



UNIVERSIDADE CATÓLICA PORTUGUESA

Business Analytics at APDL: The Determinants of Cargo Tonnage

Francisca Bastos de Magalhães Gomes

Católica Porto Business School,
March 2019



UNIVERSIDADE CATÓLICA PORTUGUESA

Business Analytics at APDL: The Determinants of Cargo Tonnage

Final Assignment in mixed
context presented to Universidade Católica
Portuguesa to obtain the
Master's Degree in Management

by

Francisca Bastos de Magalhães Gomes

Supervised by
Professora Doutora Conceição Silva
and
Professor Doutor Pedro Duarte Silva

Católica Porto Business School,
March 2019

Acknowledgments

First, I would like to express my appreciation to my supervisors, Professor Conceição Silva and Professor Pedro Duarte Silva for their support, time and guidance throughout all the process of preparing this thesis.

Secondly, I would like to thank my family and friends for all the support and friendship, specially my parents, my niece and Miguel, for the encouragement and for always taking me further, both personally and professionally.

Resumo

O crescimento da análise de dados é uma das principais forças do processo de criação de valor nas empresas. Esse processo consiste em analisar dados "parados" nas empresas, transformando-os em informação com valor para o processo de tomada de decisão. Os modelos de regressão são um método de análise preditiva, permitindo extrair insights dos dados e fazer previsões.

O objetivo deste estudo, desenvolvido na Administração dos Portos do Douro, Leixões e Viana do Castelo, S.A. (APDL), foi identificar os principais determinantes externos das toneladas de carga movimentadas no porto de Leixões.

Os dados obtidos da APDL, assim como os coletados externamente, foram utilizados para estimar quatro modelos de regressão. O primeiro modelo é mais geral, enquanto os restantes visam identificar os principais determinantes dos movimentos de carga no porto de Leixões para cada tipo de carga: contentorizada, granéis líquidos e "outros".

Os resultados mostram que o ano do movimento, o mês do movimento, o tipo de carga, as toneladas movimentadas em Lisboa, as toneladas movimentadas em Sines, as toneladas movimentadas em Aveiro, o PIB trimestral e a existência ou não de greves são os determinantes mais importantes dos movimentos de carga no porto de Leixões.

Palavras-chave: Portos, Análise de Regressão, Carga.

Abstract

Growing data analytics is one of the driving forces of companies' value-enhancement process. Data analytics consists of analysing data that is "stalled" in companies, transforming them into valuable information for the decision-making process. Regression models are a method of predictive analytics, allowing to extract insights from the data and to produce forecasts.

The aim of this study, developed in the Administration of Douro, Leixões and Viana do Castelo Ports, S.A. (APDL), was to identify the main external determinants of cargo tonnage in the port of Leixões.

The data obtained from APDL, as well as collected externally, were used to estimate four regression models. The first is a general model, while the others were aimed at identifying the main determinants of the cargo movements in the port of Leixões for each type of cargo: containerized, liquid bulk and "others".

The results show that the year of the movement, month of the movement, cargo type, tonnage moved in Lisbon, tonnage moved in Sines, tonnage moved in Aveiro, quarterly GDP and the existence or absence of strikes are the main determinants of cargo movements in port of Leixões.

Keywords: Ports, Regression Analysis, Cargo

Content

List of Figures.....	vii
List of Tables.....	ix
Introduction.....	1
1.1 Scope.....	2
1.2 Objectives.....	6
The Ports Context and Previous Studies.....	7
2.1 Ports: The national context.....	7
2.2 The company: APDL.....	8
2.3 Previous studies.....	10
Methods.....	15
3.1 Pre-processing data.....	15
3.2 Regression.....	17
Data and Sample.....	21
4.1 Non-included variables.....	31
Data analysis.....	33
5.1 Dataset.....	33
5.2 Regression model: Leixões Tonnage.....	35
5.2.1 Assumptions.....	37
5.2.2 Outliers Analysis.....	40
5.3 Regression Models: per type of cargo.....	42
5.3.1 Containerized cargo.....	42
5.3.1.1 Assumptions.....	43
5.3.1.2 Outliers Analysis.....	46
5.3.2 Liquid Bulk cargo.....	48
5.3.2.1 Assumptions.....	49
5.3.2.2 Outliers Analysis.....	51
5.3.3 Others cargo.....	53
5.3.3.1 Assumptions.....	53
5.3.3.2 Outliers Analysis.....	56
5.3.4 Comparing the various models.....	58
Conclusions.....	61
References.....	65
Appendices.....	67
Appendix 1: R script.....	67

List of Figures

Figure 1: Location and Hinterland (adapted from APDL's document)

Figure 2: Matrix representation

Figure 3: Leixões Tonnage

Figure 4: Total tonnage per quarter

Figure 5: Other ports cargo

Figure 6: Quarterly GDP of Portugal

Figure 7: Percentage of each type of cargo moved per year

Figure 8: Containerized cargo tonnage per month

Figure 9: Liquid bulk cargo tonnage per month

Figure 10: RO-RO cargo tonnage per month

Figure 11: Breakbulk cargo tonnage per month

Figure 12: Dry bulk cargo tonnage per month

Figure 13: Boxplots of the quantitative variables

Figure 14: Residual Plots

Figure 15: Normality of the model RegA

Figure 16: Influence measures

Figure 17: Influence measures

Figure 18: Residual Plots

Figure 19: Residuals

Figure 20: Normality of the model Reg1C_step

Figure 21: Influence measures

Figure 22: Influence measures

Figure 23: Residual Plots

Figure 24: Residuals

Figure 25: Normality of the model Reg1L_step

Figure 26: Influence measures

Figure 27: Influence measures

Figure 28: Residual Plots

Figure 29: Residuals

Figure 30: Normality of the model Reg1.1O

Figure 31: Influence measures

Figure 32: Influence measures

List of Tables

Table 1: Variables

Table 2: Descriptive statistics of the quantitative variables: mean per year

Table 3: Competitors by type of cargo

Table 4: Descriptive statistics of the categorical variables

Table 5: Regression model RegAI

Table 6: Regression model Reg1C_step

Table 7: Regression model Reg1L_step

Table 8: Regression model Reg1.1O

Table 9: Regression models

Chapter 1

Introduction

The present study was developed within the scope of the Master's in Management, from the Catholic Porto Business School, in the specialization of Business Analytics, having been carried out in a mixed context in the Administration of Douro, Leixões and Viana do Castelo Ports, S.A. (APDL).

Nowadays, the importance that companies give to data is on the rise. Data analytics is currently accepted as a tool with an enormous potential to increase value in companies. In this sense, we begin to perceive the emergence of analysing the "stalled" data. Many companies are already sensitive to this trend. However, we believe there is still a lot to do in this area, since a lot of companies have a multiplicity of data, but these are stored, not being used or treated to add value to them. It is not enough to have instruments for collecting or accessing information, if it is not possible to transform that information into value which, in turn, can be integrated into the decision-making process.

This thesis is divided into 6 chapters.

The first chapter presents an introduction.

Chapter 2 is a contextualisation of the study. This chapter begins by contextualizing the national ports, followed by a brief presentation of the company where the project was held. Finally, some previous studies that, like this study, considered factors external to the ports as determinants of their functioning are presented.

The third chapter describes the methodology that, in the case of this thesis, is based on regression analysis. This chapter is divided into two sections: the first explains how to prepare the data for regression analysis, and the second explains the regression method.

The fourth chapter comprises the data and sample, i.e., this chapter gives a more detailed description of the dataset we used in the study, namely descriptive statistics of the explanatory variables that were considered significant to answer the research question.

The fifth chapter corresponds to results from regression analysis. This chapter presents two distinct results: the first one is related with the identification of the best model to explain the cargo movements in the port of Leixões, while the second one is a replication of the regression analysis by type of cargo, resulting into 3 different models. In this chapter, the assumptions of multicollinearity, heteroscedasticity, autocorrelations, normality and outlier influence are verified for each model, as well as some conclusions are withdrawn.

The last chapter gives an overview of the main conclusions from the analysis of the data, identifying the limitations faced and providing perspectives for the future.

1.1 Scope

According to several authors (Acito & Khatri, 2014; OECD, 2013; Davenport, 2013), exploring data creates value for business and the emergency of business analytics is a revolution that, nowadays, is impossible to miss.

Currently, large volumes of data are being created uninterruptedly, and according to Acito and Khatri (2014, p. 567), quoting Google CEO Eric Schmitt, “There was 5 exabytes of information created between the Dawn of civilization through 2003, but that much information is now created every 2 days, and the pace is increasing”.

Collected data allows companies to identify business trends, to manage risk, and to enhance competitiveness, thus creating value for the world economy (Acito & Khatri, 2014).

According to Acito and Khatri (2014, pp. 567, 568), the emergence of Business Analytics these days is not by chance, is due to a combination of five major trends, namely:

“First, there is ready availability of large amounts of data (Manyika et al., 2011) (...) Furthermore, there is an increasing realization that data is a valuable resource (Levitin & Redman, 1998) and should be managed as an asset (Laney, 2011) Second, the maturity of business performance management in the last 5 decades has helped create a more solid bridge between business strategy and data. (...) Third, the realization that fact-based decisions are more critical at every level of the organization has resulted in the emergence of self-service analytics and business intelligence (Imhoff & White, 2011). Access to large databases and user-friendly reporting tools has contributed to the analytics revolution that is now taking place. Fourth, advanced analytics techniques have been incorporated into enterprise-level systems, making even the most sophisticated algorithms available to analysts. (...) another major driver of the business analytics phenomenon has been the sharply declining cost per performance level of three key information technologies: computing power, data storage, and bandwidth.”

This idea is corroborated by OECD (2013) that argues that it is due to the combination of various trends, both technological and socioeconomical, that business analytics are fostering the creation of Big Data, which can lead to innovation at the level of industries, processes and products.

As a result, these trends lead to a socioeconomical model based on data, in which data is the dominant asset. Data leads to “knowledge-based capital” (OECD, 2013, p. 2), which can promote innovation and sustainable growth across the economy and society.

Therefore, to have competitive advantage, companies should adopt Business Analytics. In other words, they should implement the exploitation and analysis of data as the basis for decision-making, to have valuable information to support their decisions, that otherwise would not be available. "Big data are worthless in a vacuum" (Gandoni & Haider, 2015, p. 140), i.e., it only originates value when it is inserted in the process that extracts insights and leads to decision making.

Acito and Khatri (2014) present a structural framework for Business Analytics in which it is required an alignment between the strategy of the organization and the desirable behaviours of the same to the business performance management and analytic tasks and capabilities.

According to Davenport (2013, p. 123), it is crucial for companies to "establish a culture of inquiry, not advocacy". That is, one should not try to find evidence to defend a prior idea, but rather to investigate in order to find a posteriori theory. What should be valued are the evidence and the data behind the ideas, not the opinions or the people who had them. Nevertheless, business analytics should be combined with intuition and experience. Davenport (2013) presents 6 steps for analytics-based decision making: recognizing the problem or business question and its possible alternatives; review previously applied solutions to the same or similar problem; model the solution and select the variables; collect the data; analyse the data; present the results and move on to action. The author states that, for business analytics consumers, the first and last steps are the most important. In the first step is where experience and intuition are used the most. The last step is to present and communicate the results to other executives, since "analytics is largely about telling a story with data" (Davenport, 2013, p. 122).

Big Data originates many challenges that, according to Sivarajah et al. (2017), can be divided into three main categories, which are directly linked to data life

cycle: data challenges; process challenges; management challenges. The first category is related to the type of data and its characteristics, such as volume, velocity, variety, variability, veracity, visualisation and value. The challenges of the processing phase, as the name implies, are encountered when processing and analysing data, for instance, data mining issues. Lastly, management challenges include, for example, challenges related to privacy, security and data governance.

The same authors identify three main methods that allow organizations to overcome the challenges previously mentioned. These methods are:

- i. descriptive analytics that ask "What happened in the Business?",
- ii. predictive analytics that ask "What is likely to happen in the future?",
- iii. prescriptive analytics that ask "Now what?".

Descriptive analytics seek to understand what the current situation of the business is, by describing what has already occurred. This type of analytics is the most common and consists of summarizing and describing the data, using tools such as summary statistics, dashboards and scorecards. Nevertheless, currently, the trend is to combine descriptive analytics with predictive analytics. This method analyses the data intensively as to realize, by combining several insights, why a situation has happened or identify patterns. It is also this method that includes the creation of metrics in order to monitor the performance of the business over time. Hence, these analytics contribute to a data-driven decision-making process.

Predictive analytics aims to determine future possibilities by identifying patterns and relations in the data. According to Sivarajah et al. (2017), there are two categories: machine learning techniques and regression techniques.

Prescriptive analytics are "performed to determine the cause-effect relationship between analytic results and business process optimization policies" (Sivarajah et al., 2017, p. 276).

1.2 Objectives

Our purpose is to contribute positively to the performance of the company under study, in this case Administration of Douro, Leixões and Viana do Castelo Ports, S.A. (APDL).

Given that APDL had data that were not being used, it was possible, with proper knowledge, to draw valuable information from the set of data that was available to us. Hopefully we have created the beginning of a data analytics culture in APDL, and consequently, competitive advantage. We trust that the project that was developed was important, not only for APDL, but also for us as researchers.

The main research question to be analysed is: What are the main external determinants of the monthly cargo movements in the port of Leixões?

The focus of this project is to create value from data in APDL. Hence, there are three fundamental goals:

- i. analyse the movements of cargo in the port of Leixões
- ii. evaluate the impact of each explanatory variable on the dependent variable, cargo movements in the port of Leixões.
- iii. identify the main determinants, among our explanatory variables.

Chapter 2

The Ports Context and Previous Studies

2.1 Ports: The national context

Portugal is a country with a large coastal area, which makes it a gateway into Europe across the sea. This country has a long history with maritime tradition, having been pioneer in the great navigations and discoveries.

Portugal has a geostrategic position on the intercontinental sea routes, since it has 900 kilometres of Atlantic front, possessing 10 ports.

The most recent data we found in INE¹ and PORDATA² was for 2017. In 2017, in the Portuguese ports, 93.3 million of tons of cargo were loaded and unloaded from ships and 1.8 million of passengers were transported, embarked and disembarked. Additionally, in the same year, the national ports registered the entry of 14.6 thousand vessels, which represents an increase compared to previous years. The main Portuguese ports, Leixões, Lisboa and Sines, concentrated more than half of the movements of entry of ships in Portugal, being Leixões the one that had the greater percentage (18%). As regards to the destination of goods from Portuguese ports, in the European Union, Spain was the most important destination (10.4%), followed by the Netherlands (9.3%) and the United Kingdom (6.9%). The American continent was repositioned in 2017 as the second most relevant continent in this flow, namely with the USA (12.3%), that was the main destination country this year. As for Africa, the most representative destinations were Morocco (4.8%) and Angola (3.1%). In Asia, China (1.9%) was the main country, followed by the United Arab Emirates (1.4%).

¹ https://www.ine.pt/xportal/xmain?xpgid=ine_main&xpid=INE

² <https://www.pordata.pt/>

In 2017, the movement of liquid bulk cargo (35.4 million tons) accounted for 37.9% of the total cargo movement, followed by containerized cargo (29.6 million tons), with 31.7% of the total, and dry bulk with 21.2 million tons, which corresponded to 22.7% of the total merchandise movement. The port of Sines handled 63.9% of the total liquid bulk, 59.1% of the containerized cargo and 30.0% of the solid bulk cargoes. Leixões was responsible for the movement of 24.9% of the total liquid bulk, 16.8% of the total containerized cargo and 11.1% of the solid bulk cargoes. The port of Lisbon handled 25.3% of the total solid bulk and 13.3% of the containerized cargo, while Setubal and Aveiro also registered a solid increase in total bulk, respectively, with 14.0% and 12.2%.

In the last 10 years, the Portuguese ports have been growing, above many ports of the Iberian Peninsula. (APP³)

Ports are a lever in the Portuguese economy, for instance, in 2017, maritime transport represented more than half of tonnage exported in Portugal (54.6%), which is equivalent to, approximately, 30% of the exports in value.

The Portuguese ports are present in the most varied industries, namely, petrochemical, shipbuilding and repair, agri-food, alternative energy and automobile.

2.2 The company: APDL

The project was held at the Port Authority of Douro, Leixões and Viana do Castelo (APDL) which is a company that provides services to customers and users of the port system in the north of Portugal. APDL aims to achieve a port system of excellence that induces the creation of value (APDL⁴).

³ <http://www.portosdeportugal.pt/index2018.php>

⁴ https://www.apdl.pt/pt_PT/web/apdl/header

APDL, the Administration of Douro, Leixões and Viana do Castelo Ports, S.A., has three core activities: leasing of dealership area to concessionaires, providing logistics services to ships, and acting as the port authority.

The dealers present in the Port of Leixões are: Terminal de Contentores de Leixões, S.A (TCL), Terminal de Carga Geral e de Graneis de Leixões, SA (TCGL), Petrogal, Docapesca, Marina Porto Atlântico, Silos de Leixões (SDL), SECIL and Cepsa. TCL belongs to the Yilport Leixões group and its core activity is the handling of containers. TCGL, a joint-stock company mostly owned by the ETE group, moves dry bulk and fractional general cargo. The oil terminal, Petrogal, is owned by Galp and moves liquid bulk. The Ro-Ro terminal moves cargo entering and leaving the ships by means such as wheels and ramps, therefore no cranes are needed. The companies Secil and Silos de Leixões are also located in the Port of Leixões. The first produces and sells cement, while the second provides logistic services related to the food industry, including transportation and storage of agri-food bulk.

The hinterland of the Port of Leixões corresponds to its area of influence on land. APDL divides its hinterland into local, regional and supra regional, as depicted in Figure 1.



Figure 1: Location and Hinterland (adapted from APDL's document)

On the other hand, the foreland is the set of final destinations reached, via connected ports, which, in the case of the Port of Leixões, corresponds to 184 countries. Note that this information about the foreland is provided by the concessionaires since the evidence that APDL has is not accurate.

2.3 Previous studies

According to UNCTAD (1993, p. 3): “Of the various reasons for collecting data relating to ports, two are of primary concern to port authorities: first, the data can be used for improving port operations, and secondly, they can provide an appropriate basis for planning future port development.”

Most of the previous studies focus on internal factors to assess the performance of ports, and very few on external factors. Nevertheless, our study focuses on external factors that may influence cargo movements in ports.

- **Hinterland and competition**

Understanding the concept of hinterland is crucial when it comes to ports. Ports hinterlands correspond to the geographical area that a port or terminal serve (Carvalho et al., 2015), i.e., the area covered by all the companies that use the port. Therefore, geography defines the port a company uses, and competitive markets are often characterized by the ports that share the same hinterlands.

Meersman et al. (2010, p. 218) claim that the issue that must be addressed is “competition between supply chains”, of which ports are part, and the crucial aspect to consider is the cost of these chains. Langen (2007) corroborates this idea, defending a paradigm shift in which one must consider that ports do not provide an isolated service, but rather a service that is included in a supply chain.

Therefore, according to Meersman et al. (2010), when the hinterlands coincide, the choice between ports will be based on the port that contributes to a lower cost supply chain. Consequently, the question that should be asked by the port user is: “does the port considered offer advantages compared to other ports serving the same hinterland?” (Meersman et al., 2010, p. 219).

Moreover, Langen (2007) studied the case of Austria, which constitutes the hinterland of 6 different European ports. This author demonstrated that the market share of the different ports varied over time, which corroborates the idea of competition between ports within the same geographical area.

Accordingly, cargo movements of a given port depend, not only on the conditions of this port, but also on the conditions that other ports, that have the same hinterland, offer to potential port users.

- **Climate**

The Committee on Climate Change, US Transportation, Transportation Research Board, Division on Earth, & Life Studies (2008) identifies both positive and negative impacts of climate in ports. For instance, a positive impact of an abnormal climate event is the impediment of accumulation of ice on ships by the existence of warmer winter days.

According to the same report, ports infrastructures are constructed taking into account the possibility of adverse weather conditions. Nevertheless, extreme situations may “push environmental conditions outside the range for which the system was designed.” (p. 49). Therefore, it is important to acknowledge that climate can affect the movements of cargo since it can disturb the security, normal operation of ports and maintenance of the infrastructures and systems.

It should be noted that maritime transport systems, i.e., seaports, are located in coastal areas which are very vulnerable, for instance, to storms or sea-level rise. (Becker et al., 2011)

- **Other transportation modes**

International trade consists of the exchange of goods and services between countries, allowing the sharing of new products and experiences around the world. Maritime transport is one of several modes of transport used in international trade.

Thus, maritime transport is inherently linked to international trade. In an increasingly globalized economy, in which countries are highly dependent on each other, either because of international trade or investment, ports assume importance in fostering trade (Blonigen & Wilson, 2008).

Sánchez et al. (2003) claim that improvements in the means of transport at international level promote economic globalization. According to these authors, a reduction in transport costs leads, directly, to a stimulus of imports and exports.

In the publications of INE⁵ and PORDATA⁶, in the main statistical results on the activity of the Transport and Communications sectors in Portugal, it is possible to observe the statistics on international trade by mode of transport. As far as imports and exports are concerned, maritime transport is the mode of transport responsible for the majority of the volume of goods imported in recent years. In 2017, maritime transport accounted for 61.6% of the volume of imported goods, which corresponds to 39.2 million tons, and 54.6% of the volume of exported goods. However, the value of each ton imported and exported is lower when the transport is done by sea than road, so the

⁵ https://www.ine.pt/xportal/xmain?xpgid=ine_main&xpid=INE

⁶ <https://www.pordata.pt/>

percentage of this mode of transport on imports and exports in euros is lower than the road. Thus, in 2017, maritime transport represented 26.1% and 30.8% of imports and exports in value, while road transport represented 61.9% and 60.5%, respectively.

Chapter 3

Methods

This chapter introduces the analytical method used in our empirical analysis, which is regression analysis. This is a predictive analytics method, as pointed out in chapter 1. There are various predictive models. Our choice for regression methods is due to the fact that regression analysis is the most established method for predicting continuous variables, allowing to extract insights from data, specially about the relation between predictive and explanatory variables.

In the case of our study, we have one dependent continuous variable and several explanatory variables, some of which are categorical and others quantitative. Thus, among the most appropriate analyses, we decided to adopt multiple regression.

We divided this chapter in two parts, one dealing with the methods require to pre-process the data and the second part where the regression analysis is presented.

3.1 Pre-processing data

After the problem has been formulated and the data collected, there is some steps to be taken, namely:

- i. organizing the data into an appropriate format to entry them in the software, R in our case;
- ii. data preparation.

These steps are crucial to make the data suitable for analysis and to produce more effective results.

First of all, the appropriate format, according to Afifi et al. (2011) and Faraway (2009), is matrix or rectangular format, where the columns correspond

to the variables and the rows correspond to the observations, as shown by Figure 2.

Y_1	X_{11}	X_{12}	X_{13}
Y_2	X_{21}	X_{22}	X_{23}
	
Y_n	X_{n1}	X_{n2}	X_{n3}

Figure 2: Matrix representation

Secondly, the data preparation phase must be completed before starting the analysis.

In the real world, the data that are collected are not perfect, data are often inconsistent, with outliers and missing values that make it noisy and incomplete.

According to Soibelman and Kim (2002), data preparation is the hardest and most time-consuming phase of the statistic process. Next, there is an explanation of how to handle these types of problems, through data preparation, according to the same authors.

Firstly, data cleaning consists of improving the quality of the data by solving problems, such as missing data, noisy data, outliers and contradictions.

There are several ways to deal with missing data, such as ignore the records or the attributes that include them. However, these solutions may lead to insufficient sample sizes or the elimination of important variables. Furthermore, other ways to deal with this problem is to use the most probable value, or a global constant, such as “unknown”, or the mean or median of that attribute to fill the missing value.

When it comes to atypical observations, outliers, there are more than a few methods to deal with these: binning method, which divides the data into bins and then smooth it by bins mean, medians or others; regression, by fitting the

data into regression functions; classical statistical method, by detecting outliers with, for instance, boxplots, and then removing them; clustering, by detecting and removing outliers; combined computer and human inspection.

Secondly, data integration consists in merging data from different sources in order to give the user a unified view of these data. It is important, when doing this step, to take into account that data value conflicts and redundancy problems can happen.

Thirdly, data transformation involves processes such as generalization, the possible construction of new features from old ones and the normalization of data.

Fourthly, data reduction involves the sampling of data, selecting of features, and reduction of the dimension of the data. In this step of data preparation, it can be selected a subset of objects or variables to be analysed, or new features can be created from the old ones. It is important that the information available in the new, reduced, dataset is a good representation of the original information.

Lastly, data discretization is the process of converting continuous-valued attributes into discrete variables with a small number of values.

Therefore, the data is ready, and the statistical analysis can begin. "When data are properly prepared, the analyst gains understanding and insight into the content, range of applicability, and limits of data" (Soibelman et al., 2002, p. 42).

3.2 Regression

A regression model is used to explain the relationship between a dependent variable, Y , and a set of explanatory variables, X_1, X_2, \dots, X_k .

This relationship can be represented by the equation: $Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \dots + \beta_{ik} X_{ki} + \varepsilon_i$, where $\beta_0, \beta_1, \dots, \beta_k$ are the regression coefficients, $X_1,$

X_2, \dots, X_k are the explanatory variables and ε is a random perturbation/disturbance.

The regression coefficients are constant, and each β_k represents the expected variation in Y when X_k increases one unit and the other variables remain constant. However, in linearized models, i.e., where there are logarithmic transformations, for the dependent and independent variables, then β_k represents the expected percentage change in Y when X_k increases by one percentage point and the other variables remain constant.

The random disturbance is the way of the model to take into account the randomness of choice of the model variables, as a human action. Therefore, ε disturbs and neutralizes the exact measurement of the explained variable.

The literature is unanimous regarding the idea that the initial data analysis, before estimating the models, should consist in making summary and descriptive statistics of the variables. Faraway (2009) states that, in this phase, it is important to be aware and seek for uncommon or surprising values. Additionally, histograms and boxplots can be constructed in the sense of perceiving the distribution of each of the variables individually, or scatterplots and dynamic charts to relate two or more variables.

After estimating the parameters of the model, the statistical significance of each of the explanatory variables must be analysed, in order to understand how these influence the dependent variable. Similarly, the overall measure of the quality of the adjustment, the coefficient of determination (R^2), must also be considered. The R^2 explains which percentage of the dependent variable is explained by the model.

Furthermore, many authors consider a set of key assumptions underlying regression analysis that must be verified for the inferences drawn to be accepted: multicollinearity, homoscedasticity and autocorrelation.

Multicollinearity occurs when the variables of the model are linearly associated. The diagnosis of the presence of multicollinearity can be performed through the analysis of the correlations between the variables and, formally, through the analysis of the values of Generalized Variance Influence Factors (GVIF) (Fox & Monette, 1992), which when high are indicators of multicollinearity.

The heteroscedasticity is a problem to be taken into consideration, since, in existence, it violates one of the assumptions of the least squares method, that: $\text{var}(y_i) = \text{var}(e_i) = \sigma^2$, negatively affecting estimation and inference statistic.

This can be diagnosed graphically, using residual plots, or formally, through the Breusch-Pagan test. This test has, as null hypothesis, the presence of homoscedasticity, i.e., conditional variance is constant, against the opposite alternative, heteroscedasticity, i.e., the disturbances do not have the same variance.

When heteroscedasticity is suggested, the way to continue the regression analysis in its presence is by using robust standard errors, suggested by Halbert White. This does not eliminate heteroscedasticity, but rather enables the usual techniques of statistical inference from the results of the adjustment.

Moreover, comparing the standard errors with the robust standard errors that emerged from the "remedy" applied to heteroscedasticity, the results must be analysed to see if they changed or not.

Autocorrelation happens when disturbances are dependent over time. This assumption can be verified graphically by analysing the graph with the residuals as a function of the estimated values of the dependent variable, and if a pattern is identified, it suggests the presence of autocorrelation. To test this assumption formally, the Durbin-Watson test can be performed, whose null hypothesis states that the residuals are not correlated.

To continue the estimation in the presence of autocorrelation, a heteroscedasticity and autocorrelation consistent covariance matrix, proposed by Andrews (1991), can be used. Hence, with this method, the tests of significance of the parameters are done with robust standard errors that are valid. As in the case of the remedy applied to heteroscedasticity, the results of the estimation should be compared to see if the conclusions change.

Some authors also add the assumption of normality for inference in small samples. This assumption states that the disturbances follow a normal distribution with a null mean. This assumption can be verified graphically, where, if the data are symmetrical and the points follow the straight line, the residuals follow a normal distribution. It should be noted that regression does not require this assumption, since, according to Eicker (1963) and Weisberg (2005), when the samples are large, many tests become valid.

However, when the disturbances are not normally distributed, outliers may be more common, so a proper analysis of outlier influence should be conducted. To understand if there are severe outliers, the Bonferroni Test can be performed, whose null hypothesis refers to their absence. Furthermore, if the test suggests the presence of severe atypical observations, one can identify which outliers are most influential through the analysis of the hat values, graphically. Subsequently, the model is estimated excluding these observations, in order to verify if the results change.

Chapter 4

Data and Sample

As previously stated in Chapter 1, the goal of this study is to explain the tonnage of cargo moved in the port of Leixões from 2009 until 2017. The variables used in our analysis are shown in Table 1. It should be noted that there were more variables available than those used, but in the final models they were not relevant, so we omit them for sake of simplicity.

Table 1: Variables

Variables	Description	Classification
Leixões Tonnage	Tonnage moved in the port of Leixões	Quantitative
Year	Year of the movement: from 2009 to 2017	Quantitative
Month	Number of the month of the movement: from 1 to 12	Categorical
Cargo Type	Classification of the cargo: Containers; Liquid bulk; Others	Categorical
Lisbon Tonnage	Tonnage moved in the port of Lisbon	Quantitative
Sines Tonnage	Tonnage moved in the port of Sines	Quantitative
Aveiro Tonnage	Tonnage moved in the port of Aveiro	Quantitative
Quarterly GDP	Value of the quarterly GDP, in millions of euros	Quantitative
Strikes	Strikes in that month: 0- No; 1- Yes	Categorical

Tables 2 and 4 show the descriptive statistics of the quantitative and categorical variables, respectively.

Table 2: Descriptive statistics of the quantitative variables: mean per year

Variables	Mean								
	2009	2010	2011	2012	2013	2014	2015	2016	2017
Leixões Tonnage	388105.4	404911	453293.6	461521.8	477402.2	502863.5	533989.8	519614.6	551148.9
Lisbon Tonnage	874994.7	333154.8	339197.2	307797.1	333080.4	328928.3	321730.6	283472.2	343863.1
Sines Tonnage	1809065	708758.6	723756.7	793409.8	1014271	1043971	1221293	1421815	1378927
Aveiro Tonnage	217261.6	104240.9	94786.26	92168.53	109892.1	124757.4	129336	126153.1	145008.5
Quarterly GDP	44023.29	44861.19	44062.73	42267.54	41789.85	42163.11	42931.35	43626.58	44783.64

The tonnage of cargo moved in the port the Leixões is our dependent variable. The dataset contained 31627 cargo movements of embarkation and disembarkation in the port, per 5 types of cargo, from January of 2009 to

September of 2018. The graph of Figure 3 depicts the values in table 2 and shows the increasing trend in the cargo movements in the port of Leixões in tons, from 2009 to 2017.

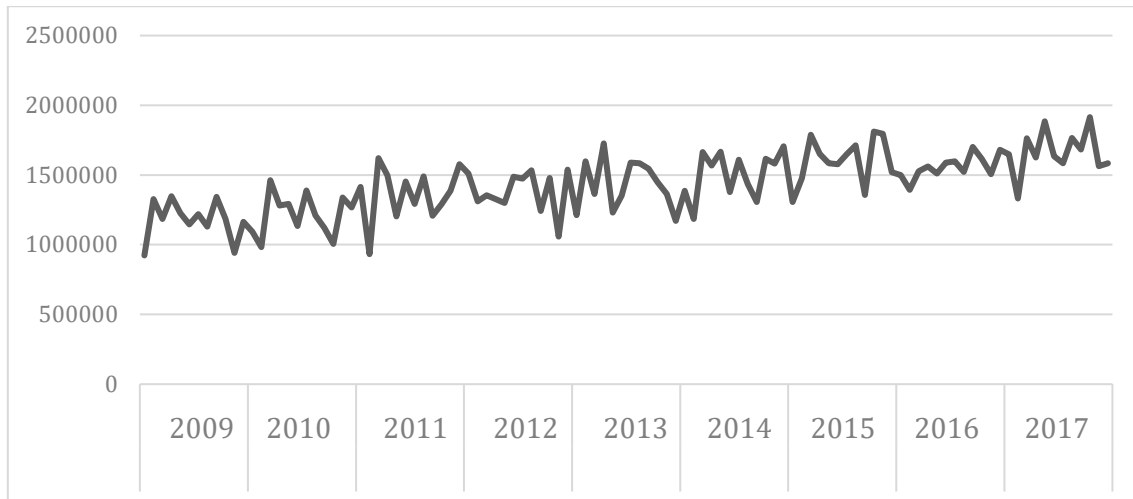


Figure 3: Leixões Tonnage

In each year of the series there are monthly fluctuations of the tons of cargo handled. However, these monthly oscillations are not regular throughout the series, i.e., there is no seasonality. This is found when comparing the cargo of each month with the average of the cargo of the corresponding year.

For example, in 2014, December has a higher than average cargo movement, being the month with the highest cargo movement in that year, while, in the previous year, its tonnage movements are below the average, being the month with the least cargo handling in 2013.

After grouping the data by quarter, in the graph of Figure 4, the monthly cargo movements are presented quarterly.

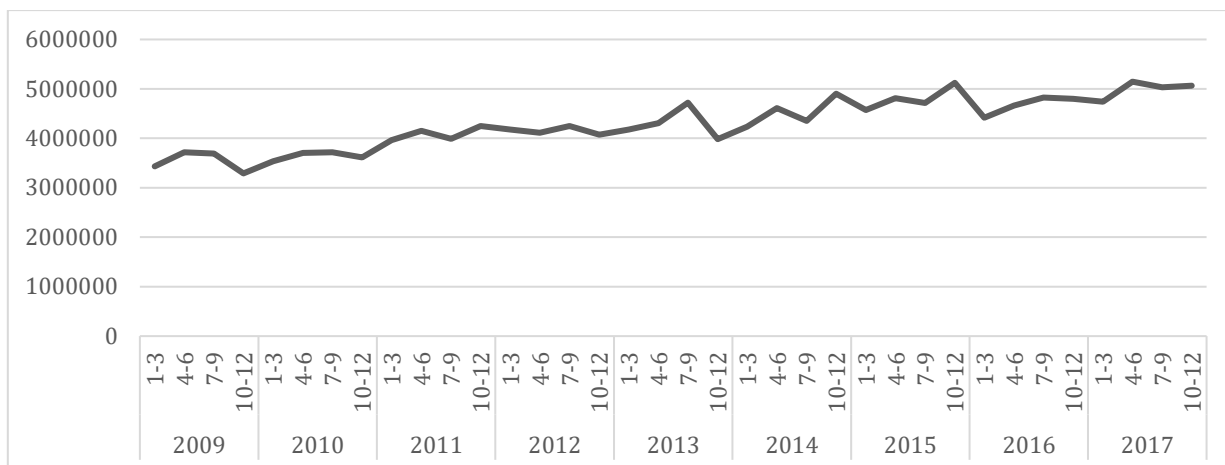


Figure 4: Total tonnage per quarter

When the analysis is done by quarter, it is still observed the trend of growth starting from 2011. It can also be noted that: starting in 2014, the quarter with the lowest amount of cargo handled is always the first; in 2014, 2015 and 2016, the quarter with the highest cargo handling is the last; there are oscillations, but these are not regular throughout the series.

Furthermore, APDL considered that competing ports were mainly those located in the geographical proximity and overlap of hinterlands of the port of Leixões.

It is important to note that the ports are monopolistic in most of its concessions, since operations involve investments in infrastructures of high economic value. Consequently, APDL explained that only in situations of suspending operations, which involve the diversion of cargo to another port, it is clear who is the competition.

As for data from other ports, the tons of cargo handled in the ports of Lisbon, Sines and Aveiro are considered explanatory variables. APDL identified its competitors, by type of cargo, as presented in Table 3.

Table 3: Competitors by type of cargo

Cargo Type	Competitors
Containerized cargo	Lisbon, Sines, Vigo and Setúbal
Liquid Bulk cargo	Sines and Coruña
Break Bulk cargo	*
Ro-Ro cargo	Setúbal and Vigo
Dry Bulk cargo	Aveiro

* information not provided

The evolution of these cargo variables in the other ports throughout the time series is shown in Figure 5 (where we only show annual values).

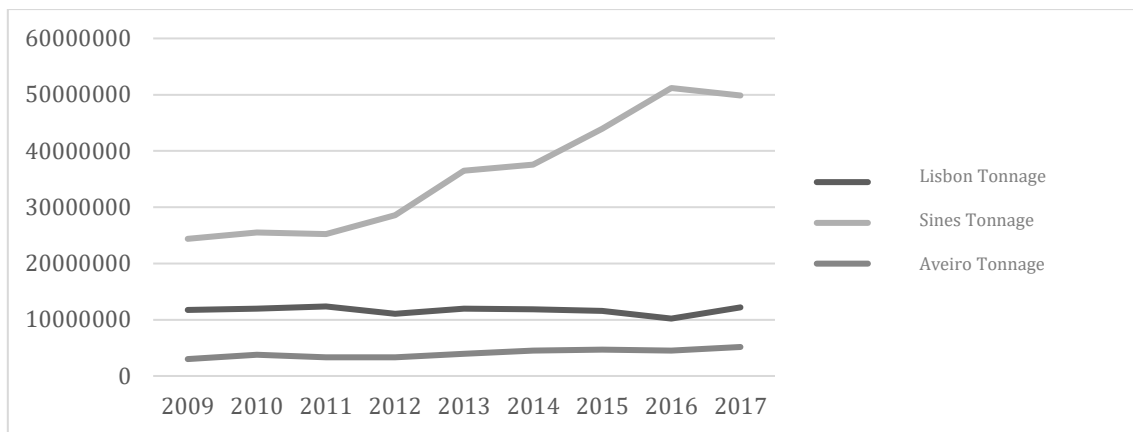


Figure 5: Other ports cargo

The evolution of the tonnage of cargo handled in the relevant ports, identified by APDL as its competitors, is stable, except in the port of Lisbon. Unlike the rest, the port of Lisbon has a higher amount of cargo moved over the years and a growing trend from 2011.

The last variable in Table 2 is GDP. This variable is calculated at market prices from the perspective of the expenditure and was obtained from the Quarterly National Accounts (“Contas Nacionais Trimestrais”) published by INE in the first trimester of 2018.

This choice of variable is justified by two reasons, according to the APDL: firstly, about 80% of the final customers are located in the metropolitan area of

Porto, Portugal; secondly, according to the company, it is expected that the movements in the port of Leixões are a reflection of the economic activity of the context where it is inserted. Thus, the industrial dynamics of the Portuguese region could determine the functioning of the port.

It should be noted that, although there is no access to data regarding the main customers of APDL, it is known that one of its main customers is GALP, which is a Portuguese company that operates in the oil terminal using the “monoboia”, whose head office is in Porto. Another major client is “Sidurgia da Maia” that imports scrap and exports iron and steel, whose location, as the name implies, is in Maia, Portugal.

The evolution of the quarterly GDP throughout the time series studied is shown in Figure 6.

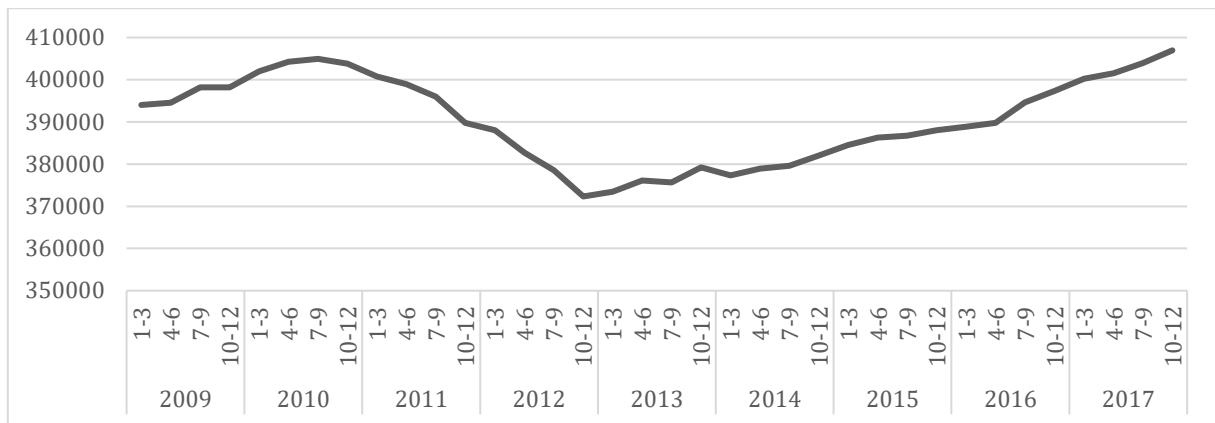


Figure 6: Quarterly GDP of Portugal

The quarterly GDP grows until the third quarter of 2010, when it reaches a high peak. From then on, it falls until the third quarter of 2012, when it assumes a growth trend until the end of the time series.

Table 4 shows the descriptive statistics of the categorical variables used.

Table 4: Descriptive statistics of the categorical variables

Variables	Values	Freq.	Perc.
Month	1	27	8.60%
	2	26	8.28%
	3	27	8.60%
	4	27	8.60%
	5	27	8.60%
	6	24	7.64%
	7	27	8.60%
	8	26	8.28%
	9	27	8.60%
	10	24	7.64%
	11	26	8.28%
	12	26	8.28%
Cargo Type	Containers	105	33.44%
	Liquid bulk	104	33.12%
	Others	105	33.44%
Strikes	0- No	284	90.45%
	1- Yes	30	9.55%

Three segments of cargo were considered in our analysis: containers, liquid bulk cargo and others.

Originally, APDL classifies them into 5 categories: containerized cargo, liquid bulk cargo, break bulk cargo, roll-on roll-off (Ro-Ro) cargo and dry bulk cargo. We reduced the number of categories, by creating the type of cargo “others” that includes the remaining 3 categories mentioned above.

Each type of cargo has its own requirements, implying an adequate infrastructure. For instance, the Ro-Ro cargo corresponds to the cargo entering and leaving the ship by means such as wheels and ramps, no cranes required. This type of cargo is moved in the Ro-Ro Terminal. Furthermore, in the oil terminal, the oceanic “monoboia” allows to move oil that belongs to the liquid bulk cargo type.

The two types of cargo that are mostly handled in the port of Leixões are liquid bulk and containerized cargo, as shown by Figure 7. The Ro-Ro cargo,

despite having a small proportion of the global movements in the port of Leixões, has been increasing since 2013.

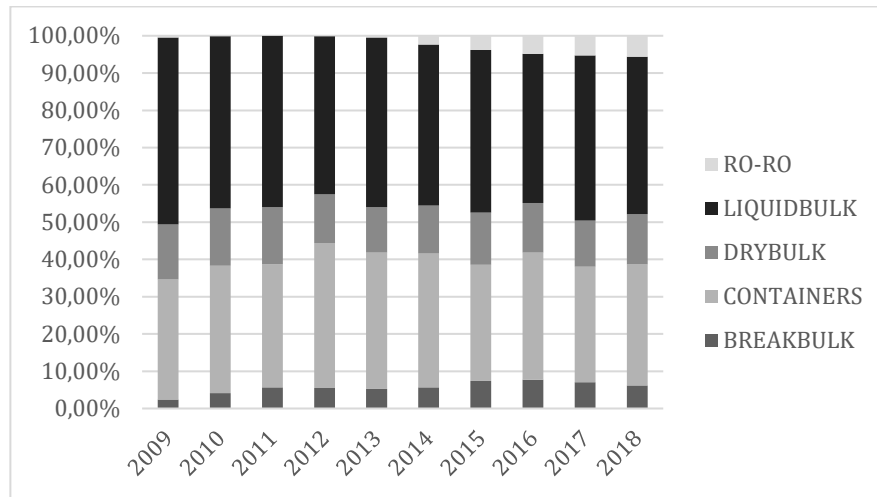


Figure 7: Percentage of each type of cargo moved per year

In addition, we made an analysis of the evolution of the tonnage moved in port of Leixões per type of cargo, namely: containerized cargo (Figure 8), liquid bulk cargo (Figure 9), RO-RO cargo (Figure 10), breakbulk cargo (Figure 11) and dry bulk cargo (Figure 12).

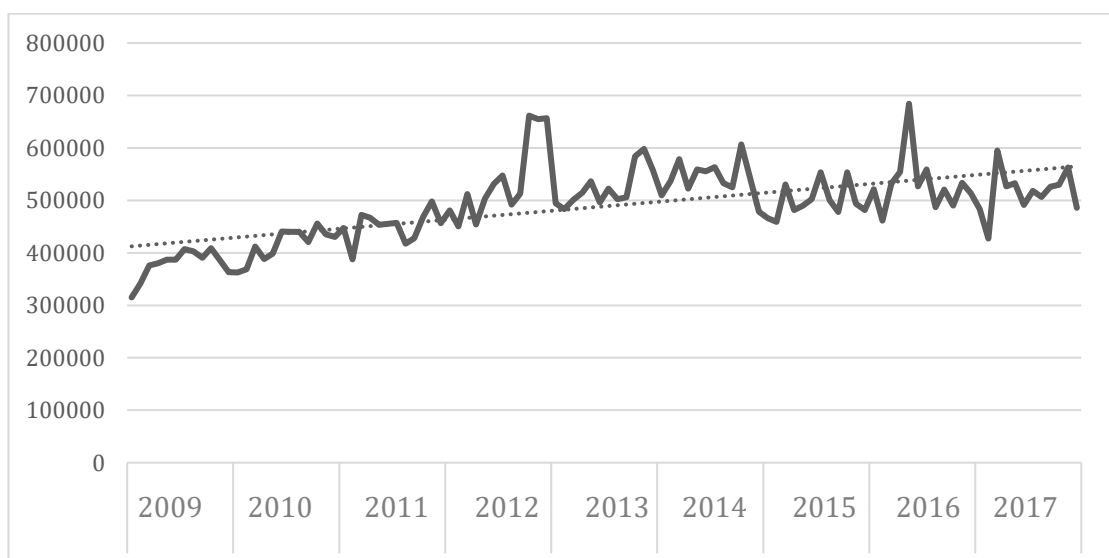


Figure 8: Containerized cargo tonnage per month

When one observes the evolution of the tons of containerized cargo moved, in Figure 8, there is a clear trend of growth from 2009 to 2014. At the end of 2014 and the beginning of 2015, there is a drop, but the trend of growth continues to occur.

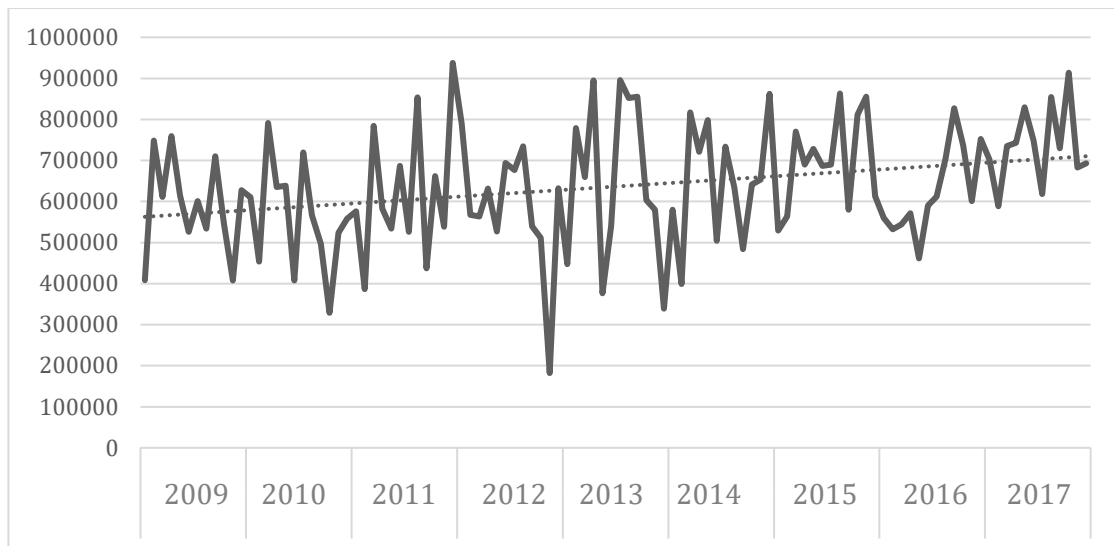


Figure 9: Liquid bulk cargo tonnage per month

As for liquid bulk, there is evidence of marked oscillations throughout the time series, as it is clear in Figure 9. In the case of this type of cargo, regular oscillations are observed throughout the series: in the years from 2009 to 2013, except in 2011, the lowest peaks of movements of cargo were at the end of the year, whereas, in the years from 2014 to 2017, these peaks were at the beginning of the year. However, there does not seem to be a pattern in the peaks of higher cargo handling.

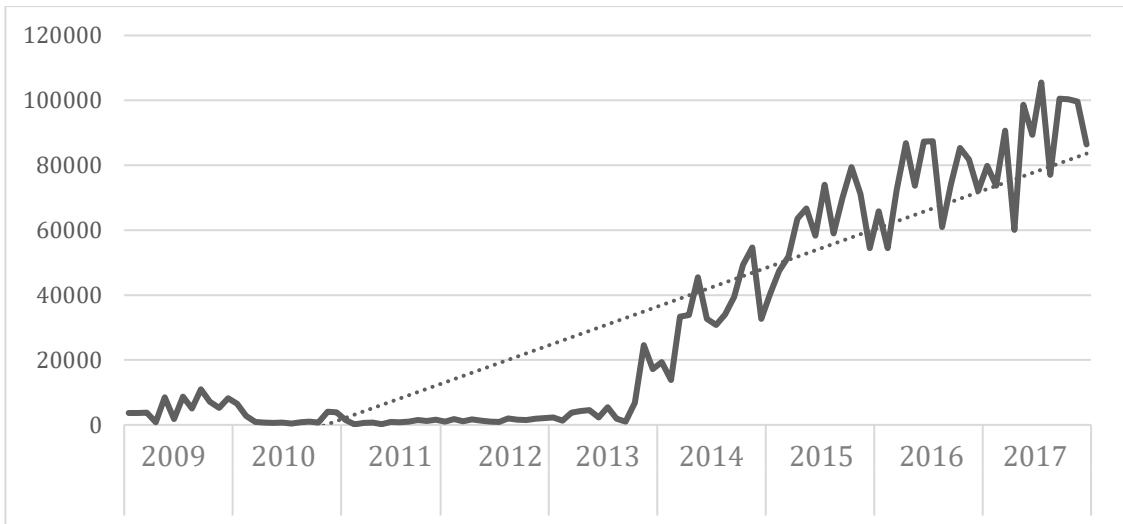


Figure 10: RO-RO cargo tonnage per month

RO-RO cargo movements, although very low at the beginning of the time series, show a strong growth trend from 2013, as depicted in Figure 10.

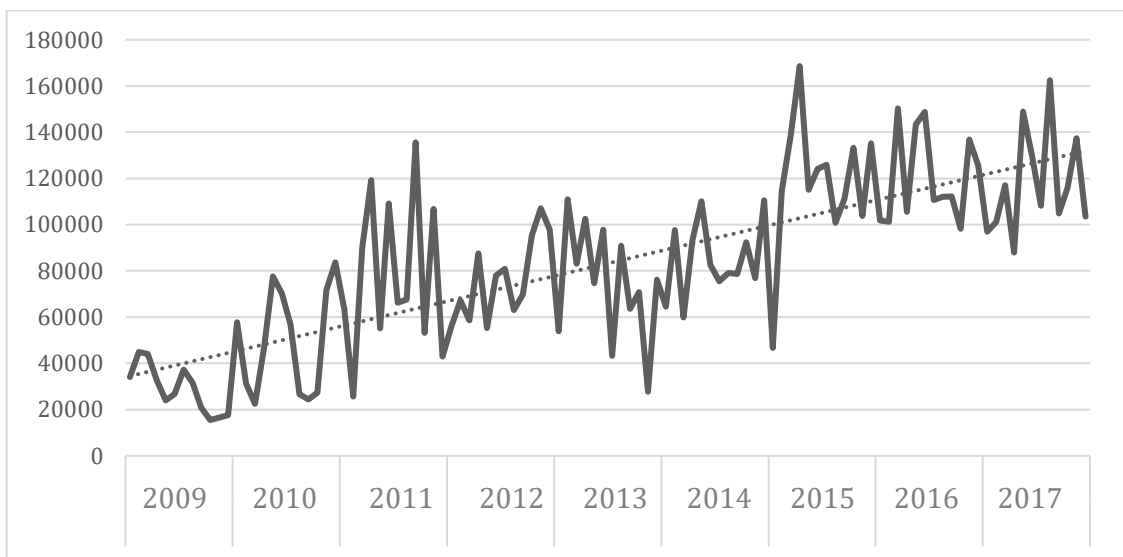


Figure 11: Breakbulk cargo tonnage per month

Figure 11 shows that the evolution of breakbulk cargo movements has a growth trend, although it has a lot of oscillations. However, it has a decrease trend from 2009 to 2010.

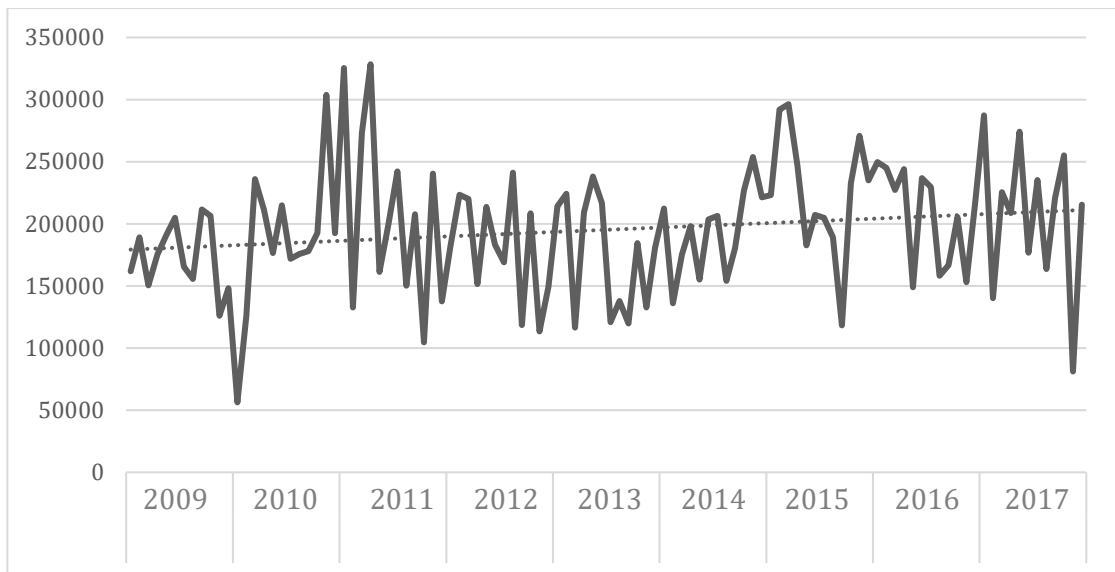


Figure 12: Dry bulk cargo tonnage per month

As for dry bulk cargo, its movements are characterized with a lot of oscillations. It can be observed an extreme low peak in 2009 and 2017, and a high peak in 2011.

Furthermore, in order to include strikes as a variable, a binary explanatory variable was created, where 0 corresponds to the months of normal workflow of all ports operations and 1 corresponds to the months where there was at least one strike in one Portuguese port.

The data regarding strikes was collected through the internet, namely, news sites.

There are several situations that lead to the paralysis of port operations, specifically strikes and construction work.

Strikes are a way for employees to negotiate working conditions. According to the Constitution of the Portuguese Republic (Constituição da República Portuguesa, Artigo 57º), workers in Portugal have the right to strike, i.e., the right to voluntarily and collectively suspend the performance of their work in order to achieve a certain goal related to their jobs.

Strikes and port operation stoppages can lead to the change of the scale of ships, diverting cargo to other ports. Thus, these events can have an impact on the cargo movements in the ports.

An extended stoppage may be favorable for the port of Leixões, when it occurs in ports and leads to a diversion of cargo to Leixões, increasing its cargo movements, or unfavorable if it occurs in Leixões, decreasing its cargo movements. For instance, a stop of 3 or more weeks at the Sines refinery is quite favorable to the port of Leixões, as it is its competitor at liquid bulk cargo level. Therefore, these cargoes will be diverted to the refinery of Leixões, implying an increase in the cargo movements of Leixões.

4.1 Non-included variables

There were other explanatory variables that we would like to have tested but were not possible to collect, or that we tested, but which were not included in the final models.

Climate, measured through weather conditions like rain, sea agitation, etc., was one possible explanatory variable suggested to be used in the analysis of the monthly movements of cargo in the port of Leixões, given that the weather conditions have impacts on transportation. However, collaborators at the Port Authority argue that in the port of Leixões there is no climate effects on the movement of cargo in the long term, only in the short term. For instance, if there is a storm that prevents a ship from entering the port to unload or upload cargo, it waits until it is possible to enter. According to the APDL, there has never been a diversion of cargo to another port, or vice versa, due to bad weather or an abnormal climate event. Another example is the tide. When there is strong sea agitation, it is not possible to operate the “monoboia”, so the activities in it are suspended. However, the refinery is already waiting for these

episodes, so they are compensated in other periods. Therefore, we did not test for climate related variables.

Nevertheless, we created an explanatory binary variable that distinguishes summer and winter months, such that we can analyse seasonality, which is, for obvious reasons, related to climate. Specifically, the pattern of seasonality that we considered was to distinguish between summer and winter months. However, this was not a statically significant variable. Note that the month was considered in our models, so in a sense we still accommodate for some seasonal effects.

The “other modes of transport” was an explanatory variable that we would like to have tested. The modes of transportation may be complementary as well as competitors of the maritime transport of the port of Leixões. Since there is no way to distinguish the two from the available data, we could conclude, by its major effects on the movement, if the majority of the other transportation modes are complementary or rival. For instance, on the one hand, a cargo can be transported to Rotterdam by ship or by road transport mode, in which case they are competing modes of transport. On the other hand, if a cargo arrives at the port of Leixões with final destination in a Vila Real company, it will have to be transported from the port to the company by road or rail, so in this case they are complementary modes of transport. Ideally, we would like to have available data regarding other transportation modes, but it was not possible to collect those data.

In addition to the variables mentioned, we also created a variable, “Time”, in order to analyse the effect of the evolution throughout the time. This variable indexes the months considered, from 1 (January 2009) to 108 (December 2017). This variable was not statistically significant as well.

Chapter 5

Data analysis

Two different studies were developed: the first one aims to explain the total tonnage of cargo moved in the port of Leixões, while the second tried to perceive the same, but by type of cargo.

We started by preparing the dataset.

5.1 Dataset

To carry out the study, a database containing 314 observations and 14 variables was constructed.

The first phase consisted of data pre-processing.

First, we started by verifying and correcting the classification of all the variables. Moreover, the original data included some negative tonnage values in the variables about the other ports. This is due to the fact that the other ports do not consider the same classifications for the cargo as APDL, so this company has to estimate the tonnage for each type of cargo from those data, and sometimes there are errors that result from the automatic way of calculation. Therefore, we thought it was best to erase the observations that contained negative tonnage. This choice is justified by the number of observations erased, 10, which is relatively small, since the original sample had 324 observations.

Also, we found that the number of missing values in each row of the dataset was zero.

As for data reduction, for the purpose of the subsequent analysis, we created three subsets for each type of cargo: “containers”, “liquidbulk” and “others”.

Additionally, to verify the existence of outliers, we used the classical statistical method: the boxplots of the numeric variables, present in Figure 13.

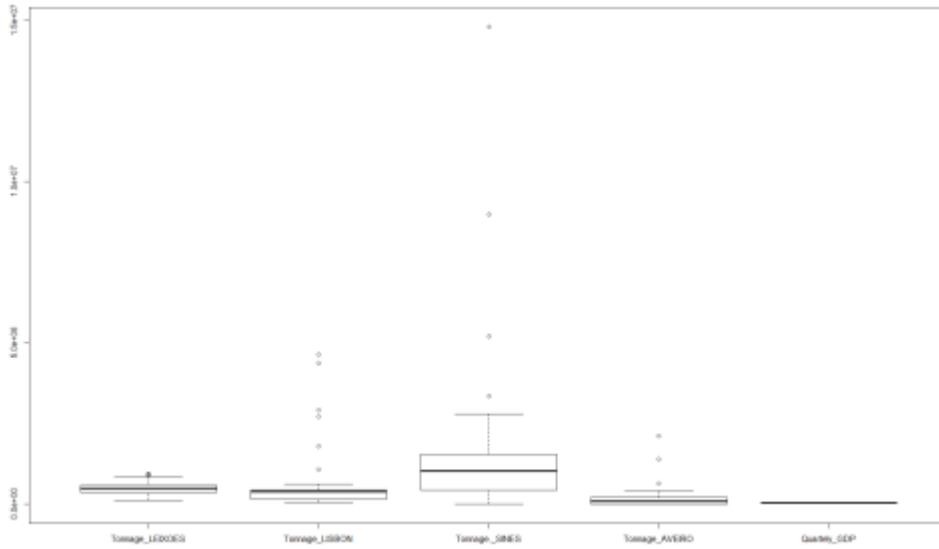


Figure 13: Boxplots of the quantitative variables

Thus, it is clear the existence of atypical values, i.e., atypical cargo movements in some months, in the ports of Leixões, Lisbon, Sines and Aveiro. Later, when studying each model, we discovered if there were severe outliers and, if so, if they were influential.

5.2 Regression model: Leixões Tonnage

As previously stated, in the context of our study, we developed a regression model that helped us to explain the movements of cargo in Leixões.

Before arriving at the final model, we estimated several models.

Firstly, we noticed that the time related variables, “month”, “time” and “seasonality” had a perfect linear relationship, originating multicollinearity, so we started by estimating three different models, including in turn one of these variables.

We estimated numerous models, until we got simpler and reduced models, with most explanatory variable coefficients as statistically significant.

We followed a general to specific strategy to search for an appropriate model, using the Akaike Information Criterion (AIC) as a guiding principle. We tested this model against a full model with all non colinear possible explanatory variable by an ANOVA F test, accepting the validity of our chosen model.

Then, we studied the effect of interactions between the tonnage of cargo handled in Sines and the existence or absence of strikes. The hypothesis that arises in this case is that the existence of strikes can lead to the diversion of cargo from one port to another.

The use of interaction terms implied the need to centre our Sines Tonnage variable to facilitate the interpretation of the coefficients.

The model with the interactions explained the cargo movements of Leixões in a larger percentage, i.e., had a larger R^2 than the model without it, which means that the interactions contribute to explain the movements in Leixões.

The final model thus estimated is shown in table 5.

Table 5: Regression model RegAI

	RegAI				
	Estimate	Std. error	t	P-value	
Intercept	-4.452e+07	4.191e+06	-10.622	0.00	***
Year	2.234e+04	2.082e+03	10.731	0.00	***
Month 2 (ref:January)	-2.909e+04	2.485e+04	-1.171	0.24273	
Month 3 (ref:January)	6.583e+04	2.461e+04	2.675	0.00789	**
Month 4 (ref:January)	6.082e+04	2.469e+04	2.463	0.01434	*
Month 5 (ref:January)	3.419e+04	2.460e+04	1.390	0.16563	
Month 6 (ref:January)	2.340e+04	2.548e+04	0.918	0.35922	
Month 7 (ref:January)	5.987e+04	2.478e+04	2.416	0.01630	*
Month 8 (ref:January)	4.771e+04	2.493e+04	1.913	0.05667	.
Month 9 (ref:January)	1.257e+04	2.468e+04	0.510	0.61075	
Month 10 (ref:January)	4.923e+04	2.542e+04	1.937	0.05373	.
Month 11(ref:January)	2.093e+04	2.528e+04	0.828	0.40843	
Month 12 (ref:January)	4.051e+04	2.492e+04	1.626	0.10508	
Cargo Type Liquid bulk (ref: Containers)	1.592e+05	1.325e+04	12.008	0.00	***
Cargo Type Others (ref: Containers)	-1.896e+05	1.289e+04	-14.711	0.00	***
Strikes (ref: No)	-4.197e+04	1.801e+04	-2.330	0.02045	*
Sines Tonnage	-9.817e-03	5.402e-03	-1.817	0.07018	.
Sines Tonnage * Strikes	-8.879e-02	2.683e-02	-3.310	0.00105	**
R ²	73.58%				

Note: significance level at 0.1% (***), 1% (**), 5% (*) and 10%(.)

The p-value of the F test of this model is very close to zero, which testifies the statistical significance of the model and the importance of the explanatory variables in explaining the Tonnage of Leixões.

Making a more concrete analysis of the impact that the chosen variables have, we can draw the following conclusions:

- In the estimated model the cargo movements in the Port of Leixões are explained in about 73.58%;
- There is a growth trend over the years;
- Although there are no statistically significant effects in most of the months, March, April and July are typically the months when more cargo movements are expected, compared to January;
- There is likely to be less movement of “others” cargo and more movement of liquid bulk cargo than of the reference type, containers;

- There is likely to be less movement of cargo in Leixões when there are strikes, compared to when there are no strikes, when Sines Tonnage is equal to its sample average;
- An increase of 1 kilogram of cargo moved in the port of Sines leads to a reduction of, approximately, 10 kilograms (9.82) in the port of Leixões, when there are no strikes;
- When there are strikes, the reduction of tons moved in Leixões, caused by each additional ton of Sines, increases by 9 kilograms, that is, it decreases by 19 kilograms, compared to when there are no strikes.

5.2.1 Assumptions

- **Multicollinearity**

To determine if the variables were linearly associated, we analysed the correlation between the numeric variables, and it was not strong.

In order to formally diagnose the presence of multicollinearity, we analysed the values of GVIF, which when high are indicators of multicollinearity. The higher GVIF was 1.19, which is not substantial. So, we did not detect multicollinearity.

- **Heteroscedasticity**

In order to determine if the model was heteroscedastic or homoscedastic, i.e., if the variance of "Tonnage_Leixoes" depended on the observations or not, we proceeded to the diagnosis. Firstly, we analysed the graphical representations, present in Figure 14.

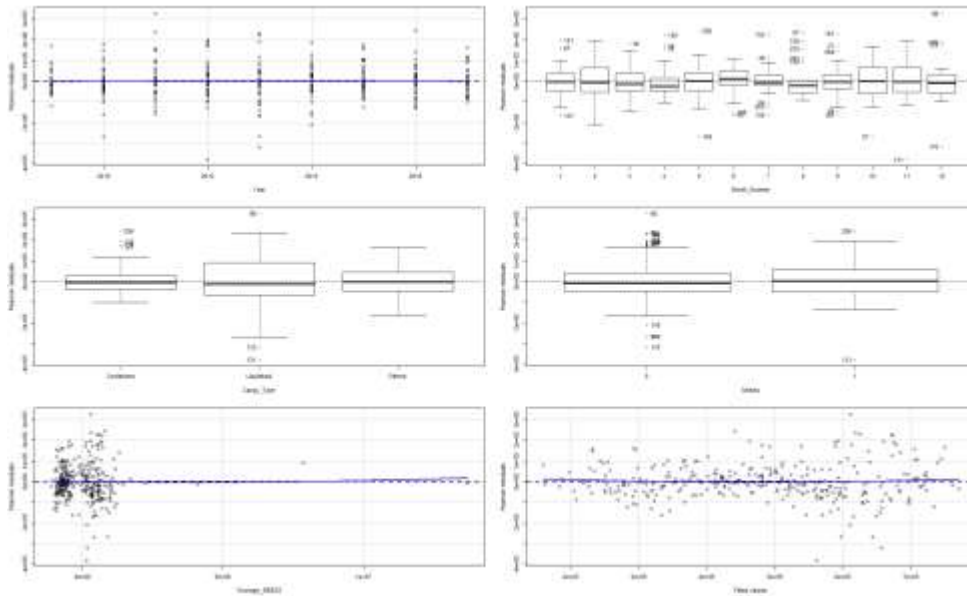


Figure 14: Residual Plots

The graphs suggest the presence of heteroscedasticity, namely through three factors: first, the residuals vary more when the type of cargo is liquid bulk; second, a greater variability of the residuals is observed in the autumn/winter months; finally, although the scale is not ideal, the residuals vary more when the tonnage moved in the port of Sines is higher.

In order to formally test the evidence for heteroscedasticity, we performed the Breusch-Pagan test. The p-value was very close to zero, so we can reject the null hypothesis.

Furthermore, in order to proceed with our analysis with the presence of heteroscedasticity, not to eliminate it, but to enable the usual techniques of statistical inference from the results of the adjustment, we proceeded to test the robust standard errors, suggested by Halbert White. Comparing the standard errors with the robust standard errors that emerged from the "remedy" applied to heteroscedasticity, we found that the results did not change significantly.

- **Autocorrelation**

Since we have 3 observations for each month, corresponding to each type of cargo, in order to verify the presence or absence of autocorrelation, we had to do the test proposed by Wooldridge (2002) for panel data.

The test detected autocorrection. One of the common measures to deal with autocorrelated residuals would be to use robust standard errors, such as those proposed by Arellano (1987). However, the best-known variant of these errors relies on large n asymptotic (where n is the number of cases by time unit) and is not applicable in our setting, where n is only equal to 3.

Therefore, in the future, these problems should be taken into account.

- **Normality**

The assumption of normality accepts that the residuals follow a normal distribution with a null mean. To verify the normality of the data, we analysed the graph of Figure 15.

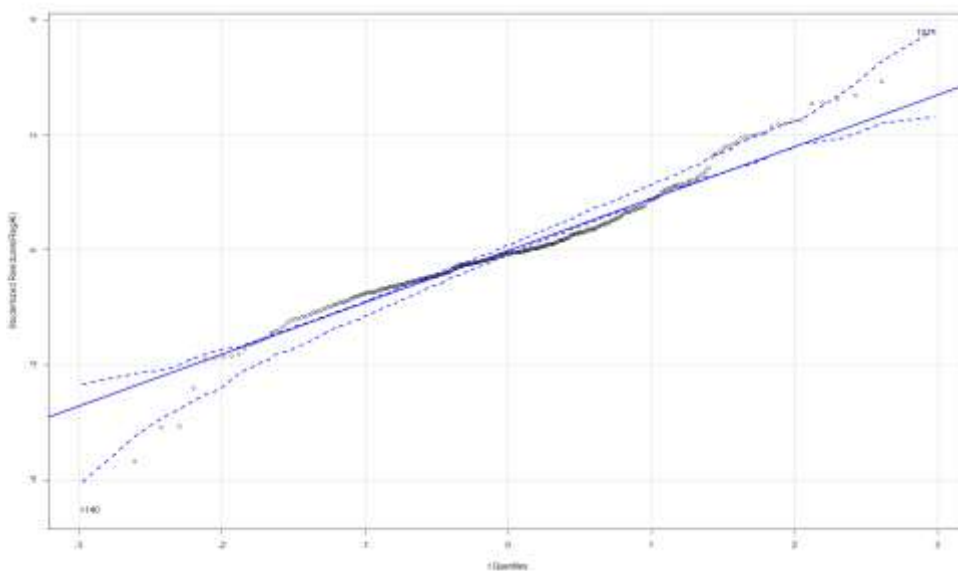


Figure 15: Normality of the model RegA

A deviation from normality is visible since the data are not symmetrical and the tails are heavy. However, since we have a relatively large sample, we relied

on asymptotic arguments to perform the analyses in the absence of normally distributed disturbances (Eicker, 1963 and Weisberg, 2005). Nevertheless, afterwards we performed an influence analysis to check possible negative effects of large outliers.

5.2.2 Outliers Analysis

To understand if the atypical observations, identified in section 5.1, skewed the chosen model, we started by identifying the severe outliers. The Bonferroni Outlier Test suggested that there were observations that were severe outliers. The graphs of Figures 16 and 17 allowed us to make the diagnosis of the influential outliers, given by observations 107, 140 and 265.

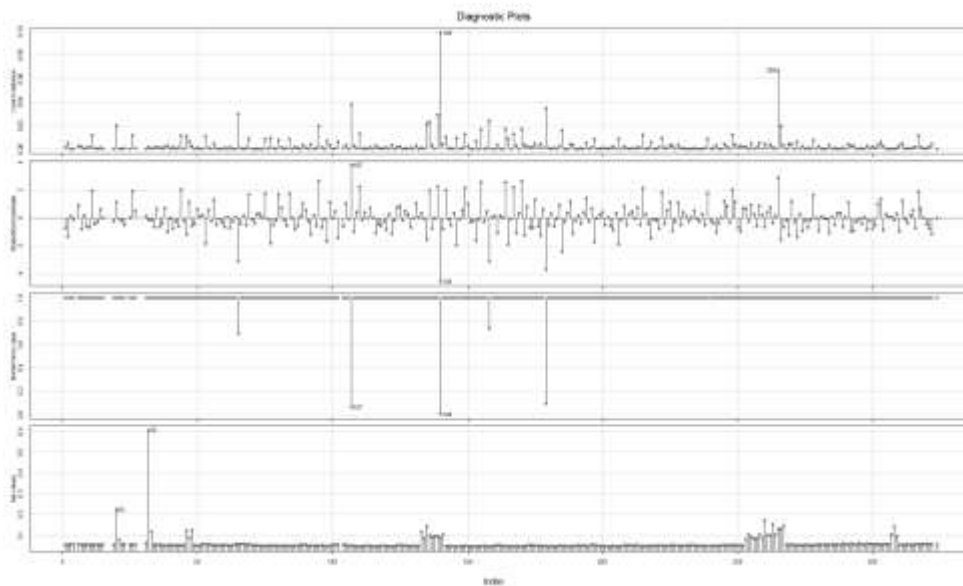


Figure 16: Influence measures

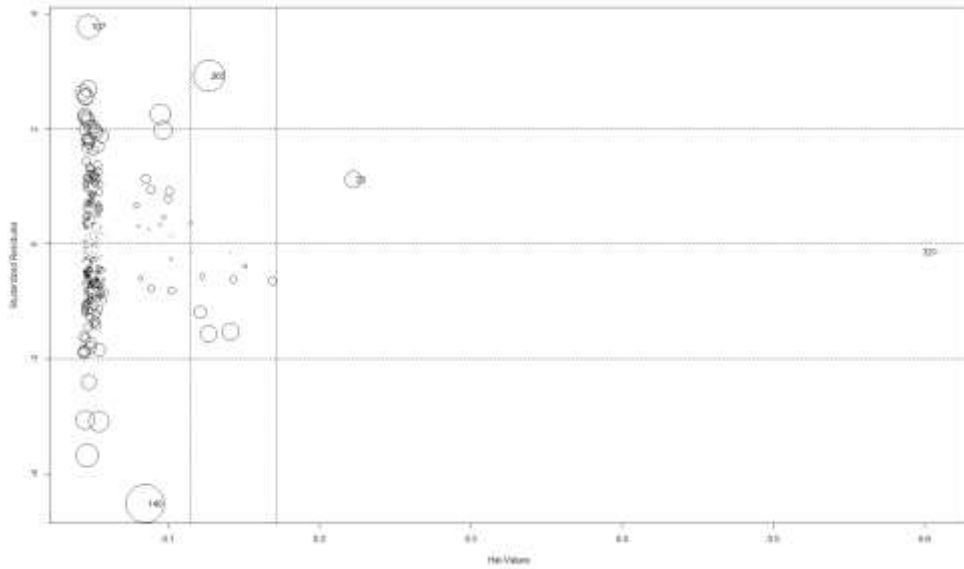


Figure 17: Influence measures

We estimated the model without the influent outliers and the conclusions did not change.

To sum up, the regression analysis allowed us to identify the main external determinants of the cargo movements in the port of Leixões: year of movement, month of movement, type of cargo, tonnage moved in the port of Sines, existence or absence of strikes and interaction between the Sines tonnage and the binary variable of the strikes.

It makes sense that the cargo movements in the port of Sines influences those in the port of Leixões, since this port was identified by APDL as one of the main competitors, namely in terms of the busiest cargoes, containers and liquid bulk.

As for the interaction, the fact that it is statistically significant to explain the cargo movements in the port of Leixões shows that the shift of cargo between Leixões and Sines during the strikes is significant. In this case, the existence of strikes causes a reduction of the tons in Leixões that is greater with the increase of the cargo moved in Sines, that is, the strikes lead to a diversion of cargo in the direction of Leixões to Sines.

5.3 Regression Models: per type of cargo

Additionally, we decided to understand the main external determinants of the movements of each type of cargo in the port of Leixões. Accordingly, as mentioned earlier, in the data reduction phase, we created 3 subsets for each type of cargo: containerized, liquid bulk and others.

Note that the processes were very similar to the processes described in the previous section.

5.3.1 Containerized cargo

In order to answer the research question for the containerized cargo, we started by estimating several models, until we had the final one. This model has the largest determination coefficient (R^2) and its explanatory variables are all statically significant.

The estimation of this model is in table 6.

Table 6: Regression model Reg1C_step

	Reg1C_step				
	Estimate	Std. error	t	P-value	
Intercept	-2.672e+07	2.884e+06	-9.266	1.16e-14	***
Year	1.407e+04	1.426e+03	9.869	6.69e-16	***
Month 2 (ref:January)	-1.748e+04	1.648e+04	-1.061	0.291690	
Month 3 (ref:January)	4.917e+04	1.649e+04	2.982	0.003703	**
Month 4 (ref:January)	2.190e+04	1.654e+04	1.324	0.188981	
Month 5 (ref:January)	5.474e+04	1.649e+04	3.320	0.001310	**
Month 6 (ref:January)	4.763e+04	1.711e+04	2.784	0.006569	**
Month 7 (ref:January)	6.546e+04	1.674e+04	3.910	0.000181	***
Month 8 (ref:January)	3.296e+04	1.656e+04	1.990	0.049655	*
Month 9 (ref:January)	2.948e+04	1.649e+04	1.788	0.077181	.
Month 10 (ref:January)	8.842e+04	1.707e+04	5.179	1.40e-06	***
Month 11(ref:January)	9.447e+04	1.813e+04	5.210	1.23e-06	***
Month 12 (ref:January)	4.650e+04	1.656e+04	2.809	0.006130	**
Lisbon Tonnage	-1.906e-02	7.915e-03	-2.409	0.018092	*
Aveiro Tonnage	-1.167e+02	4.030e+01	-2.895	0.004779	**
Quarterly GDP	-2.643e+01	3.026e+00	-8.735	1.44e-13	***
Strikes (ref: No)	3.549e+04	1.213e+04	2.925	0.004378	**
R^2	78.74%				

Note: significance level at 0.1% (***), 1% (**), 5% (*) and 10%(.)

Interpreting the coefficients of the explanatory variables of the estimated model, it is possible to conclude:

- In the estimated model the containerized cargo movements in the port of Leixões are explained in about 78.74%;
- There is a growth trend over the years;
- In most of the months, except February, it is expected more cargo movements than in January;
- An increase of 1 kilogram of containerized cargo moved in the port of Lisbon leads to a reduction of 19 kilograms in the port of Leixões;
- An increase of 1 kilogram of containerized cargo moved in the port of Aveiro leads to a reduction of 11.7 kilograms in the port of Leixões;
- An increase of 1 million euros in the quarterly GDP leads to a reduction of, approximately, 26 kilograms of containerized cargo in the port of Leixões;
- There is likely to be more movement of containers in Leixões when there are strikes, compared to when there are no strikes.

5.3.1.1 Assumptions

- **Multicollinearity**

The numerical variables do not have a very strong correlation with each other. However, there is a positive correlation above 0.5 between the variables Lisbon Tonnage and Aveiro Tonnage (0.51). Thus, the correlation between the tons of containerized cargo moved in Lisbon and those moved in Aveiro suggests that an increase in one kilo of cargo in the port of Lisbon will lead to an estimated increase in the cargo handled in Aveiro of 0.51 kilograms.

The values of Generalized Variance Influence Factors (GVIF) were not high, thus, we did not detect multicollinearity.

- **Heteroscedasticity**

To verify the existence of either homoscedasticity or heteroscedasticity, we started by analysing the graphs in Figure 18, in order to see what these graphs would suggest.

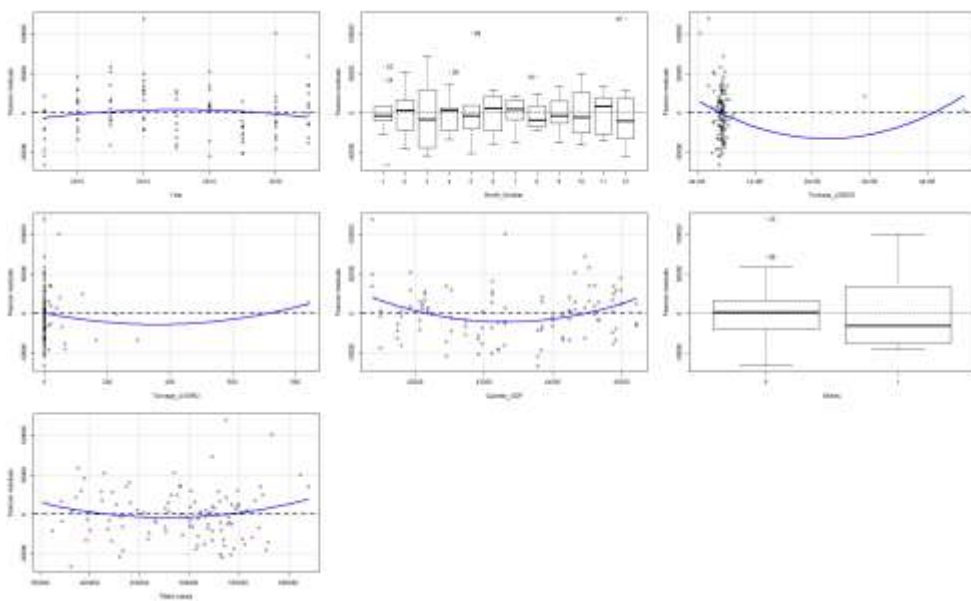


Figure 18: Residual Plots

The graphs suggest, although is not too clear, the possible existence of heteroscedasticity, namely by these fonts: there was more variability of disturbances when there were strikes and in the months of March, June, October, November and December.

However, to formally address this topic, we performed the Breusch-Pagan test that suggested that the model was homoscedastic. So, we concluded the absence of heteroscedasticity.

- **Autocorrelation**

There seems to be no autocorrelation pattern, since the disturbances seem to be independent over time, according to Figure 19.

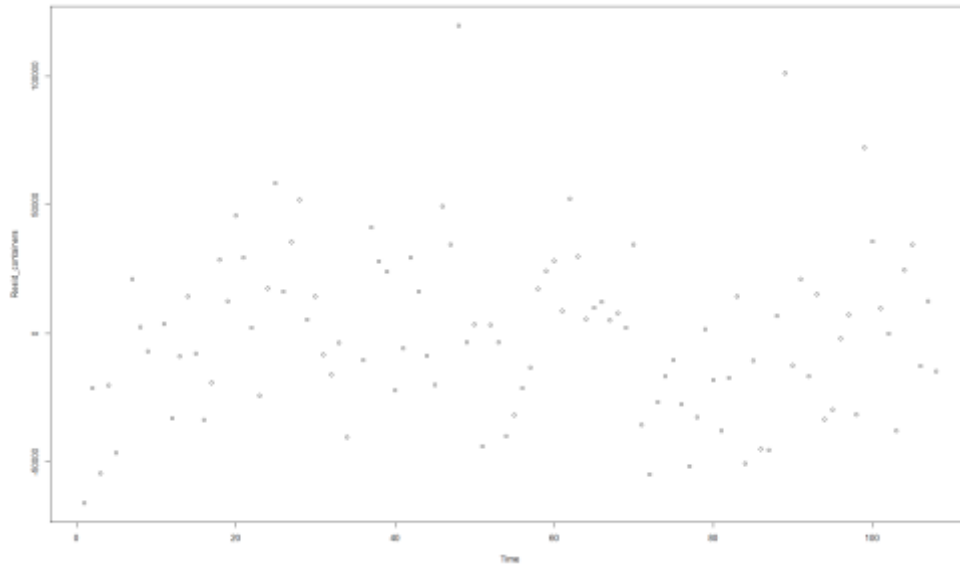


Figure 19: Residuals

Nevertheless, we decided to formally assess autocorrelation with the Durbin-Watson test that suggested the presence of autocorrelation. As a method to deal with this problem, we used the heteroscedasticity and autocorrelation consistent covariance matrix. The results did not change with this remedy.

- **Normality**

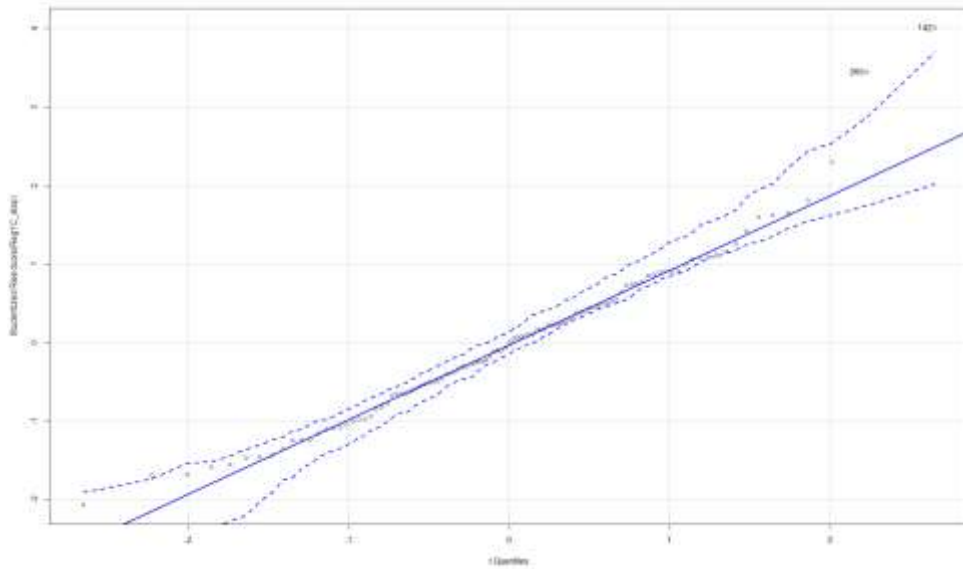


Figure 20: Normality of the model Reg1C_step

It is evident, according to the graph of the Figure 20, that there is a deviation from normality, i.e., the dots do not go along the straight line that represents the normal distribution. As it is a large sample, we relied, as in the previous section, on asymptotic arguments to perform the analyses in the absence of normally distributed disturbances.

5.3.1.2 Outliers Analysis

To understand if there were severe outliers, we performed the Bonferroni Outlier Test, that suggested that there were severe atypical observations. In Figures 21 and 22, we realized which of these outliers were influent.

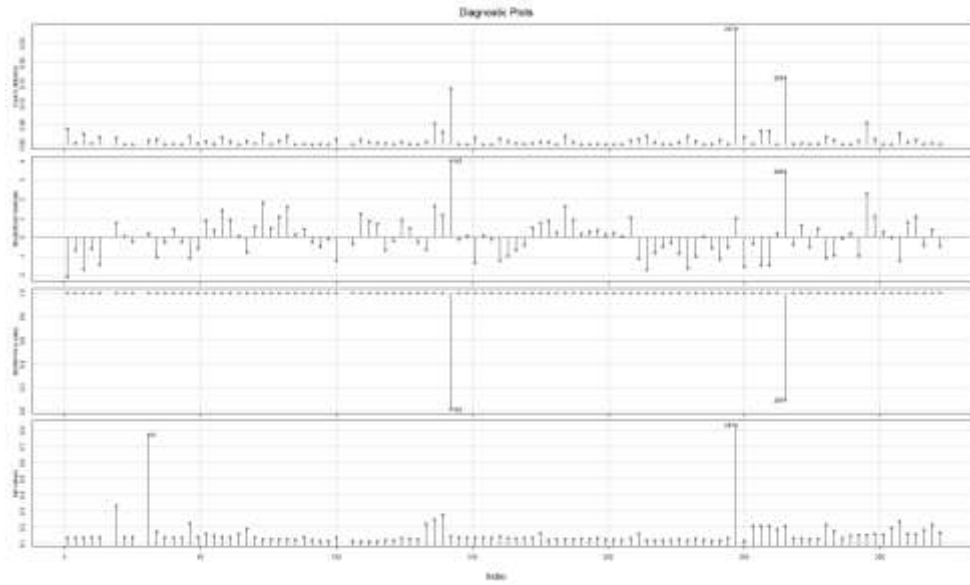


Figure 21: Influence measures

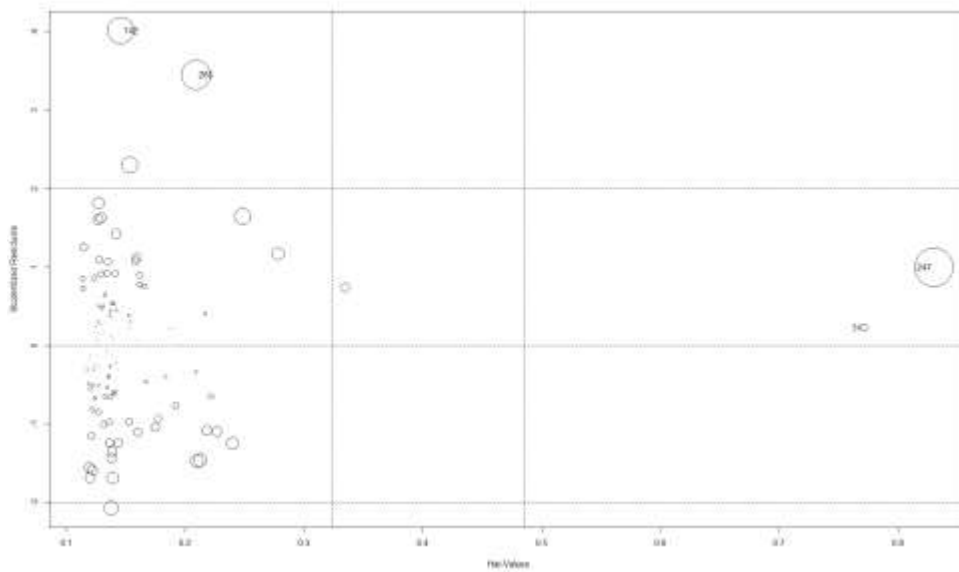


Figure 22: Influence measures

Furthermore, we estimated the model without these outliers and the results did not change.

To sum up, the main determinants of the movements of containerized cargo in the port of Leixões are year of the movement, month of the movement,

Lisbon containerized cargo tonnage, Aveiro containerized cargo tonnage, quarterly GDP and strikes.

5.3.2 Liquid Bulk cargo

Regarding the tonnage of liquid bulk handled in the port of Leixões, we estimated several models to explain it, until we reached the chosen model, estimated in Table 7.

Table 7: Regression model Reg1L_step

	Reg1L_step				
	Estimate	Std. error	t	P-value	
Intercept	-4.218e+07	1.001e+07	-4.214	5.96e-05	***
Year	2.125e+04	4.973e+03	4.273	4.78e-05	***
Month 2 (ref:January)	-5.260e+04	6.077e+04	-0.866	0.389046	
Month 3 (ref:January)	1.192e+05	5.891e+04	2.023	0.046065	*
Month 4 (ref:January)	1.322e+05	5.910e+04	2.237	0.027744	*
Month 5 (ref:January)	3.375e+04	5.891e+04	0.573	0.568154	
Month 6 (ref:January)	1.947e+02	6.099e+04	0.003	0.997460	
Month 7 (ref:January)	9.662e+04	5.891e+04	1.640	0.104455	
Month 8 (ref:January)	1.365e+05	5.910e+04	2.310	0.023165	*
Month 9 (ref:January)	5.074e+04	5.891e+04	0.861	0.391301	
Month 10 (ref:January)	6.463e+04	6.077e+04	1.063	0.290428	
Month 11(ref:January)	-1.974e+04	5.891e+04	-0.335	0.738297	
Month 12 (ref:January)	7.949e+04	6.093e+04	1.305	0.195320	
Strikes (ref: No)	-1.642e+05	4.307e+04	-3.813	0.000251	***
R ²	36.01%				

Note: significance level at 0.1% (***), 1% (**), 5% (*) and 10%(·)

We can draw some conclusions:

- The tonnage of liquid bulk cargo moved in the port of Leixões can be explain by this model in 36%;
- There is a growth trend over the years;
- There are no statistically significant effects in most of the months, except for March, April and August, which are the months when typically cargo movements are expected to be higher than in January;

- There is likely to be less movement of cargo in Leixões when there are strikes, compared to when there are no strikes.

5.3.2.1 Assumptions

- **Multicollinearity**

In order to understand if the explanatory variables were linearly associated, we analysed the GVIF. These were not high, thus, there was no indication of multicollinearity.

- **Heteroscedasticity**

The graphs from Figure 23 do not suggest the presence of serious heteroscedasticity.

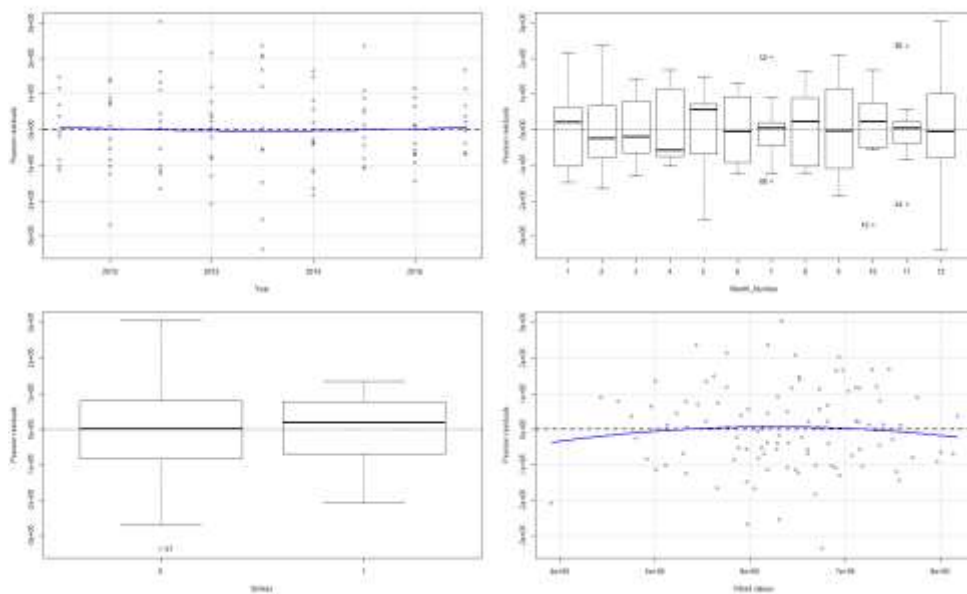


Figure 23: Residual Plots

Taking a more formal approach, the Breusch-Pagan test confirmed that there was no heteroscedasticity.

- **Autocorrelation**

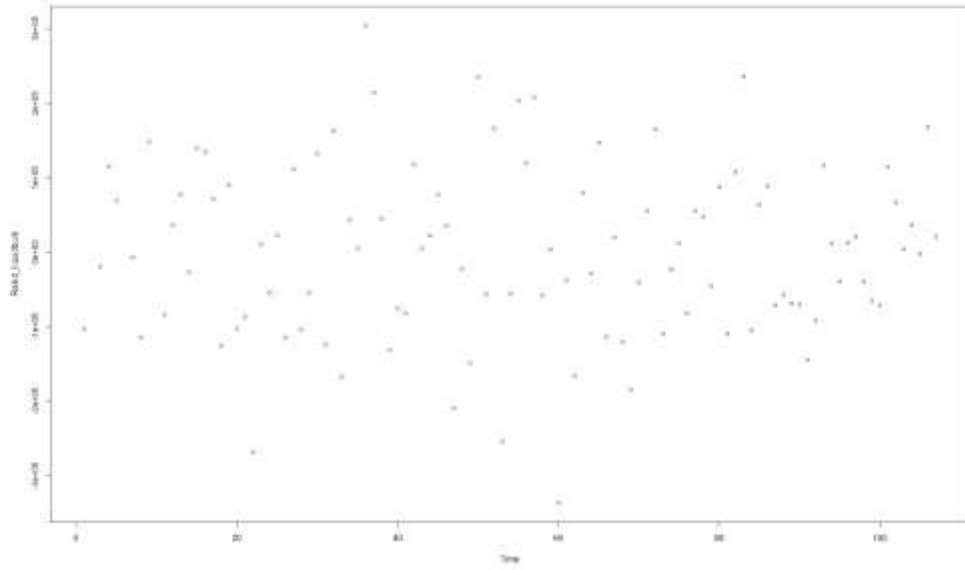


Figure 24: Residuals

It was evident, from Figure 24, that the disturbances were not correlated. This conclusion was corroborated by the Durbin-Watson test.

- **Normality**

It can be notice, by the graph in Figure 25, a deviation between the dots and the straight line, i.e., a deviation from normality. Nevertheless, like in the case of the other models, since we have a large sample, we continued the analysis, even with absence of normality.

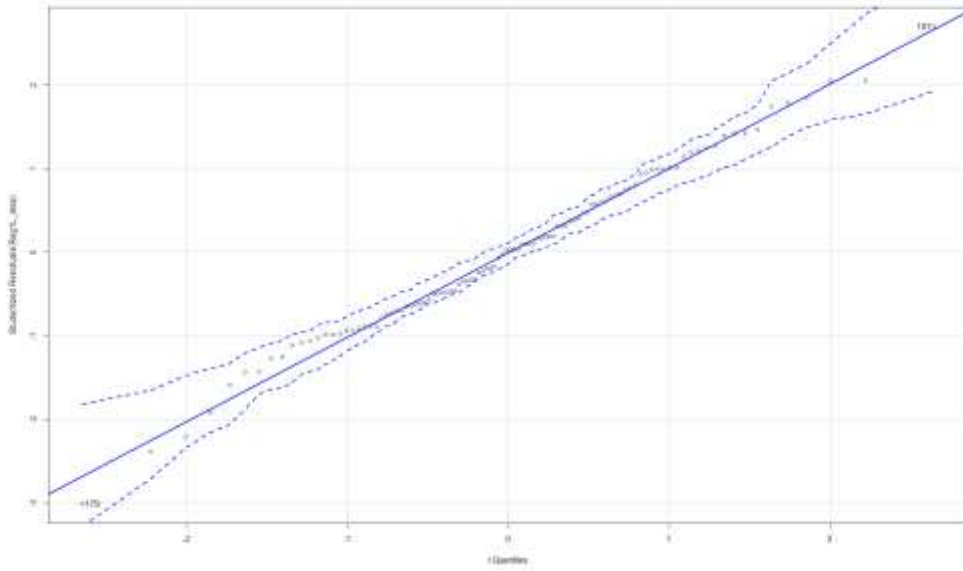


Figure 25: Normality of the model Reg1L_step

5.3.2.2 Outliers Analysis

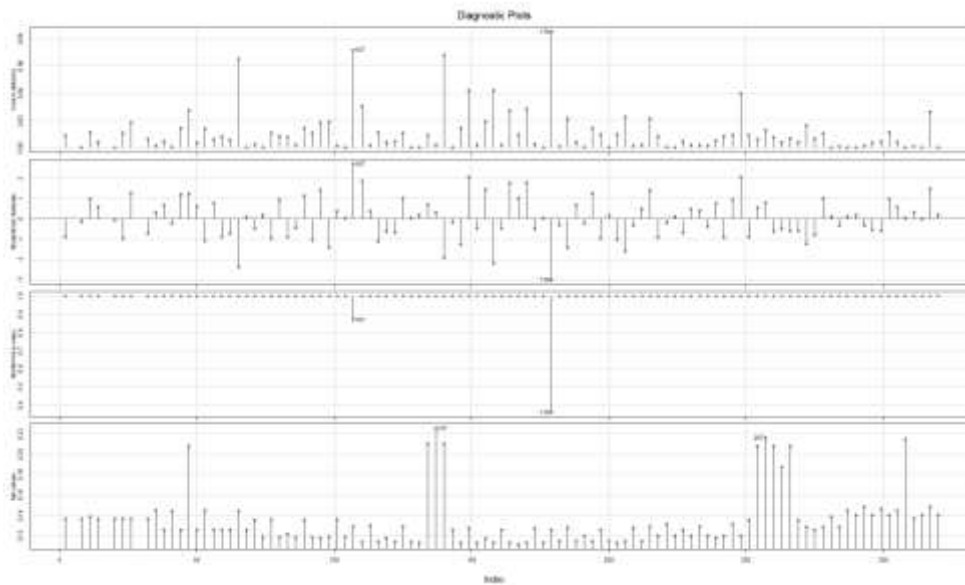


Figure 26: Influence measures

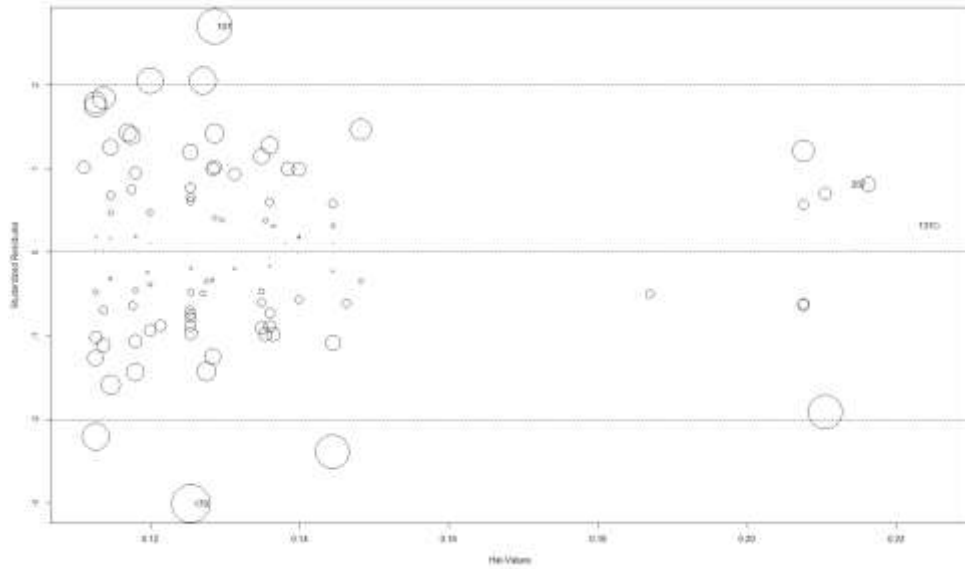


Figure 27: Influence measures

Regarding the severe outliers, suggested by the Bonferroni test, that were influential, according to Figures 26 and 27, we estimated the model without them and the results did not change.

In summary, the only external factor, other than the year and month of the movement, which influences the movements of the liquid bulk cargo in Leixões is the existence or absence of strikes.

5.3.3 Others cargo

As for the tonnage moved in Leixões from “others” cargo, i.e., dry bulk cargo, break bulk cargo and Ro-Ro cargo, the chosen model is shown in the following table:

Table 8: Regression model Reg1.1O

Reg1.1O					
	Estimate	Std. error	t	P-value	
Intercept	-5.590e+07	4.838e+06	-11.556	0.00	***
Year	2.760e+04	2.390e+03	11.546	0.00	***
Quarterly GDP	1.508e+01	5.282e+00	2.855	0.00521	**
R ²	57.28%				

Note: significance level at 0.1% (***), 1% (**), 5% (*) and 10%(.)

According to the table, we can state that:

- This model explains the movement of “others” cargo tonnage in, approximately, 57%;
- There is a growth trend over the years;
- When the Quarterly GDP increases in 1 million euros, the “others” cargo tonnage moved in the port of Leixões increases 15 tons.

5.3.3.1 Assumptions

- **Multicollinearity**

As the multicollinearity was tested in previous models, the same was done in this model. Since the value of the GVIF was not high, we concluded that there was no multicollinearity.

- **Heteroscedasticity**

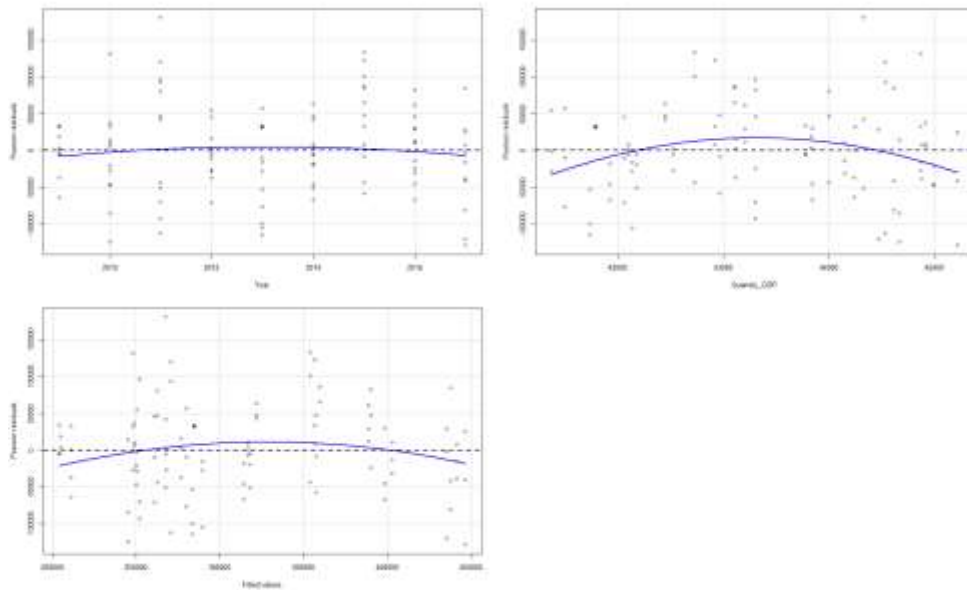


Figure 28: Residual Plots

The graphs of the Figure 28 did not suggest the presence of heteroscedasticity. Likewise, the formal test of Breusch-Pagan suggested the absence of heteroscedasticity.

- **Autocorrelation**

Regarding the independence of the disturbances over time, autocorrelation, both graphically (Figure 29) and through the formal Durbin-Watson test, the absence of autocorrelation was confirmed.

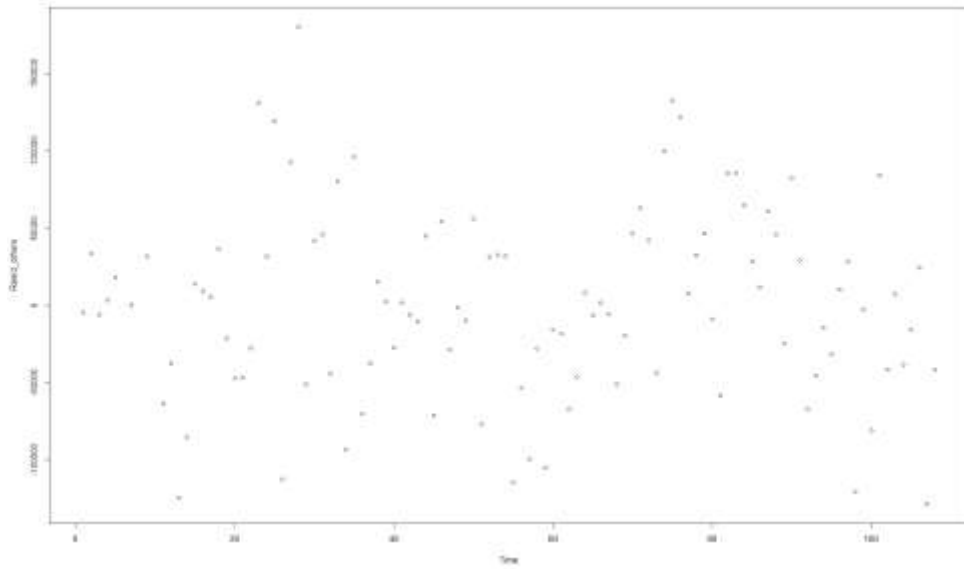


Figure 29: Residuals

- **Normality**

Regarding the distribution followed by the residuals, it is clear by the graph in Figure 30 that the distribution is not normal, as the tails are heavy, and the dots do not follow the line.

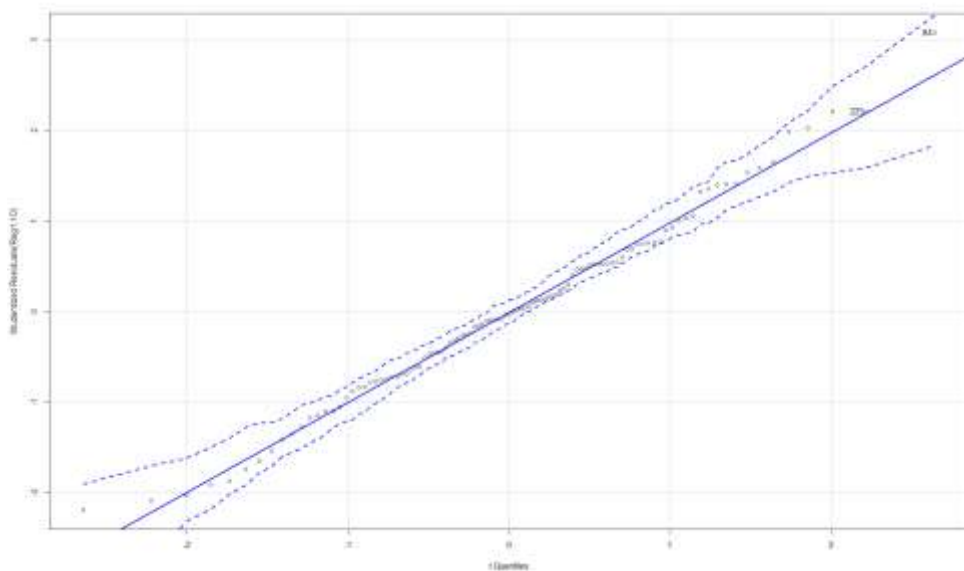


Figure 30: Normality of the model Reg1.10

However, since the sample is large, we continued the analysis in the absence of normality, relying on the asymptotic arguments.

5.3.3.2 Outliers Analysis

In order to understand if there were atypical observations that were influencing the estimation of this model, we performed the Bonferroni test, which verified the existence of severe outliers. Then, we estimated the model without the influential outliers, identified through the graphs of the Figures 31 and 32, and the results did not change.

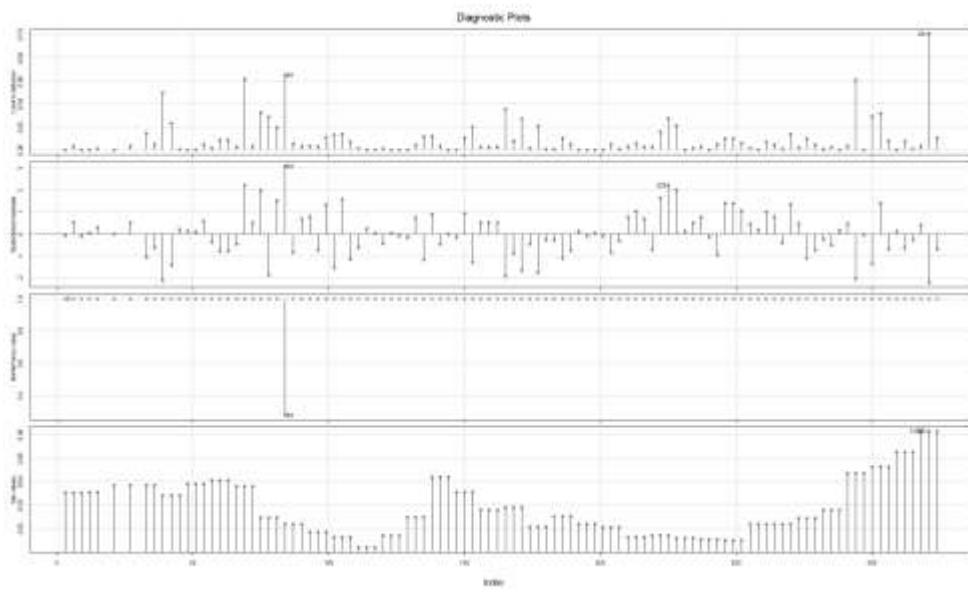


Figure 31: Influence measures

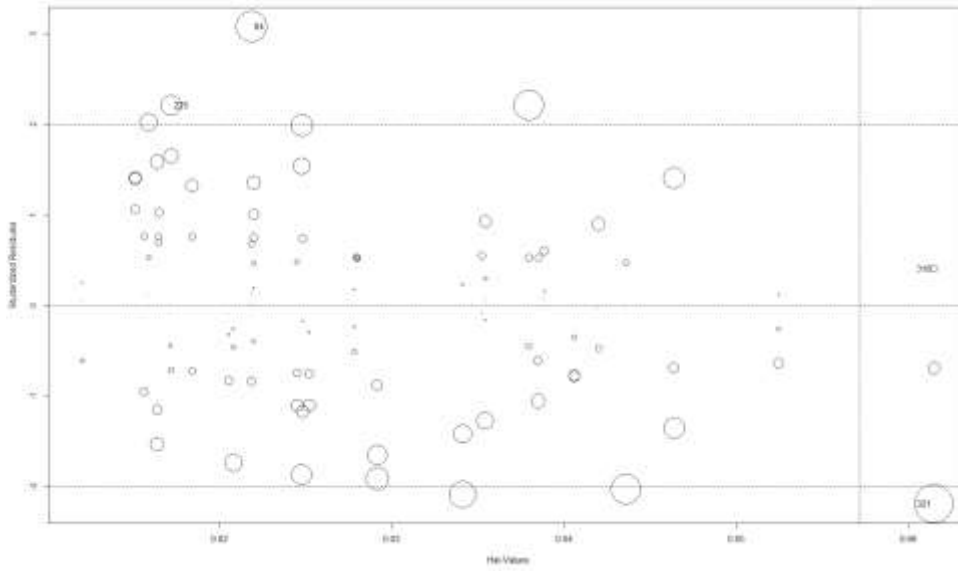


Figure 32: Influence measures

In summary, the main determinants of the “others” cargo, which include Ro-Ro cargo, dry bulk cargo and break-bulk cargo, were year of the movement and quarterly GDP.

5.3.4 Comparing the various models

Table 9: Regression models

	Containerized cargo	Liquid bulk cargo	"Others" cargo
Year	+++	+++	+++
Month 2 (ref:January)	not significant	not significant	not significant
Month 3 (ref:January)	++	+	not significant
Month 4 (ref:January)	not significant	+	not significant
Month 5 (ref:January)	++	not significant	not significant
Month 6 (ref:January)	++	not significant	not significant
Month 7 (ref:January)	+++	not significant	not significant
Month 8 (ref:January)	+	+	not significant
Month 9 (ref:January)	not significant	not significant	not significant
Month 10 (ref:January)	+++	not significant	not significant
Month 11(ref:January)	+++	not significant	not significant
Month 12 (ref:January)	++	not significant	not significant
Lisbon Tonnage	-	not significant	not significant
Sines Tonnage	not significant	not significant	not significant
Aveiro Tonnage	--	not significant	not significant
Quarterly GDP	---	not significant	++
Strikes (ref: No)	++	---	not significant
R ²	78.74%	36.01%	57.28%

Note: affects positively (+) and affects negatively (-); significance level at 0.1% (+++), 1% (++) and 5% (+); significance level at 0.1% (---), 1% (--) and 5% (-)

Table 9 shows that the percentage in which each model explains the cargo movements in the port of Leixões, as well as the explanatory variables that are statistically significant to explain them, vary according to the type of cargo.

Thus, according to the table, we have drawn the following conclusions:

- Comparing the values of the coefficients of determination (R²) of each model, we can state that there are types of cargo whose movements are better explained: the tonnage of containerized cargo handled in Leixões is explained through the greatest percentage, 78.74%, followed by the "others" cargo, 57.28%, and liquid bulk cargo, 36.01%.
- There is a trend of growth of the tonnage moved, over the years, for the three types of cargo, in the port of Leixões.
- There are statistically significant effects in most of the months for the containerized cargo, which shows the presence of some seasonality,

namely, it is expected more cargo movements, in most months, compared to the reference month, January.

- As for the liquid bulk cargo, although there are no statistically significant effects in most of the months, March, April and August are the months when, typically, more cargo movements are expected, compared to January.
- The fact that the tonnage of containerized cargo handled in the port of Lisbon is one of the main determinants of containerized cargo in Leixões, specifically having a negative correlation, is in line with what APDL thinks, since this port was identified as one of the main competitors in the containerized cargo level. The movements of the liquid bulk cargo and “others” cargo in Lisbon were not significant to explain the respective movements in Leixões.
- One aspect that caught our attention is the fact that one of the main determinants of the movements of containerized cargo identified is the Aveiro containerized cargo tonnage. This does not make sense and was unexpected, since the port of Aveiro does not move containerized cargo, having moved, throughout all the time series, only 2090 tons, thus this movements were sporadic.
- The existence of strikes having a positive effect on the movement of containerized cargo shows that the strikes lead to a larger diversion of containers to the port of Leixões, rather than in the opposite direction.
- The existence or absence of strikes is the only external factor, other than the year and month of the movement, which influences the movements of the liquid bulk cargo in Leixões. This is justified by the fact that the movement of liquid bulk in the port of Leixões is quite predictable and stable, according to the company. Therefore, it is only affected when there are mandatory stoppages of the port system, such

as maintenances, or strikes. In fact, whether the price of the raw material is high or not, oil tankers continue to reach the port of Leixões, and refineries continue to work.

- The GDP has a negative impact on the movement of containerized cargo, which is counter-intuitive. This may be due to the fact that GDP has increased as a result of the rise in the exports of services (i.e., tourism), exports of refined liquid bulk and consumption. This increase of consumption may be satisfied by other modes of transport, such as trucks, causing a diversion from containerized cargo to other modes of transport which, although more costly, may have advantages by speed.
- GDP affects positively the movements of "others" cargo in Leixões. This may be due to the investment that Leixões has made in the Ro-Ro cargo as a way to capture traffic from the roads. Therefore, there may be an increase in exports by sea taken from the road. In addition, the crisis of 2008 led the northern business community of the metal-mechanic industry to reconvert and focus on exportation, particularly of break bulk cargo, through Leixões. Therefore, this increase in GDP by the increase in exports is then positively correlated with the increase of "others" cargo.
- The port of Leixões imports the oil to be refined and, subsequently, exports the already refined product. Thus, it makes sense that GDP is not a determining factor for liquid bulk cargo, given that the majority of this type of cargo is not destined for the Portuguese market.

Chapter 6

Conclusions

We hope that this study has helped to understand the main external determinants of the tonnage of cargo handled in the port of Leixões.

Although there is autocorrelation in the first model we have estimated, which is the one that seeks to perceive the main determinants of the cargo movements in port of Leixões, the results of the three models we have estimated for each cargo type are valid.

According to the results of the models we estimated to explain the movements by type of cargo, we can draw the following conclusions:

- There is a growth tendency for all types of cargo, over the years, moved in the port of Leixões.
- In both containerized and liquid bulk cargo, the month of the movement is a statistically significant determinant. In both models, movements are likely to be higher in most months, when compared to January.
- Concerning the tons moved in other ports that were considered main determinants of the movements, these were consistent with what APDL identified as its competition, namely: containerized cargo tonnage moved in Lisbon influences the movements of the same type of cargo in Leixões.
- Strikes were also one of the significant variables to explain containerized cargo and liquid bulk cargo movements. Thus, the existence of strikes leads to a significant deviation of cargo between ports. On the one hand, at the level of containerized cargo, strikes lead to a greater diversion of cargoes to Leixões. For instance, in May

2016, more than 10000 containers were diverted to Leixões due to strikes in the port of Lisbon. On the other hand, at the level of liquid bulk, the existence of strikes is the determinant that negatively affects its movement, since the movement of this type of cargo is very constant, being only affected by mandatory stoppages, such as maintenances, and strikes.

- The hypothesis inherent to the inclusion of GDP as an external determinant of the cargo movements in the port of Leixões is that APDL expected that GDP would be reflected in the activity of the port of Leixões. On the one hand, the increase in GDP can lead to an increase in the income and consumption, which can be reflected in company's preference for a faster mode of transport than vessels, such as road transport, even though they lead to a higher cost, thus reducing the movement of cargo in Leixões. On the other hand, the increase in GDP may be linked to an increase in the exports, which is reflected in a greater movement of cargo in Leixões.

Additionally, we considered that the port of Leixões has an installed capacity and is, therefore, limited by it.

There were three main limitations to this study: firstly, limited access to the company and its data; secondly, the existence of very little literature in the context of ports that focused on external factors; thirdly, limited time to develop this study due to the change of the tutor in the company that led to a change in the research question.

A future goal will be to conduct a more in-depth study, including new variables, perhaps internal port variables, which were not considered. In addition, a future objective is to build a dashboard that can serve as an analytical tool for forecasting, based on the models we found, so that the company can predict cargo movements, taking into account different scenarios.

Furthermore, using the same data we had access to in this study, we hope to solve the autocorrelation problem that was detected in the first regression model. One possible solution would be to develop a study in which the dependent variable would be the difference between the Leixões tonnage moved in one current month and the previous month, instead of being the value of total tonnage moved in that month. Another possibility would be to estimate this model by feasible generalised least squares (Wooldridge, 2002).

References

- Acito, F. & Khatri, V. (2014). Business analytics: Why now and what next? *Business Horizons*, 57 (5), 565–570.
- Afifi, A., May, S. & Clark, V.A. (2011). *Practical Multivariate Analysis*. 5th Edition. London: Chapman & Hall. 5th Edition.
- Andrews, D. (1991). Heteroskedasticity and Autocorrelation Consistent Covariant Matrix Estimation. *Econometrica*, 59 (3), 817-858.
- Arellano, M. (1987). Practitioners'corner: Computing Robust Standard Errors for Within-groups Estimators. *Oxford bulletin of Economics and Statistics*, 49 (4), 431-434.
- Becker, A., Inoue, S., Fischer, M. & Schwelger, B. (2011). Climate change impacts on international seaports: knowledge, perceptions, and planning efforts among port administrators. *Climatic change*, 110 (1-2), 5-29.
- Blonigen, B. A. & Wilson, W. W. (2008). Port efficiency and trade flows. *Review of International Economics*, 16 (1), 21-36.
- Carvalho, A.; Coutinho, A.; Britto, P. & Lustosa, P. (2015). *2091 Texto para Discussão. Caracterização do fluxo de cargas e indicadores de concorrência entre portos brasileiros*. Brasília: Instituto de Pesquisa Econômica Aplicada.
- Clark, X., Dollar, D. & Micco, A. (2004). Port efficiency, maritime transport costs, and bilateral trade. *Journal of development economics*, 75 (2), 417-450.
- Committee on Climate Change, US Transportation, Transportation Research Board, Division on Earth, & Life Studies. (2008). *Potential Impacts of Climate Change on US Transportation: Special Report 290*. Transportation Research Board.
- Constituição da República Portuguesa, Decreto de Aprovação da Constituição, Diário da República n.º 86/1976, Série I de 1976-04-10, Artigo 57.º - (Direito à greve e proibição do lock-out)
- Davenport, T. H. (2013). Managing yourself. Keep up with your quants. An innumerate's guide to navigating big data. *Harvard Business Review, Influence, How to get, How to use*, July-August: 120-123.
- Eicker, F. (1963). Asymptotic normality and consistency of the least squares estimators for families of linear regressions. *The Annals of Mathematical Statistics*, 34 (2), 447-456.
- Faraway, J. J. (2009). *Linear Models with R*. Oxfordshire: Taylor & Francis.
- Fox, J. & Monette, G. (1992) Generalized collinearity diagnostics. *Journal of the American Statistical Association*, 417 (87), 178–183.

- Gandomi, A. & Haider, M. (2015). Beyond the hype: Big data concepts, methods, and analytics. *International Journal of Information Management*, 35 (2), 137-144.
- Langen, P. W. (2007). Port competition and selection in contestable hinterlands; the case of Austria. *European Journal of Transport and Infrastructure Research*, 7 (1), 1-14.
- Meersman, H., Van de Voorde, E. & Vanellander, T. (2010). Port Competition Revisited. *Journal of Pediatric, Maternal & Family Health-Chiropractic*, 55 (2), 210.
- OECD (2013). Exploring Data-Driven Innovation as a New Source of Growth. Mapping the Policy Issues Raised by "Big Data", *Organization for Economic Co-operation and Development*, 1-43.
- Oliveira, M. M., Santos, L. D. & Fortuna, N. (2011). *Econometria*. Lisboa: Escolar Editora.
- Sánchez, R. J., Hoffmann, J., Micco, A., Pizzolitto, G. V., Sgut, M. & Wilmsmeier, G. (2003). Port efficiency and international trade: port efficiency as a determinant of maritime transport costs. *Maritime economics & logistics*, 5 (2), 199-218.
- Savonis, M., Burkett, V. R., & Potter, J. R. (Coordinating Lead Authors) (2008). *Impacts of climate change and variability on transportation systems and infrastructure: Gulf Coast study, phase I*. Report by the U.S. Climate Change Science Program and the Subcommittee on Global Change Research.
- Sivarajah, U., Kamal, M. M., Irani, Z. & Weerakkody, V. (2017). Critical analysis of Big Data challenges and analytical methods. *Journal of Business Research*, 70, 263-286.
- Soibelman, L. & Kim, H. (2002). Data Preparation Process for Constructing Knowledge Generation through Knowledge Discovery in Databases. *Journal of Computing in Civil Engineering*, 16 (1), 39-48.
- UNCTAD (1993). *Strategic Planning for Port Authorities*. Report by UNCTAD. Geneva: United Nations Conference on Trade and Development
- UNCTAD (2017). *Evolution of the international trading system and its trends from a development perspective*. Report by UNCTAD. Geneva: United Nations Conference on Trade and Development.
- Weisberg, S. (2005). *Applied linear regression*. New Jersey: John Wiley & Sons.
- Wooldridge, J. M. (2002) *Econometric analysis of cross section and panel data*. Massachusetts: MIT press.

Appendices

Appendix 1: R script

```
##### DATA READING #####

data_apdl <- read.csv("DADOS FINAIS .csv",header = TRUE, sep = ";", dec = ",")
data_apdl

matrix(data_apdl)

summary(data_apdl)

##### PRE - PROCESSING #####

#CLASSIFICATION OF THE VARIABELS?
sapply(data_apdl, class)
#ALL NUMERIC, EXCEPT CARGO_TYPE AS FACTOR

#CATEGORICAL AS FACTORS:
data_apdl$Strikes <- as.factor(data_apdl$Strikes)
data_apdl$Seasonality <- as.factor(data_apdl$Seasonality)
data_apdl$Month_Number <- as.factor(data_apdl$Month_Number)

#ERASE THE OBS WITH NEGATIVE VALUES
neg <- data_apdl[,5:12]<0
summary(neg)

data_apdl <- data_apdl[ which(data_apdl$Tonnage_LISBON>=0),]
data_apdl <- data_apdl[ which(data_apdl$Tonnage._SINES>=0),]
data_apdl <- data_apdl[ which(data_apdl$Tonnage.SETUBAL>=0),]
data_apdl <- data_apdl[ which(data_apdl$Tonnage_AVEIRO>=0),]
data_apdl <- data_apdl[ which(data_apdl$Tonnage_VIGO>=0),]
data_apdl <- data_apdl[ which(data_apdl$Tonnage_CORUNHA>=0),]

##### DATA PREPARATION #####

#1# DATA CLEANING:

#verify the existence of missing values
#what is the number of unknown values in each row of the data_apdl dataset?

apply(data_apdl,1,function(x)sum(is.na(x)))
```

```

#there are no missing values

#verify the existence of outliers:

#using classical statistical method, BOXPLOTS:

#Tonnage Leixoes:
boxplot(data_apdl$Tonnage_LEIXOES)
outliers_Leixoes=boxplot(data_apdl$Tonnage_LEIXOES,plot = FALSE)$out
outliers_Leixoes

#Tonnage Lisbon:
boxplot(data_apdl$Tonnage_LISBON)
outliers_Lisbon=boxplot(data_apdl$Tonnage_LISBON,plot = FALSE)$out
outliers_Lisbon

#Tonnage Sines:
boxplot(data_apdl$Tonnage._SINES)
outliers_Sines=boxplot(data_apdl$Tonnage._SINES,plot = FALSE)$out
outliers_Sines

#Tonnage Setubal:
boxplot(data_apdl$Tonnage.SETUBAL)
outliers_Setubal=boxplot(data_apdl$Tonnage.SETUBAL,plot = FALSE)$out
outliers_Setubal

#Tonnage Aveiro
boxplot(data_apdl$Tonnage_AVEIRO)
outliers_Aveiro=boxplot(data_apdl$Tonnage_AVEIRO,plot = FALSE)$out
outliers_Aveiro

#Tonnage Vigo:
boxplot(data_apdl$Tonnage_VIGO)
outliers_Vigo=boxplot(data_apdl$Tonnage_VIGO,plot = FALSE)$out
outliers_Vigo

#Tonnage Corunha:
boxplot(data_apdl$Tonnage_CORUNHA)
outliers_Corunha=boxplot(data_apdl$Tonnage_CORUNHA,plot = FALSE)$out
outliers_Corunha

#GDP:
boxplot(data_apdl$Quartely_GDP)
outliers_GDP=boxplot(data_apdl$Quartely_GDP,plot = FALSE)$out
outliers_GDP

```

```

### O R ###
# DISPERSION MESURES

sapply(data_apdl[,5:12],sd,2)
CV <- function(x)
{
  sd(x)/mean(x)
}

apply(data_apdl[,c(5:12)],2,CV)

AmpIQ <- function(x)
{
  quantile(x,0.75) - quantile(x,0.25)
}

apply(data_apdl[,c(5:12)],2,AmpIQ)

#BOXPLOTS OF EVERY TONNAGE AND GDP
boxplot(data_apdl[,c(1,2,3,4,13,14)])
boxplot(data_apdl[,c(1,2,3,4,8,10,11,13,14)])

#3# DATA TRANSFORMATION:

#CREATION OF A NEW FEATURE FROM THE GIVEN ONES: "Time"
#It was done prior to the data entry in Excel

#4# DATA REDUCTION:

#CREATION OF SUBSETS FOR EACH TYPE OF CARGO:

#CONTAINERS
Containers <- data_apdl[ which(data_apdl$Cargo_Type=='Containers'),]

#LIQUIDBULK
Liquidbulk <- data_apdl[ which(data_apdl$Cargo_Type=='Liquidbulk'),]

#Others
Others <- data_apdl[ which(data_apdl$Cargo_Type=='Others'),]

#### DESCRIPTIVE STATISTICS ####

#SUMMARY()

str(data_apdl)
summary(data_apdl)

```

```

summary(data_apdl$Year)
summary(data_apdl$Time)

summary(data_apdl$Month_Number)
summary(data_apdl$Cargo_Type)
summary(data_apdl$Strikes)
summary(data_apdl$Seasonality)

summary(data_apdl$Tonnage_LEIXOES)
sd(data_apdl$Tonnage_LEIXOES)
summary(data_apdl$Tonnage_LISBON)
sd(data_apdl$Tonnage_LISBON)
summary(data_apdl$Tonnage._SINES)
sd(data_apdl$Tonnage._SINES)
summary(data_apdl$Tonnage.SETUBAL)
sd(data_apdl$Tonnage.SETUBAL)
summary(data_apdl$Tonnage_AVEIRO)
sd(data_apdl$Tonnage_AVEIRO)
summary(data_apdl$Tonnage_VIGO)
sd(data_apdl$Tonnage_VIGO)
summary(data_apdl$Tonnage_CORUNHA)
sd(data_apdl$Tonnage_CORUNHA)

summary(data_apdl$Quartely_GDP)
sd(data_apdl$Quartely_GDP)

summary(Containers)

summary(Liquidbulk)

summary(Others)

sum(Containers$Tonnage_AVEIRO)

### HISTOGRAMS FOR CARGO MOVEMENTS IN LEIXOES ###

#IN ALL THE TIME SERIES:
hist(data_apdl$Tonnage_LEIXOES, xlab = "Tonnage_Leixoes", ylab = "Frequency")

#PER YEAR:
#2009:
Histogram_LeixOes_2009 <- hist(data_apdl$Tonnage_LEIXOES[1:36], xlab =
"Tonnage_Leixoes09", ylab = "Frequency")
#2010
Histogram_LeixOes_2010 <- hist(data_apdl$Tonnage_LEIXOES[37:72], xlab =
"Tonnage_Leixoes10", ylab = "Frequency")
#2011

```

```

Histogram_LeixOes_2011 <- hist(data_apdl$Tonnage_LEIXOES[73:108], xlab =
"Tonnage_Leixoes11", ylab = "Frequency")
#2012
Histogram_LeixOes_2012 <- hist(data_apdl$Tonnage_LEIXOES[109:144], xlab =
"Tonnage_Leixoes12", ylab = "Frequency")
#2013
Histogram_LeixOes_2013 <- hist(data_apdl$Tonnage_LEIXOES[145:180], xlab =
"Tonnage_Leixoes13", ylab = "Frequency")
#2014
Histogram_LeixOes_2014 <- hist(data_apdl$Tonnage_LEIXOES[181:216], xlab =
"Tonnage_Leixoes14", ylab = "Frequency")
#2015
Histogram_LeixOes_2015 <- hist(data_apdl$Tonnage_LEIXOES[217:252], xlab =
"Tonnage_Leixoes15", ylab = "Frequency")
#2016
Histogram_LeixOes_2016 <- hist(data_apdl$Tonnage_LEIXOES[253:288], xlab =
"Tonnage_Leixoes16", ylab = "Frequency")
#2017
Histogram_LeixOes_2017 <- hist(data_apdl$Tonnage_LEIXOES[289:324], xlab =
"Tonnage_Leixoes17", ylab = "Frequency")

#PER TYPE OF CARGO:
#CONTAINERS
hist(Containers$Tonnage_LEIXOES)

#LIQUIDBULK
hist(Liquidbulk$Tonnage_LEIXOES)

#Others
hist(Others$Tonnage_LEIXOES)

# # # RELATION BETWEEN VARIABLES # # #

# DEPENDENT VARIABLE AND EXPLANATORY VARIABLES:

help("scatterplot")
library(car)

scatterplot.matrix(data_apdl[,c(5,1:4,6:7)])
scatterplot.matrix(data_apdl[,c(5,8:14)])

scatterplot.matrix(data_apdl[,c(5,1:4,12:14)], diagonal="boxplot")
scatterplot.matrix(data_apdl[,c(5,6:11)], diagonal="boxplot")

scatterplot.matrix(data_apdl[,c(5, 1:4, 6:14)], diagonal="boxplot")

scatterplot.matrix(data_apdl[,c(5, 1:4, 6:14)])

```

```

scatterplot(data_apdl$Tonnage_LEIXOES,data_apdl$Year)
scatterplot(data_apdl$Tonnage_LEIXOES,data_apdl$Time)
scatterplot(data_apdl$Tonnage_LEIXOES,data_apdl$Month_Number)

scatterplot(data_apdl$Tonnage_LEIXOES,data_apdl$Cargo_Type)

scatterplot(data_apdl$Tonnage_LEIXOES,data_apdl$Tonnage_LISBON)
scatterplot(data_apdl$Tonnage_LEIXOES,data_apdl$Tonnage._SINES)
scatterplot(data_apdl$Tonnage_LEIXOES,data_apdl$Tonnage.SETUBAL)
scatterplot(data_apdl$Tonnage_LEIXOES,data_apdl$Tonnage_AVEIRO)
scatterplot(data_apdl$Tonnage_LEIXOES,data_apdl$Tonnage_VIGO)
scatterplot(data_apdl$Tonnage_LEIXOES,data_apdl$Tonnage_CORUNHA)

scatterplot(data_apdl$Tonnage_LEIXOES,data_apdl$Quartely_GDP)

scatterplot(data_apdl$Tonnage_LEIXOES,data_apdl$Strikes)

scatterplot(data_apdl$Tonnage_LEIXOES,data_apdl$Seasonality)

library(ggplot2)
ggplot(data_apdl, aes(x=Tonnage_LEIXOES, y=Tonnage._SINES, )) +
geom_point(alpha=0.6)
ggplot(data_apdl, aes(x=Tonnage_LEIXOES, y=Time )) + geom_point(alpha=0.6)

mean(data_apdl$Tonnage_LEIXOES[data_apdl$Year == 2009])
mean(data_apdl$Tonnage_LEIXOES[data_apdl$Year == 2010])
mean(data_apdl$Tonnage_LEIXOES[data_apdl$Year == 2011])
mean(data_apdl$Tonnage_LEIXOES[data_apdl$Year == 2012])
mean(data_apdl$Tonnage_LEIXOES[data_apdl$Year == 2013])
mean(data_apdl$Tonnage_LEIXOES[data_apdl$Year == 2014])
mean(data_apdl$Tonnage_LEIXOES[data_apdl$Year == 2015])
mean(data_apdl$Tonnage_LEIXOES[data_apdl$Year == 2016])
mean(data_apdl$Tonnage_LEIXOES[data_apdl$Year == 2017])

mean(data_apdl$Tonnage_LISBON[data_apdl$Year == 2009])
mean(data_apdl$Tonnage_LISBON[data_apdl$Year == 2010])
mean(data_apdl$Tonnage_LISBON[data_apdl$Year == 2011])
mean(data_apdl$Tonnage_LISBON[data_apdl$Year == 2012])
mean(data_apdl$Tonnage_LISBON[data_apdl$Year == 2013])
mean(data_apdl$Tonnage_LISBON[data_apdl$Year == 2014])
mean(data_apdl$Tonnage_LISBON[data_apdl$Year == 2015])
mean(data_apdl$Tonnage_LISBON[data_apdl$Year == 2016])
mean(data_apdl$Tonnage_LISBON[data_apdl$Year == 2017])

mean(data_apdl$Tonnage._SINES[data_apdl$Year == 2009])
mean(data_apdl$Tonnage._SINES[data_apdl$Year == 2010])

```

```

mean(data_apdl$Tonnage._SINES[data_apdl$Year == 2011])
mean(data_apdl$Tonnage._SINES[data_apdl$Year == 2012])
mean(data_apdl$Tonnage._SINES[data_apdl$Year == 2013])
mean(data_apdl$Tonnage._SINES[data_apdl$Year == 2014])
mean(data_apdl$Tonnage._SINES[data_apdl$Year == 2015])
mean(data_apdl$Tonnage._SINES[data_apdl$Year == 2016])
mean(data_apdl$Tonnage._SINES[data_apdl$Year == 2017])

```

```

mean(data_apdl$Tonnage_AVEIRO[data_apdl$Year == 2009])
mean(data_apdl$Tonnage_AVEIRO[data_apdl$Year == 2010])
mean(data_apdl$Tonnage_AVEIRO[data_apdl$Year == 2011])
mean(data_apdl$Tonnage_AVEIRO[data_apdl$Year == 2012])
mean(data_apdl$Tonnage_AVEIRO[data_apdl$Year == 2013])
mean(data_apdl$Tonnage_AVEIRO[data_apdl$Year == 2014])
mean(data_apdl$Tonnage_AVEIRO[data_apdl$Year == 2015])
mean(data_apdl$Tonnage_AVEIRO[data_apdl$Year == 2016])
mean(data_apdl$Tonnage_AVEIRO[data_apdl$Year == 2017])

```

```

mean(data_apdl$Quartely_GDP[data_apdl$Year == 2009])
mean(data_apdl$Quartely_GDP[data_apdl$Year == 2010])
mean(data_apdl$Quartely_GDP[data_apdl$Year == 2011])
mean(data_apdl$Quartely_GDP[data_apdl$Year == 2012])
mean(data_apdl$Quartely_GDP[data_apdl$Year == 2013])
mean(data_apdl$Quartely_GDP[data_apdl$Year == 2014])
mean(data_apdl$Quartely_GDP[data_apdl$Year == 2015])
mean(data_apdl$Quartely_GDP[data_apdl$Year == 2016])
mean(data_apdl$Quartely_GDP[data_apdl$Year == 2017])

```

SELECTION OF THE REGRESSION MODEL :

#without Time and Seasonality

```

Reg1 <- lm(Tonnage_LEIXOES ~ .,data=data_apdl[,-c(2,14)])
summary(Reg1)

```

#without Time and Month_number

```

Reg2 <- lm(Tonnage_LEIXOES ~ .,data=data_apdl[,-c(2,3)])
summary(Reg2)

```

#without Month_number and Seasonality

```

Reg3 <- lm(Tonnage_LEIXOES ~ .,data=data_apdl[,-c(3,14)])
summary(Reg3)

```

SIMPLER MODELS :

#from Reg1

```

Reg1.1 <- lm(Tonnage_LEIXOES ~ .,data=data_apdl[,c(1,3,4,5,13)])
summary(Reg1.1)

```

```

#from Reg2
Reg2.1 <- lm(Tonnage_LEIXOES ~ ., data = data_apdl[,c(1,4,5, 8, 9,13)])
summary(Reg2.1)
#reducing Reg2.1 even more, excluding the variables that were not significant:
Reg2.2 <- lm(Tonnage_LEIXOES ~ ., data = data_apdl[,c(1,4,5, 13)])
summary(Reg2.2)

AIC(Reg1)
AIC(Reg2)
AIC(Reg3)
AIC(Reg1.1)
AIC(Reg2.1)
AIC(Reg2.2)

# STEP #

Reg1_step <- step(Reg1)
summary(Reg1_step)

Reg2_step <- step(Reg2)
summary(Reg2_step)

Reg3_step <- step(Reg3)
summary(Reg3_step)

#using 5% criteria --> estimating the model:
RegA <- lm(Tonnage_LEIXOES ~ .,data=data_apdl[,c(2,6,8,9,10,11,12,14)])
summary(RegA)

anova(Reg1,RegA)

##### INTERACTIONS : #####
#centre variables
data_apdl$Tonnage_SINES2 <- (data_apdl$Tonnage._SINES-
mean(data_apdl$Tonnage._SINES))

#only interactions that make sense -> tonnage_sines*greves can be interesting
RegAI <- lm(Tonnage_LEIXOES ~ .+Tonnage_SINES2*Strikes,data=data_apdl[,c(2,6,7,8,9,10,11,12,14)])

#CHOSEN MODEL:
summary(RegAI)

##### MULTICOLINEARITY #####
cor(data_apdl[,c(5,7,8,9,10, 11, 12)])
cor(data_apdl[,c(5,15)])

```

```

library(carData)
library(car)
?vif
vif(RegAI)

### HETEROSKEDASTICITY ###

#residual plots
resid <- residuals(RegAI)
resid
residualPlots(RegAI)

# Breush-Pagan test:
#To formally test the evidence of heteroskedasticity:
library(zoo)
library(lmtest)
bptest(RegAI)

#Comparing error with robust
library(lmtest)
library(sandwich)
RegA$robVCOV <- vcovHC(RegAI)
coefTest(RegAI,RegAI$robVCOV)

### AUTOCORRELATION ###

library(plm)
data_apdl_painel <- plm.data(data_apdl, index = c("Cargo_Type", "Time"))
painel_reg <- plm(Tonnage_LEIXOES ~
+Tonnage_SINES2*Strikes,data=data_apdl_painel, model = "within")
pbgttest(painel_reg, order = 2)
#there is autocorrelation

### NORMALITY ###

library(car)
qqPlot(RegAI)

#### OUTLIERS ####

outlierTest(RegAI)

boxplot(data_apdl)

influenceIndexPlot(RegAI)

```

```

influencePlot(RegAI)

#model without outliers#
data_without_outliers <- data_apdl[-c(107,140,265),]
RegAI_out <- lm(Tonnage_LEIXOES ~ .+Tonnage_SINES2*Strikes,data =
data_without_outliers[-c(2,6,7,8,9,10,11,12,14)])
summary(RegAI_out)

# # # # # ANALISYS BY TYPE OF CARGO # # # # #

#####CONTAINERS#####

#without Time and Seasonality
Reg1C <- lm(Tonnage_LEIXOES ~ .,data = Containers[-c(2,4,14,15)])
summary(Reg1C)

#without Time and Month_number
Reg2C <- lm(Tonnage_LEIXOES ~ .,data=data_apdl[-c(2,4,3,15)])
summary(Reg2C)

#without Month_number and Seasonality
Reg3C <- lm(Tonnage_LEIXOES ~ .,data=data_apdl[-c(3,4,14,15)])
summary(Reg3C)

# SIMPLER MODELS :
#from Reg1C
Reg1.1C <- lm(Tonnage_LEIXOES ~ .,data = Containers[-c(2,4,14,1,7,8,10,11,15)])
summary(Reg1.1C)
Reg1.2C <- lm(Tonnage_LEIXOES ~ .,data = Containers[-c(2,4,14,1,7,8,10,11,9,15)])
summary(Reg1.2C)

# STEP #

Reg1C_step <- step(Reg1C)
summary(Reg1C_step)

Reg2C_step <- step(Reg2C)
summary(Reg2C_step)

Reg3C_step <- step(Reg3C)
summary(Reg3C_step)

#chosen model: Reg1C_step (r2 >, all variables are statistically significant)

```

```

anova(Reg1C,Reg1C_step)

### MULTICOLINEARITY ###
cor(data_apdl[,c(5,6,9,12)])

library(carData)
library(car)
vif(Reg1C_step)

### HETEROSKEDASTICITY ###

#residual plots
resid_C <- residuals(Reg1C_step)
resid_C
residualPlots(Reg1C_step)

# Breush-Pagan test:
library(zoo)
library(lmtest)
bptest(Reg1C_step)
#the model is homoscedastic

### AUTOCORRELATION ###

plot(Containers$Time,resid_C, xlab = "Time", ylab = "Resid_containers")
durbinWatsonTest(Reg1C_step)

library(sandwich)

?vcovHAC
#vcovHAC-> heterocedastic and autocorrelated consistent
Reg1C_step$robVCOV <- vcovHAC(Reg1C_step)
#to estimate the robust standard errors
coefTest(Reg1C_step,Reg1C_step$robVCOV)

### NORMALITY ###

library(car)
qqPlot(Reg1C_step)

### OUTLIERS ###

outlierTest(Reg1C_step)

boxplot(data_apdl)

```

```

influenceIndexPlot(Reg1C_step)

influencePlot(Reg1C_step)

#model without outliers#
data_without_outliers_containers <- Containers[-c(142,247,265),]
Reg1C_step_out <- lm(Tonnage_LEIXOES ~ .,data =
data_without_outliers_containers[,c(1,3,5,6,9,12,13)])
summary(Reg1C_step_out)
summary(Reg1C_step)
#results do not change

#####LIQUID BULK#####

#without Time and Seasonality
Reg1L <- lm(Tonnage_LEIXOES ~ .,data = Liquidbulk[-c(2,4,14)])
summary(Reg1L)

#without Time and Month_number
Reg2L <- lm(Tonnage_LEIXOES ~ .,data=Liquidbulk[-c(2,4,3)])
summary(Reg2L)

#without Month_number and Seasonality
Reg3L <- lm(Tonnage_LEIXOES ~ .,data=Liquidbulk[-c(3,4,14)])
summary(Reg3L)

# SIMPLER MODELS :
#from Reg1L
Reg1.1L <- lm(Tonnage_LEIXOES ~ .,data = Liquidbulk[-c(2,4,14,1,6,7,8,9,10,11,12)])
summary(Reg1.1L)

# STEP #

Reg1L_step <- step(Reg1L)
summary(Reg1L_step)

Reg2L_step <- step(Reg2L)
summary(Reg2L_step)

Reg3L_step <- step(Reg3L)
summary(Reg3L_step)

### MULTICOLINEARITY ###
library(carData)
library(car)

```

```

vif(Reg1L_step)

### HETEROSKEDASTICITY ###
#residual plots
resid_L <- residuals(Reg1L_step)
resid_L
residualPlots(Reg1L_step)

# Breush-Pagan test:
library(zoo)
library(lmtest)
bptest(Reg1L_step)
#homoscedasticity

### AUTOCORRELATION ###
plot(Liquidbulk$Time,resid_L, xlab = "Time", ylab = "Resid_liquidbulk")
durbinWatsonTest(Reg1L_step)
#no autocorrelation

### NORMALITY ###
library(car)
qqPlot(Reg1L_step)

##### OUTLIERS #####
outlierTest(Reg1L_step)
boxplot(data_apdl)
influenceIndexPlot(Reg1L_step)
influencePlot(Reg1L_step)

#model without outliers#
data_without_outliers_liquidbulk <- Liquidbulk[-c(107,179),]
Reg1L_step_out <- lm(Tonnage_LEIXOES ~ .,data =
data_without_outliers_liquidbulk[,c(1,3,5,13)])
summary(Reg1L_step_out)

summary(Reg1L_step)

#####OTHERS#####

#without Time and Seasonality
Reg1O <- lm(Tonnage_LEIXOES ~ .,data = Others[-c(2,4,14)])
summary(Reg1O)

#without Time and Month_number
Reg2O <- lm(Tonnage_LEIXOES ~ .,data=Others[-c(2,4,3)])
summary(Reg2O)

```

```

#without Month_number and Seasonality
Reg3O <- lm(Tonnage_LEIXOES ~ .,data=Others[,-c(3,4,14)])
summary(Reg3O)

# SIMPLER MODELS :
#from Reg1O
Reg1.1O <- lm(Tonnage_LEIXOES ~ .,data = Others[,-c(2,4,14,3,6,7,8,9,10,11,13)])
summary(Reg1.1O)

# STEP #

Reg1O_step <- step(Reg1O)
summary(Reg1O_step)

Reg2O_step <- step(Reg2O)
summary(Reg2O_step)

Reg3O_step <- step(Reg3O)
summary(Reg3O_step)

### MULTICOLINEARITY ###
cor(data_apdl[,c(5,12)])

library(carData)
library(car)
vif(Reg1.1O)

### HETEROSKEDASTICITY ###

#residual plots
resid_O <- residuals(Reg1.1O)
resid_O
residualPlots(Reg1.1O)

# Breush-Pagan test:
library(zoo)
library(lmtest)
bptest(Reg1.1O)

### AUTOCORRELATION ###

plot(Others$Time,resid_O, xlab = "Time", ylab = "Resid_others")
durbinWatsonTest(Reg1.1O)
#no autocorrelation

### NORMALITY ###

```

```

library(car)
qqPlot(Reg1.1O)

### ## OUTLIERS ### ##

outlierTest(Reg1.1O)
boxplot(data_apdl)
influenceIndexPlot(Reg1.1O)
influencePlot(Reg1.1O)

#model without outliers#
data_without_outliers_others <- Others[-c(84,225,321),]
Reg1.1O_out <- lm(Tonnage_LEIXOES ~ .,data =
data_without_outliers_others[,c(1,5,12)])
summary(Reg1.1O_out)

summary(Reg1.1O)

##### PREDICTIONS #####
#Containers
predict(Reg1C_step,Containers, interval = "prediction")
pred_containers <- predict(Reg1C_step,Containers, interval = "prediction")
plot(Containers$Time[c(1:36)], pred_containers[-c(37:322)], xlab = "Time",ylab =
"Prdeictions_containers", type = "b")

pred_containers[-c(7:322)]

#Liquid bulk
predict(Reg1L_step, Liquidbulk, interval = "prediction")
pred_liquidbulk <- predict(Reg1L_step, Liquidbulk, interval = "prediction")
plot(Liquidbulk$Time[c(1:36)], pred_liquidbulk[-c(37:322)],xlab = "Time",ylab =
"Prdeictions_liquidbulk", type = "b")

pred_liquidbulk[-c(7:322)]

#Others
predict(Reg1.1O, Others, interval = "prediction")
pred_others <- predict(Reg1.1O, Others, interval = "prediction")
plot(Others$Time[c(1:36)], pred_others[-c(37:322)],xlab = "Time",ylab =
"Prdeictions_others", type = "b")

pred_others[-c(7:322)]

```