jeddah and KAP and the eastern ports of Dammam and Jubail is reflected in the AED values. Furthermore, the deviation statistics measures the deviation in the actual container flow F_{ij} compared to the predicted container F'_{ij}. It is measured in percentage where d=100% indicates that predicted flow distribution has no deviation from the actual flow. In 2011, the deviation in the estimated flow of attraction constrained model is lower (29%) compared to 56% in 2018. This indicates that even though the distance decay parameter has an impact on container flow, it weakly describes container flow in 2011, indicating that other factors may have influenced the explanatory power of distance.

While such value provides a general overview of the scenario, there might be some masking of the real degree of permeability of corresponding captive markets. This figure may mask strong captivity of some ports over closer hinterlands and overlapped hinterlands in the inland regions by major competing ports which raises the need for further consideration. Therefore, a more in-depth analysis is necessary to further investigate the impact of distance in shaping hinterland as it might reveal different judgment, especially on the explanatory power of distance on inland regions for the presence of any barriers.

Furthermore, observing the actual flows reveals a various degree of hinterland types, where local hinterland and the surrounding provinces, tend to be captive hinterland for the two major ports of Jeddah and Dammam in 2011. This picture changed in 2017 after the emergence of KAP and the developing port of Jubail penetrate the hinterland of the two major ports. Concerning the inland regions, the relative market share of Dammam port over the long-lasting captive hinterland of Riyadh province encountered a decline in 2011. The increase of freight rates during 2011 might have had a barrier effect for the Eastern ports, whereas the western ports of Jeddah benefited because of their advantageous geographic location in terms of the maritime shipping accessibility. The availability of railway services between the port of Dammam and Riyadh province provides the port an advantage over the other ports. Similarly, the other central province of Qaseem is closer to the eastern ports compared to the Western ports. Jeddah port gained higher market share of Qaseem in 2011, then the market share dropped significantly in 2017.

These observations also reveal the existence of non-homogeneous of different hinterland regions. Unfortunately, the distance decay parameter does not reflect such variability. Therefore, the varying developments in the market share at different regions highlight the need of investigating the evolution of the explanatory power of distance at the port-province level; as such, a permeability analysis is conducted to assess the competitive advantage of competing ports. The permeability index presented by Ferrari, et al. (2011) is used as well.

The index is calculated by considering the ratio between actual and predicted container flow. The index allows to measure the ability of each port in serving the hinterland region. Using the two container flow matrices: the actual flow F_{ij} and predicted flow F'_{ij} based on the attraction constrained model, the permeability index is calculated as follows:

$$I_{ij} = \frac{F_{ij}}{F'_{ij}} \tag{5.2}$$

Where I_{ij} may take the value of zero or any negative value. If I_{ij} is larger than 1, the port is able to serve the hinterland, and the impedance effect of the spatial distance is limited. If I_{ij} is less than zero, the port ability to serve the hinterland is limited or connectivity barrier exists. Figure 5.5 shows provincial permeability index by province for each port for the years 2011 and 2017.

Equally important, container traffic tends to decrease with the increase of distance where the surrounding provinces of Jeddah port in the west and Dammam in the east encountered some penetration of the small but developing ports of KAP and Jubail in 2017. On the West Coast, despite the decreased discrepancy of Jeddah port in the Northwestern provinces, the port still dominates in the Southern provinces. On the East Coast, the situation is slightly different where Dammam's ability to keep its local share has decreased. In 2011, a large proportion of containers handled in Jubail port were delivered to the local region. The port mainly serves the province of North Borders and, to a lesser extent, its province due to the dominance of Dammam port. However, this has changed in 2017, as Jubail port penetrated the hinterland of Dammam port in the Eastern province and Riyadh province. The large infrastructure development that Jubail experienced might be the reason for increasing port capability by reaching further hinterland.

Figure 5.5 also reveals interesting changes regarding the central provinces. In 2011, the Western port of Jeddah served a wider hinterland and had higher permeability index values in the central provinces of Riyadh and Qaseem compared to 2017. The port's market share was 16% higher in 2011 than in 2017, which indicates the reduction of distance impact on the container flow from Jeddah port to central provinces. The freight rates might cause this difference as shippers attempt to decrease the maritime haulage of their shipped containers when freight rates increase. Their attempt may cause longer inland transport haulage, but it can be compensated by reducing freight cost, thus, the overall cost saving. Therefore, the increasing share of the Western port of Jeddah in the central regions to the detriment of the Eastern ports might be attributed to the increase of freight rates in 2011 and the port users' preferences. However, the situation is different in Riyadh, where the permeability index of Jeddah port is relatively lower compared to that of Qaseem where Dammam port maintains a solid link to Riyadh due to not only to the close distance, since Jubail port has the same advantage, but also Dammam-Riyadh railway connection.

These results indicate that the availability of ports on different coasts, freight rates, and railway connectivity fuel the inter-port competition; thus, transforming the inland markets from captive to overlapping hinterland regions, since each port would have a competitive advantage, allows it to compete over the overlapped hinterland. However, the strength of port competition might vary in different regions based on the geographical configuration where larger countries tend to have a longer distance to inland markets that increase the impact of inland transport cost. The impact size of long-distance decreases when alternative inland transport modes are available.

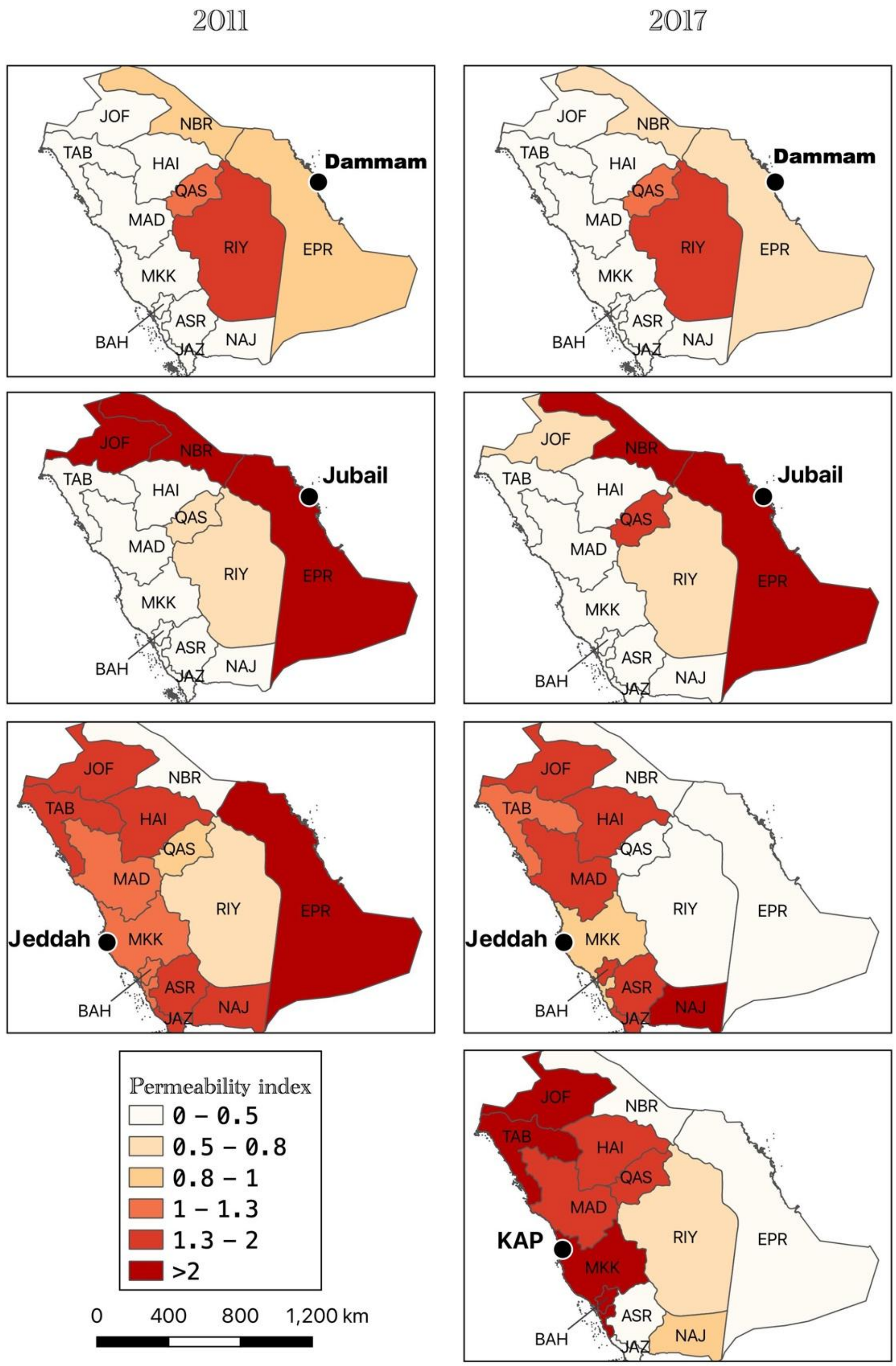


Figure 5.5 Provincial permeability index by province for each port (2011 and 2017).

5.2.4 Conclusion

The objective of this section is to determine the impact of distance on the diversion of container flow among the competing ports. This objective consists of first investigating the evolution of inland distance role on the distribution of container flow and secondly assessing the competitive position of the ports, based on the different hinterland regions. To achieve these objectives, two steps were conducted.

One, to investigate the evolution of the role of distance in two different years, the doubly constrained model is applied to estimate the distance parameters of the container flow in the years 2011 and 2018. These two parameters investigate whether distance power in explaining the container flow changed or not. Second, to assess the competitive position of the ports, the Attraction constrained model assesses the unexploited potentialities of the competing ports by analyzing the difference between the predicted and actual container flow at the port-hinterland level.

The results of the gap analysis are useful in highlighting the geographic competitive advantage of each port with respect to the other ports. The existence of other barriers that reduce the explanatory power of distance is also investigated. The analysis is conducted on the case study of Saudi Arabian ports for the years 2011 and 2017. This provides an understanding of port competitiveness within the country-level and the barriers that limit the competitive position of the ports. The analysis of this section reveals the following outcomes:

a) The stronger explanatory power of distance in 2011, indicates that the container flow is highly influenced and shaped by distance, resulting in a less contestable hinterland.

b) The role of distance in shaping the hinterland is stronger and more efficient in 2011 than in 2017.

c) The actual container flow in each province reveals the existence of non-homogeneous of different port hinterland, where the distance decay parameter does not reflect such variability. The results of gap analysis show the negative impact of poorly located ports in attracting cargo in the presence of high freight rates. The availability of railway connections in a particular port-province pair improves the port competitiveness position.

d) While foreland accessibility acts as a barrier effect in the hinterland expansion of Eastern ports, Jeddah port benefited from its geographic location in the Red Sea, especially in 2011 when freight rates reached their peak. Therefore, the impact of port location within liner shipping routes on the container flow is investigated in the next section to reveal whether the location of Jeddah port allows it to reduce the influence of distance as an impedance in reaching the central provinces of Riyadh and Qaseem.

e) Distance plays a major role in the distribution of container flow, shaping the hinterland where potential increases in container volume are limited as distance increases.

f) The findings of the applied SIM and gap analysis reveal that container flow distribution and the expansion of hinterland are influenced by (1) hinterland accessibility, (2) intermodal availability, and (3) foreland accessibility.

Further investigation on the impact of these factors in the distribution of container flow is conducted in Section 5.3.

5.3 Analyzing Port Competitiveness

The competitive position of the port depends on the port ability to attract port users and gain a higher market share compared to other ports. The port selection decision of port users is influenced by factors related to the entire transport chain in which the port is included. In the previous section, inland distance was solely considered in depicting the market share and measuring its role as an explanatory power in defining port hinterland. However, various factors related to the determinants of ports, hinterland and foreland impact the distribution of container flow and influence hinterland expansion; thus, port competitiveness.

In this section, two steps are followed to assess port competitiveness. First, investigating the impact of port geographical characteristics, and the intermodal connectivity in the inland distribution of maritime traffic at country-level. Poisson Spatial Interaction Model (SIM) is applied to analyze the impact of the impact of maritime accessibility, inland distance between ports and inland regions, railway availability between them and port location within maritime shipping routes in container flow, on port-province container flow in the case study of Saudi Arabia. Second, assessing port competitiveness by taking into account the actual and predicted container flow where they are the predicted and actual container flow are analyzed by applying gap analysis to assess the competitive position of ports at the provincial-level. The model outcomes reveal the role of these factors in shaping the hinterland of ports. This analysis provides information regarding whether the hinterland is captive or contestable.

Further, special attention is paid to the role of inland distance (port-province) maritime accessibility to determine which one has higher explanatory power in shaping the hinterland; and in turn, their impact is investigated by comparing the parameter significance of these two factors. In addition, the impact of intermodal accessibility is considered as well where Dammam port has an inland alternative to mode of railway for containers transported to the central region. Furthermore, maritime accessibility of ports is crucial especially for countries bordering two seas where ports are geographically located as in the case of Saudi Arabia. Given the availability of alternative inland mode in Dammam port and the strategic location of Jeddah port, in terms of maritime accessibility, the impact of port location and intermodal availability are of particular interest.

The analysis of this section may help port planners (e.g., port authority and terminal operators) in port planning decisions by understanding the port competitiveness and the impact of the above-mentioned factors in container traffic. Also, the analysis is of interest to policy makers and planners of inland transport since it sheds light on the distribution of container flow over the hinterland regions. This section is structured as follows: In Section 5.3.1 data used to assess hinterland connectivity is described. In Section 5.3.2, the methodology is applied. In Section 5.3.3, model calibration is conducted. Lastly, results and findings are presented in Section 5.3.4.

5.3.1 Data Description

SIM estimates the spatial interaction in the inland distribution of maritime traffic, container flow, is generated as annual port-province flow between the four major ports and the 13 provinces in Saudi Arabia during the years 2006-2018 based on data collected from Saudi Ports Authority (SPA), King Abdullah Port (KAP) website and the general authority for statistics of Saudi Arabia. The ports of Jeddah, KAP, Dammam and Jubail are used as origins. These ports accounts for more than 95% of the country's container throughput. Minor ports in the country are not assumed to impact the container flow of the four major ports and are therefore excluded from the analysis. For destinations, the 13 governmental provinces are used as destinations.

To model Spatial interaction, inland distance, Liner Shipping Connectivity Index (LSCI) are used as explanatory variables. The inland distance between the four ports and the 13 provinces, measured in kilometers, is obtained from Google maps by calculating the distance between the most populated city in each province and each port. LSCI data is collected from United Nation Conference on Trade and Development（UNCTAD). The index is used as a proxy for port performance as it measures the maritime Connectivity of the port to global liner network based on six components where these components reflect the deployed capacity of ports and the dynamic structure of global shipping network. More detailed information about LSCI is presented in Section 2.3.4.

In addition to the collected data, two dummy parameters are used in the SIM. The dummy of rail availability where a value of one indicates the availability of railway service between port-province pairs, otherwise, zero. The other dummy of port location is used to account for the advantageous port location where the dummy value is one if the port is located in the Red Sea, otherwise, zero. The total number of observations is 572. Data for KAP is available from 2014 to 2018 since the port opened in 2013, thus, the annual observations from 2006 to 2013 is 39 and 52 annual observations for the period of 2014 to 2018. Summary statistics of the variable is presented in Table 5.5.

Table 5.5 Descriptive Statistics

Variable	Obs	Mean	Std. Dev.	Min	Max
Container flow	572	86868.7	206225.6	0	1235900
LSCI	572	29.8	13.4	2.6	48.9
inland distance	572	1007.4	444.3	10	1775
Rail availability (dummy)	572	0.02	0.15	0	1
Port location (dummy)	572	0.41	0.49	0	1

The histogram distribution of container flow is depicted in Figure 5.6. There are some unique features of the data worth mentioning. About 32% of the container flows are zeros. Various studies in the literature emphasized that the presence of large zero-flow observations is the main estimation problem of gravity models. Gómez-Herrera (2013) indicated that estimating models that do not appropriately deal with the presence of zero flows perform noticeably worse than others where the excessive number of zeros leads to heteroscedasticity. Heteroscedasticity exists if the variance of the model error is not constant. The presence of heteroscedasticity indicates the violation of homoscedasticity assumption in OLS, thus, using OLS regression to estimate the parameter leads to biased estimates. As a result, having high frequency of zeros in the dependent variable (container flow) demands the use of an appropriate method that would allow for consistent estimates, which necessitates a careful consideration of the methods to prevent biased results.

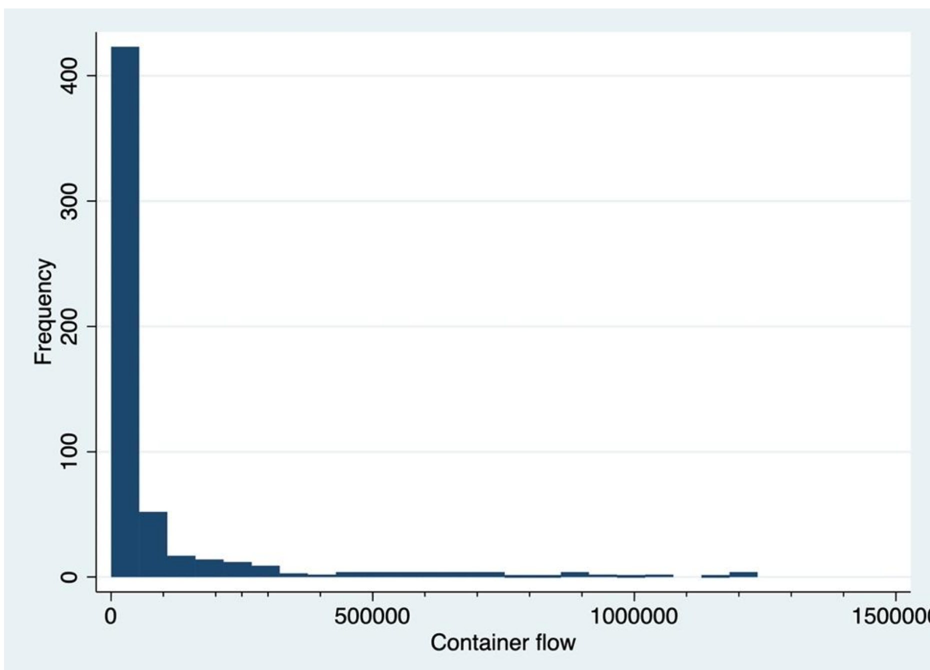


Figure 5.6 Histogram of port-province container flow (2006-2018).

Several ways are recommended in the literature to deal with heteroscedasticity. One of the ways is by taking the natural logarithm of the dependent variables. However, taking the logarithm of the dependent variable results in undefined value of observations with zero flow; thus, dropping them. Another option is to add a constant value of one to all the observations of the dependent variable. Adding a constant is not theoretically consistent as it present more variance that others (Gómez-Herrera, 2013). Alternatively, Silva and Tenreyro (2006) stated that to account for the excessive presence of zeros, it is appropriate to estimate gravity models such as spatial interaction models by using Maximum Likelihood (ML) estimator as it provides unbiased estimates in the presence of heteroskedasticity. Spatial interaction models are commonly estimated in the literature by using Poisson regression. Therefore, linear Poisson regression is applied, in this section, to estimate Spatial interaction Model.

5.3.2 Model specification

As indicated in the previous subsection, the Poisson regression method is applied to estimate the parameters of the Spatial interaction model where the impact of port characteristics, inland distance between ports and provinces, maritime accessibility, and intermodal connectivity on the inland distribution of maritime traffic is investigated. To estimate the SIM parameters, the analysis conducted in this section follows the approach proposed by Flowerdew & Aitkin (1982). The authors stated that Poisson distribution is applicable to fit SIM when the container flow between origins and destinations is assumed to be drawn from a discrete choice probabilistic process, as in this analysis of this section. Additionally, this approach is used when the mean of the response variable is a function of the independent variables. In other words, the mean of the container flow is assumed to be logarithmically linked to the linear combination of the independent variables. To converge the Poisson Pseudo-Maximum Likelihood (PPML) for the parameter estimates, the Poisson model is calibrated by using the iteratively Reweighted Least Square (IWSL) as it allows successive iteration until the value of the estimated parameters leads to convergence (Nelder & Wedderburn, 1972).

Even though this Poisson regression method provides consistent parameter estimations, the model results are not always efficient. Poisson distribution restrictively assumes that the variance of observation is equal to the mean. In practice, overdispersion occurs when observed variance is usually larger than the mean. This is due to two reasons: (1) the heterogeneity of observations causes the presence of extreme values in the dependent variable (Zeileis, 2004); and (2) the selection consists of more than one individual where their behavior varies (Flowerdew & Aitkin, 1982). In container flow analysis, both circumstances exist since an excessive number of flows may have zero value and choosing a particular port result from various strategies of port users (Moura et al., 2018). In such settings, estimating Poisson regression with the presence of over-dispersed data and the use of default standard error leads to invalid statistical inference and inefficient model results (Lindsey, 1999). The reason for this is that default standard errors can greatly overstate estimator precision, resulting in considerably small standard errors; thus, invalid statistical inferences (Hoechle, 2007). Therefore, attention must be paid to standard error to increase model efficiency.

To ensure the SIM fits well without violating Poisson regression assumptions, the clustered robust standard error is generated instead of the default standard error. The clustered standard error (covariance matrix estimator) allows the relaxation of the independently distributed residuals assumption (Froot, 1989; Rogers, 1992). In the presence of independent distributed residuals, the obtained standard errors by using the cluster estimators are consistent even if heteroscedasticity is seen across residuals (Newey & West, 1987). Therefore, standard errors are clustered by port-province pairs, in this section, to account for any intra-cluster correlations at the port-province pair-level.

Regarding choosing a suitable Spatial Interaction Model, an attraction constrained model is chosen to investigate the impact of LSCI, inland distance, railway availability, and port location in the inland distribution of container flow. This due to the model ability to consider both the spatial impedance and port-related factors. Briefly, the basic (unconstrained) spatial interaction model considers the impact of both origin and destination where origin represents port attribute and destinations represent province attribute. Since the analysis in this section is based on the factors that impacts the distribution of container flow from a transport perspective, using the unconstrained model does not allow for an analysis to be exclusively related to the distribution of container flow. Moreover, doubly constrained model only considers the impact of distance and does not allow fitting factors related to the port. Therefore, the attraction constrained model is used to fit the panel dataset, presented in Subsection 5.3.1, by including factors related to the port attribute (origin) and constraining the province attributes (destination). A detailed comparison of Attraction constrained model and the other SI models is presented in Subsection 3.2.2.

The attraction constrained model is applied to estimate container flow (F_{ijt}) between ports and provinces where ports are defined as i ($i = 1,2, \ldots, I$), provinces are defined as j ($j = 1,2, \ldots, J$), and years are defined as t ($t = 1,2, \ldots, T$). To represent factors related to port performance and geographic characteristics, the model involves fitting the variables: Liner Shipping Connectivity index (I_{it}), inland distance (d_{ij}), rail availability

(R_{ij}) and port location within shipping routes L_i. Based on the attraction constrained Equation 3.30 presented in chapter three, the model is reformulated to fit the variables of interest as follows:

$$F_{ijt} = exp\left(\alpha + \beta_1 ln\, I_{it} + \mu_j - \beta ln\, d_{ij} + \beta_3\, ln\, R_{ij} + \beta_4 ln\, L_i \right) + \varepsilon_{ij} \qquad (5.3)$$

Where:

F_{ijt} container flow from port i to province j

α constant

I_{it} liner shipping connectivity index for port i and year t

μ_j fixed effect for province j

d_{ij} inland distance between port i and province j

R_{ij} dummy for the availability of rail services between port i and province j

L_i dummy for the location of port i

ε_{ij} the clustered error on port-province pairs

Liner Shipping Connectivity Index (I_{it}) is used as an attractiveness measure of port i at year t. Obtaining a variable that represents port performance is a difficult task, Liner Shipping Connectivity Index (LSCI) is used as a proxy for port efficiency. This supported by De Oliveira & Cariou, (2015) who suggests the use of LSCI as a proxy of the competitive pressure from shippers and liner companies as the authors analyzed the relationship between port efficiency and LSCI and found that they are positively correlated. Moreover, this index is a good measure to compare different ports as it contains measures related to both port ability to serve ships and the proximity to dynamic global liner network. Thus, LSCI is used as a proxy of for port attractiveness.

The distance decay parameter d_{ij} is a repulsion measure of flow between port i and province j. Its sign is expected to be negative since the increase of distance causes a reduction in container traffic. Distance parameter of $\beta < 1$ implies that the parameter is less than proportional to distance indicating the hinterland is contestable, and conversely, $\beta > 1$ implies that the parameter is more than proportional to distance indicating to captive hinterland (Claudio Ferrari et al., 2011; Guerrero, 2018). The higher the parameter, the higher the distance impact on preventing the ports from reaching further markets, thus, limiting their hinterland scope.

The dummy variable R_{ij} is consider the availability of railway services between port i and province j where the dummy equals 1 if railway service exists, otherwise, 0. The sign of the parameter coefficient is expected to be positive since the availability of rail connection provides an advantage for port i compared to the other ports. Maloni and Jackson (2005) and Castillo-Manzano, González-Laxe, and López-Valpuesta (2013), States that rail services is widely recognized as a critical factor in the port's ability to compete for hinterland flow

As container demand at ports is likely related to the cost of shipping. the location dummy is included to account for the impact of freight rates on port location. The dummy is included for for two reasons: 1) fuel price is a major element of ship operating cost which has a direct impact on freight rates of liner shipping services (Stopford, 2009). 2) As shippers select the lowest transport cost, they prefer ports that provide the lowest freight cost. Therefore, the dummy variable L_i refers to the port location within shipping routes where it takes a value of 1 if the port i is located in the Red Sea, otherwise, 0. Its expected sign is positive as ports strategically located in the Red Sea have better accessibility to maritime shipping routes which provides a competitive advantage for the port i compared to the other ports.

To constrain the destination attribute, the model involves fitting a time-invariant fixed effect factor μ_j that allows capturing the unobserved effect of each province. Each province is presented by a dummy variable (μ_j =1, 2, …, μ_J) that has a value of one in every observation the province is part of. For example, the dummy variable (μ_1) is set to 1 for Riyadh province, otherwise, zero; similarly, the dummy variable (μ_2) is coded as 1 for Makkah province and zero for other provinces and so on. To avoid perfect multicollinearity, each province has a dummy variable except one that does not have a dummy and the attribute of this province is explained by the default intercept. In the next subsection, the model outcomes and findings are presented.

5.3.3 Findings

In this subsection, first, the attraction constrained model is applied to assess the impact of geographical characteristics, LSCI and intermodal availability in the distribution of maritime traffic. Thereafter, the gap between predicted and the actual container flow is analyzed to reveal the unexploited potentialities of ports. The outcomes of the gap analysis are used to assess port competitiveness.

The outcomes of the attraction constrained model are represented in Table 5.6. The explanatory power of the considered variables differs considerably. The obtained values of parameters allow the observed flow to fit with a 91.7% accuracy.

Table 5.6 Outcomes of the attraction constrained model

variable	coefficient	clustered s.e.	z	P-value
I_{it}	1.126	0.256	4.40	0.000
d_{ij}	-0.812	0.151	-5.38	0.000
R_{ij}	2.009	0.257	7.80	0.000
L_i	0.662	0.317	2.09	0.037
R^2	0.9098			
Obs.	572			

The distance-decay parameter (d_{ij}) is a repulsion measure that is used as measurement of trade impedance between ports and provinces since an increase in distance impacts container flow negatively, and vice versa. The distance parameter is statistically significant and has a negative effect on container flow which is in line with the apriori expectation, meaning that a rapid decrease of the intensity of inland flows with an increase in distance. The friction parameter provides an overall indicator of inland distance on the level of captivity in the hinterland. A friction parameter lower than 1 means that the friction parameter is less than proportional to distance which indicates a contestable hinterland. The higher the friction the more limited is the scope of the hinterlands (Guerrero, 2018). The model estimates a distance parameter of -0.812<1 which indicates that the hinterland is more of a contestable type.

Additionally, the impact of inland distance on the freight flow to the hinterland region depends on the geographic characteristics of the country where the friction value is higher in large countries. This can be seen in the case of the United States (Levine et al., 2009) where the friction of distance is relatively weak compared to other countries that are geographically smaller such as Italy (C. Ferrari et al., 2011), Spain (Moura et al., 2017) and France (Guerrero, 2018). One of the reasons for this impact is that intense container flow in long haulage in port-province pairs caused the small friction of distance. Low value of distance parameter is caused by the fact that long-distance inland transport is more frequently common (Thill & Lim, 2010). This can be seen in the case of this study where a high volume of container flow is transported to the inland province of Riyadh from ports of Dammam and Jeddah where the two ports are located 430 and 950 km away from Riyadh, respectively. The flow share of both ports to Riyadh province is 26% of the total container throughput in 2018. Further analysis at provincial-level is conducted later in this section to analyze the impact of distance on each port individually.

However, the share of Dammam port share on total container inflow to Riyadh is 72% in 2015. This large share is due not only to the proximate distance as Jubail port share the same advantage, but also to the availability of railway connection between Dammam and Riyadh. The impact of this railway link can be seen in the dummy variable of (R_{ij}) which represents the Dammam-Riyadh railway connection. In conformity with apriori expectation, the parameter is statistically significant and has a positive sign. This may demonstrate the impact of rail in the distance decay parameter as the friction of distance decreases due the availability of railway services (Debrie & Gouvernal, 2006; T. E. Notteboom & Rodrigue, 2005; J. P. Rodrigue et al., 2010).

Therefore, the availability of railway connection is a major determinant in container inflow to Riyadh since most of the transported containers to Riyadh are shipped through Dammam port, where 38% of total container flow between Dammam and Riyadh were transported by rail in 2015. This reflects the importance of the Railway connection as an alternative intermodal, providing Dammam port an advantage over the neighboring port of Jubail and the ports located in the Red Sea.

Considering the impact of maritime connectivity, the value for the parameter (I_{it}) is statistically significant with a positive coefficient value of 1.126. Port having larger index provide shippers with larger alternatives of shipping options to select from; thus, better port services.

It also indicates that port size is a primary variable in explaining the pattern of container flows through the ports since have larger number of port calls and have a deployed capacity are larger that allows it to reach further hinterland regions. Comparing inland distance and LSCI, the impact of LSCI greatly surpasses the impact of inland distance, indicating that maritime connectivity has higher importance in the preferences of shippers when it comes to port choice. This result is counterintuitive to what was found in the studies of (Guerrero, 2018) who investigated the impact of ship size and inland distance and found that distance outperformed ship size in explaining container flow. However, the variable used in his analysis only considers ship size. Variables, other than ship size, impact maritime connectivity such as number of port calls, liner services and the deployed capacity in the port.

The variable (L_i) represents the port location has a positive sign and appears to be statistically significant, as expected, with a coefficient of 0.662. Since the location variable represents the maritime haulage distance, having shorter distance to major routes, compared to competing ports, leads to reaching further shippers and wider hinterland, because strategic location contributes to reducing shipping cost. This is due to the fact that shipping containers through ports located closer to maritime routes tend to reduce the cost of maritime shipping, as it leads to reducing the total cost of container transport which consists of various types of costs like port charges, inland transport cost, and maritime shipping cost (Talley, 2014). Additionally, Parola et al., (2017b) state that the port selection, from the perspective of shippers, is mostly influenced by the factors related to port location. This can be seen in the case of Jeddah and KAP ports. The two ports are located in the Red Sea, within major maritime shipping routes, the parameter value proves that their strategic location is a determinant in the hinterland expansion compared to the ports located in the Persian Sea. This confirms the findings of Guerrero, (2018) who pointed out that the geographic location impacts the process of port selection, thus, expanding hinterland.

In the analysis of this section, LSCI is used as a proxy for the maritime connectivity of the ports. Other proxies were used in the literature. Meersman et al. (2010) argued that container throughput could be used as a measure for Port performance. Container throughput is used as a proxy of port performance in previous studies to analyze container flow. Therefore, a comparison is conducted between the use of LSCI and container throughput in the Spatial Interaction Model to investigate their impact on the model outcomes. The model outcomes using container throughput are presented in appendix C.1. By incorporating LSCI, its coefficient value is significant at 1.12. On the other side, fitting Imported/exported container throughput results in a coefficient value of only 0.45. The larger coefficient value of LSCI, compared to TEU, indicates that maritime connectivity has a higher impact on container flow. The comparison disclosed that maritime connectivity has a significant effect on container flow. Meaning that the ports located in the East coast are more constrained by maritime connectivity than ports located in the

Western coast. The estimated outcomes of the two models emphasize the significant impact of port accessibility since LSCI reflects both maritime and considers port efficiency.

Port efficiency is highly influenced by port location and maritime connectivity. Tongzon (2001) indicated that port efficiency could be positively influenced by the port's geographical location, port costs, infrastructure quality, and maritime connectivity. LSCI is calculated based on six components (a) the number of scheduled ship calls, (b) the total deployed annual capacity in TEU, (c) the number of liner shipping services from and to the port; (d) the number of liner companies providing shipping services from and to the port; (e) the average carrying capacity of ships in TEU; and (f) The number of other ports that are connected directly to the port. Given that LSCI considers factors related to port characteristics and connectivity to the liner shipping network, it provides a better proxy for port efficiency and maritime connectivity. Therefore, the use of LSCI provides a better proxy for the port impact on container flow than the use of container throughput.

Additionally, incorporating liner shipping connectivity into the model slightly reduced the coefficient value of port location. This is due to the impact of port location on the LSCI. However, port location is considered in the model to account for the shipping cost, which impacts freight rates. Therefore, port location influences the port choice since ports located in the Red Sea provide a lower freight rate than ports located in the Persian Gulf. Compared to LSCI, container throughput in the Spatial Interaction Model shows no substantial differences in the coefficient values of inland distance and rail availability.

5.3.3.1 Analyzing unexploited potentialities of the ports: The parameter values in the attraction constrained model showed the explanatory power of each factor in the container flow. However, additional investigation is conducted to analyze the port competitiveness at the port-province level by assessing the unexploited potentialities of competing ports. Gap analysis is applied to measure the gap between the predicted and actual container flow; that is, it takes into account the magnitude of the traffic flows. The gap measure is the percentage of the difference between predicted and actual flows. The goal is to find out whether port-province flows are overestimated or underestimated, and in turn to understand if a particular port-province pair is more competitive with respect to that of the competing ports. If actual flow is smaller than predicted flow overestimation, it is indicative of the presence of unexploited potentialities (weak port-hinterland link). On the other hand, underestimation indicates that actual flow is larger than the predicted flow, thus, the hinterland is relatively captive (Ferrari et al., 2011; Moura et al., 2018).

Figure 5.5 presents the gap percentage of the four ports in each province for 2011 and 2017. The figure reveals that ports on both coasts are able to sustain their position in the local hinterland. Still, some interesting changes appear in the surrounding provinces and inland hinterland. On the West Coast, the emergence of the port of KAP had limited impact in the local captive hinterland of Jeddah port, where Jeddah port is in the most populated city in the province, Jeddah city. The situation is different in Madinah province, where KAP has the advantage of being close to the province. The discrepancy of Jeddah port in Madinah province decreased in 2017. The massive investment in KAP port allows it to increase its market share since it started operating in 2013. The market share growth of KAP is expected to continue once the new infrastructure investments become available for operation in the future.

Regarding the surrounding provinces, Jeddah port kept its dominant share in the Southeastern provinces of Baha, Aseer, Jazan, and Najran. This situation might change once the new container terminal operator of Hutchison port holding completes the development of terminal facilities in the Southwestern port of Jazan. These developments are expected to become available in 2022 and will provide shorter inland distance to the provinces of Jazan, Aseer, and Najran. The figure also reveals interesting findings of the evolution of the contestable hinterlands in the inland provinces. It is crucial to pay attention to the central provinces of Riyadh and Qaseem.

The analysis shows that the flow from Jeddah to Riyadh is slightly underestimated for 2011 but overestimated for 2018. The significant increase in freight rates in 2011 might be the reason for the underestimate in that year, leading to an increase in the port share in the central provinces. The location of Jeddah provides it with better maritime accessibility compared to Dammam port. However, the maritime accessibility of Dammam does not impact the port significantly due to its closer inland distance to Riyadh. Also, the availability of railway connection between Dammam and Riyadh provides an inland alternative to shippers. The railway impact can be seen in the discrepancy outcome where the actual flow of Dammam port to Riyadh surpasses the predicted flow in both years, highlighting the importance of the railway connection as an alternative mode that provides an advantage for Dammam. Better inland connections provide competitive advantage for ports, thus enhancing their opportunity to increase their share (Acciaro et al., 2017; Kramberger et al., 2018). The figure also shows that the position of Jubail port in Riyadh strengthened in 2018 indicating that the impact of inland distance between them, and the port's infrastructure development has led to enhancing its competitive position.

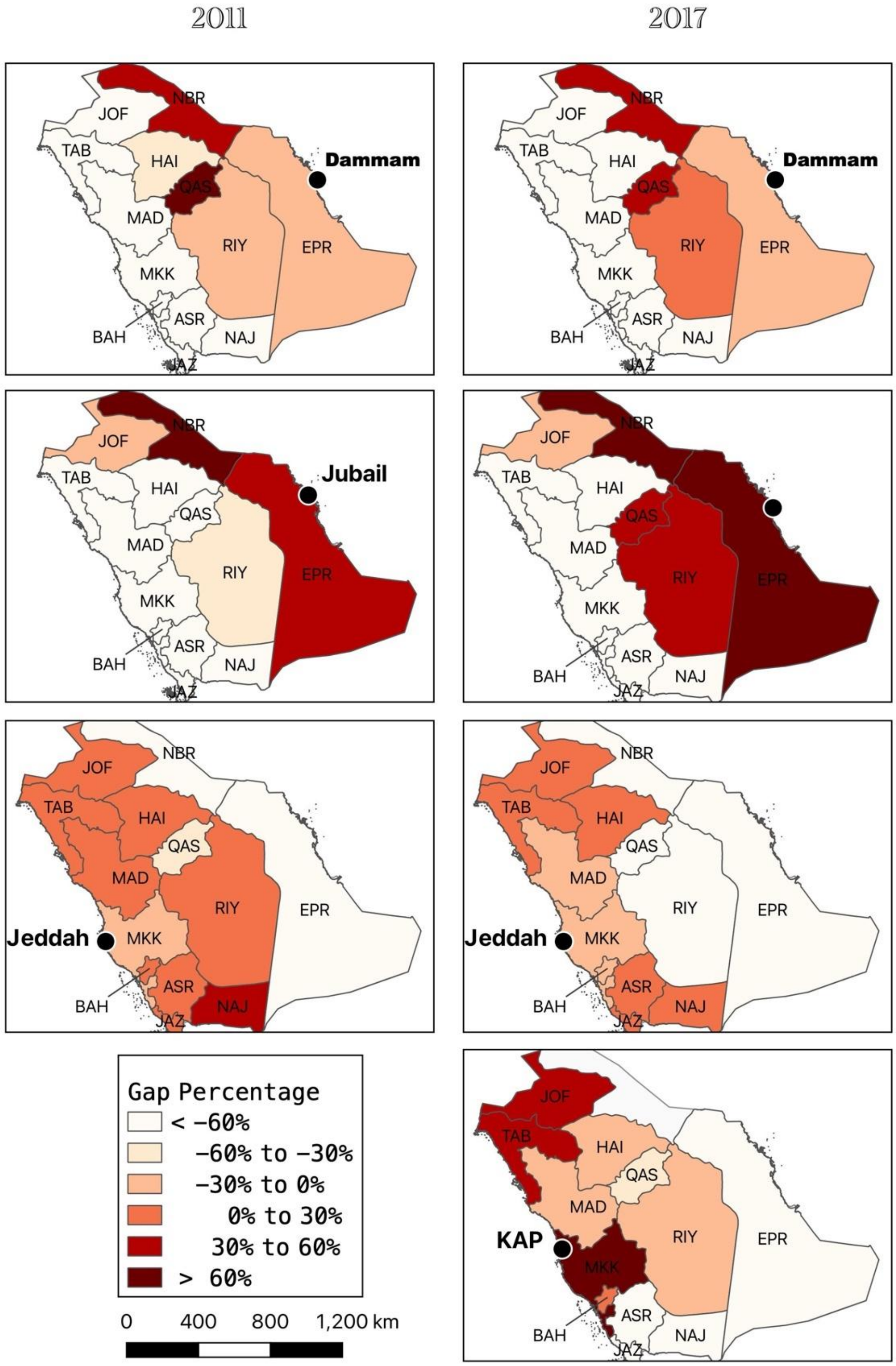


Figure 5.7 Gap percentage by province for each port based on Poisson SIM (2011-2017).

The situation is slightly different in the province of Qaseem. The figure reveals a high positive discrepancy for the Eastern ports in Qaseem province, where the two ports benefited from being close to the province. On the other hand, even though the combined share of two western ports is 34% in 2017, the linkage between the two ports and Qaseem is overestimated, indicating that unobserved factors related to the long-distance act as a key barrier such as the increased inland transport cost. This cost increase results from the new pricing policy of gasoline in Saudi Arabia, which led to the increase of gasoline prices.

The flows between the two eastern ports and the provinces of Riyadh and Qaseem have a significant positive permeability value due to their location in terms of inland accessibility, an accessibility that increases their competitiveness. The situation is different in the other two central provinces of Hail and Jouf, however. It is interesting to pay attention to the overestimation of the flow from these two provinces. The four ports have almost the same inland distance to Hail and Jouf provinces; but, the Western ports of Jeddah and KAP hold the dominant share in both provinces due to their geographic location within liner shipping routes. Despite the high permeability index value of KAP in 2017, the index value of Jeddah port remains positive in 2017. Therefore, both ports have the capability to penetrate the hinterland of Hail and Jouf. The planned expansion of the container terminal in KAP is expected to increase the port share and intensify competition.

The outcome of the gap analysis confirms the following major findings:

1) The primary role of rail availability in enlarging the market share of Dammam port in the province of Riyadh;

2) The importance of inland distance in increasing port competitiveness in the surrounding hinterland;

3) Given that inland distance acts as a barrier for the hinterland expansion, the results confirm the importance of maritime accessibility of ports in shaping the inland hinterland.

The two western ports of KAP and Jeddah benefit from being transshipment hubs due to their geographic location in terms of maritime accessibility. Their advantageous location makes them attractive destinations from mega containerships, allowing the ports to expand their hinterland beyond the surrounding region. However, maritime accessibility has a strong impact in the absence of railway connectivity since railway connections provide a lower-in-cost inland alternative for shippers, and thus reduce the impact of the costly longer maritime haulage and decrease the overall transport cost.

Moreover, to investigate the impact of the factors considered in the Attraction constrained model on the evolution of container inflow to the central provinces during the study period, Figure 5.8 presents the actual and predicted container flow of the central provinces of Riyadh, Qaseem, Hail and Jouf from 2006 to 2018. The figure reveals the evolution of the hinterlands in some of the central provinces are relatively overlapped. The container inflow to the Riyadh province shows the impact the new development in Riyadh dry terminal in 2015, where the new development caused a significant growth in Dammam-Riyadh flow, which indicates that the development of railway services in the dry terminal in Riyadh has contributed to the growing market share of Dammam port at the expense of western ports.

Unlike the case of Riyadh province, the situation is different in the other three central provinces of Qaseem, Hail, and Jouf, whose truck transport is the only inland transport. Due to the proximate distance separating the two eastern ports from the province of Qaseem, Dammam and Jubail ports are dominant over the other central province of Qaseem, where the former port gains the higher share of flow. Calling at ports located close to the major shipping routes causes considerable savings on the freight shipping cost (Ferrari et al., 2006). The contribution of inland distance to hinterlands is influenced by maritime accessibility of ports. This can be seen due to the strategic location of Jeddah port where the port share increased in the province of Qaseem during the spike of freight rates in the years 2010-2012. The other two provinces of Hail and Jouf are approximately located in equal distance from the eastern and western ports. Nevertheless, Jeddah port has a greater advantage in serving them throughout the study period due to the strategic maritime accessibility of Jeddah port. These findings reveal the importance of foreland accessibility. The choice for a port to call and the increase of the port's activity, therefore, depends on the port accessibility not only to hinterland but also to foreland.

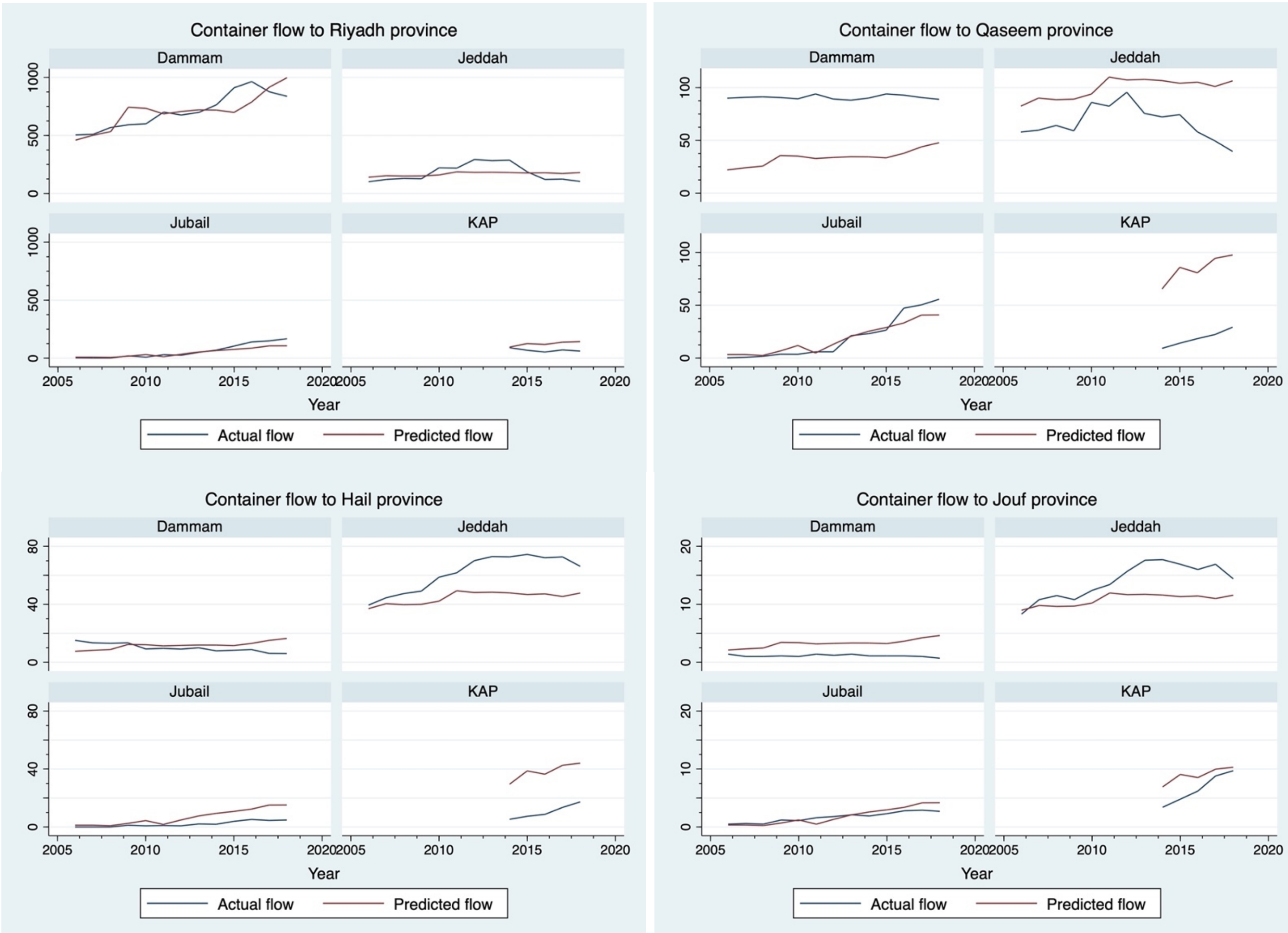


Figure 5.8 Comparison of actual and predicted inflow to the central provinces (2006-2018) in thousands.

To sum up, the results of the analysis conducted in this section reveals these major findings:

1) Inland distance between ports and provinces is a significant factor in the distribution of container traffic, and it confirms the conclusion drawn by Garcia-Alonso & Sanchez-Soriano (2009) who stated that container outflow to provinces (destinations) tends to seek the shortest inland route from ports. Thus, inland distance is a crucial factor that influences the competitiveness position of ports.

2) The port location to major liner routes is a primary variable in explaining the pattern of import container flows through the ports in the model. Western Ports of Jeddah and KAP, located strategically in a close distance to major maritime shipping routes, tend to reach further hinterland destinations, thus, serve most destinations and boast the best of hinterland connections compared to the Eastern ports of Dammam and Jubail.

3) The contribution of maritime accessibility of a port is not enough to explain container flow. The availability of intermodal connection from Riyadh province to inland hinterland is very important for the success of the port despite its weaker maritime accessibility to shipping routes. This gives the port an advantage over the western ports, thus, allowing Dammam to gain the dominant share in Riyadh province. This is in line with Acciaro et al. (2017) and Kramberger et al. (2018) who emphasized the importance of inland transport services for enhancing the capability of ports.

4) Port size is an important factor impacting container flow. The hinterland size of the port of Jubail, which is located in a proximate distance from Dammam port, has been considerably smaller than its neighboring port. The port's closeness to inland hinterland did not favor its evolution of container traffic. Similarly, the hinterland expansion of KAP port is smaller than that of Jeddah port. Consequently, the evolution of the traffic of the ports of Jubail and KAP is much more linked to their local hinterland than to distant hinterland. This is in line with Guerrero (2018) who pointed out that smaller ports have a higher proportion in markets located closer to them compared to further ones.

5.3.4 Conclusion

The major goal of this section is to explore the interactive relationship between inland distance and container flow in the hinterland by considering the geographical characteristics of competing ports and the intermodal availability in the hinterland. To achieve this goal, this study analyzed the spatial interaction of the port–hinterland container flow by investigating the impact of inland connectivity, intermodal availability, port size and its maritime accessibility on the inter-port container traffic distribution. The empirical analysis synthesized the available information on the container distribution from the four major ports of Saudi Arabia, namely, Jeddah, KAP, Dammam and Jubail, to the thirteen provinces in the country for the period (2006-2018). First, spatial interaction model is applied to investigate the explanatory power of the determinants in shaping the hinterland, then, analyzing the discrepancy between actual and predicted container flow to investigate the unexploited potentialities of ports at province-level.

The analysis showed that inland accessibility and port location within maritime shipping routes influence container flow, thus, port choice in Saudi Arabia. The study of Guerrero (2019) points out that the issues of port selection are not relevant in the case of developing countries whose high inland costs might be the reason why stronger effects of distance leads to limiting the contestable hinterland. In Saudi Arabia, this might not be the case where significant investments have been implemented in the last two decades to develop the inland connections between major cities, eventually contributing to expanding contestable hinterlands in some inland provinces.

However, the contestable hinterland might be caused by the wide geographic configuration in Saudi Arabia which lead to long-haulage as in the case of the U.S. where the friction of distance is relatively low compared to other geographically smaller countries whose friction in distance is large (Levine et al., 2009). Therefore, it would be interesting that further spatial analysis is conducted in the case of Saudi Arabia to investigate the impact of transport cost.

In addition, port characteristics contributes more than inland distance in explaining the distribution for container traffic, thus, delimiting hinterland in the case of inter-port competition at the country-level of Saudi Arabia. It is concluded that the evolution of inland distribution of container flow is greatly influenced by the strategic and geographic characteristics of ports. This is in line with Fleming and Hayuth (1994) who identified that the spatial characteristics of "centrality" and "intermediacy" play a significant role in the evolution of port activity. As a result, the inland accessibility and the port location within the shipping routes are two factors that impact the inland flow distribution of maritime traffic. Since the determinants of foreland and hinterland complement each other and are essential for the competitive position of the port, ports included in major shipping routes which have sufficient inland accessibility, are able to sustain a high share of hinterland demand.

The conducted analysis on the impact of the determinants in port-province container flow may lead to some interesting suggestions and ideas for the planning and development of not only the port but also the hinterland, as it identifies the significance of geographical conditions and transportation facilities in the port–hinterland relationship. Paying attention to the impact of infrastructure availability and spatial and location characteristics of container flow may help to avoid overcapacity and congestion of ports and inland corridors. This improves the efficiency in the allocation of resources and reinforcing the competitiveness of domestic exports (Moura et al., 2018).

Accordingly, it is essential to develop approaches to enable rapid development in a sustainable manner while maintaining economic growth through the coordination of ports and hinterlands. Therefore, the empirical analysis can help various port and inland transport stakeholders make better decisions to meet their respective objectives as follows: (1) For governments, this study can help in the evaluation of the impact of port strategies and the influence of developing inland infrastructure and changes in the economic activity in the hinterland; (2) Even though port location is fixed, by understand inland flow, port authorities can get insight into the needs of hinterland regions to improve the provided services in the port, as well as determining port competitiveness in the different inland markets to assess whether to seek cooperation with rival ports or not; (3) Terminal operators can identify the potential shippers in the hinterland, thus, prepare marketing strategies and negotiate contracts with customers.

CHAPTER 6

CONCLUSION AND RECOMMENDATIONS

This chapter presents the dissertation conclusion. Section 6.1 presents the conducted methodology and the case study. Findings and contributions are discussed in Section 6.2. In Section 6.3, the limitation of the dissertation. Suggestions for future research are presented in Section 6.4.

Again, the purpose of this dissertation is to provide tools to support the decisions of policymakers and stakeholders for short-term planning and investment decisions in the port sector. The dissertation has two main objectives: Forecasting short-term container throughput and analyzing port competitiveness. First, forecasting short-term container throughput includes identifying the structural composition of the historical pattern of container demand and modeling short-term demand forecast. Second, port competitiveness is investigated based on the impact of specified factors related to the determinants of ports, hinterland, and forelands.

6.1 Conducted Methodology

To forecast container demand, the Univariate Time Series Stochastic Model is developed based on Box and Jenkin methodology and the historical pattern of container throughput. The advantage of the model is its independence in only the historical container throughput. The Autoregressive Integrated Moving Average (ARIMA) is a model in which forecasts are obtained by regressing historical observations of the variable itself and the current value with the error terms of the past values at different lags. Box and Jenkin methodology provides an estimation framework to forecast demand than includes model identification, model estimation and diagnostic tests, and forecast generating.

The analysis of port competitiveness is applied in Sections 5.2 and 5.3. In Section 5.2, two types of SIM are used to assess the role of inland distance in the container flow from ports to hinterland regions by investigating the role of distance as explanatory power. First, the doubly constrained model is applied to assess the explanatory power of distance in the distribution of container flow between ports and provinces in Saudi Arabia. Second, attraction constrained model and gap analysis are applied to investigate unexploited potentialities of the competing ports. After using the attraction model to estimate container flow, a gap analysis is applied by comparing the gap between actual and estimated container flows.

Since the traditional SIM does not have the ability to include variables other than distance, Poisson spatial interaction model is used in Section 5.3. The statistical version of the attraction constrained model is used to investigate the competitive position of competing ports and to analyze the impact of inland distance, intermodal accessibility, and geographical proximity to the major maritime shipping routes in the container flow. Poisson distribution model is used to fit the attraction constrained model where the iteratively Reweighted Least Square (IWSL) is applied to converge the Poisson Pseudo-Maximum Likelihood (PPML) for the parameter estimates. After estimating the parameters, the impact of port size, inland distance, railway availability, and port location within the liner shipping routes are investigated. Thereafter, discrepancy analysis is applied to assess the port competitiveness at port-province container flows by using: The use of discrepancy analysis allows measuring the gap between the predicted and actual container flow. The analysis outcomes are used then to understand the hinterland type of each province and to assess the port competitiveness at port-province container flows.

6.2 Findings and Contribution

The dissertation has two major paths, forecasting container demand at the port level and assessing the competitive position of the port and its competing ones. The first path is forecasting container demand. The port of Jeddah is used in the empirical analysis to forecast container demand based on the period of January 2003 to August 2020. The empirical analysis resulted in the following findings: First, based on the data generating process and assuming the historical pattern remains the same, this approach provides a reliable forecast of demand variation for two years-period. Applying the model to forecast a third-year results in larger forecast error. Second, the model of SARIMA (1,0,1) (2,1,0) provides the most accurate forecast results. Based on the error measure of MAPE, the forecast margin of error is 6.6%. Third, the methodology provides a systematic approach to building the model and forecasting demand. This systematic approach provides the ability to understand the behavior and pattern of historical container throughput prior to forecast demand; thus, it allows to update and apply the methodology in the future.

The applied forecast is of importance to port decision-makers as it provides an instrument to obtain insight into future demand. Port authorities and terminal operators aim to avoid traffic congestion and improve container handling efficiency. Short-term forecasts help them in the planning process for operation decisions and resource allocation. Therefore, improvements and modifications of operation plans may be applied to improve productivity based on the outcomes of the forecast methodology.

The second path is focusing on port competition. Port competition in Saudi Arabia is used in the empirical analysis to assess port competitiveness and the impact of various factors in container flows in the period of 2006-2018. The results provided by the model reveal interesting findings. For instance, inland accessibility and port location within maritime shipping routes influence container flow, thus, port choice in Saudi Arabia. The study of Guerrero (2019) pointed out that the issues of port selection are not relevant in the case of developing countries where high inland costs might cause stronger effects of distance that leads to limiting the contestable hinterland.

In Saudi Arabia, this might not be the case where significant investments have been implemented in the last two decades to develop the inland connections between major cities, which eventually contributed to expanding contestable hinterlands in some inland provinces. The contestable hinterland might be caused by the wide geographic configuration of Saudi Arabia, which leads to long-haulage as in the case of the U.S., where the friction of distance is relatively low compared to other geographically smaller countries where the friction is distance is large (Levine et al., 2009). Therefore, it would be interesting that further spatial analysis is conducted in the case of Saudi Arabia and includes transport cost to investigate its impact.

In addition, Inland distance and port accessibility contribute more than port size in explaining the distribution for container traffic, thus, delimiting hinterland in the case of inter-port competition at the country-level of Saudi Arabia. It is concluded that the evolution of the distribution of container flow is greatly influenced by the strategic and geographic characteristics of ports. This is in line with Fleming and Hayuth (1994), who identified that the spatial characteristics of "centrality" and "intermediacy" play a significant role in the evolution of port activity. Therefore, the inland accessibility and the port location within the shipping routes are two factors that impact the inland flow distribution of maritime traffic. Since the determinants of foreland and hinterland complement each other and are essential for the competitive position of the port, ports included in major shipping routes and have sufficient inland accessibility are able to sustain a high share of hinterland demand.

The empirical analysis can help various port and inland transport stakeholders make better decisions to meet their respective objectives as follows:

1. For governments, the analysis can help evaluate the impact of port strategies, the influence of potential inland infrastructure, and the understanding of the changes in the economic activity in the market.
2. Even though the port location is fixed, by understanding inland flow, port authorities can get insight into the needs of hinterland regions to improve the provided services in the port, as well as determining port competitiveness in the different inland markets to assess whether to seek cooperation with rival ports or not.
3. Terminal operators can identify the potential shippers in the hinterland, thus, prepare marketing strategies.

6.3 Limitation

Maritime freight is derived from economic demand. The aggregate macroeconomic data put limitations on forecasting monthly demand due to the non-availability of monthly frequency data. In the case study, the scarcity of monthly economic data in Saudi Arabia limits the ability to use economic indicators in the short-term forecast of container demand. To forecast container demand, the historical trend of container throughput is assumed to continue in the same pattern; thus, a univariate time series model is used where historical container throughput is the only variable used to forecast demand. The unavailability of monthly economic indicators limits the univariate time series forecast to only three months.

Regarding the conducted analysis of port competition, the availability of detailed data limits the use of variables in analyzing the port competition. Several factors were not included in the spatial interaction models due to the scarcity of data. Inland distance is a repulsion parameter in the spatial interaction models. The use of distance transport cost as a repulsion factor would allow explaining the evolution of inland distance due to the technological and infrastructural developments of inland transport (Moura et al., 2017).

The cost variable has implications in the port-province container flow and including it in the spatial analysis would provide more reliable results. Therefore, not using inland transport costs imposes a limitation on the analysis. Because of the non-availability of transport cost data, transport cost is not included. Instead, inland distance is used as a proxy of the repulsion factor in the spatial constrained models. The same issue is applied to the use of a variable to explain the port location within maritime shipping routes. Including container freight rates would provide better results as it reflects the shipper preferences in regard to the port selection, thus, impacting container flow. Therefore, the non-availability of freight rates set limitations on using it; rather, a dummy that explains port location is considered.

6.4 Future Research

The results and findings discussed in this dissertation shed light on the potential topics for future research. Forecasting container demand is a complex process that includes various factors. From a methodological perspective, the forecast analysis is conducted by using a univariate time series model. If more monthly economic indicators are available, including them in the alternative model of multivariate time series will provide forecasts for a longer period.

From an empirical application perspective, the availability of monthly data to be included in the forecast modeling will allow us to examine the impact of various scenarios. An interesting variable to be considered is the total income of oil production. In the case study, Jeddah port, located in the country of Saudi Arabia, is used as a case study to forecast container demand. The GDP of the country is highly related to oil production and prices. Since the growth of the country's revenue from oil is uncertain, forecasting container demand of Jeddah port under different scenarios of oil prices and production will provide the dynamic model and reveal interesting results.

It would be desirable to extend the analysis of port competition by investigating the difference between the privately-owned port and the governmental-owned ports. Among the four ports, KAP is the only fully-owned and operated by the private sector, whereas the ports of Jeddah, Dammam, and Jubail are operated by the private sector but owned and governed by the Saudi port authority. The differentiation between governmental and private ports may result in interesting findings in the competition and complementary relationship and their impact on container flow.

Furthermore, considering the impact of implementing new transport policies in the distribution of container traffic at a national level. Changes in the gas policy in Saudi Arabia may be of interest to shippers and has an impact on container distribution. Therefore, combining the quantitative findings of this section with a qualitative interview on the impact of implementing new transport policies in Saudi Arabia (e.g., the new gas price policy) in the port choice factors from the perspective of the importer may allow determining whether the factors that impact port choice have changed or not due to changes in gasoline prices.

APPENDIX A

TOOLS AND SOFTWARES

Statistical analysis (including data analysis, estimation, validation, and calibration of models and forecasts) and graphical representations were conducted and obtained using the statistical software packages and computer programs of StataIC 17, Matlab R2021a, SIModel and Eviews11 university edition. Geographic maps were produced using the geographical information system application of QGIS (3.18.3).

APPENDIX B

POTENTIAL ARIMA MODELS AND ESTIMATION RESULTS

Dependent Variable: D(LOG(CONT))
Method: ARMA Conditional Least Squares (Newton-Raphson / Marquardt steps)
Date: 02/16/21 Time: 14:02
Sample (adjusted): 2004M04 2016M12
Included observations: 153 after adjustments
Convergence achieved after 11 iterations
Coefficient covariance computed using outer product of gradients
MA Backcast: 2003M04 2004M03

Variable	Coefficient	Std. Error	t-Statistic	Prob.
AR(1)	-0.277447	0.079843	-3.474892	0.0007
AR(2)	-0.237487	0.079286	-2.995302	0.0032
SAR(12)	0.967247	0.021128	45.78006	0.0000
MA(12)	-0.920466	0.020247	-45.46264	0.0000

R-squared	0.527838	Mean dependent var	0.003967
Adjusted R-squared	0.518332	S.D. dependent var	0.075883
S.E. of regression	0.052665	Akaike info criterion	-3.023950
Sum squared resid	0.413261	Schwarz criterion	-2.944723
Log likelihood	235.3322	Hannan-Quinn criter.	-2.991767
Durbin-Watson stat	2.049471		

Inverted AR Roots	1.00	.86-.50i	.86+.50i	.50+.86i
	.50-.86i	.00+1.00i	-.00-1.00i	-.14-.47i
	-.14+.47i	-.50+.86i	-.50-.86i	-.86+.50i
	-.86-.50i	-1.00		
Inverted MA Roots	.99	.86-.50i	.86+.50i	.50-.86i
	.50+.86i	.00+.99i	-.00-.99i	-.50+.86i
	-.50-.86i	-.86+.50i	-.86-.50i	-.99

Figure B.1 Estimation of SARIMA(2,1,0)(1,0,1) using DLCONT dataset.

Dependent Variable: LOG(CONT)-LOG(CONT(-12))
Method: ARMA Conditional Least Squares (Newton-Raphson / Marquardt steps)
Date: 02/17/21 Time: 11:17
Sample (adjusted): 2004M02 2016M12
Included observations: 155 after adjustments
Convergence achieved after 15 iterations
Coefficient covariance computed using outer product of gradients
MA Backcast: 2003M01 2004M01

Variable	Coefficient	Std. Error	t-Statistic	Prob.
AR(1)	0.967373	0.014083	68.69051	0.0000
MA(1)	-0.309078	0.076538	-4.038251	0.0001
SMA(12)	-0.920886	0.022046	-41.77034	0.0000

R-squared	0.827718	Mean dependent var	0.066696
Adjusted R-squared	0.825451	S.D. dependent var	0.126852
S.E. of regression	0.052998	Akaike info criterion	-3.017969
Sum squared resid	0.426932	Schwarz criterion	-2.959064
Log likelihood	236.8926	Hannan-Quinn criter.	-2.994043
Durbin-Watson stat	1.958734		

Inverted AR Roots	.97			
Inverted MA Roots	.99	.86+.50i	.86-.50i	.50+.86i
	.50-.86i	.31	.00+.99i	-.00-.99i
	-.50+.86i	-.50-.86i	-.86+.50i	-.86-.50i
	-.99			

Figure B.2 Estimation of SARIMA(1,0,1)(0,1,1) using D12LCONT dataset.

Dependent Variable: D(LOG(CONT),1,12)
Method: ARMA Conditional Least Squares (Newton-Raphson / Marquardt steps)
Date: 02/16/21 Time: 12:52
Sample (adjusted): 2004M04 2016M12
Included observations: 153 after adjustments
Convergence achieved after 7 iterations
Coefficient covariance computed using outer product of gradients
MA Backcast: 2003M04 2004M03

Variable	Coefficient	Std. Error	t-Statistic	Prob.
AR(1)	-0.284269	0.079353	-3.582319	0.0005
AR(2)	-0.231037	0.079449	-2.907978	0.0042
MA(12)	-0.920476	0.021203	-43.41229	0.0000

R-squared	0.523612	Mean dependent var	-0.002240
Adjusted R-squared	0.517261	S.D. dependent var	0.076180
S.E. of regression	0.052929	Akaike info criterion	-3.020313
Sum squared resid	0.420224	Schwarz criterion	-2.960893
Log likelihood	234.0540	Hannan-Quinn criter.	-2.996176
Durbin-Watson stat	2.045203		

Inverted AR Roots	-.14-.46i	-.14+.46i		
Inverted MA Roots	.99	.86-.50i	.86+.50i	.50-.86i
	.50+.86i	.00-.99i	-.00+.99i	-.50+.86i
	-.50-.86i	-.86-.50i	-.86+.50i	-.99

Figure B.3 Estimation of SARIMA(2,1,0)(0,1,1) using DD12LCONT dataset.

Dependent Variable: D(LOG(CONT),1,12)
Method: ARMA Conditional Least Squares (Newton-Raphson / Marquardt steps)
Date: 02/16/21 Time: 12:52
Sample (adjusted): 2004M03 2016M12
Included observations: 154 after adjustments
Convergence achieved after 9 iterations
Coefficient covariance computed using outer product of gradients
MA Backcast: 2003M02 2004M02

Variable	Coefficient	Std. Error	t-Statistic	Prob.
AR(1)	0.268175	0.208064	1.288904	0.1994
MA(1)	-0.564689	0.174935	-3.228001	0.0015
SMA(12)	-0.918157	0.020429	-44.94404	0.0000

R-squared	0.519416	Mean dependent var	-0.002478
Adjusted R-squared	0.513051	S.D. dependent var	0.075988
S.E. of regression	0.053026	Akaike info criterion	-3.016792
Sum squared resid	0.424570	Schwarz criterion	-2.957631
Log likelihood	235.2930	Hannan-Quinn criter.	-2.992761
Durbin-Watson stat	2.007727		

Inverted AR Roots	.27			
Inverted MA Roots	.99	.86-.50i	.86+.50i	.56
	.50-.86i	.50+.86i	.00-.99i	-.00+.99i
	-.50+.86i	-.50-.86i	-.86+.50i	-.86-.50i
	-.99			

Figure B.4 Estimation of SARIMA(1,1,1)(0,1,1) using DD12LCONT dataset.

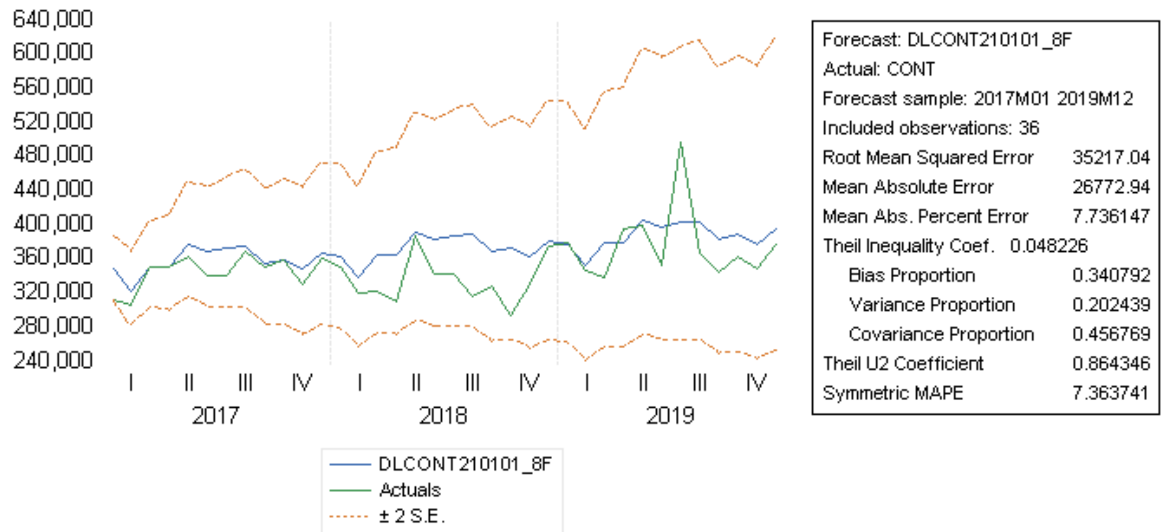


Figure B.5 Forecasts results and accuracy measures of SARIMA(2,1,0)(1,0,1).

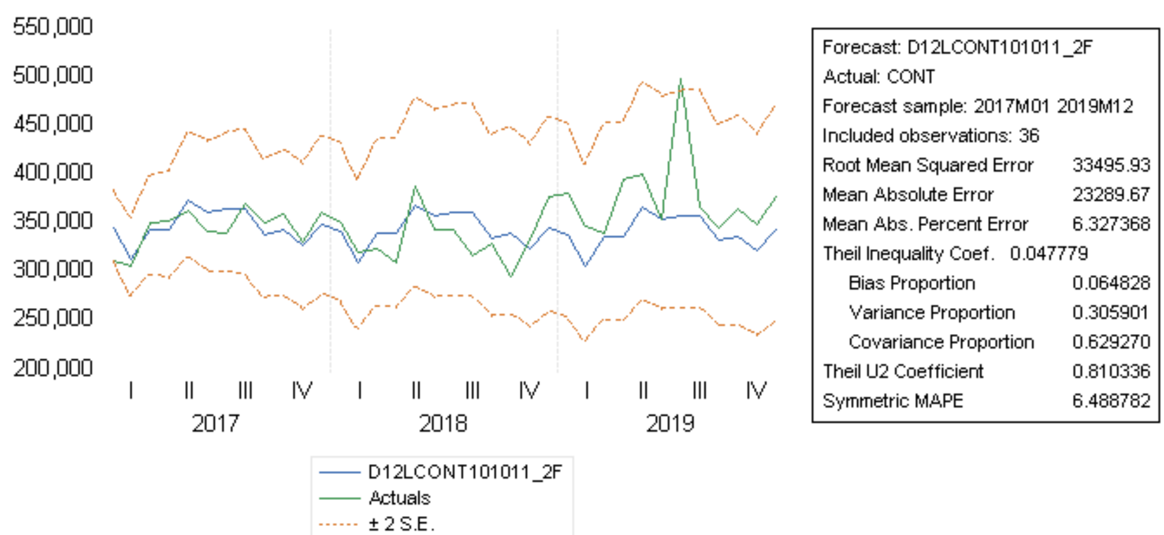


Figure B.6 Forecasts results and accuracy measures of SARIMA(1,0,1)(0,1,1).

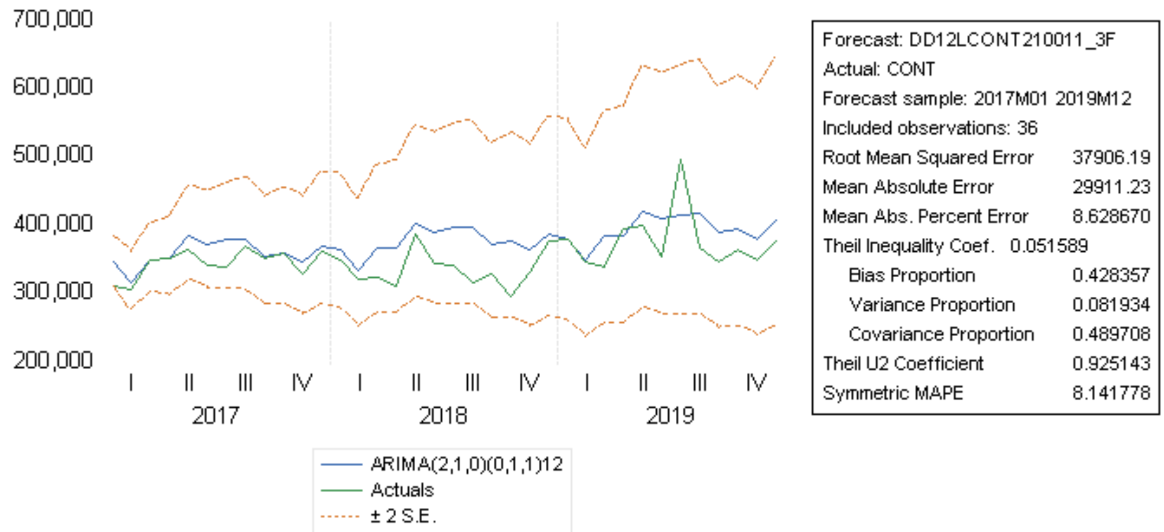


Figure B.7 Forecasts results and accuracy measures of SARIMA(2,1,0)(0,1,1).

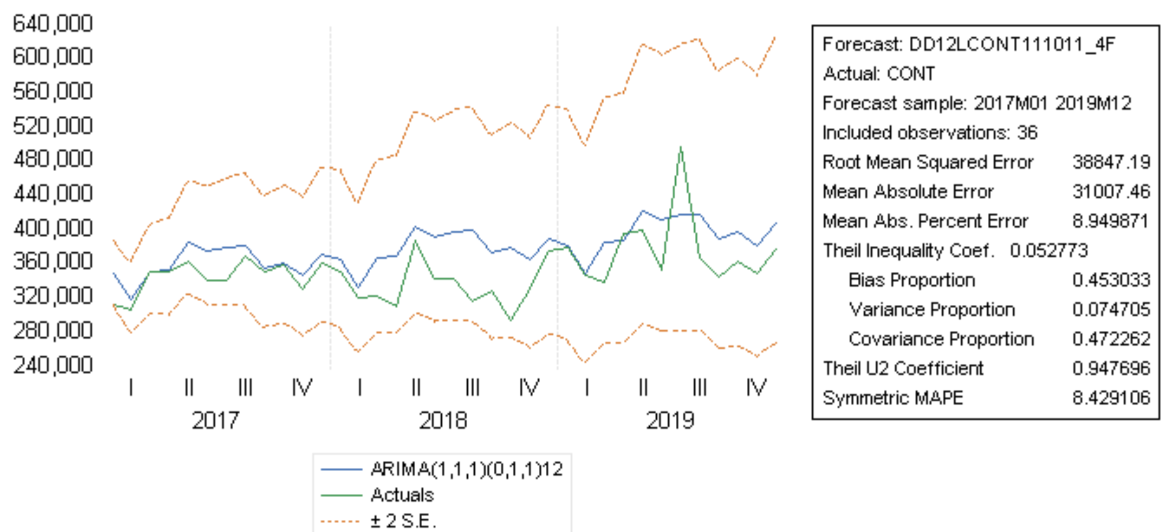


Figure B.8 Forecasts results and accuracy measures of SARIMA(1,1,1)(0,1,1).

APPENDIX C

SIM ESTIMATION RESULTS

```
PPML regression                                     No. of obs      =        572
                                                    Residual df     =         51
Statistics robust to heteroskedasticity             Wald chi2(16)   =    2695.84
Deviance                 =  12447968.25             Prob > chi2     =     0.0000
Log pseudolikelihood = -6226282.695                 Pseudo R2       =     0.9162

Number of clusters (ID)     =         52
                                   (Std. Err. adjusted for 52 clusters in ID)
```

Flow_ij	Coef.	Robust Std. Err.	z	P>|z|	[95% Conf.	Interval]
lnIEcontainerthroughput	.4997573	.1353394	3.69	0.000	.2344969	.7650177
lndistance	-.7609106	.1622089	-4.69	0.000	-1.078834	-.442987
d_Rail_ij	1.89307	.295539	6.41	0.000	1.313824	2.472316
d_Redsea_i	1.00132	.3081398	3.25	0.001	.3973768	1.605263
Province_j						
bah	-2.830642	.3614558	-7.83	0.000	-3.539082	-2.122201
epr	-.3994627	.8107122	-0.49	0.622	-1.988429	1.189504
hai	-.7761563	.3418613	-2.27	0.023	-1.446192	-.1061205
jaz	-.4236718	.4141597	-1.02	0.306	-1.23541	.3880662
jof	-1.923454	.3541278	-5.43	0.000	-2.617532	-1.229376
mad	.1560272	.3885452	0.40	0.688	-.6055074	.9175619
mkk	-1.158517	.7722253	-1.50	0.134	-2.672051	.3550171
naj	-1.739655	.453922	-3.83	0.000	-2.629325	-.8499837
nbr	-2.212659	.8306202	-2.66	0.008	-3.840645	-.5846734
qas	.0539987	.4791032	0.11	0.910	-.8850263	.9930237
riy	.6367952	.3118353	2.04	0.041	.0256093	1.247981
tab	-1.001686	.373846	-2.68	0.007	-1.734411	-.2689613
_cons	8.468888	1.720278	4.92	0.000	5.097205	11.84057

Figure C.1 SIM Model outcomes by fitting import/export container throughput.

```
PPML regression                                     No. of obs      =        572
                                                    Residual df     =         51
Statistics robust to heteroskedasticity             Wald chi2(16)   =    1839.91
Deviance              =  13390147.93                Prob > chi2     =     0.0000
Log pseudolikelihood = -6697372.536                 Pseudo R2       =     0.9098

Number of clusters (ID)     =         52
                                       (Std. Err. adjusted for 52 clusters in ID)
```

Flow_ij	Coef.	Robust Std. Err.	z	P>|z|	[95% Conf.	Interval]
lnLSCI	1.12558	.2555267	4.40	0.000	.6247566	1.626403
lndistance	-.8111572	.1507313	-5.38	0.000	-1.106585	-.5157293
d_Rail_ij	2.008617	.2573566	7.80	0.000	1.504207	2.513026
d_Redsea_i	.6624017	.3169438	2.09	0.037	.0412033	1.2836
Province_j						
bah	-2.85509	.4807019	-5.94	0.000	-3.797249	-1.912932
epr	-.5061696	.7866039	-0.64	0.520	-2.047885	1.035546
hai	-.7841595	.4643956	-1.69	0.091	-1.694358	.126039
jaz	-.4189787	.5332791	-0.79	0.432	-1.464186	.6262291
jof	-1.910937	.4371421	-4.37	0.000	-2.76772	-1.054154
mad	.0960434	.5288782	0.18	0.856	-.9405388	1.132626
mkk	-1.281193	.7658103	-1.67	0.094	-2.782154	.2197674
naj	-1.724966	.5638584	-3.06	0.002	-2.830108	-.6198236
nbr	-2.195803	.8215588	-2.67	0.008	-3.806029	-.5855771
qas	.0479488	.5279589	0.09	0.928	-.9868317	1.082729
riy	.6060486	.3900629	1.55	0.120	-.1584606	1.370558
tab	-1.00127	.4715439	-2.12	0.034	-1.925479	-.0770607
_cons	12.05145	.8985878	13.41	0.000	10.29025	13.81265

Figure C.2 SIM Model outcomes by fitting LSCI.

APPENDIX D

PORT LINER SHIPPING CONNECTIVITY INDEX (PLSCI) OF THE FOUR PORT IN SAUDI ARABIA

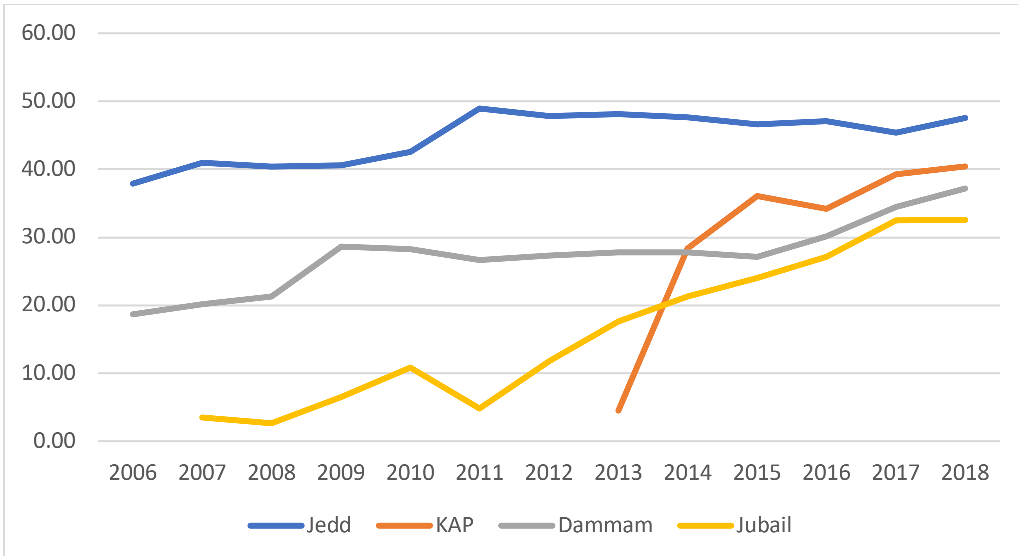


Figure D.1 Port liner shipping connectivity index (PLSCI) of the major four ports in Saudi Arabia (2006-2018).

REFERENCES

Acciaro, M., Bardi, A., Cusano, M. I., Ferrari, C., & Tei, A. (2017). Contested port hinterlands: An empirical survey on Adriatic seaports. *Case Studies on Transport Policy*. https://doi.org/10.1016/j.cstp.2017.03.006

Adhikari, R., & R., A. (2013). An Introductory Study on Time Series Modeling and Forecasting Ratnadip Adhikari R. K. Agrawal. *ArXiv:1302.6613*.

Alderton, P. M. (2008). *Port Management and Operations*. Taylor & Francis.

Arvis, J.-F., Vesin, V., Carruthers, R., Ducruet, C., & de Langen, P. (2018). Maritime Networks, Port Efficiency, and Hinterland Connectivity in the Mediterranean. In *World Bank Publications*. https://doi.org/10.1596/978-1-4648-1274-3

Benacchio, M., Ferrari, C., Haralambides, H. E., & Musso, E. (2001). On the economic impact of ports: local vs. national costs and benefits. *In Forum of Shipping and Logistics, Special Interest Group on Maritime Transport and Ports International Workshop*, 8–10.

Bichou, K. (2013). An empirical study of the impacts of operating and market conditions on container-port efficiency and benchmarking. *Research in Transportation Economics*, *42*(1), 28–37. https://doi.org/10.1016/j.retrec.2012.11.009

Bichou, K., & Bell, M. G. H. (2007). Internationalisation and consolidation of the container port industry: Assessment of channel structure and relationships. *Maritime Economics and Logistics*, *9*(1), 35–51. https://doi.org/10.1057/palgrave.mel.9100170

Blonigen, B., & Wilson, W. (2006). *International trade, transportation networks*.

Box, G. E. P., Jenkins, G. M., & Reinsel, G. C. (2013). Time series analysis: Forecasting and control: Fourth edition. In *Time Series Analysis: Forecasting and Control: Fourth Edition*. https://doi.org/10.1002/9781118619193

Brooks, M. R., Puckett, S. M., Hensher, D. A., & Sammons, A. (2012). Understanding mode choice decisions: A study of Australian freight shippers. *Maritime Economics and Logistics*, *14*(3), 274–299. https://doi.org/10.1057/mel.2012.8

Bruyne, K. De, Magerman, G., & Hove, J. Van. (2012). Multilateral Gravity – A network approach. *1st Annual CIRANO Workshop on Networks in Trade and Finance*, 1–12.

Candace Deans, P. (2011). The Impact of Mega-Ships. In *ITF OECD* (Issue 4).

Chambers, J. C., Mullick, S. K., & Smith, D. D. (1971). How to choose the right forecasting technique. In *Harvard Business Review* (Vol. 49, Issue 4).

Chen, S. H., & Chen, J. N. (2010). Forecasting container throughputs at ports using genetic programming. *Expert Systems with Applications*, *37*(3), 2054–2058. https://doi.org/10.1016/j.eswa.2009.06.054

Chlomoudis, C., Karalis, A., Pallis, A., & Pallis, A. (2003). Ports organizations and the worlds of production theory. *European Journal of Transport and Infrastructure Research*, *3*(1), 77–94.

Chou, M. T., H. S. Lee, and K. L. (2003). A study of forecasting the volume of trans and the harbor operation for port of Kaohsiung. *Journal of Maritime Sciences*, *12*(2), 235-250.

Chou, C. C., Chu, C. W., & Liang, G. S. (2008). A modified regression model for forecasting the volumes of Taiwan's import containers. *Mathematical and Computer Modelling*, *47*(9–10), 797–807. https://doi.org/10.1016/j.mcm.2007.05.005

Chou, C., Chu, C. W., & Liang, G. S. (2003). Competitiveness Analysis of Major Ports in Eastern Asia. *Journal of the Eastern Asia Society for Trasnportation Studies*, *5*(January 2003), 682–697.

Clarke, G. P., Clarke, M., & Wilson, A. G. (1986). Reexamining old problems with new methods: Portbury revisited. *Environment & Planning C: Government & Policy*, *4*(3). https://doi.org/10.1068/c040353

Coto-Millán, P., Baños-Pino, J., & Castro, J. V. (2005). Determinants of the demand for maritime imports and exports. *Transportation Research Part E: Logistics and Transportation Review*, *41*(4), 357–372. https://doi.org/10.1016/j.tre.2004.05.002

Cullinane, K., & Wang, Y. (2009). A capacity-based measure of container port accessibility. *International Journal of Logistics Research and Applications*, *12*(2). https://doi.org/10.1080/13675560902749340

de Langen, P. . (2003). Forecasting container throughput- a method and implications for port planning.pdf. *Journal of Internation Logistics and Trade*, *1*(1), 29–39.

Debrie, J., & Gouvernal, E. (2006). Intermodal rail in Western Europe: Actors and services in a new regulatory environment. *Growth and Change*, *37*(3). https://doi.org/10.1111/j.1468-2257.2006.00331.x

Dekker, S. (2005). *Port investment: towards an integrated planning of port capacity* [Delft University]. http://www.narcis.nl/publication/RecordID/oai:tudelft.nl:uuid:7224341e-cd4d-4d4e-a5ba-f262ee089305

Dicken, P. (2015). *Global Shift: Mapping the Changing Contours of the World Economy* (7th ed.). The Guilford Press.

Ee, J. Y. C., Kader, A. S. A., Ahmad, Z., & Beng, L. K. (2014). Univariate throughput forecasting models on container terminal equipment planning. *Jurnal Teknologi (Sciences and Engineering)*, *69*(7), 163–171. https://doi.org/10.11113/jt.v69.3283

Elbayoumi, O., & Dawood, A. (2016). Analysis of the Competition of Ports in the Middle East Container Ports Using HHI. *Journal of Shipping and Ocean Engineering*, *6*(6), 339–347. https://doi.org/10.17265/2159-5879/2016.06.003

Fagerholt, K. (2000). Evaluating the trade-off between the level of customer service and transportation costs in a ship scheduling problem. *Maritime Policy and Management*, *27*(2). https://doi.org/10.1080/030888300286563

Farhan, J., & Ong, G. P. (2018). Forecasting seasonal container throughput at international ports using SARIMA models. *Maritime Economics and Logistics*, *20*(1), 131–148. https://doi.org/10.1057/mel.2016.13

Ferrari, C., Parola, F., & Gattorna, E. (2011). Measuring the quality of port hinterland accessibility: The Ligurian case. *Transport Policy*, *18*(2), 382–391. https://doi.org/10.1016/j.tranpol.2010.11.002

Ferrari, Claudio, Parola, F., & Morchio, E. (2006). Southern European ports and the spatial distribution of EDCs. *Maritime Economics and Logistics*, *8*(1). https://doi.org/10.1057/palgrave.mel.9100150

Ferrari, Claudio, Parola, F., & Tei, A. (2011). North Africa and International Trade Networks: Port Investments and Market Opportunities. *Current Issues in Shipping, Ports and Logistics*, 203–220.

Figueiredo De Oliveira, G., & Cariou, P. (2015). The impact of competition on container port (in)efficiency. *Transportation Research Part A: Policy and Practice*, *78*, 124–133. https://doi.org/10.1016/j.tra.2015.04.034

Fleming, D. K., & Hayuth, Y. (1994). Spatial characteristics of transportation hubs: centrality and intermediacy. *Journal of Transport Geography*. https://doi.org/10.1016/0966-6923(94)90030-2

Flowerdew, R., & Aitkin, M. (1982). A METHOD OF FITTING THE GRAVITY MODEL BASED ON THE POISSON DISTRIBUTION. *Journal of Regional Science*. https://doi.org/10.1111/j.1467-9787.1982.tb00744.x

Froot, K. A. (1989). Consistent Covariance Matrix Estimation with Cross-Sectional Dependence and Heteroskedasticity in Financial Data. *The Journal of Financial and Quantitative Analysis*, *24*(3). https://doi.org/10.2307/2330815

Fung, M. K. (2002). Forecasting Hong Kong's container throughput: An error-correction model. *Journal of Forecasting*, *21*(1), 69–80. https://doi.org/10.1002/for.818

Garcia-Alonso, L., Martinez-Pardo, A., & Vallejo-Pinto, J. A. (2016). Analysis of the spatial development of the hinterland of ports: A case study. *International Journal of Shipping and Transport Logistics*, *8*(2), 111–128. https://doi.org/10.1504/IJSTL.2016.075007

Garcia-Alonso, L., & Sanchez-Soriano, J. (2009). Port selection from a hinterland perspective. *Maritime Economics and Logistics*, *11*(3), 260–269. https://doi.org/10.1057/mel.2009.9

Gaur, P. (2005). *Port Planning as a Strategic Tool : A Typology* (Issue April) [University of Antwerp]. http://lnweb90.worldbank.org/exteu/SharePapers.nsf/($all)/F1A12C37A9384D528525722B005B767A/$File/PORT+PLANNING+AS+A+STRATEGIC+TOOL.PDF

Gautam, A., & Singh, V. (2020). Parametric versus non-parametric time series forecasting methods: A review. *Journal of Engineering Science and Technology Review*, *13*(3), 165–171. https://doi.org/10.25103/JESTR.133.18

Ge, J., Zhu, M., Sha, M., Notteboom, T., & Shi, W. (2019). Towards 25 , 000 TEU vessels ? A comparative economic analysis of ultra - large containership sizes under different market and operational conditions. *Maritime Economics & Logistics*. https://doi.org/10.1057/s41278-019-00136-4

Gerrits, W. A. (2007). *Dynamic Port Planning under Competition*. Delft University of Technology.

Gosasang, V., Chandraprakaikul, W., & Kiattisin, S. (2011). A comparison of traditional and neural networks forecasting techniques for container throughput at bangkok port. *Asian Journal of Shipping and Logistics*, *27*(3). https://doi.org/10.1016/S2092-5212(11)80022-2

Grant, R. (2016). Contemporary Strategy Analysis. In *John Wiley & Sons Ltd*.

Grossmann, H., Otto, A., Stiller, S., & Wedemeier, J. (2007). Growth Potential for Maritime Trade and Ports in Europe. *Intereconomics*, *42*(4), 226–232. https://doi.org/10.1007/s10272-007-0223-x

Guerrero, D. (2014). Deep-sea hinterlands: Some empirical evidence of the spatial impact of containerization. *Journal of Transport Geography*, *35*, 84–94. https://doi.org/10.1016/j.jtrangeo.2014.01.010

Guerrero, D. (2018). Impacts of transport connections on port hinterlands. *Regional Studies*, *53*(4), 540–549. https://doi.org/10.1080/00343404.2018.1474192

Guerrero, D., Laxe, F. G., Seoane, M. F., Montes, C. P., & A, J. E. L. C. (2015). *a Geographical Analysis of the Relationship Between Inland Accessibility and Maritime Transport Supply*. *41*, 33–46.

Haezendonck, E., Verbeke, A., & Coeck, C. (2006). Strategic Positioning Analysis for Seaports. *Research in Transportation Economics*, *16*(December 2006), 141–169. https://doi.org/10.1016/S0739-8859(06)16007-2

Halim, R. A., Kwakkel, J. H., Tavasszy, L. A., Shi, X., Li, H., Garcia-Alonso, L., Monios, J., Vallejo-Pinto, J. Á., Anggraini, A. R., & Oliver, J. (2016). A strategic model of port-hinterland freight distribution networks. *Transportation Research Part E: Logistics and Transportation Review*, *21*(9), 1689–1699. https://doi.org/10.1016/j.rtbm.2016.05.004

Haralambides, H. (2017). Globalization, public sector reform, and the role of ports in international supply chains. In *Maritime Economics and Logistics* (Vol. 19, Issue 1). Palgrave Macmillan UK. https://doi.org/10.1057/s41278-017-0068-6

Hausman, W. H., Lee, H., & Subramanian, U. (2005). Global Logistics Indicators, Supply Chain Metrics, and Bilateral Trade Patterns. *The World Bank*. https://doi.org/10.2139/ssrn.869999

Hayuth, Y. (1981). Containerization and the Load Center Concept. *Economic Geography*, *57*(2), 160. https://doi.org/10.2307/144140

Heaver, T. D. (1995). The implications of increased competition among ports for port policy and management. *Maritime Policy and Management*. https://doi.org/10.1080/03088839500000045

Hoechle, D. (2007). Robust standard errors for panel regressions with cross-sectional dependence. *Stata Journal*, *7*(3), 281–312. https://doi.org/10.1177/1536867x0700700301

Host, A., Pavlić Skender, H., & Zaninović, P. A. (2019). Trade Logistics – the Gravity Model Approach. *Proceedings of Rijeka Faculty of Economics, 37*(1), 327–342. https://doi.org/10.18045/zbefri.2019.1.327

Iwanow, T., & Kirkpatrick, C. (2009). Trade Facilitation and Manufactured Exports: Is Africa Different? *World Development*. https://doi.org/10.1016/j.worlddev.2008.09.014

Janssens, S., Meersman, H., & de Voorde, E. (2003). Port throughput and international trade: have port authorities any degrees of freedom left? In R. Loyen, E. Buyst, & G. Devos (Eds.), *Struggling for Leadership: Antwerp-Rotterdam Port Competition between 1870 --2000* (pp. 91–113). Physica-Verlag HD.

Kerkman, K., Martens, K., & Meurs, H. (2017). A multilevel spatial interaction model of transit flows incorporating spatial and network autocorrelation. *Journal of Transport Geography*. https://doi.org/10.1016/j.jtrangeo.2017.02.016

Kim, C.-B. (2007). Forecasting the Trading Volumes of Marine Transport and Ports Logistics Policy-Using Multiplicative Seasonal ARIMA Model. In *Journal of Korea Port Economic Association*.

Kim, J.-H. (2008). The Forecast of the Cargo Transportation and Traffic Volume on Container in Gwangyang Port, using Time Series Models. *Journal of Korean Navigation and Port Research*. https://doi.org/10.5394/kinpr.2008.32.6.425

Knudsen, D. C., & Fotheringham, A. S. (1986). Matrix Comparison, Goodness-of-Fit, and Spatial Interaction Modeling. *International Regional Science Review*, *10*(2). https://doi.org/10.1177/016001768601000203

Kramberger, T., Rupnik, B., Strubelj, G., & Monios, J. (2018). Using dry ports for port co-opetition: the case of Adriatic ports. *International Journal of Shipping and Transport Logistics*, *10*(1), 18. https://doi.org/10.1504/ijstl.2018.10008533

Lam, W. H. K., Ng, P. L. P., Seabrooke, W., & Hui, E. C. M. (2004). Forecasts and Reliability Analysis of Port Cargo Throughput in Hong Kong. *Journal of Urban Planning and Development*, *130*(3). https://doi.org/10.1061/(asce)0733-9488(2004)130:3(133)

Lemke, C. (2010). Combinations of time series forecasts: when and why are they beneficial? *PhD Thesis*, *January*. http://eprints.bournemouth.ac.uk/14609/4/licence.txt

Levine, B., Nozick, L., & Jones, D. (2009). Estimating an origin-destination table for US imports of waterborne containerized freight. *Transportation Research Part E: Logistics and Transportation Review*. https://doi.org/10.1016/j.tre.2008.11.001

Liu, M., Kang, S. H., & Ahn, W. C. (2016). Analysis of the Market Structure and Shift-effects in North China Ports. *Asian Journal of Shipping and Logistics*, *32*(3), 179–186. https://doi.org/10.1016/j.ajsl.2016.09.007

Luo, M., & Grigalunas, T. A. (2003). A spatial-economic multimodal transportation simulation model for US coastal container ports. *Maritime Economics and Logistics*, *5*(2), 158–178. https://doi.org/10.1057/palgrave.mel.9100067

Malchow, M. B., & Kanafani, A. (2004). A disaggregate analysis of port selection. *Transportation Research Part E: Logistics and Transportation Review*, *40*(4), 317–337. https://doi.org/10.1016/j.tre.2003.05.001

Meersman, H., & Van de Voorde, E. (2013). Port management, operation and competition: a focus on North Europe. In *The handbook of maritime economics and business.* (pp. 921–936). Informa Law from Routledge.

Meersman, Hilde, Voorde, E. Van De, & Vanelslander, T. (2010). Port Competition Revisited. *Riview of Business and Economics Literature*, *55*(2), 210–233.

Merk, O., Busquet, B., & Aronietis, R. (2015). The Impact of Mega-Ships. In *Int Transp Forum OECD*.

Merkel, A. (2017). Spatial competition and complementarity in European port regions. *Journal of Transport Geography*. https://doi.org/10.1016/j.jtrangeo.2017.04.008

Moshe, B.-A. (2008). Discrete Choice Analysis of Shippers' Preferences. In B. Denis, M. Ben-Akiva, H. Meersman, & E. van de Voorde (Eds.), *Recent Developments in Transport Modelling* (pp. 135–155). Emerald Group Publishing Limited. https://doi.org/10.1108/9781786359537-008

Moura, T. G. Z., Garcia-Alonso, L., & del Rosal, I. (2018). Influence of the geographical pattern of foreign trade on the inland distribution of maritime traffic. *Journal of Transport Geography*, *72*(December 2017), 191–200. https://doi.org/10.1016/j.jtrangeo.2018.09.008

Moura, T. G. Z., Garcia-Alonso, L., & Salas-Olmedo, M. H. (2017). Delimiting the scope of the hinterland of ports: Proposal and case study. *Journal of Transport Geography*, *65*(September), 35–43. https://doi.org/10.1016/j.jtrangeo.2017.09.012

Nelder, J. A., & Wedderburn, R. W. (1972). Generalized linear models. *Journal of the Royal Statistical Society: Series A (General)*, *135*(3), 370–384.

Newey, W. K., & West, K. D. (1987). A Simple, Positive Semi-Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Author(s): Whitney K. Newey and Kenneth D. West Source: *Econometrica*, *55*(3).

Ng, A. K. Y., Ducruet, C., Jacobs, W., Monios, J., Notteboom, T., Rodrigue, J. P., Slack, B., Tam, K. chai, & Wilmsmeier, G. (2014). Port geography at the crossroads with human geography: Between flows and spaces. In *Journal of Transport Geography*. https://doi.org/10.1016/j.jtrangeo.2014.08.012

Notteboom, T. (2009). *The relationship between seaports and the intermodal hinterland in light of global supply chains* (pp. 25–75). https://doi.org/10.1787/9789282102251-3-en

Notteboom, T. E. (1997). Concentration and load centre development in the European container port system. *Journal of Transport Geography*, *5*(2), 99–115. https://doi.org/10.1016/s0966-6923(96)00072-5

Notteboom, T. E. (2004). Container Shipping And Ports: An Overview. *Review of Network Economics*, *3*(2). https://doi.org/10.2202/1446-9022.1045

Notteboom, T. E., & Rodrigue, J. P. (2005). Port regionalization: Towards a new phase in port development. *Maritime Policy and Management*, *32*(3). https://doi.org/10.1080/03088830500139885

Notteboom, T. E., & Winkelmans, W. (2001). Structural changes in logistics: How will port authorities face the challenge? *Maritime Policy and Management*. https://doi.org/10.1080/03088830119197

Pak, J. K. J., & Yeo, Y. S. P. G. (2011). A Study on forecasting container volume of port using SD and ARIMA. *Journal of Navigation and Port Research International Edition*, *35*(4), 343–350.

Parola, F., Risitano, M., Ferretti, M., & Panetti, E. (2017a). The drivers of port competitiveness: a critical review. *Transport Reviews*, *37*(1), 116–138. https://doi.org/10.1080/01441647.2016.1231232

Parola, F., Risitano, M., Ferretti, M., & Panetti, E. (2017b). The drivers of port competitiveness: a critical review. *Transport Reviews*. https://doi.org/10.1080/01441647.2016.1231232

Patil, G. R., & Sahu, P. K. (2016). Estimation of freight demand at Mumbai Port using regression and time series models. *KSCE Journal of Civil Engineering*, *20*(5), 2022–2032. https://doi.org/10.1007/s12205-015-0386-0

Peng, W. Y., & Chu, C. W. (2009). A comparison of univariate methods for forecasting container throughput volumes. *Mathematical and Computer Modelling*, *50*(7–8), 1045–1057. https://doi.org/10.1016/j.mcm.2009.05.027

Pham, T. Y., Jeon, J. W., Dang, V. L., Cha, Y. D., & Yeo, G. T. (2016). A Longitudinal Analysis of Concentration Developments for Container Terminals in Northern Vietnam. *Asian Journal of Shipping and Logistics*, *32*(3), 157–164. https://doi.org/10.1016/j.ajsl.2016.09.004

Rashed, Y. (2016). *Container Throughput Modelling and Forecasting: An Empirical Dynamic Econometric Time Series Approach*. University of Antwerp.

Rashed, Y., Meersman, H., Sys, C., Van de Voorde, E., & Vanelslander, T. (2015). The role of leading economic indicators in forecasting container demand: A dynamic econometric approach. In *World Conference on Transport Research Society (WCTRS)*.

Robinson, R. (2002). Ports as elements in value-driven chain systems: The new paradigm. *Maritime Policy and Management*, *29*(3), 241–255. https://doi.org/10.1080/03088830210132623

Rodrigue, J.-P., & Notteboom, T. (2017). Re-assessing port-hinterland relationships in the context of global commodity chains. *Ports, Cities, and Global Supply Chains*, 67–82.

Rodrigue, J. (2012). supply chain management, logistics changes and the concept of friction. In *Cities, Regions and Flows* (pp. 2842–2843).

Rodrigue, J. P., Comtois, C., & Slack, B. (2017). The geography of transport systems. In *The Geography of Transport Systems* (4th ed.). Routledge.

https://doi.org/10.4324/9781315618159

Rodrigue, J. P., Debrie, J., Fremont, A., & Gouvernal, E. (2010). Functions and actors of inland ports: European and North American dynamics. *Journal of Transport Geography*, *18*(4). https://doi.org/10.1016/j.jtrangeo.2010.03.008

Rodrigue, J. P., & Notteboom, T. (2010). Comparative North American and European gateway logistics: The regionalism of freight distribution. *Journal of Transport Geography*. https://doi.org/10.1016/j.jtrangeo.2010.03.006

Rogers, W. (1992). Regression standard errors in clustered samples. *Stata Technical Bulletin Econometrica Stata Technical Bulletin*, *9*(9).

Sanders, F., Verhaeghe, R., & Dekker, S. (2007). Investment dynamics for a congested transport network with competition: application to port planning. *The Proceedings of the 23th International Conference of The System Dynamics Society*, *March*, 17–21. http://www.systemdynamics.org/conferences/2005/proceed/papers/VERHA347.pdf

Santos Silva, J. M. C., & Tenreyro, S. (2006). The log of gravity. *Applied Economics*, *45*(3), 311–327. https://doi.org/10.1080/00036846.2011.599786

Schmukler, S. L., Hall, B. H., Pindyck, R. S., & Daniel, L. (1997). *Econometric Models and Economic Forecasts*. *August*.

Seabrooke, W., Hui, E. C. M., Lam, W. H. K., & Wong, G. K. C. (2003). Forecasting cargo growth and regional role of the port of Hong Kong. *Cities*, *20*(1), 51–64. https://doi.org/10.1016/S0264-2751(02)00097-5

Shi, X., & Li, H. (2016). Developing the port hinterland: Different perspectives and their application to Shenzhen Port, China. *Research in Transportation Business and Management*, *19*, 42–50. https://doi.org/10.1016/j.rtbm.2016.05.004

Shin, Chang-Hoon, and S.-H. J. (2011). A study on application of ARIMA and neural networks for time series forecasting of port traffic. *Journal of Navigation and Port Research*, *35*(1), 83–91.

Stopford, M. (2009). Maritime Economics. In *Maritime Economics* (3rd ed.). Routledge. https://doi.org/10.4324/9780203891742

Syafi'i, Kuroda, K., & Takebayashi, M. (2005). *Forecasting The Demand of Container Throughput in Indonesia*. *47*.

Talley, W. K. (2014). Maritime transport chains: Carrier, port and shipper choice effects. *International Journal of Production Economics*. https://doi.org/10.1016/j.ijpe.2013.10.009

Talley, W. K., & Ng, M. W. (2018). Hinterland transport chains: A behavioral examination approach. *Transportation Research Part E: Logistics and Transportation Review*, *113*(February), 94–98. https://doi.org/10.1016/j.tre.2018.03.001

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

https://doi.org/10.1016/j.jtrangeo.2010.03.010

Tiller, K. C., & Thill, J. C. (2017). Spatial patterns of landside trade impedance in containerized South American exports. *Journal of Transport Geography*, *58*, 272–285. https://doi.org/10.1016/j.jtrangeo.2017.01.001

Tiwari, P., Itoh, H., & Doi, M. (2003). Containerized Cargo Shipper's Behavior in China: A Discrete Choice Analysis. *Journal of Transportation and Statistics*, *6*(January), 71–86.

Tsamboulas, D., & Kapros, S. (2000). Decision-making process in intermodal transportation. *Transportation Research Record 1707*, *1707*(1), 86–93.

UNCTAD. (2011). Review of Maritime Transport. In *United Nations Publications*. https://doi.org/10.4324/9781315070629-5

UNCTAD. (2013). Review of Maritime Transport, 2013. In *United Nations Publications*.

UNCTAD. (2017). Review of Maritime Transporte, 2017. In *Review of Maritime Transport*.

UNCTAD. (2018). *Review of Maritime Transport*.

Vanelslander, T. (2008). Expansion in cargo handling: Geographical and functional issues. *Maritime Policy & Management*, *35*(2), 193–214. https://doi.org/10.1080/03088830801956847

Veldman, S., Garcia-Alonso, L., & Vallejo-Pinto, J. Á. (2011). Determinants of container port choice in Spain. *Maritime Policy and Management*, *38*(5), 509–522. https://doi.org/10.1080/03088839.2011.597450

Wang, G. W. Y., Zeng, Q., Li, K., & Yang, J. (2016). Port connectivity in a logistic network: The case of Bohai Bay, China. *Transportation Research Part E: Logistics and Transportation Review*, *95*, 341–354. https://doi.org/10.1016/j.tre.2016.04.009

Wang, J. J. (1998). A container load center with a developing hinterland: a case study of Hong Kong. *Journal of Transport Geography*, *6*(3), 187–201. https://doi.org/10.1016/S0966-6923(98)00011-8

Wang, L., Goodchild, A., & Wang, Y. (2018). The effect of distance on cargo flows: a case study of Chinese imports and their hinterland destinations. *Maritime Economics and Logistics*. https://doi.org/10.1057/s41278-017-0079-3

Williams, P. A., & Fotheringham, S. (1984). *The Calibration of Spatial Interaction Models by Maximum Likelihood Estimation with Program SIMODEL* (Vol. 7). Geographic Monograph Series.

Wilson, A. G. (1967). A statistical theory of spatial distribution models. *Transportation Research*. https://doi.org/10.1016/0041-1647(67)90035-4

Winkelmans, W., Meersman, H., Van de Voorde, E., Van Hooydonk, E., Verbeke, A., & Huybrechts, M. (2002). Port competitiveness: an economic and legal analysis of the factors determining the competitiveness of seaports. In *Antwerp, Editions De Boeck Ltd*.

Zeileis, A. (2004). Econometric computing with HC and HAC covariance matrix estimators. *Journal of Statistical Software*, *11*. https://doi.org/10.18637/jss.v011.i10

Zhao, Z., Zhang, C., & Huang, L. (2013). Research on combination forecast of port cargo throughput based on time series and causality analysis. *Journal of Industrial Engineering and Management*, *6*(1), 124.