Logistics in Emerging Economies: Forecasts and Analysis for Mexico and Turkey

Bachelor's thesis in International Logistics

JOAKIM ERLANDSSON
OTTO PIHL

DEPARTMENT OF TECHNOLOGY MANAGEMENT AND ECONOMICS
DIVISION OF SERVICE MANAGEMENT AND LOGISTICS

CHALMERS UNIVERSITY OF TECHNOLOGY
Gothenburg, Sweden 2023
www.chalmers.se
Report number E2023:008

Report number E2023:08

Logistics in Emerging Economies: Forecasts and Analysis for Mexico and Turkey

JOAKIM ERLANDSSON
OTTO PIHL

DEPARTMENT OF TECHNOLOGY MANAGEMENT AND ECONOMICS
Division of Service Management and Logistics
CHALMERS UNIVERSITY OF TECHNOLOGY
Gothenburg, Sweden 2023

Logistics in Emerging Economies: Forecasts and Analysis for Mexico and Turkey

JOAKIM ERLANDSSON
OTTO PIHL

© JOAKIM ERLANDSSON, 2023
© OTTO PIHL, 2023

Report number E2023:08
Department of Technology Management and Economics
Chalmers University of Technology
SE-412 96 Gothenburg
Sweden
Telephone + 46 (0)31-772 1000

Cover: Word cloud featuring logistical terms and concepts.
(Word Cloud by *Epic Top 10*)

Gothenburg, Sweden 2023

PREFACE

We are two students from the International Logistics program at Chalmers University of Technology, in Gothenburg, Sweden, who have authored this report. The program covers various aspects of the logistics industry, including road, rail, air, and sea transportation, as well as elements of jurisprudence, economics, and sustainability. While this report mainly focuses on the logistics and economic components, it also acknowledges the significance of sustainability and jurisprudence in the analysis of logistics.

We would like to express our great gratitude towards our supervisor, Mostafa Parsa, for the invaluable lectures, guidance, dedication, and patience throughout the entire project. Without the support of Mostafa, this report would not have been possible.

We would also like to thank our examiner, Dan Andersson, for his valuable comments and guidance on the report.

Gothenburg, Sweden 2023

Logistics in Emerging Economies: Forecasts and Analysis for Mexico and Turkey

JOAKIM ERLANDSSON
OTTO PIHL

Department of Technology Management and Economics
Chalmers University of Technology

SUMMARY

This project focuses on examining the factors that can impact the logistics sector in emerging economy countries, specifically Mexico and Turkey. The main objective is to find variables impacting the logistics sector, and then forecast the future trends of logistics in the short-term and mid-term horizons. To accomplish this, we utilize freight transportation data obtained from the Organization for Economic Cooperation and Development (OECD) database, categorized by different modes of transportation. Additionally, we gather various factors that can influence logistics and freight transportation from the World Bank database of world development indicators.

The initial step in our analysis involves data preparation and cleaning, ensuring that the information is accurate and ready for further analysis. Next, we employ principal component analysis (PCA) to reduce the dimensionality of the data. The outcomes of the PCA reveal some differences when comparing the two countries. However, we identify demographic, economic, and logistical factors as influential in predicting future increases in freight transportation for both Mexico and Turkey.

To identify the factors that significantly affect logistics trends, we perform exploratory data analysis and employ a time series regression model. We utilize Exponential Smoothing (ETS) and Autoregressive Integrated Moving Average (ARIMA) methods for forecasting logistics trends in the short-term and mid-term horizons. The time series regression model, with an Adjusted R-squared value of 0.963, indicates that the variables affecting the logistics sectors of both emerging economy countries can be classified into demographic, economic, and logistical factors. Specifically, we find that the young population, GDP, and export values are statistically influential factors in shaping the future of logistics in emerging markets.

In analyzing the influential factors affecting the logistics sector in Mexico, several key aspects have emerged. Manufactured exports, exports to high-income countries, and the increase in consumption among the population are identified as significant factors driving logistics development. Moreover, the presence of a young population in Mexico is crucial for the logistics growth. Another noteworthy variable is negative rural growth, indicating a shift of people from rural to urban areas. This trend is particularly important for emerging economies, often serving as an indicator of overall logistics development.

Similarly, in the case of Turkey, various variables from different categories exerts a significant impact on the development of freight transportation. Demographic factors, such as age distribution and geographical positioning of the population, play a key role in logistics trend. Additionally, economic factors, including GDP, interest rate, and debt stocks, have proven influential in understanding the dynamics of freight transportation. Furthermore, logistical factors, such as merchandise imports and exports, have emerged as crucial drivers of growth in Turkey's logistics sector.

Moreover, the forecasts generated by ETS and ARIMA models demonstrate an overall upward trend in freight transportation for both Mexico and Turkey. However, variations are observed between the modes of transport and the two countries. The choice of the best-fit models differs across modes of transport. In the case of Mexico, the ARIMA model outperforms the others for all modes except when forecasting total freight. For Turkey, the ETS model is more accurate in predicting air and total freight, while ARIMA fits better for the other modes of transport. The ETS forecast for rail transportation in Turkey suggests a relatively stable trend, showing neither a significant increase nor decrease in the upcoming years. On the other hand, the ETS forecast for air freight transportation in Mexico indicates a general increasing trend for the next few years, followed by a slight decline in the final two years of the forecast, implying a potential shift in the trend. To validate the accuracy of our forecasts, we assess the residual properties and conduct statistical tests.

In conclusion, this project aims to uncover the factors that impact the logistics sector in emerging economy countries, focusing on Mexico and Turkey. By employing data analysis techniques, such as PCA, exploratory data analysis, and time series regression models, we identify influential factors and generate forecasts for the short-term and mid-term horizons. The results highlight the importance of demographic, economic, and logistical factors in shaping the future trends of logistics in emerging markets. The forecasted trends indicate an overall upward trajectory in freight transportation for both countries, with variations observed across modes of transport. The reliability of our forecasts is supported by residual analysis and statistical tests.

Key words: Emerging Market, Forecast, Logistics, Freight Transportation, Mexico, Turkey, Principal Component Analysis, Times Series Regression Model, Exponential Smoothing, Autoregressive Integrated Moving Average.

TABLE OF CONTENTS

List of Figures:

List of Tables:

1. INTRODUCTION

Emerging markets are countries experiencing rapid economic growth, increasing large populations, and an expanding middle-class population (Manners-Bell et al., 2014a). These key points have made emerging markets a focus point for global industries and organizations. This, because they represent both challenges and opportunities. One known vital focal point is the population growth in these countries. The population increase is expected to drive the demand for goods and services, increasing demand for different logistical measures. Services such as transportation, warehousing, and supply chain management, are all services and aspects which will be subject to an increase in demand (Manners-Bell et al., 2014a). Emerging markets possess a very large portion of the world’s population, further increasing the relevance of the countries classified as such and cements them as important actors in the global landscape of the future (European Central Bank, 2010).

In the last decade (2012-2022), emerging markets have contributed to half of the world's Gross Domestic Product (GDP) and 66% of its growth (World Economics, n.d.). This growth has been further boosted by China, which has led the expansion of trade volume generated by emerging markets. As a result, emerging markets’ share of global trade volume has increased from 32% in 2000 to 46% in 2019 (OECD, 2021). Emerging markets are hence key for the future international landscape.

Urbanization and a change in consumer behavior and patterns are factors that can, in addition to population growth and other demographic factors, affect the evolution of the logistics landscape for these emerging countries (Manners-Bell et al., 2014a). The demand for services regarding urban logistics is estimated to increase as more and more people move into larger cities and urban areas. Changes in consumption, and the rapid increase of global e-commerce are crucial points providing global industries opportunities (Manners-Bell et al., 2014). For example, as more people in these markets gain access to the internet and mobile devices, e-commerce and digital services are experiencing significant growth.

Based on these considerations, it is widely important for industries and global corporations to be aware of different demographic trends, in addition to economic and financial conditions in emerging markets. And how they might impact the demand for logistics services and different flows of goods between countries and continents (Hirschinger et al., 2015a).

For these emerging economies, logistics and supply chain management play a crucial role into helping them fully develop as countries (Hirschinger et al., 2015a). To fully capitalize on these opportunities, businesses and investors must be able to navigate the unique challenges of operating in these markets. One of the most significant of these challenges is the need for efficient transportation and logistics systems to support the movement of goods and services across regional borders, and to accommodate, as previously mentioned, the constant increase in population. Improving the transportation and logistics systems of emerging economies can significantly impact their global value chains, given that these countries play essential roles in global supply and distribution networks. An increase in global connectivity will also be key for the development of the countries (Hirschinger et al., 2015a). An effective logistics system enhances trade volumes and attracts export and import-oriented investments, promoting trade capacities and market development. Contrarily, logistical constraints and inefficiencies can substantially increase lead times and the cost of logistics services. Inadequate transportation infrastructure and logistical barriers, such as deficient customs procedures, can worsen the markets' susceptibility to logistical uncertainties and capacity limitations, negatively affecting trade and globalization (Hirschinger et al., 2015a).

In conclusion, the importance of logistics in emerging economies cannot be overstated. As these countries experience rapid economic growth and urbanization, the demand for logistics services is expected to increase significantly. Global corporations and organizations need to understand demographic trends, economic and financial conditions, and changes in consumer behavior to meet the changing demand for logistics services and adapt to new market conditions.

This project aims to identify the factors that impact the logistics sector in emerging economies and forecast the future of logistics for the short term and mid-term. The freight transportation data, categorized by modes, is sourced from the Organization for Economic Cooperation and Development (OECD) database. Similarly, the World Bank database is used to gather information on various factors that influence logistics and freight transportation. We use several important statistical methods, which include but are not limited to, principal component analysis (PCA), Exponential Smoothing (ETS), and AutoRegressive Integrated Moving Average (ARIMA) to achieve our purpose.

1.1. Background

This report aims to provide an extensive overview of the logistics trend in two emerging markets: Mexico and Turkey. We will identify potential variables that influence the logistics sector in these emerging market countries through data and statistical analysis, such as time series regression analysis. Finally, we forecast the future trend of logistics in these countries for the short- and mid-term, corresponding to 1–5 years ahead. The chosen forecast (dependent) variable is freight transportation measured in million ton-kilometers, showcasing the logistics within a country.

1.2. Aim of the study

This report further aims to investigate and analyze emerging markets through the perspective of development and the logistics sector within the countries. We analyze it through different vital factors such as demography, economic and financial indicators, consumption patterns, and different flows of goods, to name a few. We intend to increase the knowledge of different emerging markets and show how their development will affect the future of logistics in the coming years. To increase knowledge, determining what factors will influence the logistics sector in an emerging market is the first step of the report. Thus, forecasting the future of logistics in emerging markets through the determined factors will hopefully provide extensive insight into the future of the logistics sector of both Mexico and Turkey.

1.3. Research questions

This project sheds light on the following two research questions:

1) What factors can influence the logistics sector in an emerging economy country?

2) How is the future of logistics forecasted for short-term and mid-term in the emerging economy country?

1.4. Delimitations

The scope of this study is limited to two selected emerging economy countries, Mexico, and Turkey. These two countries are chosen based on their status as emerging markets and importance in the global economy. The study focuses on the logistics industry in these countries, how logistics can help with development, and the challenges and opportunities faced by companies operating in these markets. Moreover, it shows how the growth within these emerging economy countries will affect the future landscape regarding logistics. These delimitations help define the study's scope and focus and ensure that the results are relevant and meaningful in the context of the research questions and goals. For our research questions, we define logistics as the dependent variable of freight transport by different modes of transportation. These modes include road transportation, Inland transportation, rail transportation, and air transportation. Regarding data collection, this project is limited to data from the World Bank databank of world indicators, and data regarding freight transportation from the Organization for Economic Cooperation and Development (OECD). However, delimitations regarding not incorporating maritime logistics were made because of the absence of relevant data, thus risking an inaccurate forecast.

2. FRAME OF REFERENCE

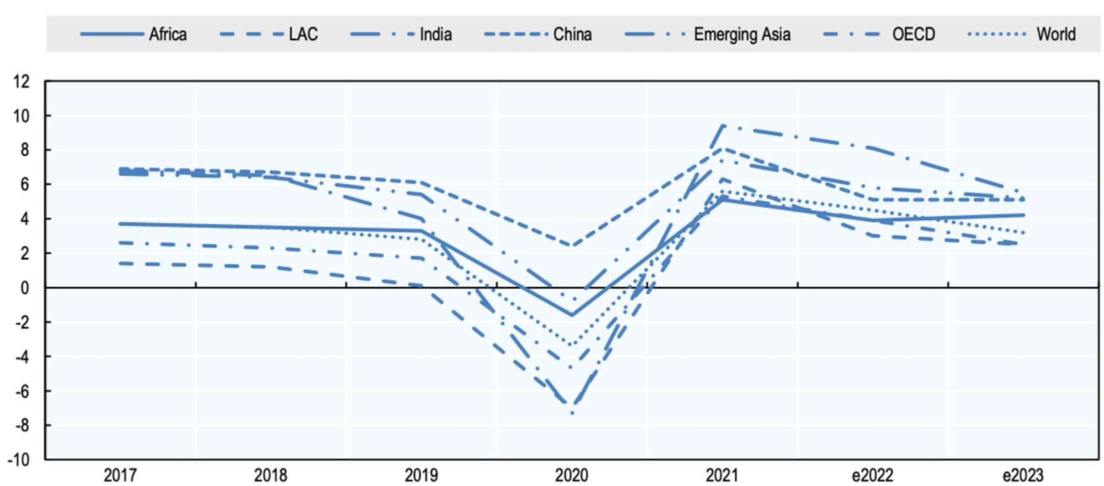
Emerging markets are countries that are developing at a rapid pace and showing substantial growth. These countries are defined by high economic growth rates, rising living standards, increasing levels of urbanization, growing demand for goods and transport, and often a shift from agriculture to manufacturing services. This section will discuss the main features of emerging markets, their importance to the global economy, potential challenges, and the role of logistics in emerging countries.

2.1. Emerging markets

Emerging markets are growing faster than already-developed markets and are expected to continue their growth in the coming years. This is due to several factors, such as rising populations, urbanization, investments, and infrastructure improvement. The increased productivity through investments in infrastructure and the growth of the workforce via the increasing population can be reviewed as a springboard for developing the economic situation in less developed countries (World Bank & International, 2011). However, different elements must be considered when reviewing the potential growth of other markets worldwide. Trade, mainly the export of goods and services from a specific country, contributes to the economic development of the emerging market (World Bank & International, 2011). Therefore, lowering trade barriers to and from the world towards an emerging market is crucial for continuing financial growth.

Related to trade and highly correlated to the development of a country is the yearly measure of the Gross Domestic Product (GDP) (Marquis & Raynard, 2015). GDP represents the total value of all goods and services produced within a country's borders in a specific period, usually a year. The GDP measures the size and health of a country's economy (Callen, n.d.). The Annual GDP growth had a significant decrease in 2020 because of the pandemic (see Figure 1).

Figure 1: *Annual Real GDP Growth*



Note: The annual GDP growth in percentage for the different regions of the world (OECD, 2022).

Emerging markets are a hot topic for investors worldwide. Those investors also play a significant role in the country's development. The institutions and companies in an emerging market must engage in political strategies and infrastructure development that affect their business to continue the market's growth and the company's possible success (Marquis & Raynard, 2015). In addition, the rapid urbanization in emerging markets presents good opportunities for business, taking advantage of the increasing demand for a smaller geographical area.

Demographic statistics and trends are also impacting the development of the markets. For example, emerging markets are characterized by a growing younger population, leading to an increase in the available workforce (Marquis & Raynard, 2015). In addition, other demographic trends such as life expectancy, median age, and age dependency influence the country's developing potential.

With a wealthier population, consumption amounts and patterns change (Piotrowicz & Wojciech, 2015). Urbanization, people seeking better living conditions, and increased consumption create a greater demand, which provides opportunities for businesses worldwide.

Emerging markets provide opportunities for investors and businesses; however, these opportunities also come with significant risks. Non-developed markets are usually characterized by political instability, hence the importance of engaging in the political strategies of the country (Marquis & Raynard, 2015). Despite the significant economic growth in emerging markets, the financial stability is usually not at the required level, proving substantial risks for investors and businesses seeking opportunities. Internal strategies mitigating these risks are needed to succeed in emerging markets.

2.1.1. Mexico

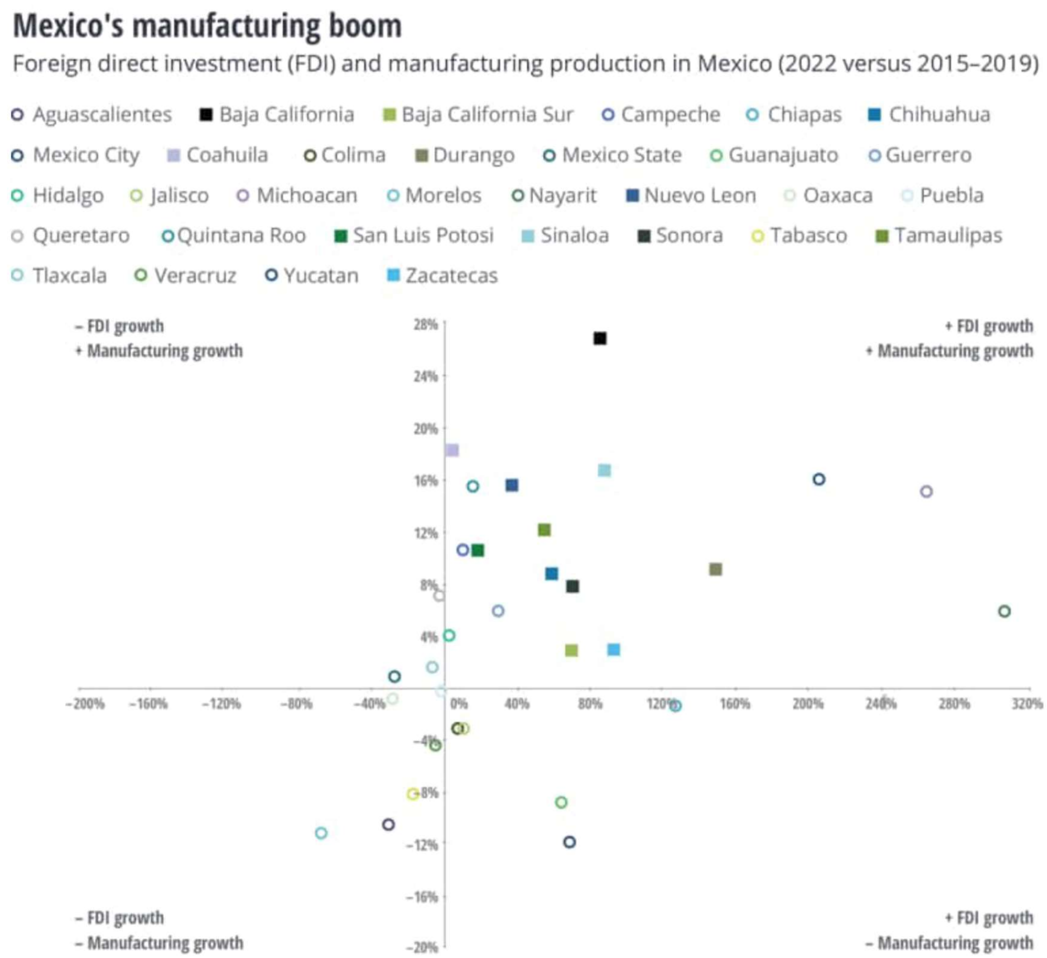
Mexico is a diverse and extensive country with a population of approximately 130 million people. It is the 11th most populous country in the world. Mexico is considered a middle-income country, and it possesses the 15th largest economy in today's world (The World Bank, 2023c)**.** However, this considered, Mexico is still considered an emerging economy country. Mexico's rich and complex history spans many centuries, from the pre-Columbian era to the Spanish conquest and beyond. The country is also home to diverse cultures and traditions shaped by the many civilizations that have inhabited the region over time.

The population of Mexico can be considered young, with the country's median age being 29 years. However, the proportion of elderly people is expected to increase in the coming years, which will present some healthcare and social security challenges. The gender distribution of the population is evenly distributed throughout the population, with 51% of the population being female and 49% being male. Mexico has, in recent decades, experienced a growth in urbanization, with approximately three-quarters of the total population living in more urbanized areas (The World Bank, 2022). This has brought forth challenges regarding urban infrastructure and development, including traffic congestion, pollution, and access to affordable housing (Kim & Zangerling, 2016). Nevertheless, urbanization has contributed greatly to the economic development of the country, where large cities such as Mexico City and Guadalajara serve as hubs for innovation, technology, and cultural development (Kim & Zangerling, 2016).

Industries of high importance for Mexico include manufacturing, agriculture, tobacco, and petroleum production (UNIDO, n.d.-a). The country has also experienced significant economic growth and development in these areas in recent years, but the country still faces challenges such as inequality, poverty, and mass unemployment. Mexico's industrial landscape is diverse and has experienced rapid exponential growth in recent decades. It plays a vital role in the country's economy, and the industrial landscape employs millions of Mexicans. Some more critical sectors of industry in Mexico are manufacturing, electronics, energy, and construction (UNIDO, n.d.). Thanks to its highly skilled workforce in these industries and favorable geographical position, Mexico is attractive for foreign investment and international workforce.

One of the most important industries in Mexico is industrial manufacturing, which accounts for a large and significant portion of the country's total GDP and is a vital sector of employment for the country. The sector has also experienced growth in various cities and areas around Mexico, such as Mexico City and Queretaro, as Figure 2 illustrates. Automotive, aerospace, electronics, and textiles are just a few industries that comprise a large portion of the nation's manufacturing sector (UNIDO, n.d.). Numerous significant automakers, including General Motors, Ford, and Volkswagen, all have a significant presence in Mexico, which is also the seventh-largest auto producer in the world (International Trade Administration, 2023). The nation's aerospace industry is also expanding quickly, with extensive global aerospace firms established in the country.

Figure 2: *Mexico's manufacturing boom.*



Note: *FDI* is measured as the percentage variation between the average investment made in the first half of 2021 versus the first half of 2015-2019. *Manufacturing production* is measured as the percentage change between Q2 2022 and the annual average from 2015 to 2019. Squares represent states in the north of the country (Deloitte Insights, 2022a).

Mexico as a country is also wealthy in the aspect of natural resources. Natural resources which are especially important for the nation are resources such as minerals, oil, and gas. One sector which plays a key part in the economy of the country is the nation's mining sector. With large deposits of gold, silver, copper, and other minerals, it contributes greatly to the economic development of Mexico (International Trade Administration, 2023). The energy division of Mexico is also experiencing growth and development, with key investments in wind power and solar power being key areas of interest for investors.

In addition to Mexico's manufacturing and natural resources sectors, the construction part of the industrial landscape is also crucial to the country's economic development, as previously mentioned. New infrastructure, commercial real estate, as well as residential estates, are all factors that have helped grow the construction business. However, in recent years due to events such as the COVID-19 pandemic, demand for construction has decreased (International Trade Administration, 2023).

In conclusion, Mexico's industrial landscape is diverse and contributes significantly to the development of the nation. The industrial sector has, since the pandemic, and as per Figure 3, been able to grow further in areas such as manufacturing output, exports and imports for industrial products and goods, and capital investments. Furthermore, the country's prominence in the aforementioned areas makes it an attractive destination for foreign investment and trade.

Figure 3: *Mexico's manufacturing*

Note: Since August 2020, Mexico's manufacturing industry has increased both in 2021, and 2022. Mainly through both manufacturing exports and intermediate and capital goods import (Deloitte Insights, 2022b).

Regarding global trade, Mexico is an important actor. With a widely diverse economy, trade patterns, and a beneficial geographical position, Mexico serves as a gateway to North and South America.

Mexico is one of the most prominent exporting countries in the world, with products exported mainly comprised of industrial products (UNIDO, n.d.). Even though the increase of manufactured exports is down compared to 2021, Figure 3 still illustrates an increase of 17.2% in this area. Products that are being exported include electronics, petroleum, agricultural goods, and automobiles. Mexico was a key player in NAFTA, the North American Free Trade Agreement, with its leading trading partner being the United States of America. The United States accounts for approximately 80% of the total export of Mexico. Other notable partners of the trade include China, Japan, and Canada, as well as several countries in the European Union. NAFTA has now been replaced by a new bilateral trade agreement between Mexico, The United States, and Canada called the United States-Mexico-Canada Agreement (UMSCA) (International Trade Administration, 2023).

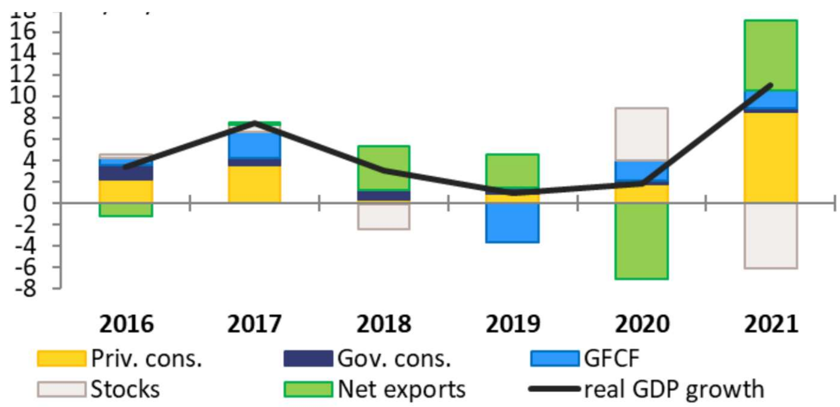
The import of goods also plays a vital role in the economy of Mexico, particularly in areas such as machinery, petroleum products, and electronics (Santander, 2023). Same as with exports, the United States is the country that accounts for the most significant percentage of Mexico's imports—accounting for around 44% of the total imports of Mexico (Santander, 2022). Notable trading partners here include Germany, South Korea, and China.

As previously mentioned, the North American Free Trade Agreement (NAFTA) has played a crucial role in how regional trade in North America has developed. With reduced tariffs and increased market access for its goods and services, regional trade for Mexico has also been able to experience rapid growth (UNIDO, n.d.). Mexico is also working on further strengthening its relationship with other Latin American Countries.

2.1.2. Turkey

Turkey is the 19th largest economy in the world, with a GDP of approximately $720 million (The World Bank, 2020a). With a large population of 70 million people, the GDP per capita measures at $9626. Turkey experienced significant economic growth in the early years of the century, following substantial reforms that reduced poverty and increased the overall wealth of its population. As a result, despite the impact of the COVID-19 pandemic, Turkey has fared relatively well compared to many neighboring countries. Notably, Turkey saw economic growth in 2020, uncommon in a year characterized by widespread economic downturns. (The World Bank, 2020a). As an effect of the above, Turkey has seen substantial GDP growth in recent years (see Figure 4)

Figure 4: *Turkey Real GDP Growth and Contributions*

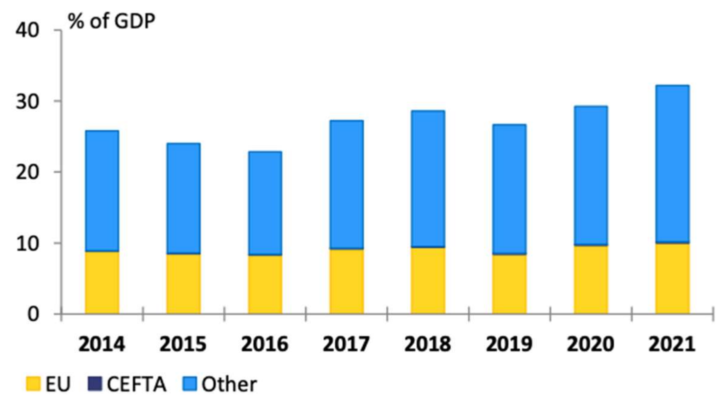


Note: Turkey's GDP growth and the contribution of different logistical and economic areas (European Commission, 2022).

Turkey is a member of the international economic cooperation called the G20, consisting of 20 of the biggest economies in the world. The G20 members represent around 85% of the world's GDP and around two-thirds of the world's total population and play an essential role in the governance of major economic activities worldwide (G20, 2022). Turkey's membership in the G20 reflects its growing economic influence in the global arena. Turkey was included in the G20 in 1999 as a non-permanent guest and became a permanent member in 2008 as the G20 was upgraded to Heads of State/Government following the financial crisis the previous year (G20, 2022).

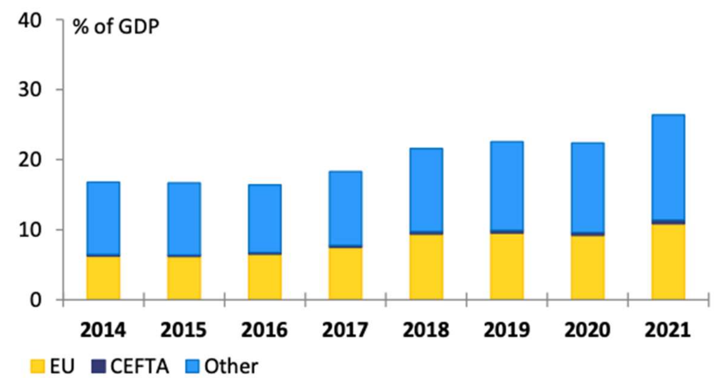
The G20 membership has significantly impacted Turkey's engagement with other major economies, increasing foreign trade (Kesgingöz et al., 2017). Before 2012, Turkey's trade was primarily dominated by G20 countries, but since then, the country has been attempting to diversify its trade partners (Kesgingöz et al., 2017). Overall, Turkey's membership in the G20 indicates its growing economy and impact on the global economy. See Figures 5 and 6 for Turkey's trade with countries in the European Union and members of the Central European Free Trade Agreement (CEFTA) compared to the rest of the world.

Figure 5: *Turkey* – Import of Goods



Note: Turkey’s import of goods as a percentage of its GDP, categorized by the trade partner (European Commission, 2022).

Figure 6: *Turkey* – Export of Goods



Note: Turkey’s export of goods as a percentage of its GDP, categorized by the trade partner (European Commission, 2022).

Turkey is also a North Atlantic Treaty Organization (NATO) member and has had strong connections to the US, both political and military (International Trade Administration, 2022). In 2021, the total trade between the US and Turkey was around $28 billion, an increase compared to the previous year by 33% (International Trade Administration, 2022).

Turkey's strategic geographical positioning also benefits the country's growth regarding logistical benefits and trade. Turkey is often viewed as the crossroads of world trade near Europe, Asia, and the middle east, and the bridge between the East and the West. Because of this, Turkey is popular amongst investors, and the country invests heavily in its infrastructure to facilitate and increase its status as a great logistical actor (The World Bank, 2023a). As a result, Turkey has made positive changes in recent years regarding the country's logistics, including reforms to modernize and streamline customs work in hopes of lowering both cost and time correlated to the import and export of goods (International Trade Administration, 2022). Furthermore, tax exemptions and simplified procedures for licenses and permits regarding trade have been made to attract foreign investments (International Trade Administration, 2022).

Turkey has established itself as a prominent player in global trade, with a diverse range of products and industries contributing to its position as a major exporting country. Turkey’s primary export industries include automotive products, basic metals, machinery & appliances, and textiles. In addition, however, food & beverages, chemicals, and rubber & plastics are also exported in large numbers from Turkey (UNIDO, n.d.-c). Though, it is worth noting that when ranked by their value-added, the top export industries in Turkey are led by the tobacco sector, followed by the automotive, textiles, and chemicals industries, respectively (UNIDO, n.d.-b).

Demographical factors are also beneficial in Turkey, with a growing population of almost 1 million yearly (The World Bank, 2020b). The country also has a young population and a large available workforce. The total population of Turkey is around 80 million, with a median age of 31, significantly lower than other European countries, according to the World Bank (2020). The country also has a rapidly growing urban population, with more than 75% of the people living in urban areas as of 2020 (The World Bank, 2023d). Rapid urbanization has led to a growing demand for housing, infrastructure, and public services, creating business opportunities in these sectors.

Political and social stability is often used to determine a country's potential growth. However, Turkey has had a relatively low number on the political stability scale, showing that the country is below the world average regarding this topic (The Global Economy, 2022). Furthermore, despite maintaining economic growth in 2020, Turkey has had problems with inflation. In addition, concerns have been regarding the government, its authoritarianism, and the handling of the economy and its fluctuations (Human Rights Watch, 2021).

Turkey is a promising country, according to many, because of the earlier mentioned factors, but there are still many risks challenging its status as an emerging market:

- Political tensions like corruption scandals, coup attempts, and protests threaten the country's reputation.
- The Turkish currency fluctuations are heavily affected by the change in global interest rates, political uncertainty, and economic imbalances that have made it difficult for possible investors to plan and invest in the country.
- Inflation has also been a significant problem in Turkey, affected by currency fluctuations and political instability.

2.2. Logistics in emerging markets

Logistics and supply chain management are increasingly important to the growth of countries and different economies (Manners-Bell et al., 2014b). Global supply chains are more complex than ever, with actors included being spread around the world. Due to the increase of globalization and the ease of international business, countries and corporations have been able to experience rapid growth in their economies and finances with the opportunity to source manufacturing and logistics to emerging economy countries (Hirschinger et al., 2015b). However, as actors have previously seen emerging markets as a source of materials and inexpensive labor, a paradigm shift is now occurring (Hirschinger et al., 2015b). Emerging market countries are experiencing rapid development in their economies, and this, combined with a steady increase in population, will, in turn, affect how global corporations and supply chains act. Emerging markets are future market opportunities for international industries. As a result of economic development, the populations in these countries are getting wealthier as well, resulting in a change in consumer behavior which directly affects the demand for logistics services. Considering these aspects, the logistics industry inside these emerging markets will experience rapid development and an increase in aspects such as freight movement and the need for transportation (Manners-Bell et al., 2014b).

Different factors which can influence the development of the logistics sector in emerging markets include, but are not limited to:

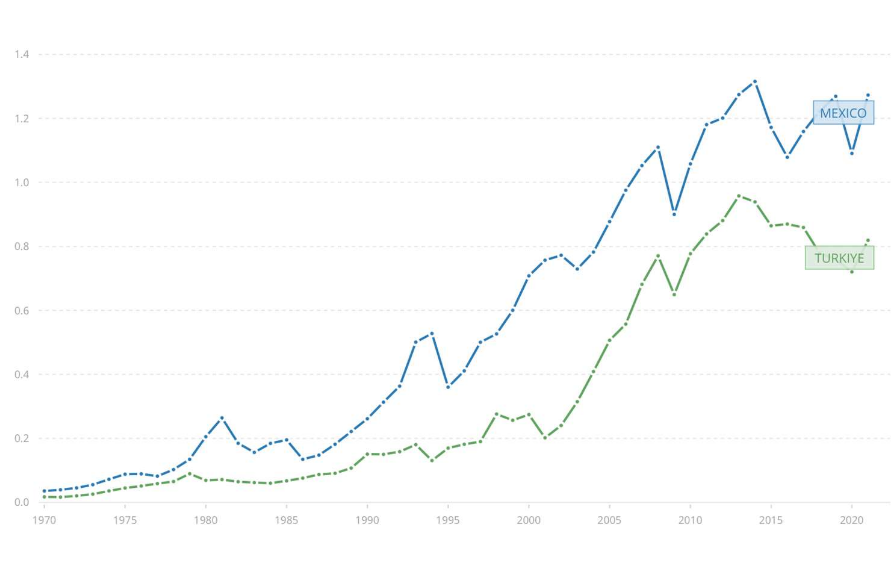
- *Economic Growth*: In rising economy nations, the logistics sector is primarily driven by economic growth. The need for logistics services grows as these nations' economies and GDPs expand quickly. Consumer demand growth, industry expansion, and increased global trade are some factors fueling this growth of GDP. For instance, economic development has increased demand for logistics services, particularly in the e-commerce industry. Both Mexico and Turkey have, over a 50-year period, experienced significant growth in their GDP. As per Figure 7, there has been a steady increase; however, since 1990, there has been an even more rapid increase than from 1970 to 1990.

- *Trade Liberalization*: In rising economy nations, trade liberalization is a significant factor boosting the development of the logistics sector. The need for logistics services has expanded because of the policies that many of these nations have implemented to promote international trade.

- *Technological Development:* Technological development has also significantly influenced the development of the logistics sector in developing economies. Technologies for communication and transportation have advanced, making it simpler to track shipments, manage logistics operations, and increase supply chain visibility.

- *Infrastructure Development*: The efficiency of the logistics sector in emerging economy nations has greatly benefited from infrastructure development. Reduced transportation costs, quicker deliveries, and greater connectivity have all been made possible by improved infrastructure, such as better roads, ports, and airports. To enhance logistics operations, governments, and private sector participants have made significant investments in developing infrastructure in emerging economy nations.

Figure 7: *GDP Increase for Mexico and Turkey between 1970-2021*



Note: GDP increase is a good indicator of a nation's economic growth and development. GDP increase is often connected to logistics infrastructure development.
X-axis: years, Y-axis=Current USD$(billions) (The World Bank, 2023b).

There are also challenges related to development that must be resolved for and by these emerging economy countries. Infrastructure that still is inadequate, a shortage of experienced labor, a complex regulatory environment, and political instability are just a few of the industry's many problems in emerging markets. These issues must be resolved for the logistics sector in developing economy nations to continue to expand and flourish. Logistics and freight movement in emerging markets are also increasingly influenced by sustainability. In response to increased demand for environmentally friendly transportation options from governments, consumers, and businesses, logistics companies and industries are investing in alternative fuels, electric vehicles, and other green technologies. However, as previously mentioned, the lack of adequate infrastructure and funding can make it challenging to implement these sustainability initiatives.

2.2.1. Current logistics infrastructure and development in Mexico

Mexico has a well-developed logistics infrastructure with a comprehensive network of highways, railways, ports, and airports. As previously stated, the country has an advantageous geographical position between Latin America and the United States—something the country has benefitted tremendously from and continues to use.

Road transportation is the most used mode of transport, and the highway system in Mexico is extensive, with over 200,000 kilometers of paved roads connecting major cities and industrial areas (Boske & Harrison, 2013). The key advantage of road transportation in Mexico is its ability to be able to transport goods into areas which limited accessibility and inadequate infrastructure, making it an all-around helpful mode of transportation. There are, however, challenges regarding road vehicle transportation in the country, with high-value cargo sometimes being reported stolen or missing, and cross-border customs procedures between Mexico and the U.S. often occur with long delays and waiting times (Rodrigue & Notteboom, 2010).

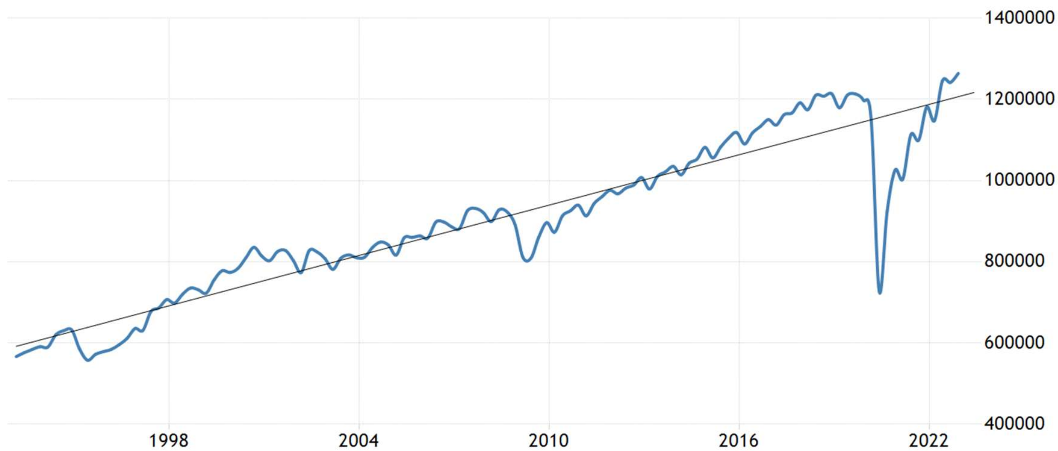
Rail as a mode of transportation is also common, often used as a more cost-efficient mode of transport for specific cargo and longer distances. The railway system is also developed, with over 26,000 kilometers of track operated by private and state-owned companies (International Trade Administration, 2023). However, due to the infrastructure not being as developed as that of the trucking industry, railway transportation still falls behind road transportation in volume and usage frequency (Manners-Bell et al., 2014a). Rail transportation does, however, enable accessibility between major cities and regions in the country, as well as connect many of Mexico's ports to crucial logistics hubs that are located more inland.

Mexico has over 100 ports located around the country's coastal regions, with critical ports connected to major ocean freight networks. Ocean freight is the second most used mode of transport used by the country (Manners-Bell et al., 2014a), and important ports include Manzanillo, Veracruz, and Altamira, among others (International Trade Administration, 2023).

Mexico's warehousing and distribution sector is growing linearly with the growing e-commerce development. It is expected to continue to grow further with the increase of digitalization in the retail industry and the increased willingness to use digital channels from a consumer perspective (Lopez & Lee, 2018). The warehouse-infrastructure range from small storage facilities to sizeable regional distribution centers. Logistics parks that house companies that also offer more services than only warehousing are also increasing. These parks are generally located near important industrial areas such as Mexico City and Guadalajara or other logistics hubs such as ports, airports, or other transportation hubs (American Industries, 2022). The cost of warehousing in Mexico can vary depending on which region the warehouse is in, the size of the warehouse, and what services are offered by the company operating the warehouse. In general, the cost of warehousing in Mexico is lower than that of other countries in North America, which can further attract global corporations to the country, and further increases the foreign investments being made.

The transportation sector plays a vital role in the infrastructure development of Mexico and contributes to the GDP of the country significantly (Trading Economics, 2023b), which is also illustrated in Figure 8, where you can see the transport sector's contribution to the GDP of Mexico from 1990 to the end of 2022.

Figure 8: *Transport sector contribution to the GDP of Mexico*



Note: The transportation sector of Mexico is a pivotal contributor to the GDP—currency in MXN (Trading Economics, 2022).

Mexico's success story with the increase in manufacturing productivity as well as logistics in the past three decades is partly attributed to its increased integration into U.S. supply chains, a feat made possible by the establishment of the North American Free Trade Agreement and the current trade agreement of UMSCA (Manners-Bell et al., 2014). Between 1993 and 2012, the value of goods traded between the two countries increased from 290 billion USD to 1.1 trillion USD (Manners-Bell et al., 2014). Mexico's rapid growth, described previously in the report, has sparked predictions that it could eventually surpass China as the second-largest trading partner of the United States. Factors contributing to this include Mexico's proximity to the U.S., access to natural resources, and low labor cost (Boske & Harrison, 2013).

With the integration of low-cost Asian production locations into the global trading community, a considerable portion of Mexican manufacturing was, however, outsourced to the Pacific region. Nevertheless, in recent years, this trend has flipped. The increasing expenses in China and the enhanced productivity in Mexico have joined forces to establish Mexico as a valuable manufacturing source (Manners-Bell et al., 2014).

2.2.2. Current logistics infrastructure and development Turkey

The infrastructure related to logistics in Turkey is under constant development, and as mentioned earlier, lots of investments are made into the sector. Turkey's geographical position provides access to the major sea routes between Asia and Europe and has made Turkey an important hub for international logistics. The government has prioritized modernizing logistics-related infrastructure to increase capacity and efficiency, such as ports, terminals, railways, and airports (International Trade Administration, 2022).

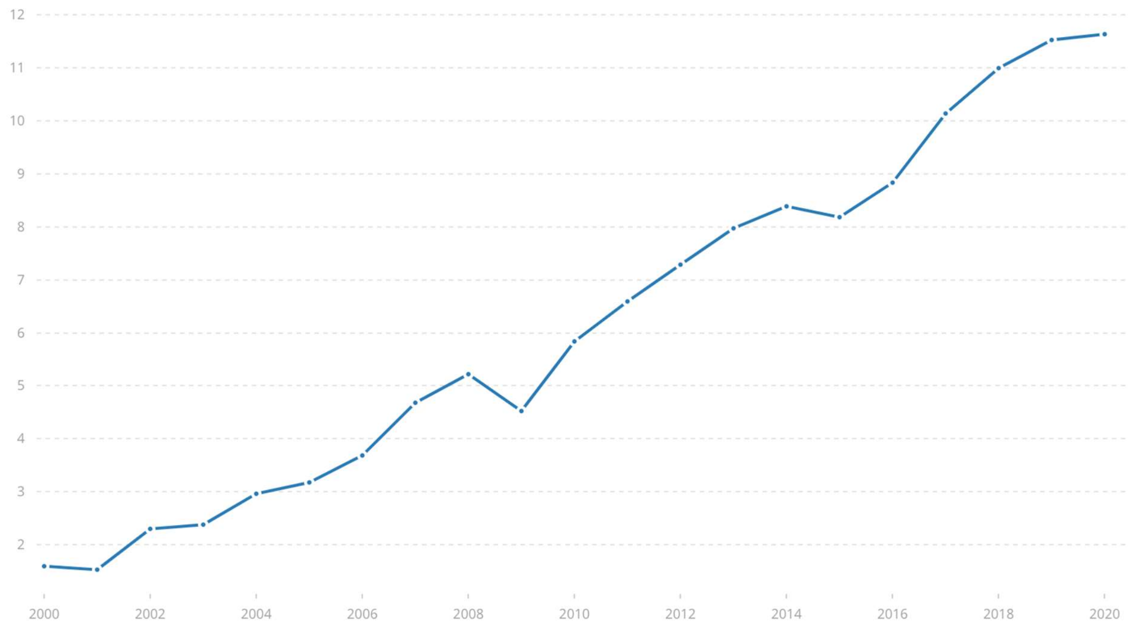
Some critical investments have been made in the Istanbul Airport. $6.5 billion have been invested in the airport, which is set to become the biggest airport in the world, with a capacity of around 150 million passengers yearly (Naz Inci et al., 2022). The total number of airports in Turkey has gone from 26 in 2002 to 56 in 2021, with a few more currently under construction.

Significant investments have also been made into the road transport segment. Over 90 percent of the transport within the borders of Turkey is made by road (Manners-Bell et al., 2014c). Yavuz Sultan Selim Bridge has been built, connecting the east and west sides of Istanbul, linking a 95-kilometer-long motorway to the Asian side of the country. The Osmangazi bridge has also been built, connecting the 427-kilometer-long road between Istanbul and Izmir. Providing better efficiency within the country (Naz Inci et al., 2022). The double-deck Marmaray & Eurasia tunnels currently under construction will also benefit road transport in Turkey, providing more alternatives regarding passing the Bosporus River and potentially decreasing the bottleneck effect of road transportation through Istanbul (Naz Inci et al., 2022). The total range of dual carriageways has increased from around 6000 kilometers in 2002 to 28000 kilometers in 2021. When looking at road infrastructure currently under development, this number will increase in the coming years (Manners-Bell et al., 2014c).

The total range of high-speed railways in Turkey has gone from non-existent to 1200 kilometers between 2002 to 2021, showcasing the investments in this transportation department. The middle corridor railway route passing through Turkey, stretching from northern Europe to China, has been constructed and improved to benefit both the capacity, efficiency, and availability of rail transport to and from Turkey (Naz Inci et al., 2022). A 5000-kilometer increase in the conventional railway is also under construction to meet the goal set by the government to increase the total ton-kilometers handled by rail by 15% (Manners-Bell et al., 2014c).

As mentioned earlier, Turkey's geographical position is precious regarding sea transport, with ports connected to the Mediterranean, Marmara, Black, Aegean, and Levantine Seas. According to a market study by Flanders Investment and Trade (2020), sea transport is the most prominent mode of transport for international trade regarding import and export. It is increasing its distance from other alternatives every year. The total twenty-foot-equivalent-unit (TEU) containers handled yearly have showcased a vast increasing trend in the last few decades, as seen in Figure 9. Investments in seaport infrastructure have increased in the last decade, with around €340 million invested between 2010 and 2017 (Statista, 2022). Turkey also plans to increase the dry-bulk handling capacity in its seaports to around 500 million tons and increase the liquid bulk handling capacity to approximately 350 million tons (Manners-Bell et al., 2014c).

Figure 9: *Container port traffic - Turkey*

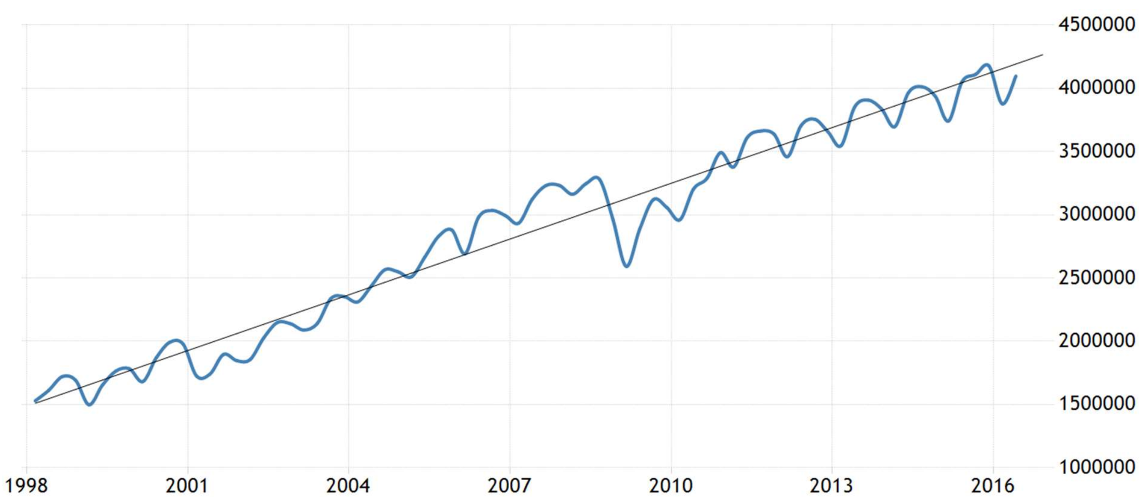


Note: Increasing trend of containers handled in the seaports of Turkey. Y-axis: number of TEUs handled (in millions). X-axis: year. (World Bank, 2022)

Turkey is well known for its expertise in the automotive industry, one of Europe's largest industries. Turkey is home to 17 vehicle manufacturers and is one of the few countries able to produce over a million cars yearly (Manners-Bell et al., 2014c). The general trend of outsourcing logistics solutions has also benefited Turkey, where global actors within the logistics industry have moved parts of their operations to the country (Manners-Bell et al., 2014c). The increased demand from manufacturers to move their production to warehouses closer to the market, also known as near-sourcing, will potentially increase the importance of Turkey's capacity and efficiency regarding logistic-related infrastructure (Manners-Bell et al., 2014c).

The transport sector plays a vital role in the infrastructure development of Turkey, just as of Mexico as earlier described, and has contributed to the country's total GDP increasingly since 1998 (see Figure 10).

Figure 10: *Transport sector contribution to the GDP of Turkey*



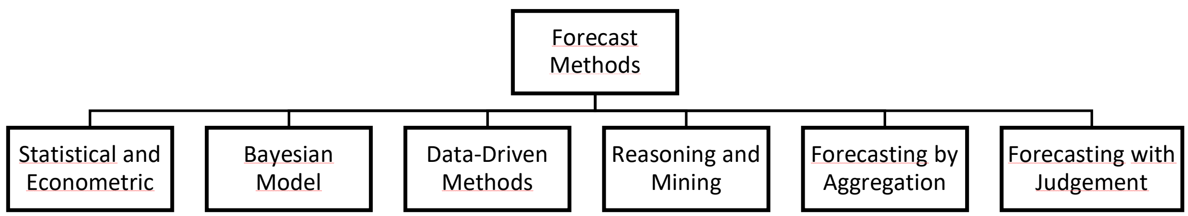
Note: The transportation sector of Turkey is a pivotal contributor to the GDP—currency in TRY (Trading Economics, 2023a).

2.3. Forecast methods

Forecasting methods refer to a set of techniques and approaches used to predict future outcomes based on historical data and other relevant information (Petropoulos et al., 2022a). Forecasting aims to estimate the most likely future values of a variable, such as sales, demand, and in the case of this report, freight transported in different modes of transport.

There are several types of forecasting methods, each with its own assumptions, some of which will be further described in this chapter (see Figure 11).

Figure 11: *Forecast methods relevant to the project.*



Note: The six different types of forecasting methods will be described further in this section.

2.3.1. Statistical and econometric models

Statistical and econometric models are mathematical representations of relationships between different variables in a given system. These models are used to describe, explain, or predict different behaviors of the system being studied via forecasting. Different models that are classified as statistical models are, for instance, linear regression models and logistic regression models.

Statistical models are used to analyze data and draw conclusions about populations from different sample data. They typically involve estimating the parameters of a probability distribution that describes the data and evaluating the significance of the estimated parameters using statistical tests (Davison, 2003).

Econometric models are used to analyze economic metrics and the relationship between different economic variables. These models frequently also involve estimating parameters and testing hypotheses about economic relationships and causal effects (Hymans, n.d.).

Statistical and econometric models are commonly incorporated and used in forecasting methods to predict future values of variables of interest. The process of using these models for forecasting involves estimating the model parameters based on historical data and then using the estimated model to predict future values of the dependent variables (Hymans, S.H. n.d.).

Different types of statistical and econometric models can be used for forecasting, depending on the nature of the data and the relationships being studied. For example, time series models such as ARIMA (AutoRegressive Integrated Moving Average) as well as exponential smoothing models are commonly used to forecast time series data, such as stock prices, GDP, or inflation rates (Hyndman & Athanasopoulos, 2018). These models rely on the patterns and trends observed in past data to make predictions.

In contrast, cross-sectional models such as regression analysis are used to predict the value of a dependent variable based on the values of one or more independent variables. These models can forecast the impact of changes in the independent variables on the dependent variable, such as predicting sales based on changes in marketing spend or predicting housing prices based on changes in interest rates (Hyndman & Athanasopoulos, 2018). In addition to estimating the model parameters, forecasting methods typically involve evaluating the accuracy and reliability of the forecasted values. This can be done using a variety of metrics such as mean squared error or R-squared. Overall, statistical, and econometric models are powerful tools for forecasting and can be used in a wide range of applications to predict the future behavior of complex systems.

This report uses data that is indexed in time order, with a time span of the data being between 1970-2021. Because of this, the most appropriate data-driven techniques for forecasting the data are the time-series methods. For this report, three time series methods are used. These three are time-series regression analysis, exponential smoothing (ETS), and Auto-Regressive Integrated Moving Average (ARIMA).

Time series regression models are statistical methods that analyze the relationship between a dependent variable and one or more independent variables. It is used to forecast the future values of the dependent variable with the help of historical data regarding the independent variables.

Exponential smoothing, abbreviated ETS, is a time-series forecasting method that uses a weighted average of past observations to forecast future values. The principle behind this model is that more recent data is more important or gives more weight to the forecast than older data. The model includes smoothing parameters that decide how much weight is given to newer observations. There are several variations of ETS models, and the difference between them is in how they capture seasonality and trends in the data.

AutoRegressive Integrated Moving Average, more commonly called the ARIMA model, is another forecasting model which forecasts future observations with the help of historical data. The key aspect of the ARIMA model is to differentiate data. Meaning you convert the data being forecasted to a stationary time series. A stationary time series has constant statistical attributes such as a constant mean and no seasonality. Hence resulting in a more conservative forecast, whereas the ETS model's values fluctuate over time and are more volatile in its forecast.

2.3.2. Bayesian forecasting

The Bayesian approach to forecasting naturally produces a probabilistic forecast, reflecting the uncertainty about future values of the phenomenon of interest based on all available information, including integrating all unknown quantities (Petropoulos et al., 2022b). To generate such forecasts, the Bayesian approach requires a predictive model for the relevant phenomenon, a model for the observed data, and a distribution representing prior beliefs regarding unknowns (Petropoulos et al., 2022b).

By applying Bayes' theorem and the calculus of probability distributions, the Bayesian predictive density function can be calculated, which describes the behavior of the future random variable given observed data up to a certain point in time (Petropoulos et al., 2022b). Unlike frequentist approaches to forecasting, the Bayesian approach factors in parameter uncertainty through integration concerning the posterior probability density function for the unknowns. This posterior probability density function is derived from Bayes' theorem. It is proportional to the likelihood function, which defines the assumed model for the data, and the prior probability density function, which captures prior beliefs about the unknowns (Petropoulos et al., 2022b). Bayesian model averaging can also accommodate uncertainty about the assumed predictive model by taking a weighted average of model-specific predictions with posterior model probabilities as the weights (Petropoulos et al., 2022b).

2.3.3. Data-driven methods

Data-driven forecasting methods leverage historical data to automatically detect patterns and correlations between variables and utilize them to forecast future values. These methods differ from model-based approaches that require defining a mathematical model to describe the relationships between variables. An example of a primary type of data-driven forecasting is the time-series method (Petropoulos et al., 2022)

Data-driven forecasting offers advantages over other forecasting methods, including the ability to model complex and nonlinear relationships between variables and adapt to changing patterns over time. Data-driven methods can be computationally demanding because they require extensive data for reliable predictions. However, as data availability increases and machine learning technology further develops, data-driven forecasting methods are gaining popularity (Hyndman & Athanasopoulos, 2018). These techniques provide a powerful tool for making predictions in diverse fields such as finance, economics, logistics, and marketing.

2.3.4. Reasoning and mining

Reasoning and mining are two critical aspects of forecasting. Mining involves using machine learning to get an understanding of large data sets and be able to identify correlations. Mining can help identify trends and patterns in the data through different statistical techniques. Reasoning is the opposite of mining and focuses more on logical inference, with the help of expertise within the forecasted segment and possible judgment.

2.3.5. Forecasting by aggregation

Forecasting with aggregation is a method that combines the forecasts of multiple individuals or models to arrive at a single aggregate forecast. Aggregation aims to produce a more accurate forecast than any individual forecast by leveraging the strengths of different forecasters or models while mitigating their weaknesses (Zotteri et al., 2005).

Aggregation can be achieved in various ways. Examples of achieving aggregation can be through the following:

1. Simple averaging: taking the average of the forecasts provided by different individuals or models.
2. Weighted averaging: giving more weight to specific forecasts based on past accuracy or other factors. For instance, more recent data is more relevant than older data.
3. Combining probability distributions: using statistical techniques to connect the probability distributions generated by different forecasters or models.

Aggregation forecasting is often used when there is a high degree of uncertainty or when multiple models are available to generate forecasts (Zotteri et al., 2005). Overall, aggregation is a powerful technique for improving the accuracy and reliability of forecasts.

2.3.6. Forecasting with judgment

It's common to use judgmental forecasting in practice. Often, it's the only viable option, mainly when there are no historical data available, a new product is being introduced, a new competitor enters the market, or when facing entirely unprecedented market conditions (Hyndman & Athanasopoulos, 2018).

Judgmental forecasting is used in three primary scenarios: when there is no available data, making statistical methods infeasible and judgmental forecasting the only viable approach, or when data is available and statistical forecasts are generated, which are then adjusted using judgment. and finally, when both data and judgmental forecasts are available independently, and they are combined (Hyndman & Athanasopoulos, 2018).

One type of judgmental forecasting is the Delphi method of forecasting. The Delphi method is designed to develop forecasts in a structured and iterative manner using a group of experts. To implement and manage the process, a facilitator is appointed. The Delphi method typically consists of the following steps:

1. Assembling a panel of experts.
2. Distributing forecasting tasks or challenges to the experts.
3. Collecting initial forecasts and justifications from the experts and summarizing them to provide feedback.
4. Providing feedback to the experts, who review their forecasts considering the feedback. This step may be repeated until a satisfactory level of consensus is achieved.
5. Constructing final forecasts by aggregating the experts' forecast

3. METHODS

This chapter studies the prediction methods used in this project. First, we investigate the main concepts in the introduction section. Then, we elaborate on the three main methods we use to deal with the research questions (RQ), i.e., time series regression analysis, exponential smoothing (ETS), and ARIMA (AutoRegressive Integrated Moving Average).

3.1. Theoretical framework

This section elaborates on the fundamental concepts we use and links them with our project.

3.1.1. Forecasting goal

In business, forecasting is a frequently encountered statistical task that aids decision-making regarding production scheduling, transportation, and personnel and offers valuable insight for long-term strategic planning. Forecasting strives to make the most precise predictions about future outcomes by utilizing all available information, including historical data and insights into future events that could impact the forecasts.

This project aims at forecasting the logistics trend in two emerging economy countries, Turkey, and Mexico. When forecasting, there are three different time horizons in which the forecast can be made. The three ranges are short-term, medium-term, and long-term. In this project, we consider time intervals ranging from a few months to one year for short-term, from one year up to five years for mid-term, and up to ten years for long-term forecasting.

3.1.2. Determining what to forecast

Determining accurately what to forecast is a crucial step in the process. It involves identifying the problem, clarifying the scope of the forecast, and collecting relevant data for the method. Following these steps dramatically improves the ability to make informed decisions and achieve better outcomes thanks to data-driven insights.

This project aims to study influential factors on (RQ1) and forecast (RQ2) freight transported for different modes for short-term and mid-term horizons. These variables which will be forecasted are air, road, rail, and inland freight transported in million-ton-kilometers. To answer RQ1, we use time series regression analysis to study several independent variables retrieved from the world bank's website (databank.worldbank.org/source/world-development-indicators). These independent variables vary from logistics, population growth, GDP increase, manufacturing, financial, educational, and import and export of goods, to name a few. To answer RQ2, we use ETS and ARIMA to forecast for short-term and mid-term for freight transported data retrieved from the OECD's website (Organization for Economic Co-operation and Development) (https://data.oecd.org/transport/freight-transport.htm).

3.1.3. The basic steps in a forecasting task

There are five key stages when doing a forecast analysis.

Step I) Problem definition is, as previously mentioned, the key to the forecasting analysis. Having a concrete and definite problem where all analysis factors are apparent is crucial. It is also essential to know how the forecast will be utilized, for whom, or what part of the organization the forecast is intended. This requires that the forecaster synchronizes the relevant actors involved in the forecast and ascertains that data collection and database maintenance are kept up to date for future planning and forecasting. The problem for this project is defined in two separate research questions, RQ1: What factors can influence the logistics sector of the emerging economy country? And RQ2: How is the future of logistics forecasted for short-term and mid-term in the emerging economy country? These questions have been developed to accurately analyze the logistics industry and its development in Mexico and Turkey. This forecast is aimed to provide further information regarding the development of these countries.

Step II) Gathering correct data is essential. Statistical databases, including the World Bank and the OECD (Organization for Economic Co-operation and Development), have been used in this research to gather large datasets about Mexico and Turkey, combined with articles and publications relevant to the topic. The World Bank dataset includes but is not limited to historical data for financial, economic, population, educational, investment, and environmental independent variables. The OECD dataset includes freight transported for different transport modes: air, road, rail, and inland.

Step III) Data preparation includes data cleaning, dealing with missing data, preliminary analyses, and data reduction, which are critical when analyzing the data. Data cleaning is a crucial aspect before the analyses can be made. Data cleaning is the process of identifying and removing errors and inconsistencies from the raw data. Afterward, we will deal with missing data in the dataset because there are some missing data in the World Bank dataset. This project performs an interpolation technique to impute missing values in a time series. Before doing a preliminary analysis, data reduction is vital in our study since the World Bank dataset is extremely large. We use the principal component analysis (PCA), which is popular for data reduction (Johnson & Wichern, 2007). We refer to Appendix A.1 for the R-studio code related to the data preparation.

Then, an appropriate way to begin is by preliminary analyses such as time series graphs and scatter plot matrix including plots of two-dimensional graphs, histograms, and calculation of correlations between variables. It helps us to see patterns in the dataset. In the next chapter, the patterns will be depicted. It is also essential to investigate whether seasonality affects the dataset, whether any outliers need to be further investigated, how strong the relationship between the variables in the data collection is, and how it affects the analysis.

Step IV) Choosing and fitting models appropriate for your dataset is vital for forecasting and analyzing. Because logistics and world bank data are indexed in time order, the most appropriate data-driven forecasting technique is the time-series method. We use three time series models: regression models, ETS, and ARIMA. Statistical analysis and these data-driven time series methods are conducted to provide a more comprehensive understanding of the datasets and the prediction forecast.

Step V) Evaluating the forecasting model is the final stage of the forecasting process. This step evaluates the error of different forecast methods and compares the results of the different models.

3.1.4. Forecasting principles and concepts

This section will describe the different theoretical concepts used in this report.

Principal Component Analysis

Principal component analysis (PCA) is a method in statistics used to reduce the number of variables or features that are used to describe a dataset. PCA works by transforming the original data into a new system, which is then divided into several different principal components. The first principal component captures most of the variance in the dataset, and the second principal component captures the second most variance, and so forth. The principal components with the highest variance are essential to keep because the variance tells us about the underlying structure of the data.

For this project, PCA will be used to reduce the number of variables in the independent variable dataset. This to easier be able to forecast the dependent variables. By reducing the number of variables in the dataset with the help of PCA, unnecessary variables are removed while preserving the most important ones. After the data reduction, our variables will be divided into principal components compiled of only the most essential variables for research question one.

Time series object

In statistical analysis and forecasting, a time series object is a collection of chronologically ordered observations over time. Time series data is mainly structured according to time intervals, such as monthly, quarterly, or yearly. And in context of this report represents the freight development for Mexico and Turkey.

For instance, an example of time series objects is in Table 1, where the number of observations is divided into five years. It partly shows road freight transported between 2015 and 2020 for Mexico (extracted from OECD).

Table 1: *Time Series Object – Location, mode of transport, year, value*

Location	**Mode of Transport**	**Year**	**Freight transported (million ton-km)**
Mexico	Road	2015	245136
Mexico	Road	2016	251122
Mexico	Road	2017	256136
Mexico	Road	2018	260642
Mexico	Road	2019	258684
Mexico	Road	2020	240394

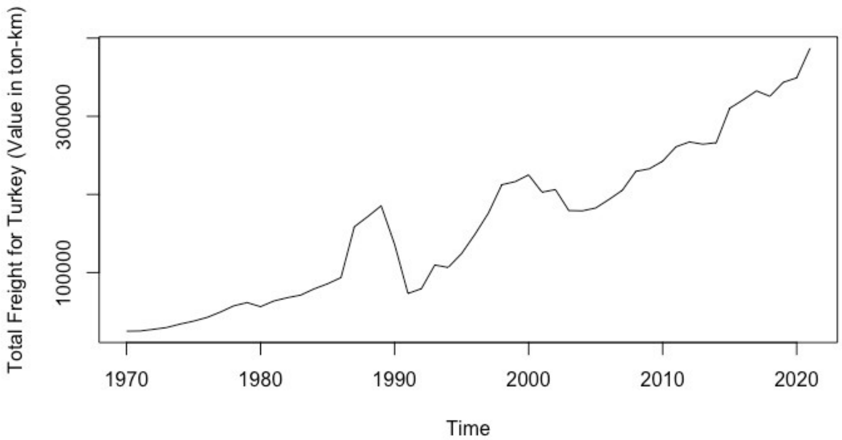
Time Series Patterns

Different patterns need to be analyzed and understood when looking at and describing time series. Patterns are usually named and categorized under three different sections:

1. Trend
2. Seasonal
3. Cyclic

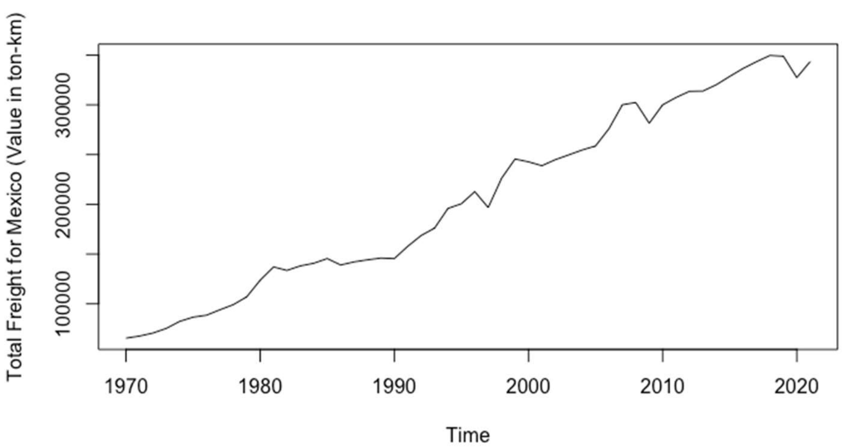
A trend in a time series can be defined as a consistent pattern that occurs over a long period of time. Three general types of trend patterns occur in a time series, namely upward-, downward-, and horizontal trends. Figures 12 and 13 depict a consistent upward trend from 1970 to 2021. In contrast, a downward trend would indicate a decrease in values over time, representing the opposite pattern to that which is observed in the depicted figures. A horizontal trend occurs when a time series is somewhat stationary, depicting neither an increase nor a decrease over time. Identifying the trend of a time series is essential regarding predictions of future values. An increasing trend over a long period of time leads to an increase in the probability that future predicted points will take a higher value.

Figure 12: *Logistics trend in Turkey*



Note: The logistics general trend in Turkey indicates a strong increase from 1970 to 2021.

Figure 13: *Logistics trend in Mexico*



Note: The logistics trend in Mexico indicates an even increase over the time period of 1970 to 2021.

Seasonality in a time series refers to a recurring pattern influenced by seasonal factors such as the time of the year or the day of the week. A fixed and known frequency characterizes it. Cyclic patterns in a time series are fluctuations that do not follow a fixed frequency. These fluctuations usually align with the "business cycle" (expansion, peak, contraction, and trough). The general length of these fluctuations is at least two years. According to Figures 12 and 13, Turkey and Mexico have no seasonality and cyclic patterns in logistic time series data.

Autocorrelation

Autocorrelation (ACF) measures the linear correlation between the lagged values of a time series. Over successive time intervals, autocorrelation represents the degree of similarity between a given time series and its lagged version. In terms of conceptualization, it's similar to the correlation between two different time series. However, autocorrelation uses the same time series twice: first in its original form, then lagged by one or more periods. Figures 14 and 15 respectively show ACF for logistic time series in Turkey and Mexico.

Figure 14: *Autocorrelation Function for Logistic Time Series for Turkey*

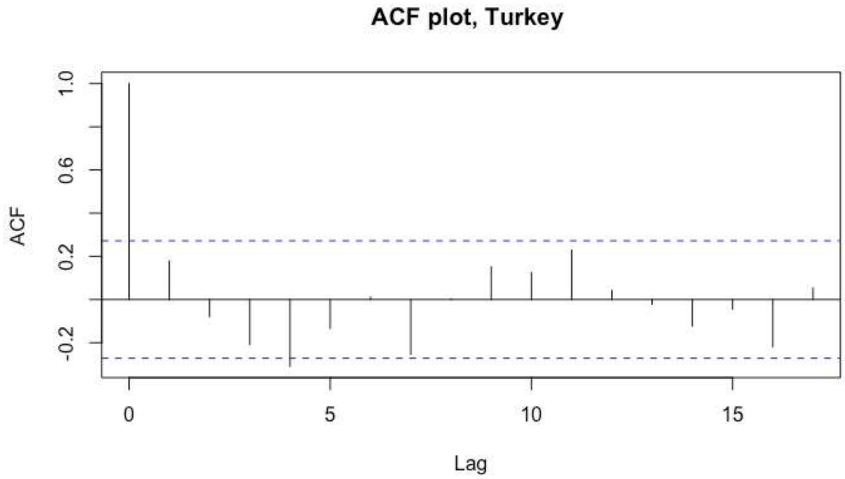
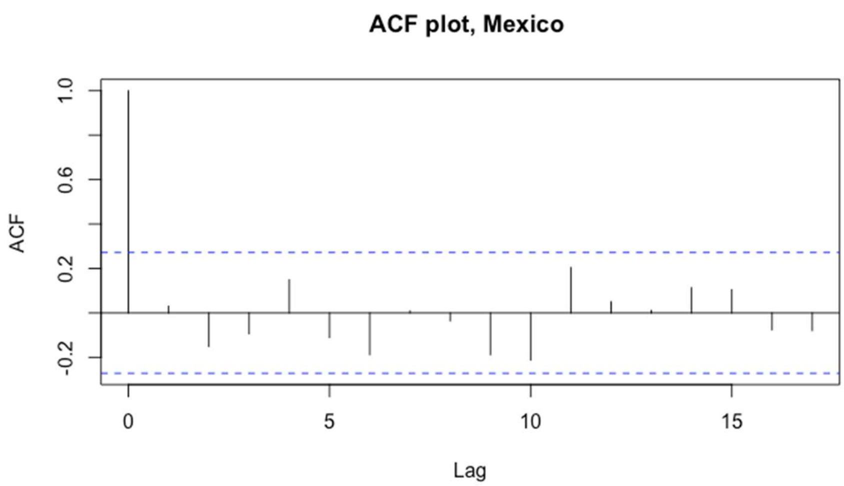


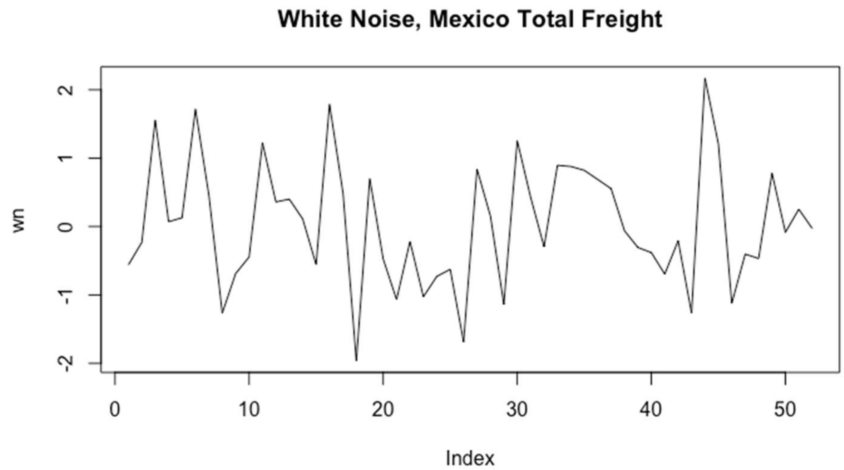
Figure 15: *Autocorrelation Function for Logistic Time Series for Mexico*



White Noise

When a time series shows no autocorrelation, it is called white noise. In the case of a white noise series, we anticipate that each autocorrelation will be near zero. However, some variation will be expected; hence the autocorrelations will never be zero. When analyzing a white noise series, it is typically expected that around 95% of the spikes in the autocorrelation function will fall within the boundaries $\pm 2/\sqrt{T}$. While it is not unusual for some random variation, this range is an indicator for assessing the statistical significance of the autocorrelation function values. In a white noise time series plot, the data points appear randomly scattered around a horizontal line, with no clear trend or pattern (see Figure 16).

Figure 16: *White Noise, Mexico Total Freight*



Note: White noise in a time series plot means the different data plots appear randomly spread around zero

Fitted Values

The fitted variables in a time series are denoted by the following formula:

$$\hat{y}_t | t-1$$

Meaning the forecast of $\hat{y}_t$, which is our dependent variable, is based on observations for $y_1, \ldots, y_{t-1}$, are obtained by using all prior observations to forecast each individual observation in the series.

While fitted values are based on all available past observations in a time series, including future observations, they are not necessarily accurate forecasts since the parameters utilized in the forecasting method are estimated using all data available in the series. As an example, in a forecast where the average method is being used, which uses the average of historical data to make future predictions, the fitted values are given by the formula: $\hat{y}_t = \hat{C}$. Where $\hat{c}$ is computed by taking the average of all observations in the time series, including those that come after time t. In this scenario, a parameter must be estimated based on the data. The symbol above the C tells us that this is an estimation. When observations beyond time t are utilized to estimate C, the resulting fitted values are not considered accurate forecasts.

Residuals

Residuals in a time series model refer to the differences between the model's observed- and predicted values. The residuals represent the variability that the model does not explain or capture. When evaluating the model, the residuals can help detect remaining patterns in the data, improve the model's accuracy, and evaluate the overall goodness of fit of the model. A good forecasting method will yield residuals meeting Assumptions 1 and 2 which are essential, and Assumptions 3 and 4 which are recommendatory.

Assumption 1: The mean of residuals is zero. The forecast is biased when the mean of residuals is not zero.

Assumption 2: The residuals must be uncorrelated, which we can find by their ACF plot. When the residuals are correlated, information remains in the residuals, which must be captured in prediction. Plot ACF of the residuals is used to see if there are any significant correlations between the residuals at different lag values.

Assumption 3: The variance of residuals is almost constant (heteroscedasticity). By plotting the residuals (the difference between the predicted value and actual value) over time, it is easy to see if there are any patterns or trends in the data. One can also look for heteroscedasticity in the data, which occurs when the variance in the residuals changes over time. If there are patterns or heteroscedasticity in the residuals and data, it may indicate that the model does not capture the relevant information.

Assumption 4: The residuals are normally distributed. Plotting a histogram makes it easy to see whether the residuals are normally distributed. A normal distribution of residuals is desirable because it indicates that the model captures the relevant information in the data. If the residuals are skewed, it may indicate that the model is not accounting for some variability in - the data.

Ljung-Box Test

The Ljung-Box test is a statistical test to find out if there is significant autocorrelation or not in the residuals of the time series. The test evaluates a null hypothesis that the autocorrelation of a time series up to a certain lag is equal to zero. If the test result has a very low p-value, one which is below a certain threshold, it indicates that the null hypothesis should be rejected, therefore indicating that there is autocorrelation in the time series. If the p-value of the test, however, is larger than the threshold, it can be concluded that there is no autocorrelation in the residuals of the time series. In addition to testing for autocorrelation in the residuals, the Ljung-Box test also test if the residuals are behaving as a white noise series or not. Meaning that there is no remaining pattern in the data that the forecast model has not captured. In the context of this report, the value threshold for the p-value of the Ljung-Box test is set at 0.05.

Evaluating Forecast Accuracy

Evaluating the accuracy of the forecast is a crucial step when forecasting large amounts of data and using various models. The goal of evaluating the various models is to determine the reliability of the models and how well the models predict and fit the data.

It is a standard procedure to split the data into two parts: the training and test data while selecting models. This approach uses the training data to estimate any parameters of a forecasting technique and reserves the test data to assess the model's accuracy.

Cross-validation

One further step to evaluate forecast accuracy is through *time series cross-validation.* Time series cross-validation is the concept of predicting current data and comparing it with the actual observed result. For instance, if we want to evaluate the forecast model with the help of time series cross-validation, we can do it by predicting freight transported in 2021. We start by predicting the freight transported for 2021 with the help of various variables, and then we compare the prediction result with what the freight transported in 2021 was. By doing this, we can see whether the forecast model is accurate.

Aikake Information Criterion (AIC)

The Aikake Information Criterion (AIC) method evaluates the models and rejects the ones using the most independent variables. The reason for this is that too many independent variables, which in practice indicate the same factors, cause overfitting, and skew the dataset. A fewer number of variables gives a lower AIC score which is preferred over a higher score.

The mathematical formula for the AIC model is the following:

$$AIC = T\,log\left(\frac{SSE}{2}\right) + 2(k+2)$$

T: Number of observations

k: Number of predictors

SSE: Sum Squared Error

Aikake Information Criterion Corrected (AICc)

When the AIC method is based on a minimal number of observations, with few data points available for analysis, the number of predictors selected by the AIC tends to be too many. The AICc model- has been developed as a correction method for situations where the AIC method may not be as reliable due to a small number of observations or sample size.

The mathematical formula for the AICc model is the following:

$$AIC_c = AIC + \frac{2(K+2)(K+3)}{T-K-3}$$

Bayesian Information Criterion (BIC)

The BIC method provides a quantitative measure of the goodness of fit of a model, with lower values of BIC showing a better tradeoff between the fit of the model and its complexity. The BIC method penalizes too many independent variables in the same way as the AIC method, although BIC has even stricter thresholds for the parameters. When comparing two or more models, the one with the lower BIC is preferred which is the same as the AIC method.

The Bayesian Information Criterion (BIC) is particularly helpful in selecting an appropriate model that strikes a balance between the goodness of fit and complexity. On the other hand, the Akaike Information Criterion (AIC) is better suited for identifying the best model for forecasting future observations.

The mathematical formula for the BIC model is the following:

$$BIC = T\log\left(\frac{SSE}{T}\right) + (k+2)\log(T)$$

T = Number of observations

k = Number of predictors

SSE = Fit of the model

Prediction Intervals

A prediction interval is an interval in which the predicted value of a predicted variable is probable to fall within. Prediction intervals are typically made with different degrees of confidence, usually set at 95% or 99%. A 95% prediction interval is less broad than a 99% prediction interval since the 99% prediction interval is 99% certain that the predicted value of the variable falls within the interval.

Compared to a point forecast, prediction intervals provide additional information, indicating the range of the possible outcome. Having an interval provides the level of certainty or uncertainty of the prediction. In contrast, a wider interval generally shows that the prediction might be less conclusive, which can result from a lack of data or insufficient patterns of the data used. This project uses 80% and 95% prediction interval as follows:

$$\hat{y}_{T+H|T} \pm 1{,}96\hat{\sigma}_h$$

Where $\hat{\sigma}_h$ is the standard deviation estimate, and h is the h-step ahead forecast. The value 1.96 is the coverage probability of 95% under the standard normal distribution.

Collinearity

Collinearity is a statistical term that refers to a situation with a significant linear correlation between two or more predictor variables. The correlation between two or more predictor variables can lead to potential model overfitting, increasing standard errors of the regression coefficients, reducing the model's power, and increasing the difficulty of detecting significant effects. Collinearity also affects the possibility of interpreting a predictor variable's effects on the forecasted variable.

There are several different methods regarding dealing with collinearity. In this project, in the case of a significant correlation between two predictor variables, one is removed from the dataset to reduce the risk of potential overfitting.

3.2. Time series regression models

A time series regression model is a statistical method that utilizes historical values of a variable and other relevant factors, such as economic indicators or seasonal patterns, to examine and estimate the variable's future behavior over time. This technique is commonly employed in numerous fields, including finance, economics, and engineering, to generate predictions and aid decision-making processes. The entire R code used for the time series regression models in this project can be found in Appendix A.2.

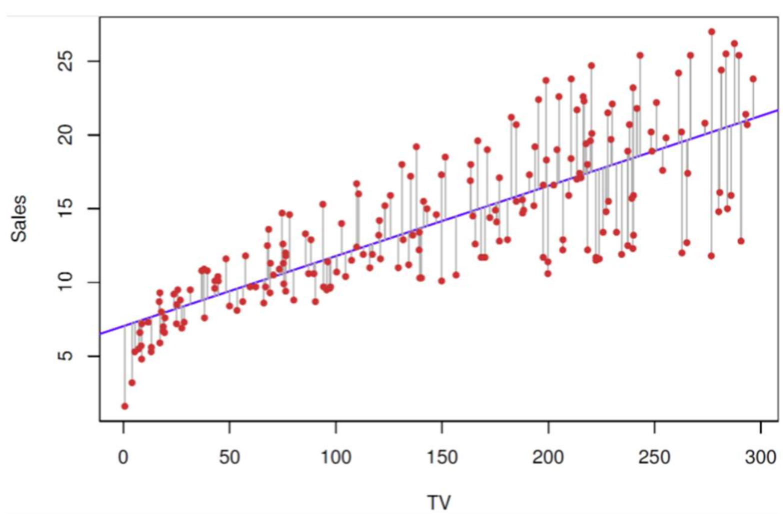
3.2.1. The linear model

In this subsection, the linear- and multiple linear regression models will be explained.

Simple Linear Regression

The basic linear regression model aims to find a linear relationship between only one independent variable (predictor variable) and the dependent variable (forecast variable). The equation is specified as $\gamma_t = \beta_0 + \beta_1 x_1 + \mathcal{E}_t$, where β_0 is the coefficient of the intercept displaying the value on the Y-axis when x_1 is 0, while β_1 is the coefficient displaying the slope. The $\mathcal{E}_t$, or the error, is the random error displaying the variance in γ_t that is not explained by x_t , and the deviation of the predictors from the line. Figure 17 shows the error as the vertical lines from the points to the regression line for each observation point.

Figure 17: *Linear regression model*



Note: A linear regression model that showcases the deviations of the predictor (James et al., 2021)

Multiple linear regression

When there is more than one predictor variable, we call it a multiple regression model. The multiple regression model is usually specified as:

$$y_t = \beta_0 + \beta_1 x_{1,t} + \beta_2 x_{2,t} + \cdots + \beta_k x_{k,t} + \varepsilon_{t,}$$

where β_0 is the intercept. $\beta_1 x_{1,t}$is the regression coefficient of the first independent variable, and following on, B_{2x2} is the regression coefficient of the second independent variable, and so forth. We use the multiple linear regression model in this project to find influential factors on logistic trends in Mexico and Turkey.

3.2.2. Selecting predictors

The process of selecting significant predictors is crucial for forecast accuracy. Predictors are variables that may impact the forecasted variable. Selecting predictors involves identifying candidate variables likely to be associated with the outcome variable and evaluating their usefulness in predicting. When forecasting the growth of logistics in an emerging economy country, demographic trends, financial aspects, political stability, consumption patterns, and import and export volumes are predictors that most likely will be significant. However, guessing what predictors are associated with the forecasted variable is precarious. Instead, using statistical modeling with mathematical calculations to determine the significance and correlation of the predictors will provide a more accurate forecast without the risk of misjudgment of the significance of predictors. Therefore, we use the time series regression model along with the PCA to find effective factors on logistic trends.

3.3. Exponential smoothing (ETS)

Exponential smoothing, abbreviated ETS, is a statistical method used for time series forecasting that calculates a weighted average of past observations, with more recent observations being given greater weight than older ones. The basic idea behind exponential smoothing is to provide a more accurate forecast by emphasizing recent observations and reducing the influence of older observations. The smoothing factor of the level, also known as the smoothing coefficient or alpha, $a,$ is a parameter used in exponential smoothing to control the weight given to the most recent observation. Another smoothing parameter called beta, $b,$ is also included in the model. Beta is the smoothing parameter for the trend in the model, with a higher value of beta indicating that a more recent trend in the data is being given more weight in the forecast. The smoothing factors determine how quickly the weights of the past observations decay as the time series moves forward. A smoothing factor of 0 gives equal weight to all past observations, while a smoothing factor of 1 gives even more weight to the most recent observation.

Exponential smoothing is one of the models that will be used most prominently in this project, and the code used can be found in Appendix A.3. The dataset which is to be forecasted consists of data from 1970 until 2021, and even though the data from, for instance, 1970 is important for the forecast and indirectly affects the forecast, the more recent data, such as freight transport in 2020 will have a larger effect on the forecast than freight transported in 1970 or 1980. With this stated, exponential smoothing is key for this project.

3.3.1. Simple exponential smoothing

Single exponential smoothing is the most basic variation of the three exponential smoothing models. Simple exponential smoothing does not consider trend or seasonality when forecasting, making it the most basic.

The formula for simple exponential smoothing is the following:

$$F_t = \alpha \, d_{t-1} + (1-\alpha) f_{t-1} \; 0 < \alpha < 1$$

Where:

- Alpha 'α' is the smoothing factor in value between 0 and 1. If alpha is close to 0, more weight will be given to older data, whereas more weight will be given to more recent data if alpha is closer to 1.
- F_t is the forecast variable F over time-period t. Which for our dataset will be freight transported for the forecasted time period.
- d_{t-1} is the actual value for the previous period t-1.
- f_{t-1} is the forecast for the previous period t-1 which represents the predicted level of time series at time t-1.

This formula is used to update the forecast for the current period based on the actual value for the previous period and the previous forecast. Making it a viable option for forecasting for the short- and mid-term.

3.3.2. Trend methods

Taking trends into consideration when forecasting with exponential smoothing increases the complexity of the forecast. An example of this method is Holt's linear trend method, more commonly known as 'double exponential smoothing'. Double exponential smoothing is a time series forecasting method that extends the simple exponential smoothing method to handle time series with a linear trend.

Double exponential smoothing uses two different components to make predictions: the level, which is also called the intercept, and the trend. The level component represents the expected value of the time series, while the trend component represents the expected rate of change of the data in the time series.

Double exponential smoothing incorporates two new equations into the ETS-formula. The level equation and the trend equation.

Level equation = $\ell_t = \alpha y_t + (1-\alpha)(\ell_{t-1} + b_{t-1})$

Trend equation = $b_t = \beta * (\ell_t - \ell_{t-1}) + (1 - \beta *) \, b_{t-1}$

The level equation represents the level of alpha from simple exponential smoothing and is the long-term average of the time series model. With more weight being given to older or newer data. The trend, however, represents the change in the slope of the time series model. The trend which is indicated by β, determines how much the slope of the model changes over time. Where a small value of β indicates that the slope does not vary much over time, and a high value of β indicates a larger variation of change. The trend can either be upwards or downwards, with the point of change in the trend denominated as the breakpoint.

3.3.3. A taxonomy of exponential smoothing

There are three trend components and three seasonal components, creating a taxonomy of the different combinations, as seen in Table 2. *None (N)*, in the trend department of the taxonomy, means no trend. *Additive (A)* is when there is, in fact, a trend in the data. The third one is *additive damped (A_d,)* implicating that the data has a damped trend, meaning it flattens over time.

Table 2: *Taxonomy of exponential smoothing – Trend and Seasonal components.*

Trend Component	Seasonal Component		
	N (None)	A (Additive)	M (Multiplicative)
N (None)	(N,N)	(N,A)	(N,M)
A (Additive)	(A,N)	(A,A)	(A,M)
Ad (Additive Damped)	(Ad,N)	(Ad,A)	(Ad,M)

3.3.4. Innovations state space models for exponential smoothing

The earlier described ways of using the trend and seasonal components when using exponential smoothing create a point forecast. However, it is also beneficial to create a forecast interval, depicting the possible range of the predicted outcome.

State space models comprise a measurement equation that characterizes the observed data and state equations that describe the evolution of unobserved components or states (such as level, trend, and seasonal variation) over time. This framework of a measurement equation and state equations gives rise to the term "state space models”.

State space modeling is a mathematical framework that considers underlying time-varying systems. The state space model consists of a measurement equation that relates the observed data to the underlying state variables and a set of state equations that describe how the state variables evolve over time. Exponential smoothing combined with the state space model can incorporate different types of exponential smoothing methods while also considering the potential presence of missing data or errors. This results in more accurate predictions with trend and seasonality patterns in the data.

To include the error in the equation of state space models, the error component is added to Tables 3 and 4. The error component can be additive, as illustrated in Table 3, or multiplicative, as illustrated in Table 4, thus adding two equations to each taxonomy section. A third letter is added to each taxonomy Table to differentiate between a model with additive and multiplicative errors and differentiate the models from the methods.

Table 3: *Additive Error Models*

Additive Error Models			
Trend	**Seasonal**		
	N	***A***	***M***
N	$y_t = \ell_{t-1.} + \varepsilon_t$ $\ell_t = \ell_{t-1} + \alpha\varepsilon_t$	$y_t = \ell_{t-1} + s_{t-m} + \varepsilon_t$ $\ell_t = \ell_{t-1} + \alpha\varepsilon_t$ $s_t = s_{t-m} + y\varepsilon_t$	$y_t = \ell_{t-1}s_{t-m} + \varepsilon_t$ $\ell_t = \ell_{t-1} + \alpha\varepsilon_t/s_{t-m}$ $s_t = s_{t-m} + y\varepsilon_t/\ell_{t-1}$
A	$y_t = \ell_{t-q} + b_{t-1} + \varepsilon_t$ $\ell_t = \ell_{t-1} + b_{t-1} + \alpha\varepsilon_t$ $b_t = b_{t-1} + \beta\varepsilon_t$	$y_t = \ell_{t-1} + b_{t-1} + s_{t-m} + \varepsilon_t$ $\ell_t = \ell_{t-1} + b_{t-1} + \alpha\varepsilon_t$ $b_t = b_{t-1} + \beta\varepsilon_t$ $s_t = s_{t-m} + y\varepsilon_t$	$y_t = (\ell_{t-1} + b_{t-1})s_{t-m} + \varepsilon_t$ $\ell_t = \ell_{t-1} + b_{t-1} + \alpha\varepsilon_t/s_{t-m}$ $b_t = b_{t-1} + \beta\varepsilon_t/s_{t-m}$ $s_t = s_{t-m} + y\varepsilon_t/(\ell_{t-1}+b_{t-1})$
$\boldsymbol{A_d}$	$y_t = \ell_{t-1} + \phi b_{t-1} + \varepsilon_t$ $\ell_t = \ell_{t-1} + \phi b_{t-1} + \alpha\varepsilon_t$ $b_t = \emptyset b_{t-1} + \beta\varepsilon_t$	$y_t = \ell_{t-1} + \phi b_{t-1} + s_{t-m} + \varepsilon_t$ $\ell_t = \ell_{t-1} + \phi b_{t-1} + \alpha\varepsilon_t$ $b_t = \phi b_{t-1} + \beta\varepsilon_t$ $s_t = s_{t-m} + y\varepsilon_t$	$y_t = (\ell_{t-1} + \phi b_{t-1})s_{t-m} + \varepsilon_t$ $\ell_t = \ell_{t-1} + \phi b_{t-1} + \alpha\varepsilon_t/s_{t-m}$ $b_t = \phi b_{t-1} + \beta\varepsilon_t/s_{t-m}$ $s_t = s_{t-m} + y\varepsilon_t/(\ell_{t-1} + \phi b_{t-1})$

Table 4: *Multiplicative Error Models*

Multiplicative Error Models			
Trend	**Seasonal**		
	N	***A***	***M***
N	$Y_t = \ell_{t-1}(1+\mathcal{E}_t)$ $\ell_t = \ell_{t-1}(1+\alpha\mathcal{E}_t)$	$y_t = (\ell_{t-1}+s_{t-m})(1+\varepsilon_t)$ $\ell_t = \ell_{t-1}+\alpha(\ell_{t-1}+s_{t-m})\mathcal{E}_t$ $s_t = s_{t-m}+y(\ell_{t-1}+s_{t-m})\mathcal{E}_t$	$y_t = \ell_{t-1}s_{t-m}(1+\mathcal{E}_t)$ $\ell_t = \ell_{t-1}(1+\alpha\mathcal{E}_t)$ $s_t = s_{t-m}(1+y\varepsilon_t)$
A	$y_t = (\ell_{t-1}+b_{t-1})(1+\mathcal{E}_t)$ $\ell_t = (\ell_{t-1}+b_{t-1})(1+\alpha\mathcal{E}_t)$ $b_1 = b_{t-1}+\beta(\ell_{t-1}+b_{t-1})\mathcal{E}_t$	$y_t = (\ell_{t-1}+b_{t-1}+s_{t-m})(1+\varepsilon_t)$ $\ell_t = \ell_{t-1}+b_{t-1}+\alpha(\ell_{t-1}+b_{t-1}+s_{t-m})\mathcal{E}_t$ $b_1 = b_{t-1}+\beta(\ell_{t-1}+b_{t-1}+s_{t-m})\mathcal{E}_t$ $s_t = s_{t-m}+y(\ell_{t-1}+b_{t-1}+s_{t-m})\mathcal{E}_t$	$y_t = (\ell_{t-1}+b_{t-1})s_{t-m}(1+\varepsilon_t)$ $\ell_t = (\ell_{t-1}+b_{t-1})(1+\alpha\varepsilon_t)$ $b_1 = b_{t-1}+\beta(\ell_{t-1}+b_{t-1})\mathcal{E}_t$ $s_t = s_{t-m}(1+y\mathcal{E}_t)$
A_d	$y_t = (\ell_{t-1}+\phi b_{t-1})(1+\mathcal{E}_t$ $\ell_t = (\ell_{t-1}+\phi b_{t-1})(1+\alpha\mathcal{E}$ $b_1 = \phi b_{t-1}+\beta(\ell_{t-1}+\phi b_{t-1})$	$y = (\ell_{t-1}+\phi b_{t-1}+s_{t-m})(1+\varepsilon_t)$ $\ell_t = \ell_{t-1}+\phi b_{t-1}+\alpha(\ell_{t-1}+\phi b_{t-1}+s_{t-m}$ $b_1 = \phi b_{t-1}+\beta(\ell_{t-1}+\phi b_{t-1}+s_{t-}$ $s_t = s_{t-m}+y(\ell_{t-1}+\phi b_{t-1}+s_{t-m})$	$y_t = (\ell_{t-1}+\phi b_{t-1})s_{t-m}(1+\varepsilon_t)$ $\ell_t = (\ell_{t-1}+\phi b_{t-1})(1+\alpha\varepsilon$ $b_1 = \phi b_{t-1}+\beta(\ell_{t-1}+\phi b_{t-1})$ $s_t = s_{t-m}(1+y\mathcal{E}_t)$

3.4. AutoRegressive Integrated Moving Average (ARIMA)

In this section the properties and attributes of the ARIMA forecasting model will be described. ARIMA models are efficient and common for forecasting time series. ARIMA models fit the project background and enables for high quality quantitative analyses. The entirety of the code used for this project can be found in Appendix A.4.

3.4.1. Stationarity and differencing

A stationary time series is a time series with values that is not dependent on time. Patterns like trends or seasonality within the time series cannot be considered stationary; hence the time components impact the value. A stationary time series has a constant mean value, which is necessary for ARIMA modeling.

Differencing is another crucial concept in the ARIMA model, used to make a non-stationary time series stationary. This is done by taking the difference between the observations in the time series. The differencing can be done several times to remove patterns showcasing a trend or seasonality. The number of times differencing is applied is determined by the autocorrelation plots of the series. Differencing is especially important regarding the ARIMA model related to the autoregressive and moving average components, which will be discussed in the upcoming sections of the report.

3.4.2. Autoregressive models

The general description of an autoregressive model is that the forecast of a variable is done by past values, or lag values, of itself. In the ARIMA model notation, the autoregressive component is denoted as “p” showcasing the total number of preceding values included in the model. The formula for autoregressive models with the order of p can be written as:

$$y_t = c + \emptyset_1 y_{t-1} + \emptyset_2 y_{t-2} + \cdots + \emptyset_p y_{t-1} + \varepsilon_{t,}$$

Autoregressive models are similar to multiple regression models. However, in the case of autoregressive models, they include lagged values of y_t as predictors. Because this formula is an autoregressive model with the order P, it can be referred to as an AR(p) model.

Autoregressive models are considered valuable when the time series depicts autocorrelation. The autoregressive component captures the correlation of the predicted variable to its past values. By incorporating the autoregressive component into the ARIMA model, trends and patterns can be captured in the time series that are not explained by the other components.

3.4.3. Moving average models

Unlike the autoregressive model, the moving average model considers the residual errors of the forecasted variable and its past values as follows:

$$y_t = c + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \cdots + \theta_q \varepsilon_{t-q},$$

In this formula, the denotation ε_t is referred to as the white noise of the model. This model is also commonly called a moving average (MA(q) Model) of order q. The values of ε_t are not observed, therefore, it is not considered a regression model.

In other words, the relationship of the forecast variable and its lagged values are captured by the lagged values of the error. The moving average component in the ARIMA model is denoted as “q”, and the value it takes is based on the number of lagged residuals considered in the time series. The moving average is considered a better option than the autoregressive model when handling non-stationary time series. Moreover, the moving average component has a significant ability to capture short-term variations in the time series.

3.4.4. Non-seasonal ARIMA models

Non-seasonal ARIMA models combine the two previously described models, namely the autoregressive and the moving average, using the differencing concept.

Non-seasonal ARIMA models are used when there is no clear seasonality in the data, These models can be specified using the notation ARIMA(p,d,q) by the following formulation:

$$y'_t = c + \phi_1 y'_{t-1} + \cdots + \phi_p y'_{t-p} + \theta_1 \varepsilon_{t-q} + \epsilon_t,$$

In this formula, the y'_t is the time series after being differenced. One-time differencing is the most common variant; however, it is possible to difference a time series several more times. The values afterward, also denoted as the predictors, include lagged values and errors. The formula above is an example of the ARIMA (p,d,q) model. The p is denoted as the order of the autoregressive part (AR), the d indicates the degree of differencing which is involved in the model (I), and the q is the order of the moving average (MA).

In the context of this project, in addition to ETS method, the non-seasonal ARIMA model will be used to forecast freight transport for a short- and mid-term perspective for the two emerging economy countries with the help of the historical data from the OECD.

3.4.5. Initial states in ARIMA

The initial states, often given by forecasting programs, is a crucial part of the forecasting concept. The initial states are the starting values that are used by the model when generating the entire forecast. These values are based on the available data and are used to institute the internal state of the model. These values are also critical for the model’s ability to capture any form of patterns in the data and are a sign of the reliability of the forecast. We let the RStudio optimize the initial states of ARIMA method.

4. RESULTS

This section of the report will present the results for both research questions one (RQ1) and two (RQ2).

4.1. Results for "What factors can influence the logistics sector in an emerging economy country?"

Our first research question is what factors can influence the logistics sector in an emerging economy country. To address it, we use PCA method and time series regression analysis. First, we reduce the number of independent variables through data preparation. Before implementing PCA for data reduction in R-studio, superfluous and irrelevant variables that would have no effect on the logistics industries in the emerging economy countries were also removed, resulting in the current dataset of 250 variables.

After completing the data preparation, the next step is to proceed with the principal component analysis (PCA) of the dataset. As described in the methodology section, PCA measures the variance-covariance relationship of variables in the dataset, with the goal of reducing the dimensionality of large datasets. Before PCA, each variable in the cleaned dataset (250 variables) is normally standardized which has a variance of one. A common rule of thumb is to select the number of principal components that explain at least 70-80% of the total variation in the data. This can be visually assessed by plotting a scree plot, which shows the variance explained by each component in decreasing order. The number of components to keep is generally chosen at the "elbow" of the plot, where adding more components results in only a marginal increase in the total variation explained.

For Turkey, three principal components are found after proceeding with principal component analysis, namely PC1, PC2, and PC3, containing a total of 43 variables. As for Mexico, two principal components were found, PC1 and PC2 containing a total of 12 variables. These principal components are chosen because they cover about 70% of the total variance (Figures 18 and 19) as well as the "elbow" of the scree plot plots (Figures 20 and 21). Furthermore, all the various variables are designated unique metadata indicators, which will be explained more in the upcoming section. The following section cover the composition of each principal component for both Mexico and Turkey, as well as the formulations for them.

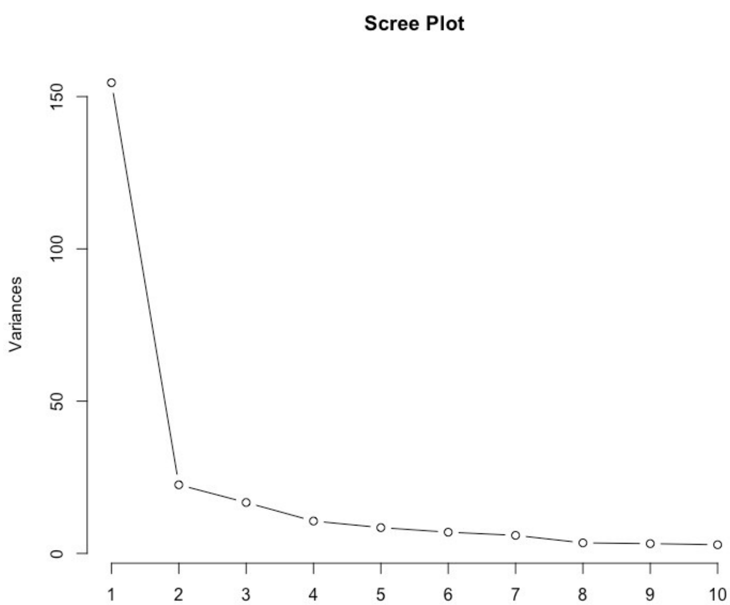
Figure 18: *Principal Components for Mexico's cleaned dataset*

```
Importance of components:
                           PC1     PC2     PC3    PC4     PC5     PC6     PC7     PC8     PC9    PC10    PC11    PC12    PC13    PC14    PC15    PC16
Standard deviation     12.4326 4.74771 4.08945 3.2604 2.91112 2.64015 2.43547 1.86123 1.79243 1.68843 1.64777 1.49639 1.43719 1.30864 1.26955 1.16472
Proportion of Variance  0.5991 0.08737 0.06482 0.0412 0.03285 0.02702 0.02299 0.01343 0.01245 0.01105 0.01052 0.00868 0.00801 0.00664 0.00625 0.00526
Cumulative Proportion   0.5991 0.68647 0.75129 0.7925 0.82534 0.85236 0.87535 0.88877 0.90123 0.91228 0.92280 0.93148 0.93949 0.94612 0.95237 0.95763
                          PC17   PC18    PC19   PC20    PC21    PC22    PC23    PC24    PC25   PC26    PC27    PC28    PC29    PC30    PC31    PC32
Standard deviation     1.09876 1.0162 0.94908 0.9369 0.86266 0.84484 0.81591 0.76413 0.74843 0.7183 0.62640 0.61817 0.56607 0.54463 0.52536 0.52371
Proportion of Variance 0.00468 0.0040 0.00349 0.0034 0.00288 0.00277 0.00258 0.00226 0.00217 0.0020 0.00152 0.00148 0.00124 0.00115 0.00107 0.00106
Cumulative Proportion  0.96231 0.9663 0.96980 0.9732 0.97609 0.97885 0.98144 0.98370 0.98587 0.9879 0.98939 0.99087 0.99211 0.99326 0.99433 0.99540
                          PC33    PC34    PC35    PC36    PC37   PC38    PC39   PC40    PC41      PC42
Standard deviation     0.46435 0.41860 0.40008 0.37325 0.36502 0.3596 0.31335 0.2786 0.24335 5.521e-15
Proportion of Variance 0.00084 0.00068 0.00062 0.00054 0.00052 0.0005 0.00038 0.0003 0.00023 0.000e+00
Cumulative Proportion  0.99623 0.99691 0.99753 0.99807 0.99859 0.9991 0.99947 0.9998 1.00000 1.000e+00
```

Figure 19: *Principal Components for Turkey's cleaned dataset*

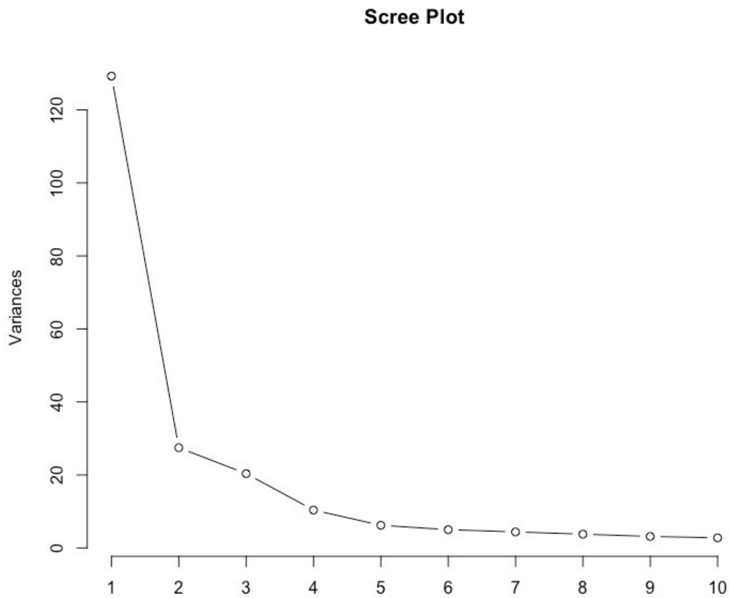
```
Importance of components:
                           PC1    PC2     PC3     PC4     PC5     PC6     PC7     PC8     PC9    PC10    PC11    PC12    PC13    PC14    PC15    PC16
Standard deviation     11.3689 5.2407 4.51227 3.22240 2.49138 2.23720 2.09483 1.94156 1.77474 1.66448 1.57351 1.47251 1.36756 1.30818 1.25803 1.15768
Proportion of Variance  0.5524 0.1174 0.08701 0.04438 0.02653 0.02139 0.01875 0.01611 0.01346 0.01184 0.01058 0.00927 0.00799 0.00731 0.00676 0.00573
Cumulative Proportion   0.5524 0.6697 0.75674 0.80112 0.82764 0.84903 0.86778 0.88389 0.89735 0.90919 0.91977 0.92904 0.93703 0.94435 0.95111 0.95684
                          PC17    PC18    PC19    PC20    PC21    PC22   PC23    PC24    PC25    PC26    PC27    PC28    PC29    PC30   PC31    PC32
Standard deviation     1.13861 1.03891 0.98566 0.95339 0.85838 0.84554 0.7799 0.70739 0.67745 0.65587 0.58550 0.54416 0.50238 0.47807 0.4599 0.44112
Proportion of Variance 0.00554 0.00461 0.00415 0.00388 0.00315 0.00306 0.0026 0.00214 0.00196 0.00184 0.00147 0.00127 0.00108 0.00098 0.0009 0.00083
Cumulative Proportion  0.96238 0.96699 0.97114 0.97503 0.97817 0.98123 0.9838 0.98597 0.98793 0.98977 0.99123 0.99250 0.99358 0.99455 0.9955 0.99629
                          PC33    PC34   PC35   PC36    PC37    PC38   PC39    PC40    PC41      PC42
Standard deviation     0.39969 0.35790 0.3432 0.3416 0.31294 0.28675 0.2629 0.25246 0.18223 5.049e-15
Proportion of Variance 0.00068 0.00055 0.0005 0.0005 0.00042 0.00035 0.0003 0.00027 0.00014 0.000e+00
Cumulative Proportion  0.99697 0.99752 0.9980 0.9985 0.99894 0.99929 0.9996 0.99986 1.00000 1.000e+00
```

Figure 20: *Scree Plot for PCA Mexico*



Note: Of the total variance, PC1, and PC2 correspond to 70% of the total variance. Total variance equals to 250, where PC1 and PC2 make up 175 of those 250, i.e., 70%. The elbow shape of the graph indicates a quick decline in variance importance after PC2.

Figure 21: *Scree Plot for PCA Turkey*



Note: Of the total variance, PC1, PC2, and PC3 correspond to 70% of the total variance. Total variance equals to 250, where PC1, PC2, and PC3 make up 175 of those 250, i.e., 70%. The elbow shape of the graph indicates a quick decline in variance importance after PC3.

For Mexico, the following two principal components are extracted. PC1 and PC2. These two are chosen because they cover 70% of the variance.

Independant variable	PC1: GDP Component
NE.CON.TOTL.KD	0.0801045490446743
NE.CON.TOTL.KN	0.0800870368731901
NY.GDP.MKTP.KD	0.0801540311468885
NY.GDP.MKTP.KN	0.0801697253238944

The above variable abbreviations are further explained in Table 5.

Formulation for PC1

$$PC1 = 0.0801045490446743 * NE.CON.TOTL.KD + 0.0800870368731901 * NE.CON.TOTL.KN + 0.0801540311468885 * NY.GDP.MKTP.KD + 0.0801697253238944 * NY.GDP.MKTP.KN$$

Index	PC2: Demographic and Logistical Component
NE.GDI.STKB.CD	0.134117368619207

NE.CON.TOTL.ZS	0.1263646000657
NV.MNF.FBTO.ZS.UN	0.150690382837373
TX.VAL.MANF.ZS.UN	0.126489608669901
TX.VAL.MRCH.HI.ZS	0.181420426475349
SP.POP.0014.TO	0.115807546437612
PX.REX.REER	0.101550222262554
SP.RUR.TOTL.ZG	0.103106794943553

The above variable abbreviations are further explained in Table 5.

Formulation for PC2:

$$PC2 = 0.134117368619207 * NE.GDI.STKB.CD + 0.1263646000657 * NE.CON.TOTL.ZS + 0.150690382837373 * NV.MNF.FBTO.ZS.UN + 0.126489608669901 * TX.VAL.MANF.ZS.UN + 0.181420426475349 * TX.VAL.MRCH.HI.ZS + 0.115807546437612 * SP.POP.0014.TO + 0.101550222262554 * PX.REX.REER + 0.103106794943553 * SP.RUR.TOTL.ZG$$

The output of the regression model for Mexico in Figure 22 from R indicates that PC1 and PC2 both are statistically significant, with both coefficient estimates having very low p-values in the column furthest to the right, as well as three stars which indicate high relevance to the dependent variable.

Figure 22: *Regression Model for Mexico using PCs*

```
Residuals:
   Min     1Q Median     3Q    Max 
-25196 -11597  -3937  10421  28815 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept) 235657.7     2267.3 103.939  < 2e-16 ***
PC1          19032.8      595.8  31.947  < 2e-16 ***
PC2           2757.6      322.3   8.556 1.74e-10 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 14690 on 39 degrees of freedom
Multiple R-squared:  0.9648,	Adjusted R-squared:  0.963 
F-statistic: 535.2 on 2 and 39 DF,  p-value: < 2.2e-16
```

Regression Formula: $Y = 19032.8 * PC1 + 2757.6 * PC2 + 235657.7 + Residuals$,

where Y is the prediction variable of Mexico's total freight transportation measured in million ton-km.

Notes: PC1: GDP Component, PC2: Mixed Demographic and Logistical Component

The regression model for Mexico also has the following attributes:

- A value of 0.9648 for Multiple R-squared which is considered high. It indicates that the variation of the dependent variable, i.e., Freight transported, can be explained up to 96.48% by PC1 and PC2
- An Adjusted R-squared value of 0.963 i.e., 96.3%. The adjusted r-squared value is somewhat lower than the multiple r-squared because the adjusted r-squared is more sensitive to the correlation of PC2 regarding the dependent variable of freight transportation. The adjusted R-squared also takes degree of freedom of the model into consideration, whereas multiple R-squared does not. Further explaining why the values differ.
- A model p-value of $2.2 * 10^{-16}$ indicates a very good fit of model.
- An F-statistic value of 535.2 on 2 and 39DF, meaning an analysis of variance of two independent variables with 39 observations. This value is above the significance level of 0.05, meaning that we can conclude that the model is significant.

For Turkey, the following three principal components are extracted successfully. PC1, PC2, and PC3. These three are chosen because they cover 70% of the total variance.

Independent variable	PC1: Logistical Component
TX.VAL.MRCH.WL.CD	0.0866453862481206
TX.QTY.MRCH.XD.WD	0.0867519059783125
DT.DOD.DLXF.CD	0.0863595669598187
TX.VAL.MRCH.XD.WD	0.0866511279069433
TX.VAL.MRCH.CD.WT	0.0865498834804234
NY.GDP.PCAP.KD	0.0861917953215151
NV.IND.TOTL.CD	0.086109398552134
TM.VAL.MRCH.XD.WD	0.0860799454349721
TM.VAL.MRCH.CD.WT	0.086417322593013
SP.POP.1564.TO	0.0861596128607706
EN.URB.MCTY	0.0872148801399272
BM.GSR.TOTL.CD	0.0867367967672794
EN.URB.LCTY	0.0873873666234652
BX.GSR.TOTL.CD	0.0870354862139321
BM.GSR.NFSV.CD	0.0865949211105418
BM.GSR.FCTY.CD	0.0861025433975739

TM.VAL.SERV.CD.WT	0.0863464900064815
TM.VAL.MRCH.WL.CD	0.086014945670998
TM.QTY.MRCH.XD.WD	0.0874358168802998
NY.GDP.PCAP.KN	0.0861832271123425
NV.IND.MANF.CD	0.0869528521711468
NE.EXP.GNFS.CD	0.0870619246673187
DT.DOD.DPPG.CD	0.0866111901918288
NE.IMP.GNFS.CD	0.0866033856700947
BX.GSR.MRCH.CD	0.0868572666550789
BM.GSR.GNFS.CD	0.0865537533503046
DT.DOD.DECT.CD	0.0867943504182405
BX.GSR.GNFS.CD	0.0869723354296957
BM.GSR.MRCH.CD	0.0863083772733233

The above variable abbreviations are further explained in Table 6.

Formulation for PC1:

$$PC1 = 0.0866453862481206 * TX.VAL.MRCH.WL.CD + 0.0867519059783125 * TX.QTY.MRCH.XD.WD + 0.0863595669598187 * DT.DOD.DLXF.CD + 0.0866511279069433 * TX.VAL.MRCH.XD.WD + 0.0865498834804234 * TX.VAL.MRCH.CD.WT + 0.0861917953215151 * NY.GDP.PCAP.KD + 0.086109398552134 * NV.IND.TOTL.CD + 0.0860799454349721 * TM.VAL.MRCH.XD.WD + 0.086417322593013 * TM.VAL.MRCH.CD.WT + 0.0861596128607706 * SP.POP.1564.TO + 0.0867367967672794 * BM.GSR.TOTL.CD + 0.0873873666234652 * EN.URB.LCTY + 0.0870354862139321 * BX.GSR.TOTL.CD + 0.0865949211105418 * BM.GSR.NFSV.CD + 0.0861025433975739 * BM.GSR.FCTY.CD + 0.0863464900064815 * TM.VAL.SERV.CD.WT + 0.086014945670998 * TM.VAL.MRCH.WL.CD + 0.0874358168802998 * TM.QTY.MRCH.XD.WD + 0.0861832271123425 * NY.GDP.PCAP.KN + 0.0869528521711468 * NV.IND.MANF.CD + 0.0870619246673187 * NE.EXP.GNFS.CD + 0.0866111901918288 * NE.IMP.GNFS.CD + 0.0868572666550789 * BX.GSR.MRCH.CD + 0.0865537533503046 * BM.GSR.GNFS.CD + 0.0867943504182405 * DT.DOD.DECT.CD + 0.0869723354296957 * BX.GSR.GNFS.CD + 0.0863083772733233 * BM.GSR.MRCH.CD$$

Independent variable	PC2: Economic Component
TM.VAL.MRCH.HI.ZS	0.143413330052424
TM.VAL.MANF.ZS.UN	0.173071996028362

TM.VAL.AGRI.ZS.UN	0.121196392943651
SP.POP.0014.TO	0.152979696767252
FR.INR.DPST	0.112178917042538
BG.GSR.NFSV.GD.ZS	0.1363516067692
NE.CON.TOTL.ZS	0.136504682155388
TX.VAL.MRCH.HI.ZS	0.160398783258507
TX.VAL.MANF.ZS.UN	0.138281928108262
NY.ADJ.DKAP.GN.ZS	0.11311832351479
NY.TRF.NCTR.CD	0.101566671387288
BN.TRF.CURR.CD	0.101563619070735

The above variable abbreviations are further explained in Table 6.

Formulation for PC2:

$$PC2 = 0.143413330052424 * TM.VAL.MRCH.HI.ZS + 0.173071996028362 * TM.VAL.MANF.ZS.UN + 0.121196392943651 * TM.VAL.AGRI.ZS.UN + 0.152979696767252 * SP.POP.0014.TO + 0.112178917042538 * FR.INR.DPST + 0.1363516067692 * BG.GSR.NFSV.GD.ZS + 0.136504682155388 * NE.CON.TOTL.ZS + 0.160398783258507 * TX.VAL.MRCH.HI.ZS + 0.138281928108262 * TX.VAL.MANF.ZS.UN + 0.11311832351479 * NY.ADJ.DKAP.GN.ZS + 0.101566671387288 * NY.TRF.NCTR.CD + 0.101563619070735 * BN.TRF.CURR.CD$$

Independent variable	PC3: Travel Services Component
TX.VAL.TRVL.ZS.WT	0.165874906486119
BX.GSR.TRVL.ZS	0.164060353834971

The above variable abbreviations are further explained in Table 6.

Formulation for PC3:

$$PC3 = 0.165874906486119 * TX.VAL.TRVL.ZS.WT + 0.164060353834971 * BX.GSR.TRVL.ZS$$

The output of the regression model for Turkey in Figure 23 from R indicates that PC1, PC2, and PC3 are statistically significant, with both coefficient estimates having very low p-values in the column furthest to the right, as well as three stars which indicate high relevance to the dependent variable.

Figure 23: *Regression Model for Turkey using PCs*

```
Residuals:
   Min     1Q Median     3Q    Max
-70045 -12621  -1091  25266  55417

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   193401       4910  39.390  < 2e-16 ***
PC1            26739       1886  14.176  < 2e-16 ***
PC2            -6301       1384  -4.555 5.28e-05 ***
PC3           -18194       3942  -4.615 4.38e-05 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 31820 on 38 degrees of freedom
Multiple R-squared:  0.885,     Adjusted R-squared:  0.8759
F-statistic: 97.47 on 3 and 38 DF,  p-value: < 2.2e-16
```

Regression Formula: $Y = 26739 * PC1 - 6301 * PC2 - 18194 * PC3 + 193401 + Residuals$, where Y is the prediction variable of Turkey's total freight transportation measured in million ton-km.

Notes: PC1: Logistical Component, PC2: Economic Component, PC3: Travel Services Component

The regression model contains the listed attributes:

- A large multiple R-squared value of 0.885, implies that PC1, PC2, and PC3 can explain 88.5% of the variation in freight transportation.
- The adjusted R-squared value of 0.8759 indicates that PC1, PC2, and PC3 explain 87.59% of the variation in freight transportation while accounting for the degrees of freedom in the model.
- A P-value of $2.2 * 10^{-16}$ implies a very good fit of the model.
- The F-statistic of 97.47 on 3 and 38 DF, meaning an analysis of the three predictors, or independent variables, with 38 observations. The value is above the threshold of 0.05, indicating the significance of the model.

4.1.1. Mexico

In this section, the result for RQ1 for Mexico is presented.

Interpretation of the PCs

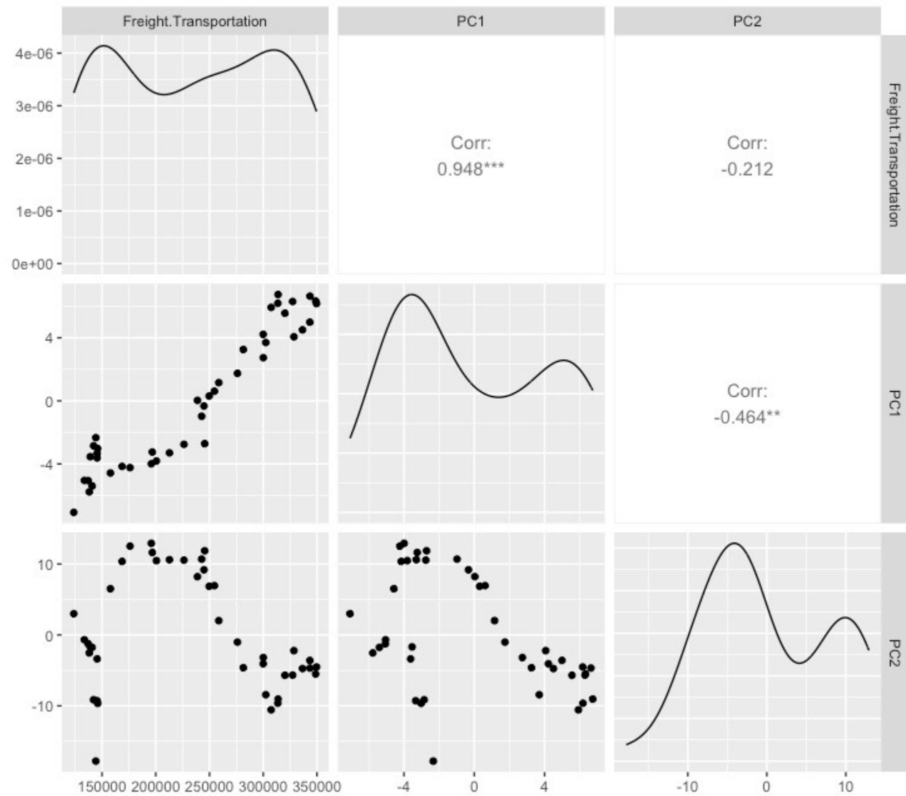
Using PCA and time-series regression analysis, as previously explained in the result section, the key variables affecting the logistics sector in Mexico are divided into two principal components. Namely PC1 and PC2. Table 5 illustrates the composition of each PC. The result of the principal component analysis show that key factors (variables) affecting the logistics sector in Mexico are factors such as economic factors, demographic factors, and factors regarding logistics. For PC1 it is exclusively economic factors such as *GDP*. For PC2 it is a combination all three factors such as *rural population growth* for demographic, *changes in inventories* for logistics, as well *final consumption expenditure* for economic. Table 5 further illustrates the total variable composition of the PCs, with corresponding indicators with their description also available.

Table 5: *PCA Variable Composition Mexico*

Indicator	**Indicator Description**
PC1: GDP Component	
NE.CON.TOTL.KD	*Final consumption expenditure (constant 2015 US$)*
NE.CON.TOTL.KN	*Final consumption expenditure (constant LCU)*
NY.GDP.MKTP.KD	*GDP (constant 2015 US$)*
NY.GDP.MKTP.KN	*GDP (constant LCU)*
PC2: Mixed Demographic and Logistical Component	
NE.GDI.STKB.CD	*Changes in inventories (current US$)*
NE.CON.TOTL.ZS	*Final consumption expenditure (% of GDP)*
NV.MNF.FBTO.ZS.UN	*Food, beverages and tobacco (% of value added in manufacturing)*
TX.VAL.MANF.ZS.UN	*Manufactures exports (% of merchandise exports)*
TX.VAL.MRCH.HI.ZS	*Merchandise exports to high-income economies (% of total merchandise exports)*
SP.POP.0014.TO	*Population ages 0-14, total*
PX.REX.REER	*Real effective exchange rate index (2010 = 100)*
SP.RUR.TOTL.ZG	*Rural population growth (annual %)*

Furthermore, the scatterplot matrix in Figure 24 depicts each PCs correlation with the dependent variable of freight transportation. PC1 clearly exhibits a very strong correlation with the freight transportation variable. However, interesting to note is that PC2 does not have a strong correlation to the freight transportation variable. It does, however, possess a very strong correlation with PC1, which indicates that the variables in PC2 have a very strong impact on how the variables in PC1 behaves. A way to notice this is to either look at the correlation coefficients shown in Figure 25, but also by looking at each scatterplot. By doing this, it is clearly shown that PC1 has a more defined correlation to the dependent variable than that of PC2.

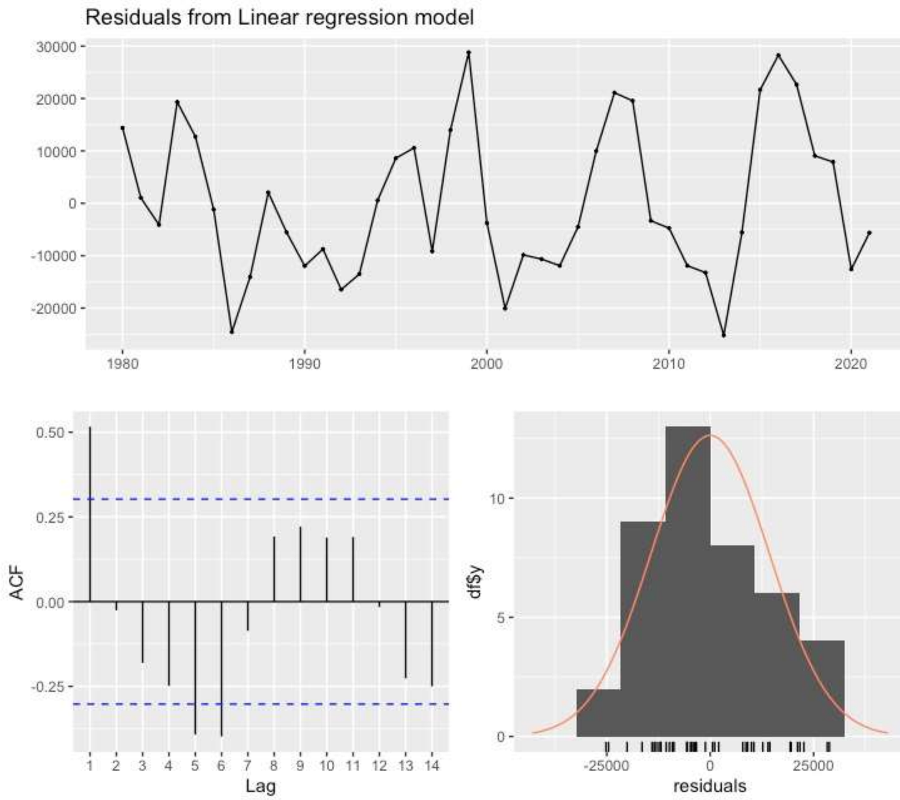
Figure 24: *Scatterplot Matrix for Mexico*



Residual properties

Figure 25 illustrates the characteristics of the residuals for the linear regression model. The residuals are distributed evenly around zero, with a mean of zero, indicating that the model is unbiased. The histogram depicts that the residuals are normally distributed, further suggesting that the model is a good fit. However, the ACF plot indicates a potential threat of autocorrelation in the data, as three lag values exceed the boundaries.

Figure 25: *Residuals from Linear Regression model for Mexico*



4.1.2. Turkey

In this section, the results regarding research question one for Turkey will be presented.

Interpretation of the PCAs

After the data reduction process and the categorization of the variables into the three different Principal components, the variables in Table 6 are left. The table also contains a complete description of the indicators. The key indicators can be categorized into three major areas: economic factors, demographic factors, and logistic-related factors.

PC1 contains a mix of all three areas and a total of 29 variables, with economic factors such as *GDP per capita*, *Primary income payments,* and *external debt stocks,* demographic factors such as *Population in largest cities, population ages 15-64, and Population in urban agglomerations of more than 1 million,* and logistic-related factors such as *export volume index, export value index, and import of goods and services.*

PC2 consists of 12 variables, also containing all the major categorization areas. PC2 includes economic factors such as *Deposit interest rate and Net secondary income*, demographic factors such as *population ages 0-14,* and logistical factors like *Merchandise exports to high-income economies* and *Manufactures exports.*

PC3 consists of only two variables related to logistics: Travel *services (% of commercial service exports)* and *Travel services (% of service exports, BoP).*

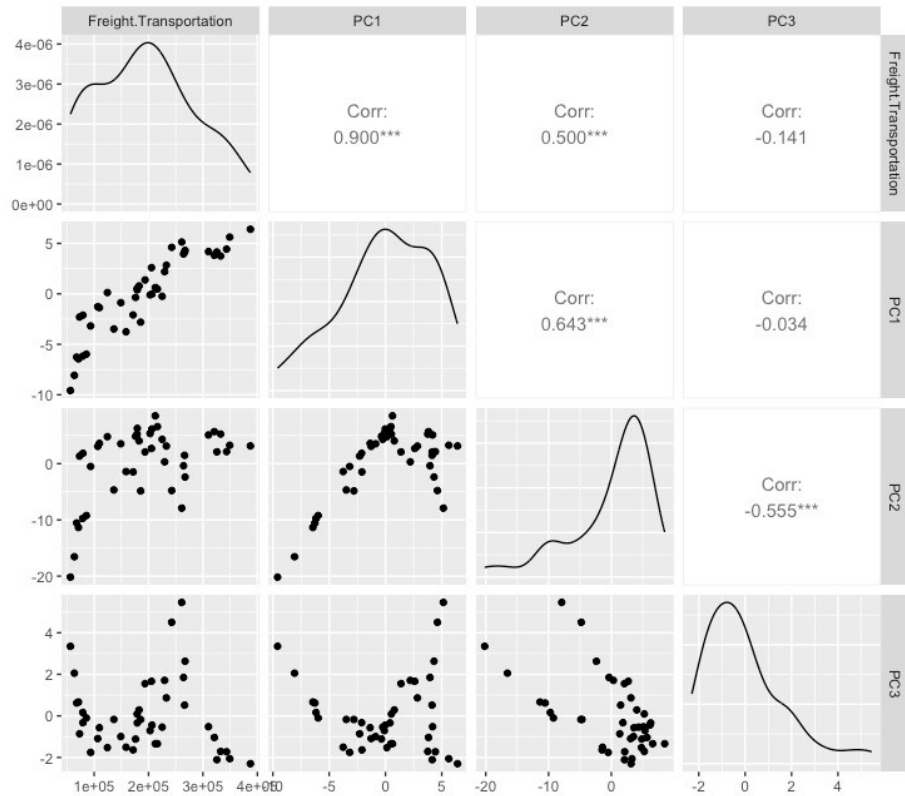
Table 6: *PCA Variable Composition Turkey*

Indicator	Indicator Description
PC1: Logistical Component	
TX.VAL.MRCH.WL.CD	*Merchandise exports by the reporting economy (current US$)*
TX.QTY.MRCH.XD.WD	*Export volume index (2000 = 100)*
DT.DOD.DLXF.CD	*External debt stocks, long-term (DOD, current US$)*
TX.VAL.MRCH.XD.WD	*Export value index (2000 = 100)*
TX.VAL.MRCH.CD.WT	*Merchandise exports (current US$)*
NY.GDP.PCAP.KD	*GDP per capita (constant 2015 US$)*
NV.IND.TOTL.CD	*Industry (including construction), value added (current US$)*
TM.VAL.MRCH.XD.WD	*Import unit value index (2015 = 100)*
TM.VAL.MRCH.CD.WT	*Merchandise imports (current US$)*
SP.POP.1564.TO	*Population ages 15-64, total*
EN.URB.MCTY	*Population in urban agglomerations of more than 1 million*
BM.GSR.TOTL.CD	*Imports of goods, services and primary income (BoP, current US$)*
EN.URB.LCTY	*Population in largest city*
BX.GSR.TOTL.CD	*Exports of goods, services and primary income (BoP, current US$)*
BM.GSR.NFSV.CD	*Service imports (BoP, current US$)*
BM.GSR.FCTY.CD	*Primary income payments (BoP, current US$)*
TM.VAL.SERV.CD.WT	*Commercial service imports (current US$)*
TM.VAL.MRCH.WL.CD	*Merchandise imports by the reporting economy (current US$)*
TM.QTY.MRCH.XD.WD	*Import volume index (2000 = 100)*
NY.GDP.PCAP.KN	*GDP per capita (constant LCU)*
NV.IND.MANF.CD	*Manufacturing, value added (current US$)*
NE.EXP.GNFS.CD	*Exports of goods and services (current US$)*
DT.DOD.DPPG.CD	*External debt stocks, public and publicly guaranteed (PPG) (DOD, current US$)*
NE.IMP.GNFS.CD	*Imports of goods and services (current US$)*

BX.GSR.MRCH.CD	*Goods exports (BoP, current US$)*
BM.GSR.GNFS.CD	*Imports of goods and services (BoP, current US$)*
DT.DOD.DECT.CD	*External debt stocks, total (DOD, current US$)*
BX.GSR.GNFS.CD	*Exports of goods and services (BoP, current US$)*
BM.GSR.MRCH.CD	*Goods imports (BoP, current US$)*
PC2: Economic Component	
TM.VAL.MRCH.HI.ZS	*Merchandise imports from high-income economies (% of total merchandise imports)*
TM.VAL.MANF.ZS.UN	*Manufactures imports (% of merchandise imports)*
TM.VAL.AGRI.ZS.UN	*Agricultural raw materials imports (% of merchandise imports)*
SP.POP.0014.TO	*Population ages 0–14, total*
FR.INR.DPST	*Deposit interest rate (%)*
BG.GSR.NFSV.GD.ZS	*Trade in services (% of GDP)*
NE.CON.TOTL.ZS	*Final consumption expenditure (% of GDP)*
TX.VAL.MRCH.HI.ZS	*Merchandise exports to high-income economies (% of total merchandise exports)*
TX.VAL.MANF.ZS.UN	*Manufactures exports (% of merchandise exports)*
NY.ADJ.DKAP.GN.ZS	*Adjusted savings: consumption of fixed capital (% of GNI)*
NY.TRF.NCTR.CD	*Net secondary income (Net current transfers from abroad) (current US$)*
BN.TRF.CURR.CD	*Net secondary income (BoP, current US$)*
PC3: Travel Services Component	
TX.VAL.TRVL.ZS.WT	*Travel services (% of commercial service exports)*
BX.GSR.TRVL.ZS	*Travel services (% of service exports, BoP)*

The correlation between freight transportation and PC1, PC2, and PC3 is illustrated in Figure 26. The strongest correlation to the freight transportation variable is seen concerning PC1 with a correlation of 0.9 and three stars, showing a significant correlation between the two. The correlation between PC2 and the freight transportation variable is 0.5, followed by three stars, indicating a strong positive relationship. The relationship between PC3 and the freight transported variable is -0.141 followed by no stars. However, a correlation of -0.555, followed by three stars between PC3 and PC2, demonstrates the significance of the PC3 for the model.

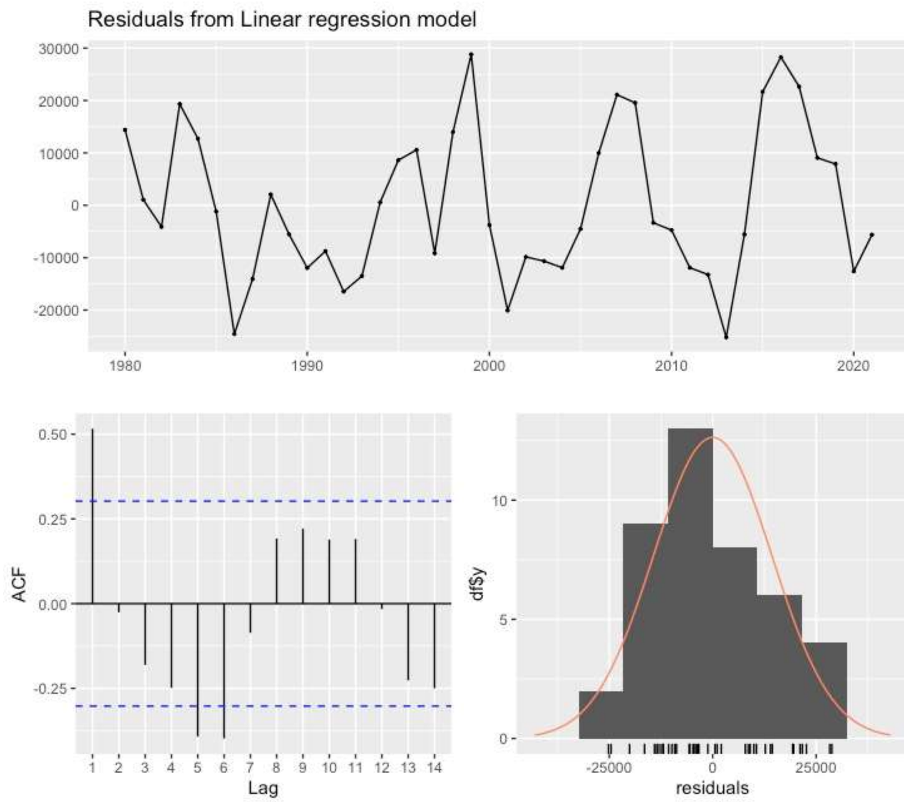
Figure 26: *Scatterplot Matrix for Turkey*



Residual properties

The properties of the residuals for the linear regression model are depicted in Figure 27. The values are equally spread around the zero, resulting in a mean value of zero, indicating that the model is unbiased. The histogram resembles somewhat of a normal distribution, implying a good fit of the model. However, the spikes at lag values 5 and 6 exceed the boundaries in the ACF plot, showcasing a threat of potential autocorrelation in the data.

Figure 27: *Residuals from Linear Regression model for Turkey*



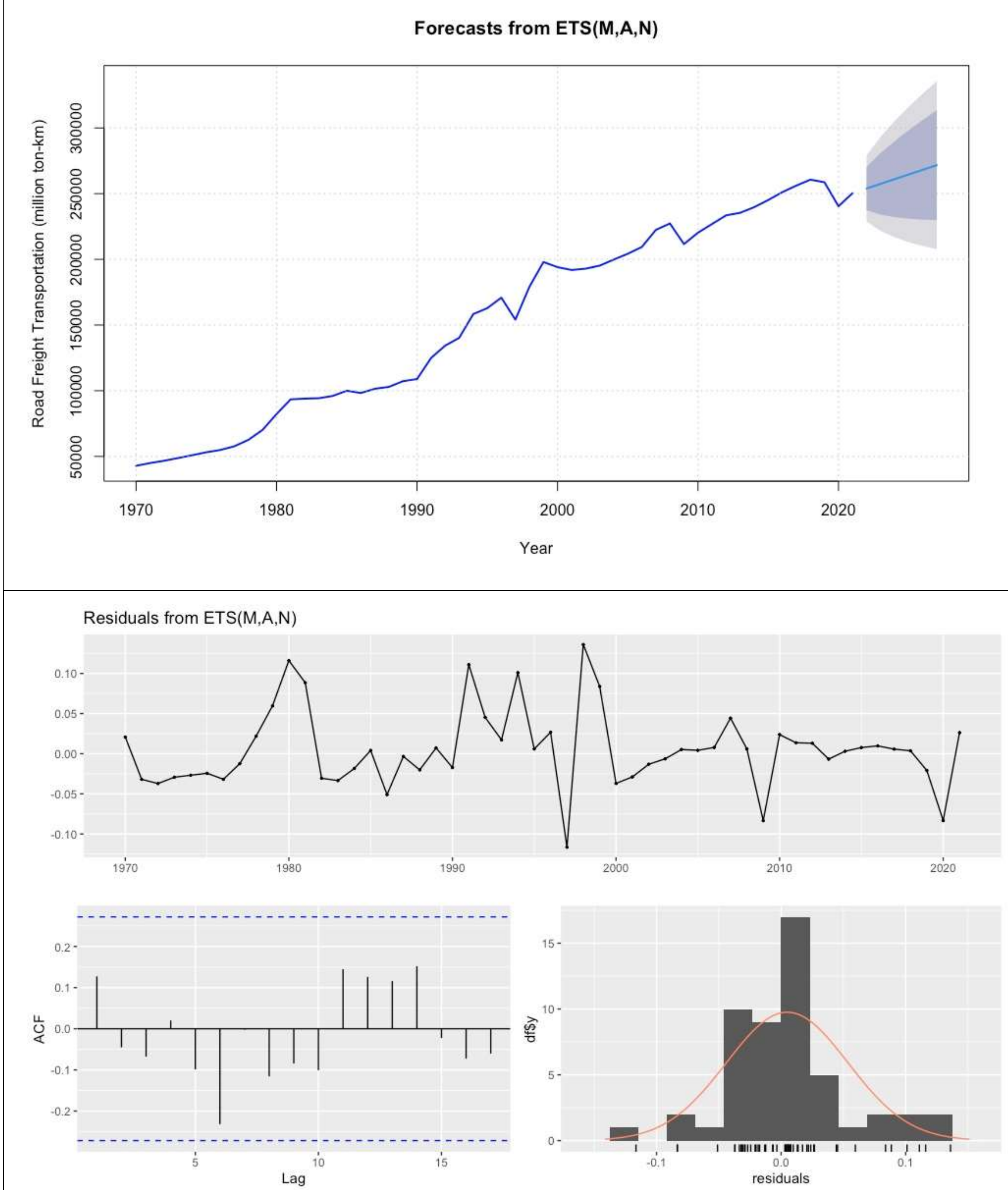
4.2. Results for "How is the future of logistics forecasted for short-term and mid-term in the emerging economy country?"

This section address to research question 2 about how the future of logistics is forecasted for short-term and mid-term in the emerging economy country. This section presents the results of the forecasts by the ETS and ARIMA methods, for both Mexico and Turkey.

4.2.1. Mexico

This section covers the forecasted results of ETS and ARIMA for Mexico for road, rail, inland, and air modes of freight transportation.

Figure 28: *Road Freight ETS Forecast for Mexico*

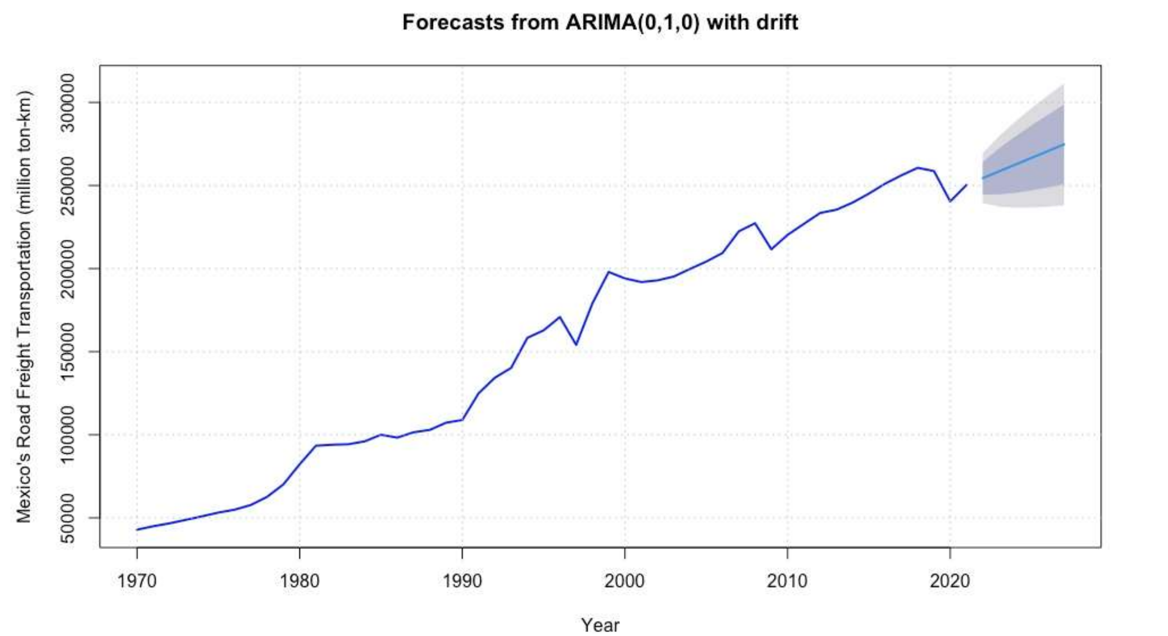


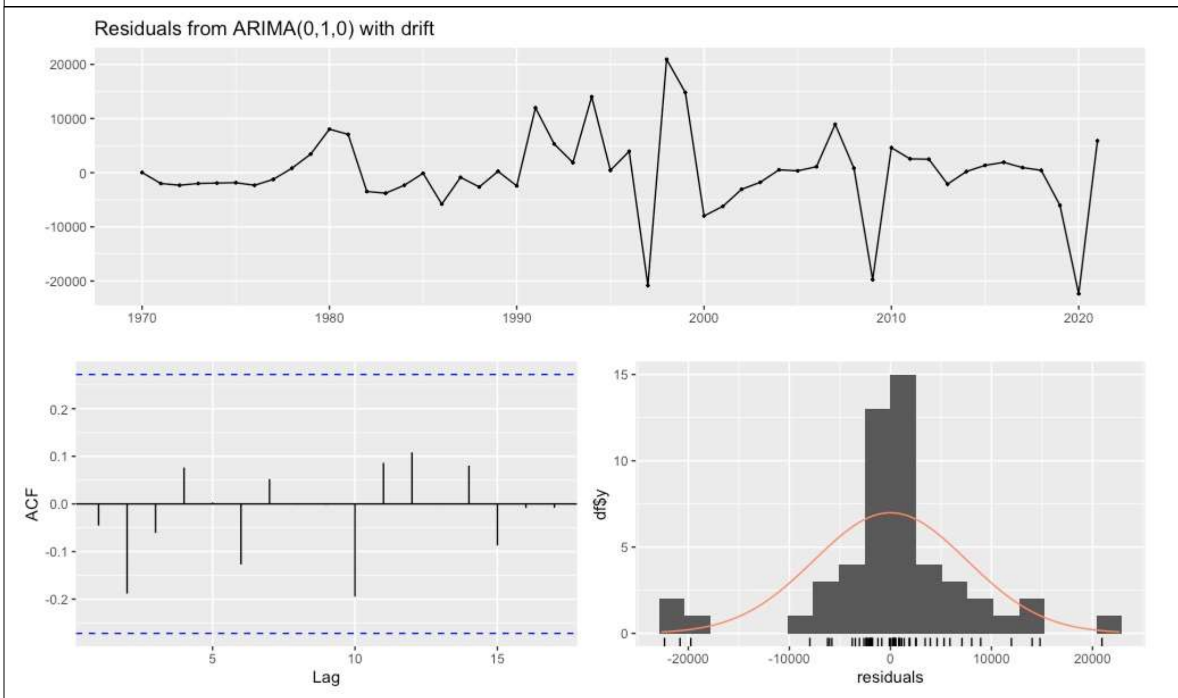
```
ETS(M,A,N)

Call:
 ets(y = timeseries2)

  Smoothing parameters:
    alpha = 0.9999
    beta  = 1e-04

  Initial states:
    l = 38437.9125
    b = 3556.8074

  sigma:  0.0506

     AIC     AICc      BIC
1128.262 1129.567 1138.019
> # Check the residuals
> checkresiduals(fit)

        Ljung-Box test

data:  Residuals from ETS(M,A,N)
Q* = 7.1574, df = 10, p-value = 0.7105

Model df: 0.   Total lags used: 10

> # forecast for the next 6 years
> forecast_result <- forecast(fit, h=6)
> print(forecast_result)
     Point Forecast    Lo 80    Hi 80    Lo 95    Hi 95
2022       253908.9 237453.8 270363.9 228743.0 279074.7
2023       257468.4 234018.9 280917.8 221605.5 293331.2
2024       261027.9 232088.2 289967.5 216768.5 305287.3
2025       264587.4 230915.7 298259.0 213091.0 316083.7
2026       268146.9 230214.9 306078.9 210134.9 326158.9
2027       271706.4 229839.9 313572.9 207677.1 335735.7
```

The best-fit model of ETS for Mexico's road freight transportation is $ETS(M, A, N)$, meaning multiplicative error, additive trend, and no seasonality with estimated smoothing parameters $\alpha = 0.999\ and\ \beta = 10^{-4}$. The output also returns the estimates for the initial states $l_0 = 38437.9125\ and\ b_0 = 3556.8074$. The forecasted future of road freight transportation in Mexico shows an increasing trend for both the short-term and mid-term perspective when considering the point forecast, as illustrated in Figure 28. We can also say that with 80% (dark-shaded) and 95% (light-shaded) probability the mean of forecasted road freight will be within the corresponding confidence intervals. The point forecast and prediction intervals are illustrated in the table in Figure 28**.**

The residuals of the ETS forecast for the road freight transportation of Mexico display a mean of approximately zero, meaning that the data has a good fit for the model. There are no lag values outside the boundaries in the ACF plot, meaning there is no autocorrelation in the data. The histogram of the residuals can be considered as a normal distribution. Moreover, the p-value of the Ljung-Box test for residuals is relatively large. Thus, we can conclude that there is no autocorrelation for the residuals, and they are not distinguishable from a white noise series. With these points considered, the model is suitable for the dataset.

Figure 29*: *Road freight ARIMA Forecast for Mexico





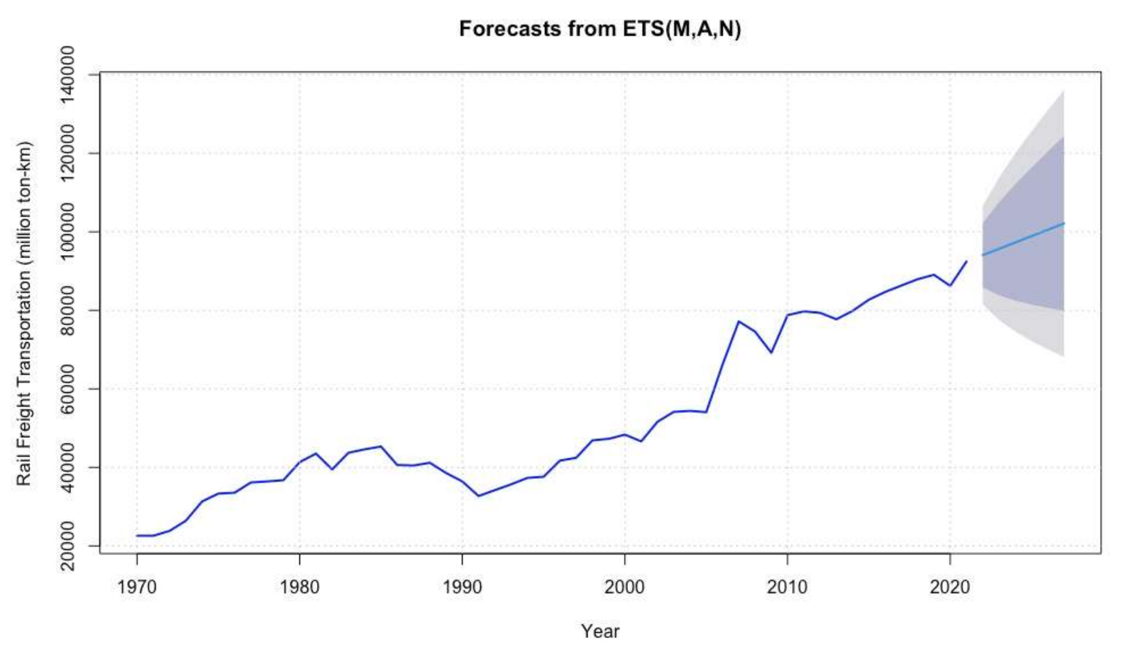
```
Series: timeseries1
ARIMA(0,1,0) with drift

Coefficients:
         drift
      4068.353
s.e.  1058.635

sigma^2 = 58299128:  log likelihood = -527.83
AIC=1059.66   AICc=1059.91   BIC=1063.52
> # Check the residuals
> checkresiduals(fit)

        Ljung-Box test

data:  Residuals from ARIMA(0,1,0) with drift
Q* = 6.3476, df = 10, p-value = 0.7853

Model df: 0.   Total lags used: 10

> # forecast for the next 6 years
> forecast_result <- forecast(fit, h=6)
> print(forecast_result)
     Point Forecast    Lo 80    Hi 80    Lo 95    Hi 95
2022       254418.4 244633.2 264203.5 239453.3 269383.4
2023       258486.7 244648.4 272325.0 237322.9 279650.5
2024       262555.1 245606.7 279503.4 236634.8 288475.3
2025       266623.4 247053.1 286193.7 236693.2 296553.6
2026       270691.8 248811.5 292572.0 237228.8 304154.7
2027       274760.1 250791.5 298728.7 238103.3 311416.9
```

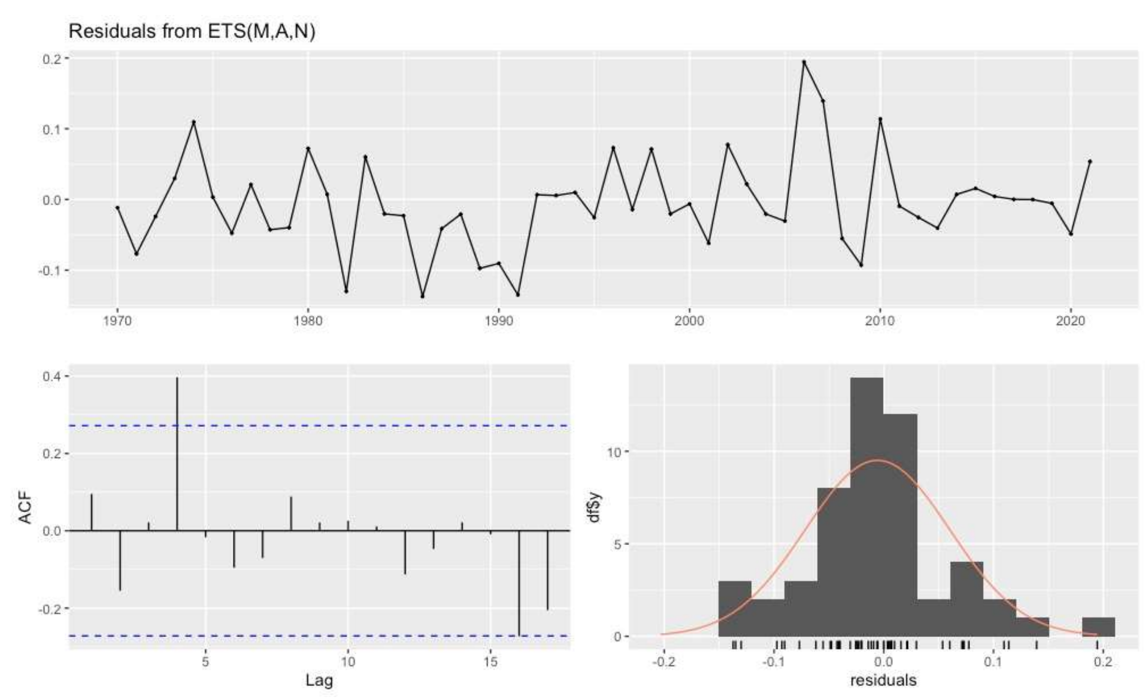
The optimal model of ARIMA for Mexico's road freight transportation is the $ARIMA(0,1,0)$ model, meaning there is no autoregressive or moving average component ($p = 0, q = 0$), however there is first order differencing ($d = 1$) which mean that the time series has been differenced once to make it stationary.

The forecasted future of road freight transportation in Mexico shows an increasing trend for both the short-term and mid-term perspective when considering the point forecast, as illustrated in Figure 29. We can also say that with 80% (dark-shaded) and 95% (light-shaded) probability the mean of the forecasted road freight will be within the corresponding confidence intervals. The values of the point forecast as well as prediction intervals for 2023 to 2027 are given in Figure 29**.**

The residuals of the ARIMA forecast for road freight transportation of Mexico display a mean of approximately zero, meaning that the data has a good fit for the ARIMA forecast model. Furthermore, there are no lag values outside the boundaries in the ACF plot, meaning there is no autocorrelation in the data. The residuals in the histogram can be considered normally distributed. Moreover, the p-value of the Ljung-Box test for residuals is relatively large, meaning we can conclude that the residuals are behaving like white noise. With these facts assert that the selected ARIMA forecast model is suitable and fits the data.

Figure 30: *Rail Freight ETS Forecast for Mexico*





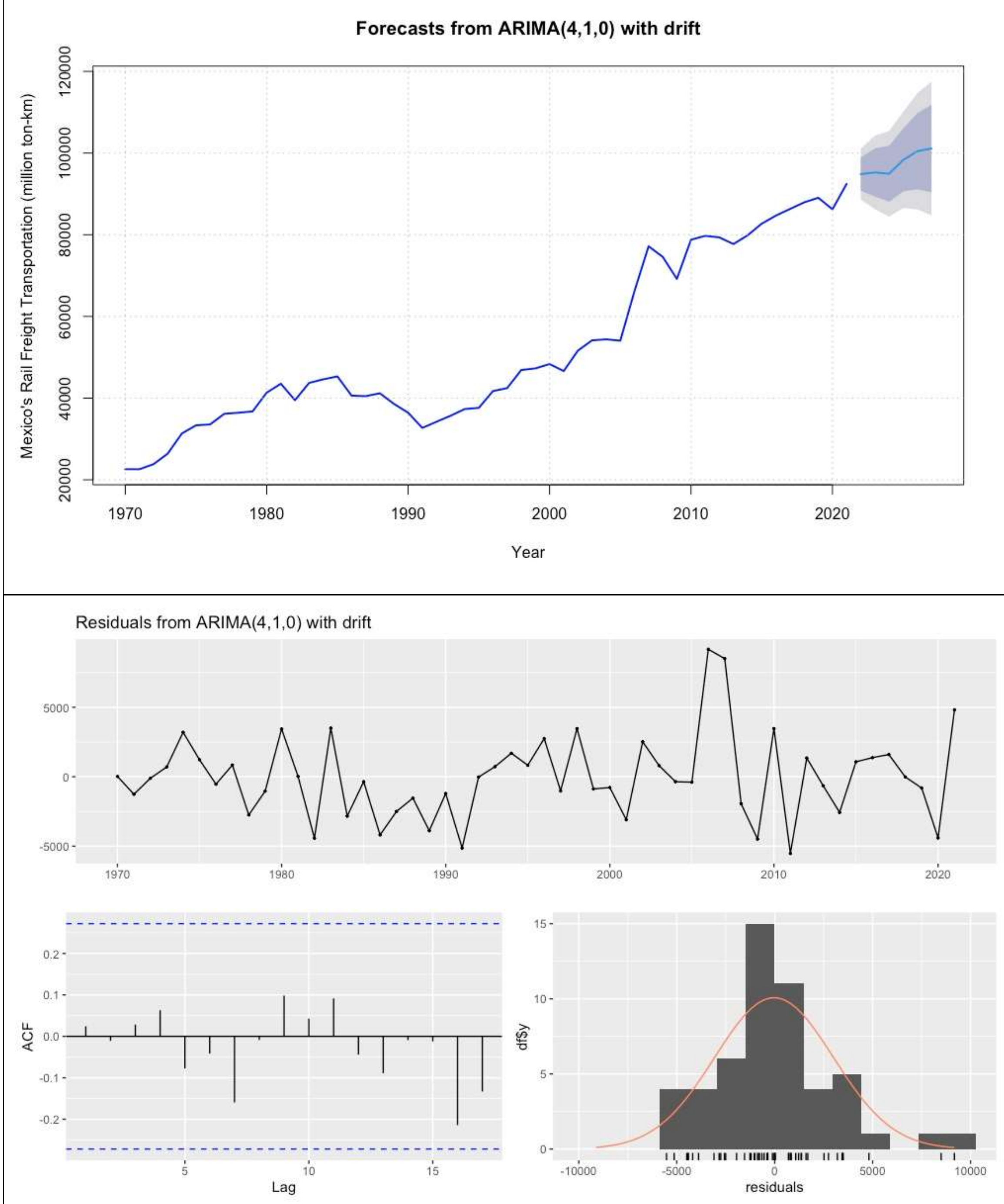
```
ETS(M,A,N)

Call:
 ets(y = timeseries2)

  Smoothing parameters:
    alpha = 0.9999
    beta  = 0.0222

  Initial states:
    l = 20984.3484
    b = 1873.7797

  sigma:  0.0681

     AIC     AICc      BIC
1054.578 1055.882 1064.334
> # Check the residuals
> checkresiduals(fit)

        Ljung-Box test

data:  Residuals from ETS(M,A,N)
Q* = 12.554, df = 10, p-value = 0.2497

Model df: 0.   Total lags used: 10

> # forecast for the next 6 years
> forecast_result <- forecast(fit, h=6)
> print(forecast_result)
     Point Forecast    Lo 80    Hi 80    Lo 95    Hi 95
2022       94051.62 85841.19 102262.1 81494.85 106608.4
2023       95666.71 83813.64 107519.8 77539.00 113794.4
2024       97281.80 82465.66 112097.9 74622.47 119941.1
2025       98896.89 81440.12 116353.7 72199.06 125594.7
2026      100511.98 80601.56 120422.4 70061.62 130962.3
2027      102127.07 79881.86 124372.3 68105.96 136148.2
```

The optimal model for Mexico's rail freight transportation regarding ETS is once again the $ETS(M, A, N)$ model. M, A, N means it has a multiplicative error, an additive trend, and no seasonality. For this model, the estimated smoothing parameters are $\alpha = 0.999\ and\ \beta = 0.0222$. The output also illustrates the estimates for the initial states, $l_0 = 20984.3484\ and\ b_0 = 1873.7797$.

The forecasted future of rail freight transportation in Mexico depicts an increasing trend for both a short-term and mid-term perspectives, as illustrated in Figure 30**,** when considering the point forecast. We can also say that with 80% (dark-shaded) and 95% (light-shaded) certainty, the mean value of forecasted rail freight will be within the corresponding confidence intervals. The values of the point forecast as well as prediction intervals for 2023 to 2027 are given in Figure 30**.**

The residuals of the ETS forecast for the rail freight transportation of Mexico display a mean of approximately zero, indicating that the data has a good fit for the ETS forecast model. Furthermore, there is one lag value outside the boundaries in the ACF plot, meaning that there is trace of autocorrelation in the data. The histogram of the residuals can be considered as a normal distribution. Moreover, the p-value of the Ljung-Box test for residuals is relatively large. With these points taken into consideration, the model criteria are not fulfilled, and therefore the model is not optimally suitable for the data.

Figure 31: *Rail Freight ARIMA Forecast for Mexico*



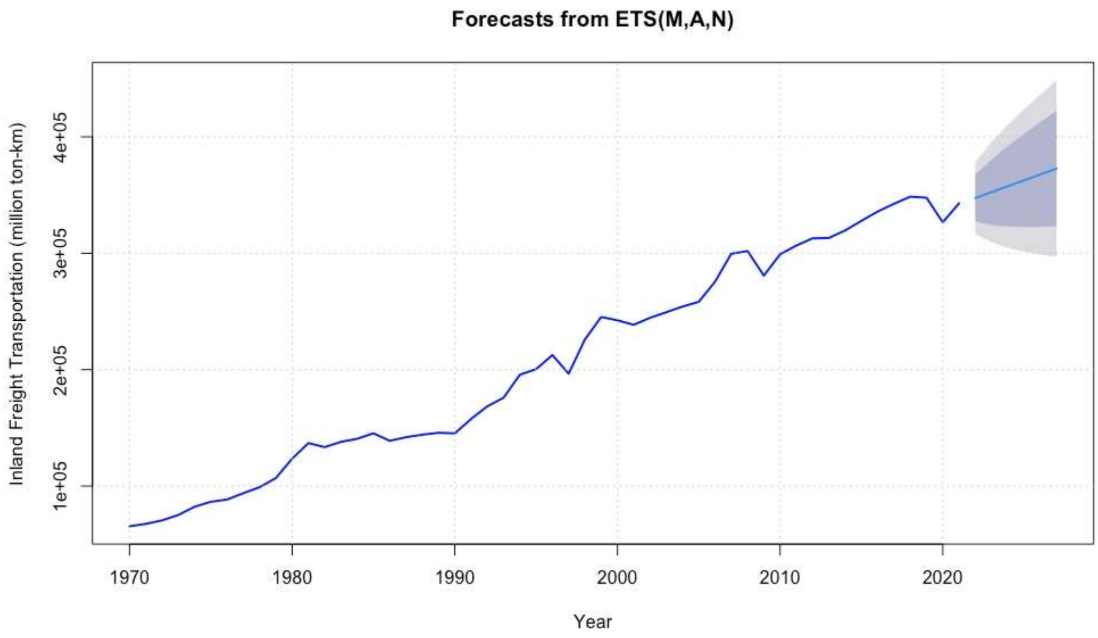
```
Series: timeseries1
ARIMA(4,1,0) with drift

Coefficients:
         ar1      ar2     ar3     ar4      drift
      0.0360  -0.1802  0.0196  0.3779  1412.5809
s.e.  0.1309   0.1293  0.1298  0.1294   550.8997

sigma^2 = 10169910:  log likelihood = -481.57
AIC=975.15   AICc=977.06   BIC=986.74
> # Check the residuals
> checkresiduals(fit)

        Ljung-Box test

data:  Residuals from ARIMA(4,1,0) with drift
Q* = 3.1291, df = 6, p-value = 0.7925

Model df: 4.   Total lags used: 10

> # forecast for the next 6 years
> forecast_result <- forecast(fit, h=6)
> print(forecast_result)
     Point Forecast    Lo 80     Hi 80    Lo 95    Hi 95
2022       94854.08 90767.18  98940.99 88603.70 101104.5
2023       95246.26 89361.61 101130.90 86246.47 104246.0
2024       94933.45 88085.19 101781.71 84459.93 105407.0
2025       98301.74 90597.25 106006.23 86518.74 110084.7
2026      100455.27 91154.48 109756.06 86230.94 114679.6
2027      101122.71 90414.37 111831.05 84745.71 117499.7
```

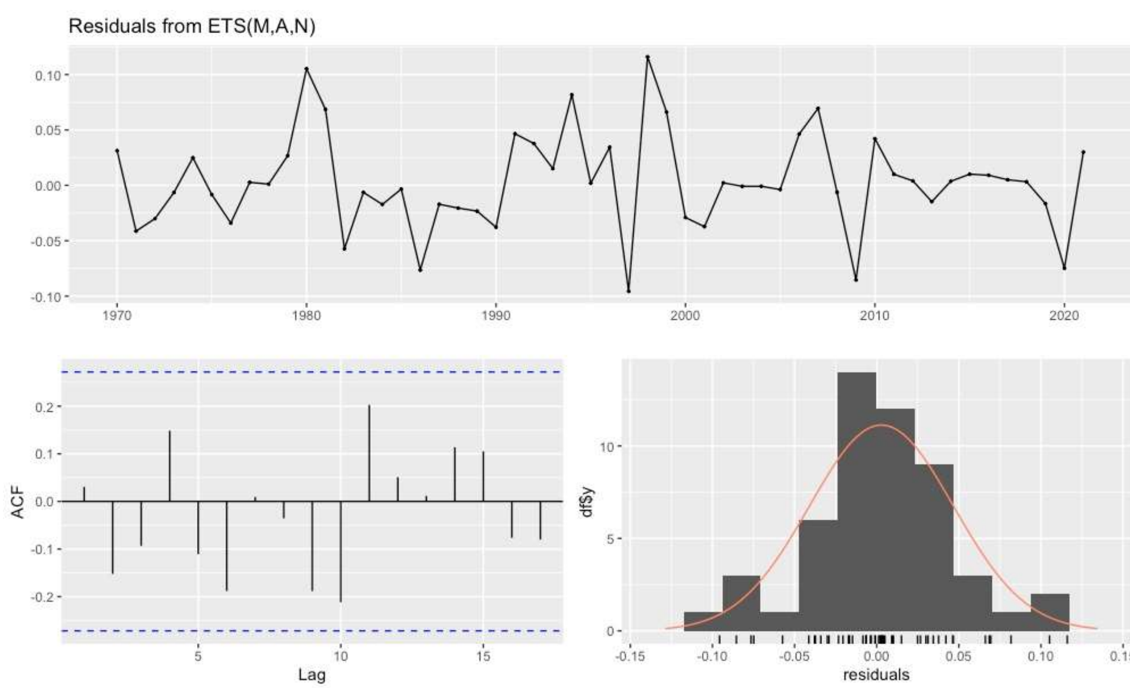
The ARIMA forecast for rail freight development for Mexico shows a different sort of development when compared to the previous ETS forecast. The optimal ARIMA model for the rail freight forecast is $ARIMA(4,1,0)$. $ARIMA$ (4,1,0) includes four lag observations in the autoregressive component, one degree of differencing, and no moving average window.

Starting with the criteria for the model, we can see that in the residual graph, the mean for the residuals is approximately zero. Furthermore, there are no lag values outside the boundaries in the ACF plot which indicates no autocorrelation. Lastly, the histogram shows that the data is normally distributed. The Ljung-box test indicates a large p-value which means that we can't distinguish the residuals from white noise, and that there is no autocorrelation in the residuals. All the model criteria are fulfilled, and therefore the ARIMA model can be considered a better fit for the data compared to the ETS model.

When looking at the point forecast in Figure 31**,** there is more spread in development than what we have previously seen in the other forecasts. The forecast illustrates both an increase and decrease in freight in the period being forecasted. However, in total, according to the point forecast, the freight is will increase significantly from a mid-term perspective. We can also say that with 80% and 95% confidence that the forecasted rail freight will be within the corresponding confidence intervals, showing both a possible decrease and increase in rail freight transported.

Figure 32: *Inland Freight ETS Forecast for Mexico*





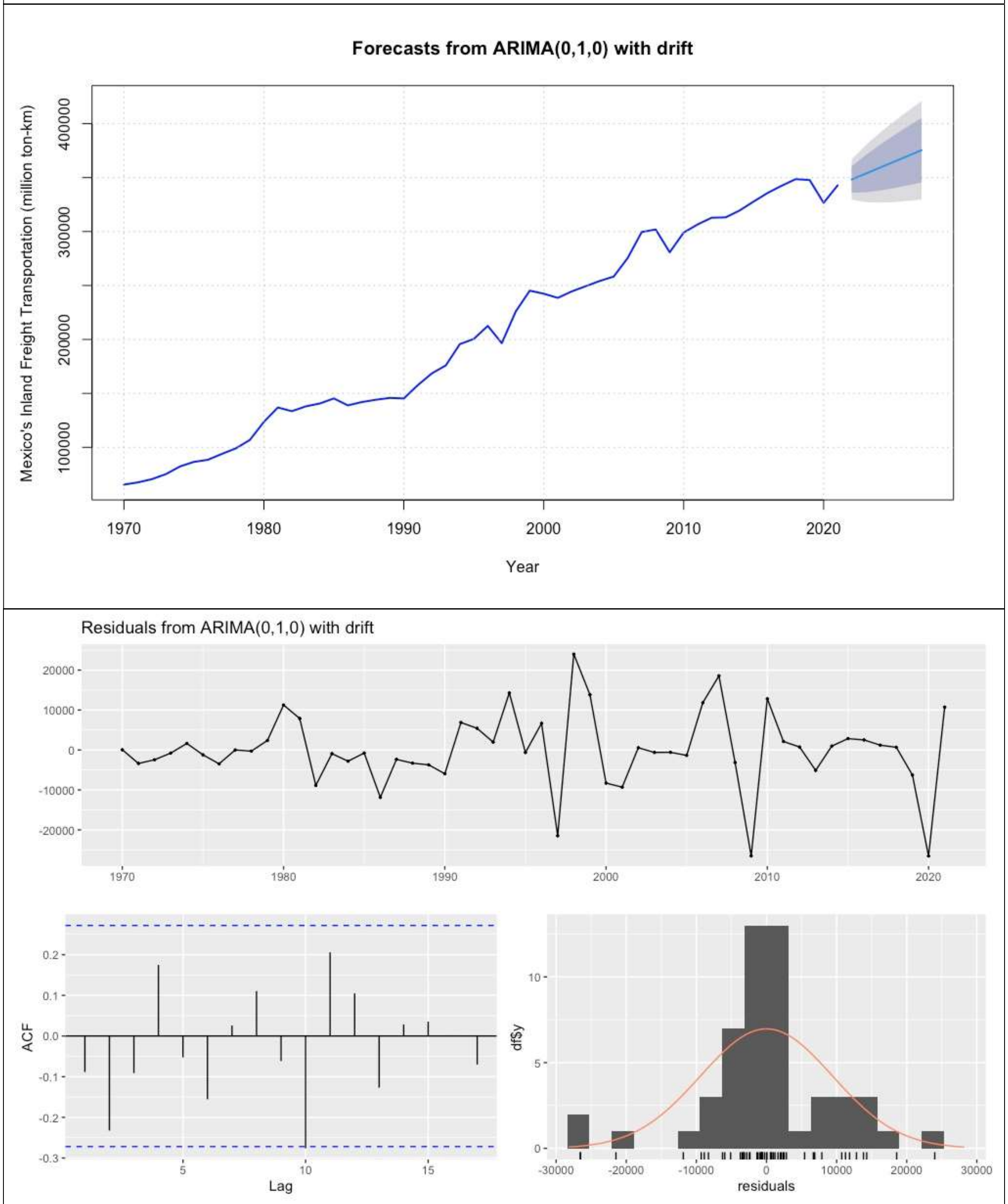
```
ETS(M,A,N)

Call:
 ets(y = timeseries2)

  Smoothing parameters:
    alpha = 0.9573
    beta  = 1e-04

  Initial states:
    l = 58419.5712
    b = 5055.3336

  sigma:  0.0453

     AIC     AICc      BIC
1149.960 1151.264 1159.716
> # Check the residuals
> checkresiduals(fit)

	Ljung-Box test

data:  Residuals from ETS(M,A,N)
Q* = 11.437, df = 10, p-value = 0.3245

Model df: 0.   Total lags used: 10

> # forecast for the next 6 years
> forecast_result <- forecast(fit, h=6)
> print(forecast_result)
     Point Forecast    Lo 80    Hi 80    Lo 95    Hi 95
2022       347418.3 327248.4 367588.2 316571.1 378265.6
2023       352475.8 324325.6 380626.0 309423.8 395527.9
2024       357533.4 323037.6 392029.1 304776.7 410290.0
2025       362590.9 322596.3 402585.5 301424.4 423757.4
2026       367648.4 322687.7 412609.1 298886.9 436409.9
2027       372705.9 323149.7 422262.2 296916.2 448495.7
```

The ETS forecast graph for inland freight in Mexico shows a steady increase via the point forecast in the period being forecasted. With an increase in inland freight transported every year between 2022 and 2027. The Best ETS model for inland freight is the $ETS(M, A, N)$ model, meaning it has a multiplicative error, additive trend, and no seasonality. Furthermore, the residual graph show that the mean is around zero for the residuals of the data. There are also no lag values outside the boundaries in the ACF plot, indicating that there is no autocorrelation in the data. The residuals are also evenly distributed in the normal distribution histogram. All model criteria are being fulfilled with these points stated, and the model is therefore a good fit for the dataset.

The ljung-box test in the forecast output has a large p-value above 0.05, indicating that there is no autocorrelation in the residuals, and they are behaving like a white noise series. Furthermore, the smoothing factors for the ETS model are $\alpha = 0.9573\ and\ \beta = 10^{-4}$. The initial states for the model are $l_0 = 58419.5712\ and\ b_0 = 5055.3336$.

As previously stated, the point forecast in Figure 32 & also in the forecast table in Figure 32 illustrates that there will be a steady increase in inland freight transport for a short and mid-term perspective. The confidence forecast also indicate that we can with 80% (dark Area) and 95% (light area) certainty claim that the forecasted inland freight will be within the corresponding confidence intervals. The confidence forecasts however show us both a possible decrease and increase in the freight transport.

Figure 33: *Inland Freight ARIMA Forecast for Mexico*



```
Series: timeseries1
ARIMA(0,1,0) with drift

Coefficients:
         drift
      5437.804
s.e.  1316.998

sigma^2 = 90227624:  log likelihood = -538.97
AIC=1081.93   AICc=1082.18   BIC=1085.8
> # Check the residuals
> checkresiduals(fit)

        Ljung-Box test

data:  Residuals from ARIMA(0,1,0) with drift
Q* = 13.542, df = 10, p-value = 0.1949

Model df: 0.   Total lags used: 10

> # forecast for the next 6 years
> forecast_result <- forecast(fit, h=6)
> print(forecast_result)
     Point Forecast    Lo 80    Hi 80    Lo 95    Hi 95
2022       348224.8 336051.6 360398.0 329607.5 366842.2
2023       353662.6 336447.1 370878.2 327333.7 379991.5
2024       359100.4 338015.8 380185.1 326854.2 391346.6
2025       364538.2 340191.8 388884.7 327303.5 401772.9
2026       369976.0 342755.8 397196.2 328346.4 411605.7
2027       375413.8 345595.6 405232.0 329810.8 421016.8
```

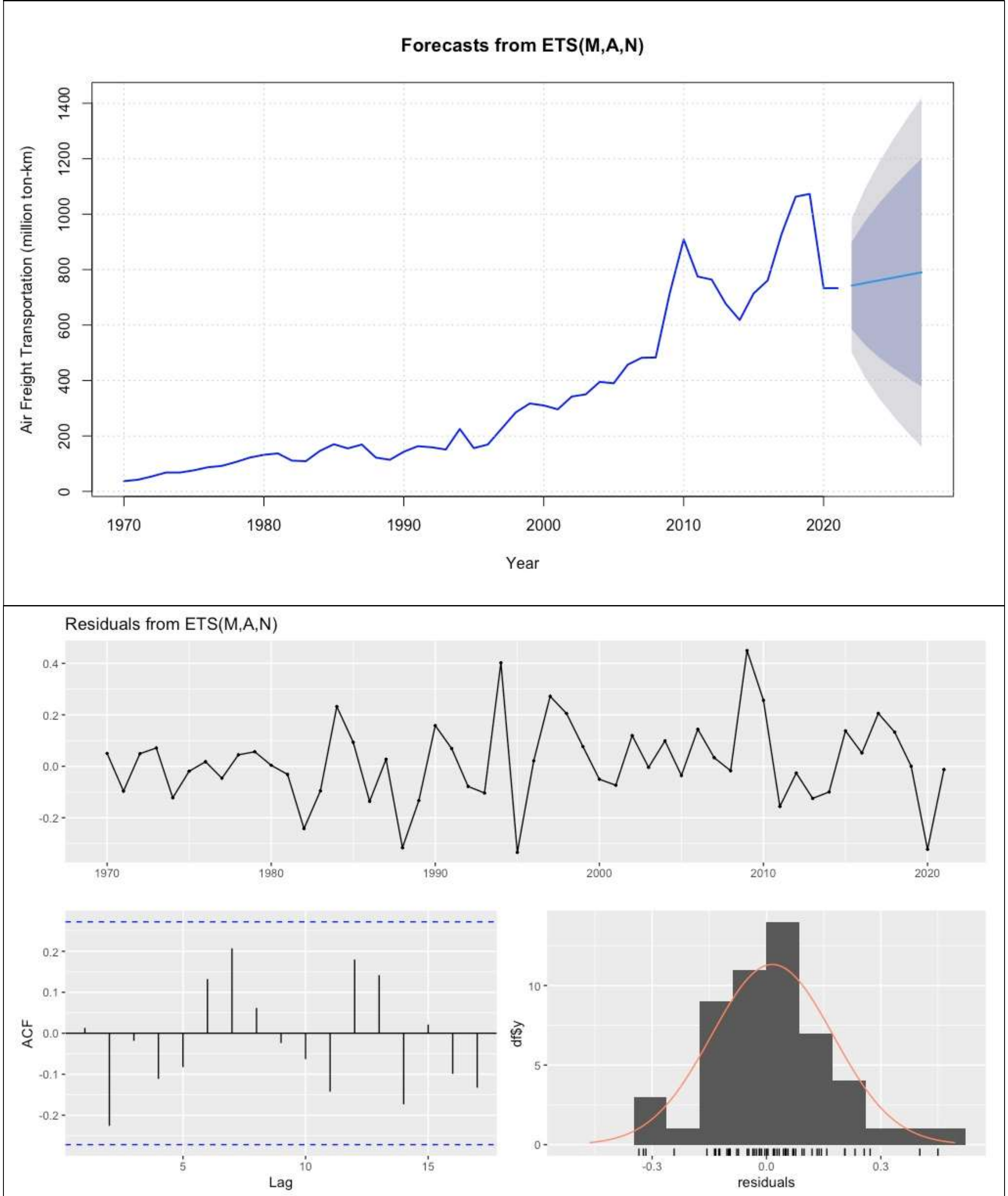
The ARIMA forecast graph for inland freight transported shows through the point forecast that there will be a steady increase in freight transported for the short and mid-term perspective. For inland freight for Mexico, the $ARIMA(0,1,0)$ model is the most optimal ARIMA model. The zero indicates no autoregressive parts, the one indicates first order differencing to make the data stationary, and the second zero indicates no moving average component.

Regarding the model criteria we can see that the mean is perhaps a bit lower than zero, meaning that the residuals do not fit the data optimally. Furthermore, there is one lag value which touches the boundary in the ACF plot, however it does not exceed it. Therefore, there is no autocorrelation in the dataset. The normal distribution histogram shows us that the data is unevenly spread in the histogram. The criteria can be considered unmet, and when comparing these results with the ETS model, the ETS model fits the data better because of the criteria being met for that model. Furthermore, the Ljung-Box test value is lower than previously seen, however still above the threshold of 0.05 meaning there is no autocorrelation in the residuals, and they behave as a white noise series.

The point forecast table in Figure 33 show us that there will be an increase in freight transported via inland, and the confidence forecasts indicate that we can with 80% (dark area) and 95%(light area) certainty assure that the freight will developed within these intervals.

However, since the model does not optimally fit the data, the ETS model can be considered superior for inland freight.

Figure 34: *Air Freight ETS Forecast for Mexico*



```
ETS(M,A,N)

Call:
 ets(y = timeseries2)

  Smoothing parameters:
    alpha = 0.9999
    beta  = 1e-04

  Initial states:
    l = 25.761
    b = 9.4801

  sigma:  0.1654

     AIC     AICc      BIC
592.9308 594.2351 602.6870
> # Check the residuals
> checkresiduals(fit)

	Ljung-Box test

data:  Residuals from ETS(M,A,N)
Q* = 8.3301, df = 10, p-value = 0.5966

Model df: 0.   Total lags used: 10

> # forecast for the next 6 years
> forecast_result <- forecast(fit, h=6)
> print(forecast_result)
     Point Forecast    Lo 80     Hi 80    Lo 95     Hi 95
2022       742.5024 585.0820  899.9229 501.7486  983.2563
2023       752.0039 526.4404  977.5675 407.0342 1096.9736
2024       761.5054 481.6083 1041.4025 333.4397 1189.5711
2025       771.0069 443.5662 1098.4476 270.2296 1271.7842
2026       780.5084 409.6297 1151.3870 213.2984 1347.7183
2027       790.0099 378.4380 1201.5817 160.5650 1419.4547
```

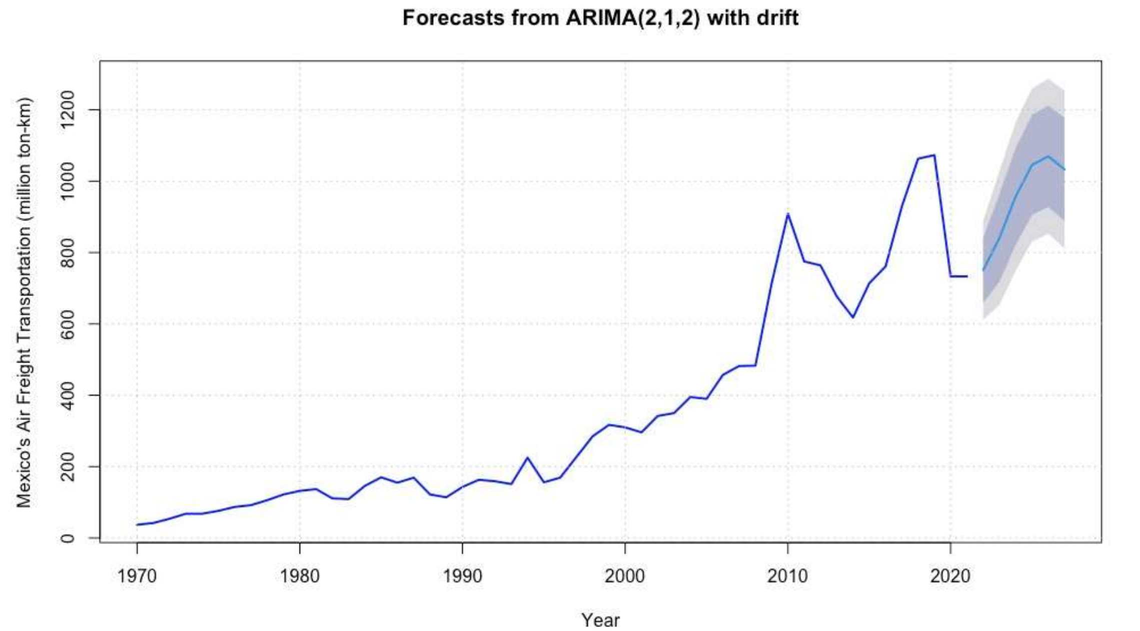
For air freight transported in Mexico, the best ETS model for forecasting is once again the $ETS(M, A, N)$ model. To explain it once more, it indicates that the model has a multiplicative error, additive trend, and no seasonality. The smoothing parameters are $a = 0.999 \; and \; \beta = 10^{-4}$.

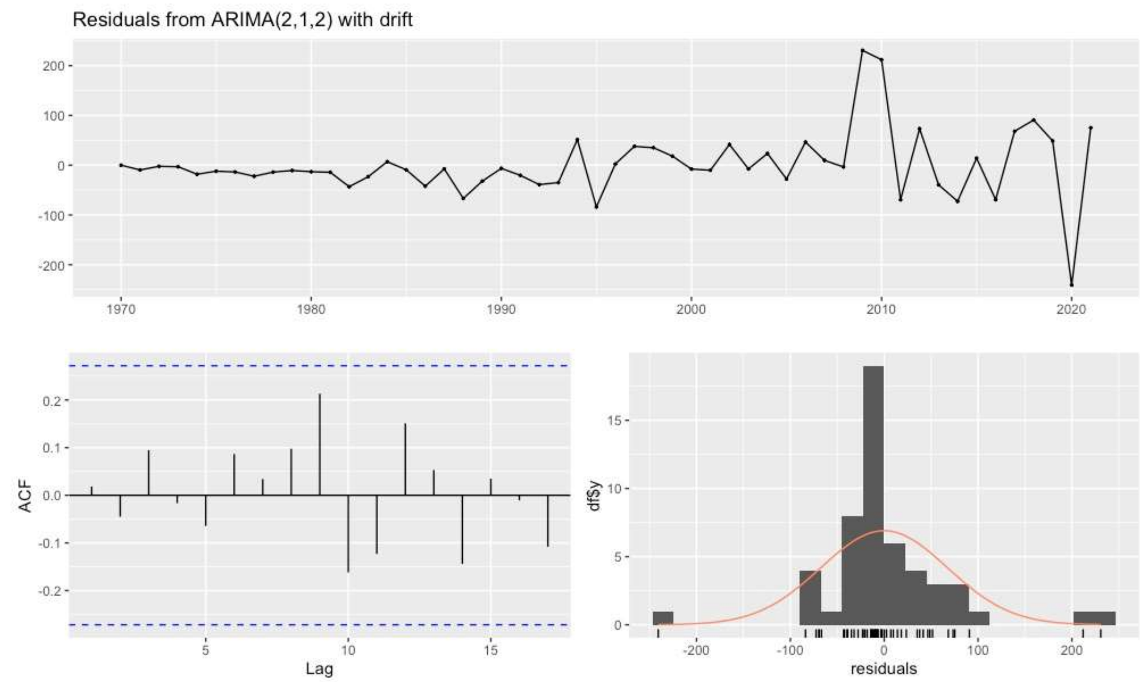
The residuals have a mean of approximately zero when looking at the residual graph, all the lag values are also inside the ACF boundaries which once again indicate that there is no autocorrelation in the data. Lastly the residuals have a normal distribution. With all these three criteria fulfilled, the model fits the data in a good way.

Furthermore, the Ljung-Box illustrates a relatively large value, which once again indicates that there is no autocorrelation between the residuals, and they are not distinguishable from a white noise series. The initial states for the model are $l_0 = 25.761 and \; b_0 = 9.4801$.

Furthermore the prediction intervals in Figure 34 tells us that we can with 80% (dark area) and 95% (light area) certainty expect the freight development for air freight to fall within these intervals. The prediction intervals show us both a possible decrease and increase in air freight transported for Mexico.

Figure 35: *Air Freight ARIMA Forecast for Mexico*





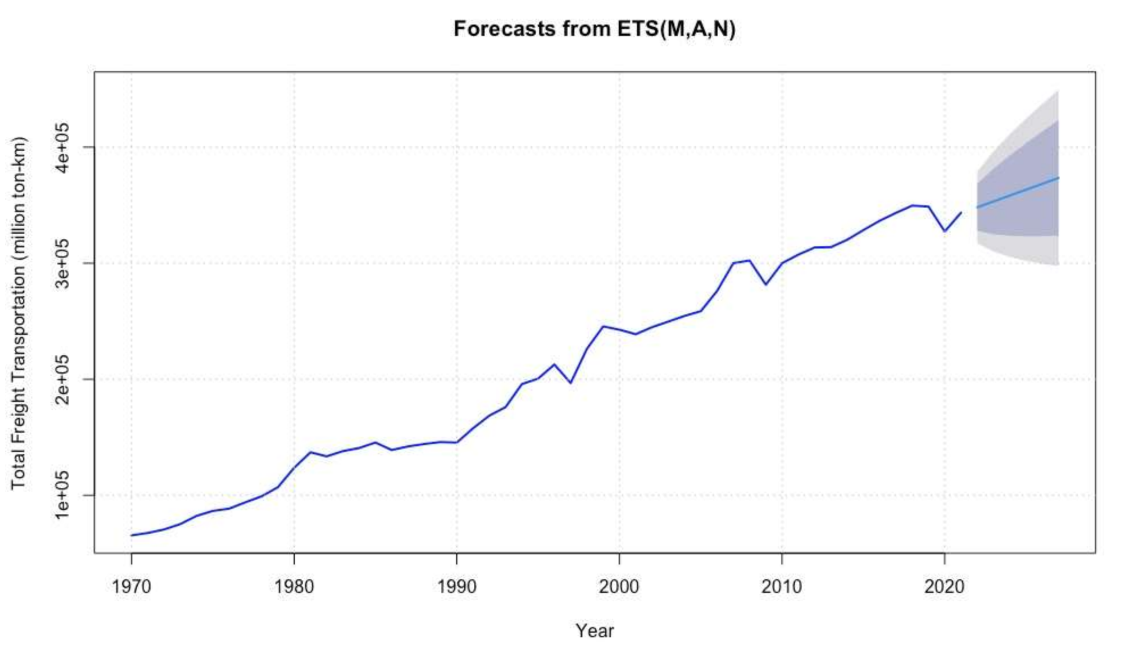
```
Series: timeseries1
ARIMA(2,1,2) with drift

Coefficients:
         ar1      ar2      ma1     ma2    drift
      1.3686  -0.8941  -1.4656  0.7677  16.5139
s.e.  0.1324   0.0898   0.1978  0.1644   5.5515

sigma^2 = 5052:  log likelihood = -288.03
AIC=588.07   AICc=589.97   BIC=599.66
> # Check the residuals
> checkresiduals(fit)

        Ljung-Box test

data:  Residuals from ARIMA(2,1,2) with drift
Q* = 6.7723, df = 6, p-value = 0.3424

Model df: 4.   Total lags used: 10

> # forecast for the next 6 years
> forecast_result <- forecast(fit, h=6)
> print(forecast_result)
     Point Forecast    Lo 80     Hi 80    Lo 95     Hi 95
2022       750.5481 659.4599  841.6363 611.2407  889.8555
2023       841.0017 718.2745  963.7288 653.3067 1028.6967
2024       957.7837 821.7689 1093.7985 749.7670 1165.8004
2025      1045.4145 905.1609 1185.6681 830.9151 1259.9139
2026      1069.6094 927.6532 1211.5656 852.5062 1286.7127
2027      1033.0507 888.5656 1177.5358 812.0798 1254.0216
```

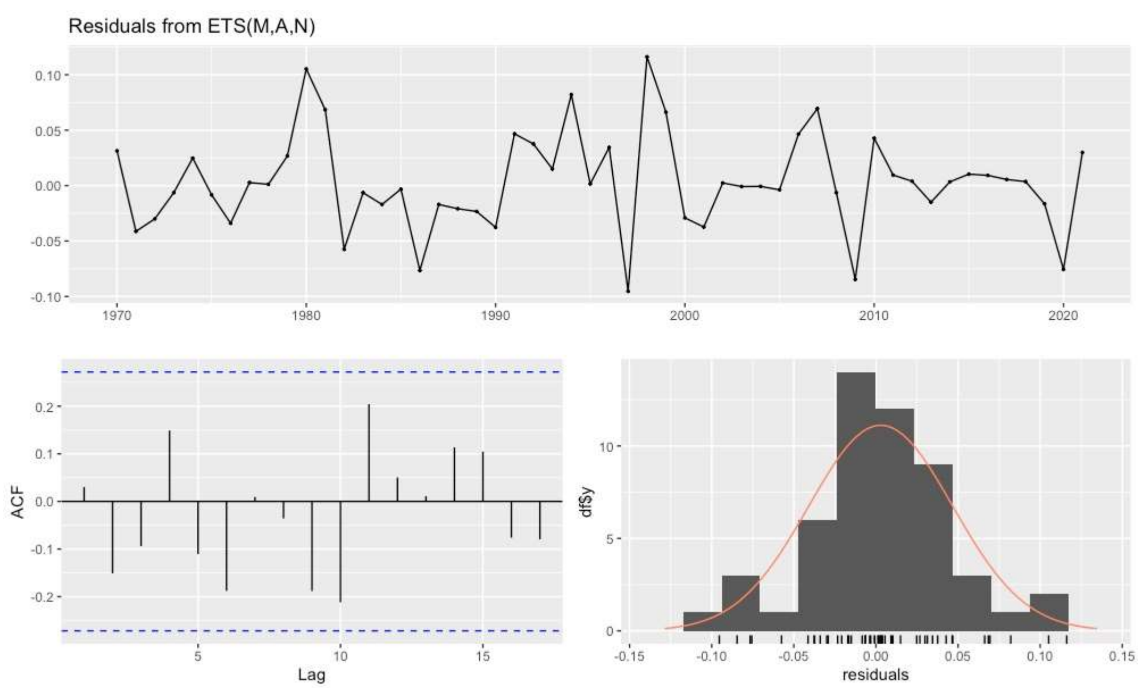
The optimal ARIMA model for forecasting air freight transported for Mexico is the $ARIMA(2,1,2)$ model, meaning two autoregressive parts, first order differencing, and 2 moving average components.

The residuals have a mean of approximately zero, they are normally distributed, and the ACF plot shows no lag values outside the boundary indicating that there is no autocorrelation in the data. The Ljung-Box test of the model has a rather large value compared to the threshold of 0.05, indicating that the residuals are behaving like white noise, and there is no autocorrelation in the residuals.

Regarding the point forecast and the prediction intervals, the point forecast estimates a total increase from 2022 to 2027, however from 2025 to 2027 there is a decrease in air freight transported. We can also say that with 80% (dark area) and 95% (light area) certainty, the freight development will be within the corresponding intervals, as also illustrated in Figure 35.

Figure 36: *Total Freight ETS Forecast for Mexico*





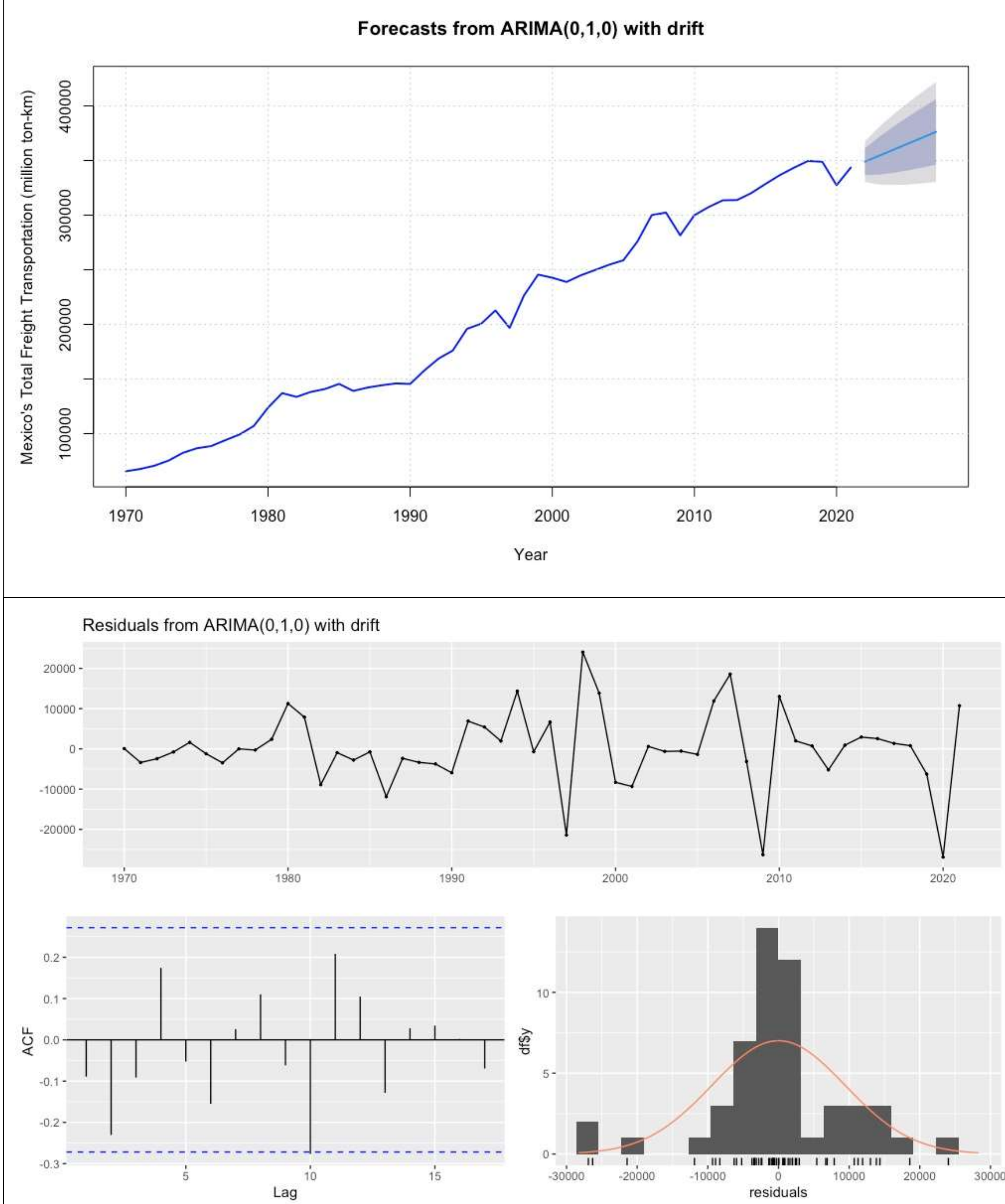
```
ETS(M,A,N)

Call:
 ets(y = timeseries2)

  Smoothing parameters:
    alpha = 0.9573
    beta  = 1e-04

  Initial states:
    l = 58445.8366
    b = 5061.5939

  sigma:  0.0453

     AIC     AICc      BIC
1150.120 1151.424 1159.876
> # Check the residuals
> checkresiduals(fit)

        Ljung-Box test

data:  Residuals from ETS(M,A,N)
Q* = 11.416, df = 10, p-value = 0.326

Model df: 0.   Total lags used: 10

> # forecast for the next 6 years
> forecast_result <- forecast(fit, h=6)
> print(forecast_result)
     Point Forecast    Lo 80    Hi 80    Lo 95    Hi 95
2022       348158.5 327942.8 368374.3 317241.2 379075.8
2023       353222.4 325008.4 381436.4 310072.8 396371.9
2024       358286.2 323712.5 392859.9 305410.3 411162.1
2025       363350.0 323265.2 403434.8 302045.6 424654.4
2026       368413.8 323352.1 413475.6 299497.8 437329.9
2027       373477.7 323810.3 423145.0 297518.0 449437.3
```

When forecasting the total freight development for Mexico for a short- and mid-term perspective with ETS, it is once again the $ETS(M, A, N)$ model which is the most optimal.

The estimated smoothing parameters are $\alpha = 0.999$ *and* $\beta = 10^{-4}$, while the initial states are $l_0 = 38437.9125$ *and* $b_0 = 3556.8074$. The point forecast shows an increasing trend for both the short-term and mid-term outlooks, as illustrated in Figure 36. Additionally, confidence intervals of 80% (dark area) and 95%(light area) are included in the plot to provide an estimate of the forecasted mean. The point forecast and prediction intervals are also displayed in the table in Figure 36.

The residuals of the ETS forecast for total freight transportation in Mexico show a good fit for the model, with a mean close to zero, no significant autocorrelation, and a histogram that can be approximated by a normal distribution. Furthermore, the p-value of the Ljung-Box test for residuals is relatively high, indicating no significant autocorrelation, and no distinguishment from a white noise series. Therefore, we can conclude that the ETS model is appropriate for the total freight transportation dataset in Mexico, and that we can se an increase in total freight transported.

Figure 37: *Total Freight ARIMA Forecast for Mexico*



```
Series: timeseries1
ARIMA(0,1,0) with drift

Coefficients:
         drift
      5451.451
s.e.  1319.898

sigma^2 = 90625572:  log likelihood = -539.08
AIC=1082.16   AICc=1082.41   BIC=1086.02
> # Check the residuals
> checkresiduals(fit)

        Ljung-Box test

data:  Residuals from ARIMA(0,1,0) with drift
Q* = 13.498, df = 10, p-value = 0.1972

Model df: 0.   Total lags used: 10

> # forecast for the next 6 years
> forecast_result <- forecast(fit, h=6)
> print(forecast_result)
     Point Forecast    Lo 80    Hi 80    Lo 95    Hi 95
2022       348971.5 336771.4 361171.5 330313.1 367629.8
2023       354422.9 337169.4 371676.4 328036.0 380809.8
2024       359874.4 338743.3 381005.5 327557.1 392191.6
2025       365325.8 340925.7 389725.9 328009.1 402642.5
2026       370777.3 343497.1 398057.4 329055.9 412498.6
2027       376228.7 346344.8 406112.6 330525.2 421932.2
```

For the total freight transport forecast with the ARIMA model, the $ARIMA(0,1,0)$ has once again the best fit. There are no autoregressive parts in the model, there is first order differencing, and there is no moving average component.

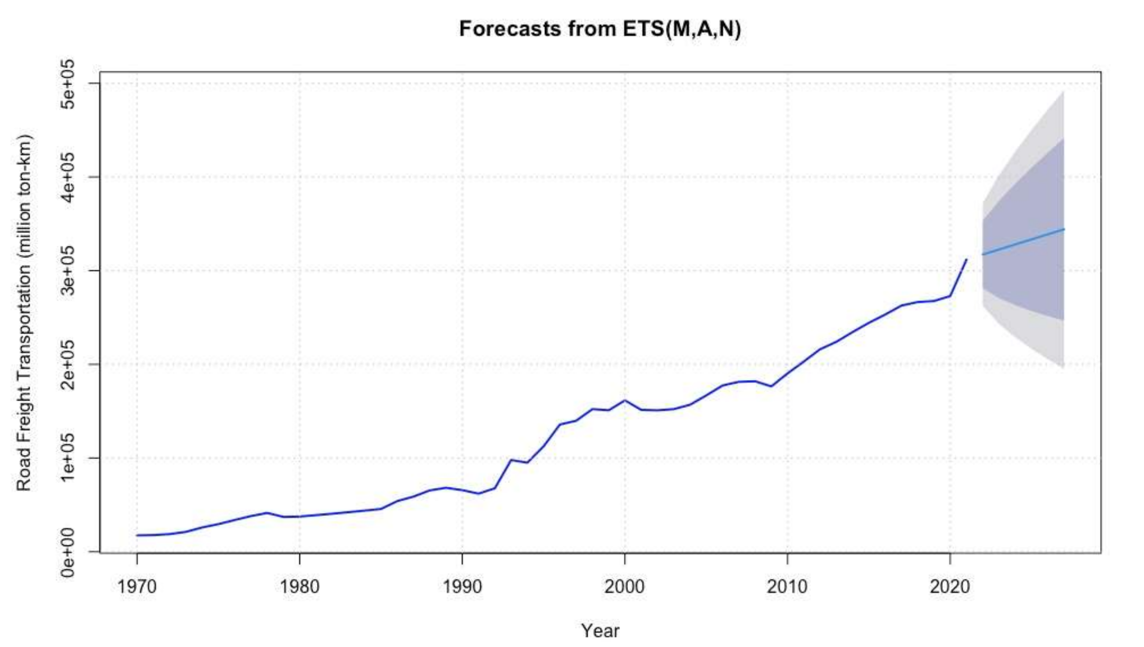
The criteria for the model are fully met with the residuals having a mean of approximately zero, no autocorrelation, and being normally distributed in the histogram in Figure 37. The Ljung-Box test p-value is once more above the threshold of 0.05 indicating that there is no autocorrelation in the residuals, and that they are indistinguishable from a white noise series.

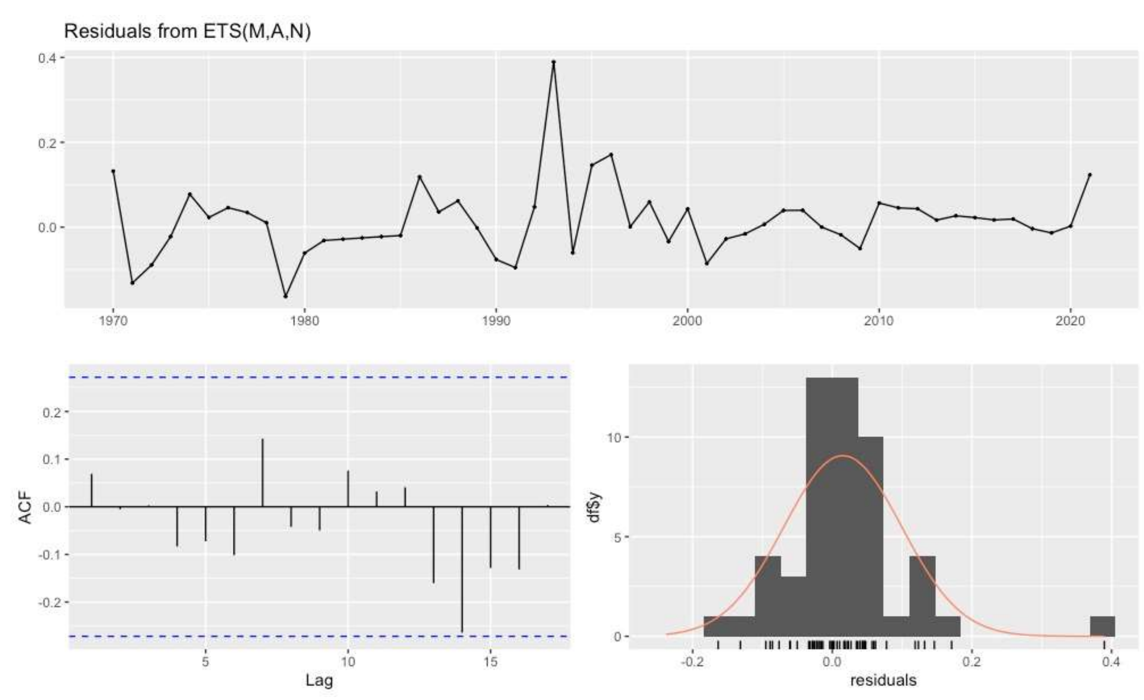
For the point forecast, the model forecasts a total growth of freight transported for Mexico, with 80% (dark area) and 95% (light area) certainty that it will be within the prediction intervals as illustrated in the table in Figure 37, as well as the forecast graph in Figure 37.

4.2.2. Turkey

This section covers the forecasted results of ETS and ARIMA for Mexico for road, rail, inland, and air modes of freight transportation.

Figure 38: *Road freight ETS Forecast for Turkey*





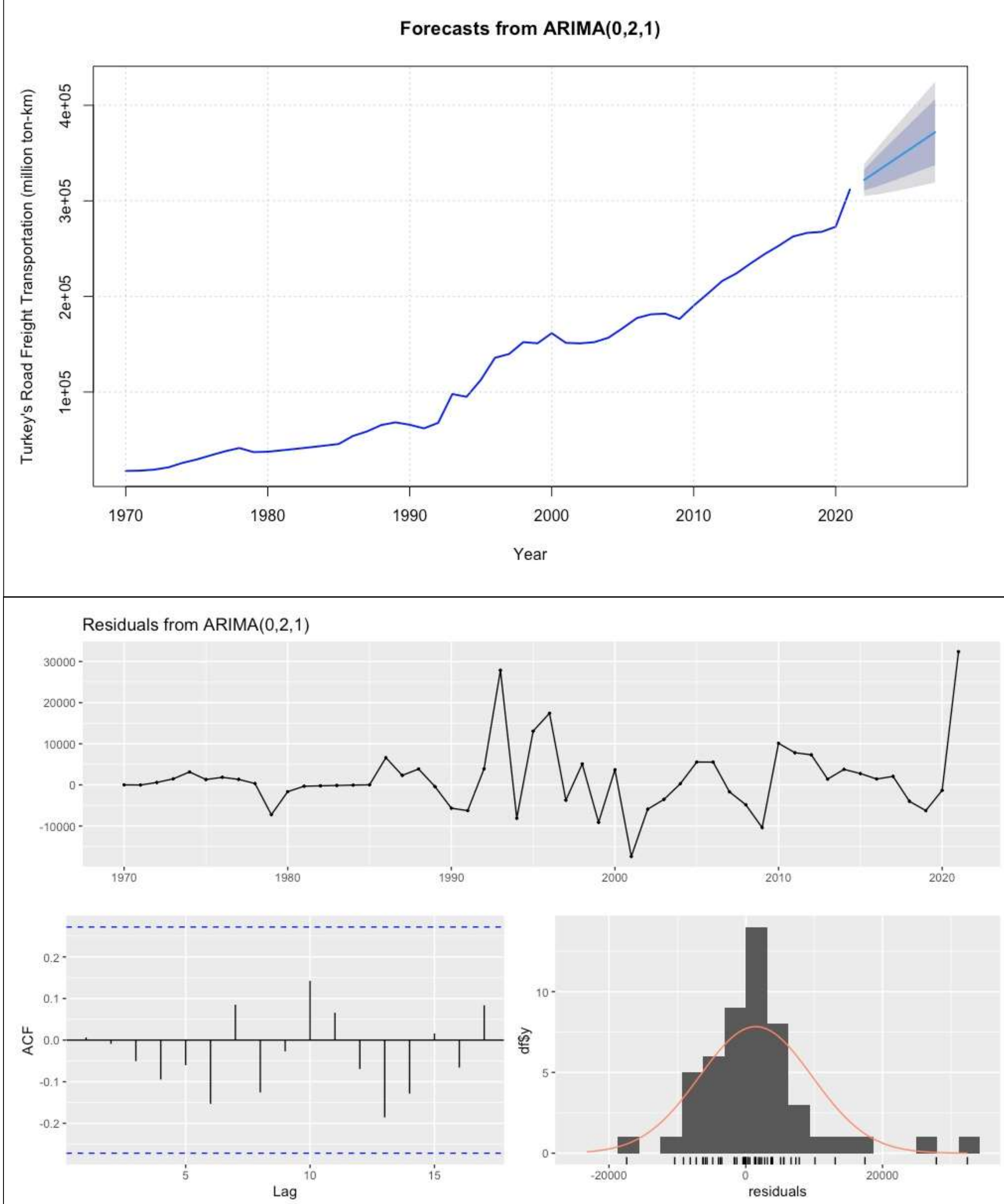
```
ETS(M,A,N)

Call:
 ets(y = timeseries2)

  Smoothing parameters:
    alpha = 0.9999
    beta  = 0.0208

  Initial states:
    l = 12497.3665
    b = 2913.0259

  sigma:  0.0884

     AIC     AICc      BIC
1146.068 1147.372 1155.824
> # Check the residuals
> checkresiduals(fit)

	Ljung-Box test

data:  Residuals from ETS(M,A,N)
Q* = 3.555, df = 10, p-value = 0.9652

Model df: 0.   Total lags used: 10

> # forecast for the next 6 years
> forecast_result <- forecast(fit, h=6)
> print(forecast_result)
     Point Forecast    Lo 80    Hi 80    Lo 95    Hi 95
2022       317202.1 281274.4 353129.7 262255.5 372148.6
2023       322589.5 270721.4 374457.7 243264.1 401915.0
2024       327977.0 263140.2 392813.9 228817.6 427136.5
2025       333364.5 256967.3 409761.7 216525.1 450204.0
2026       338752.0 251609.6 425894.4 205479.2 472024.8
2027       344139.5 246768.3 441510.7 195223.2 493055.8
```

The best-fit model of ETS on Turkey's road freight transportation is the $ETS\ (M, A, N)$ model, standing for: multiplicative error, an additive trend, and no seasonality with estimated smoothing parameters of $\alpha = 0.9999\ and\ \beta = 0.0208$.

The residuals from the ETS forecast for the road transportation of Turkey measured in ton-km do display a mean of more than zero. However, the mean is not significantly higher than zero, depicting an accurate fit of the forecast model. The ACF plot has no points exceeding the boundaries, meaning there is no autocorrelation in the data. The residuals in the histogram are considered somewhat normally distributed, and the p-value from the Ljung-box test is high, thus depicting that the ETS forecast model is suitable. The initial states of the model taken from the output data is the following: $l = 12497.3665\ and\ b = 2913.0259$.

The projected future of freight transportation by road in Turkey does show an increase for the mid-term perspective when considering the point forecast. We can also say that with 80% (dark-shaded) and 95% (light-shaded) confidence, the forecasted road freight will be within the corresponding confidence intervals, showing both a decrease and increase in road freight transported. The above values from the data output show the minimum and maximum values of the forecast for 2023 to 2027 with regard to the different confidence intervals.

Figure 39: *Road freight ARIMA Forecast for Turkey*



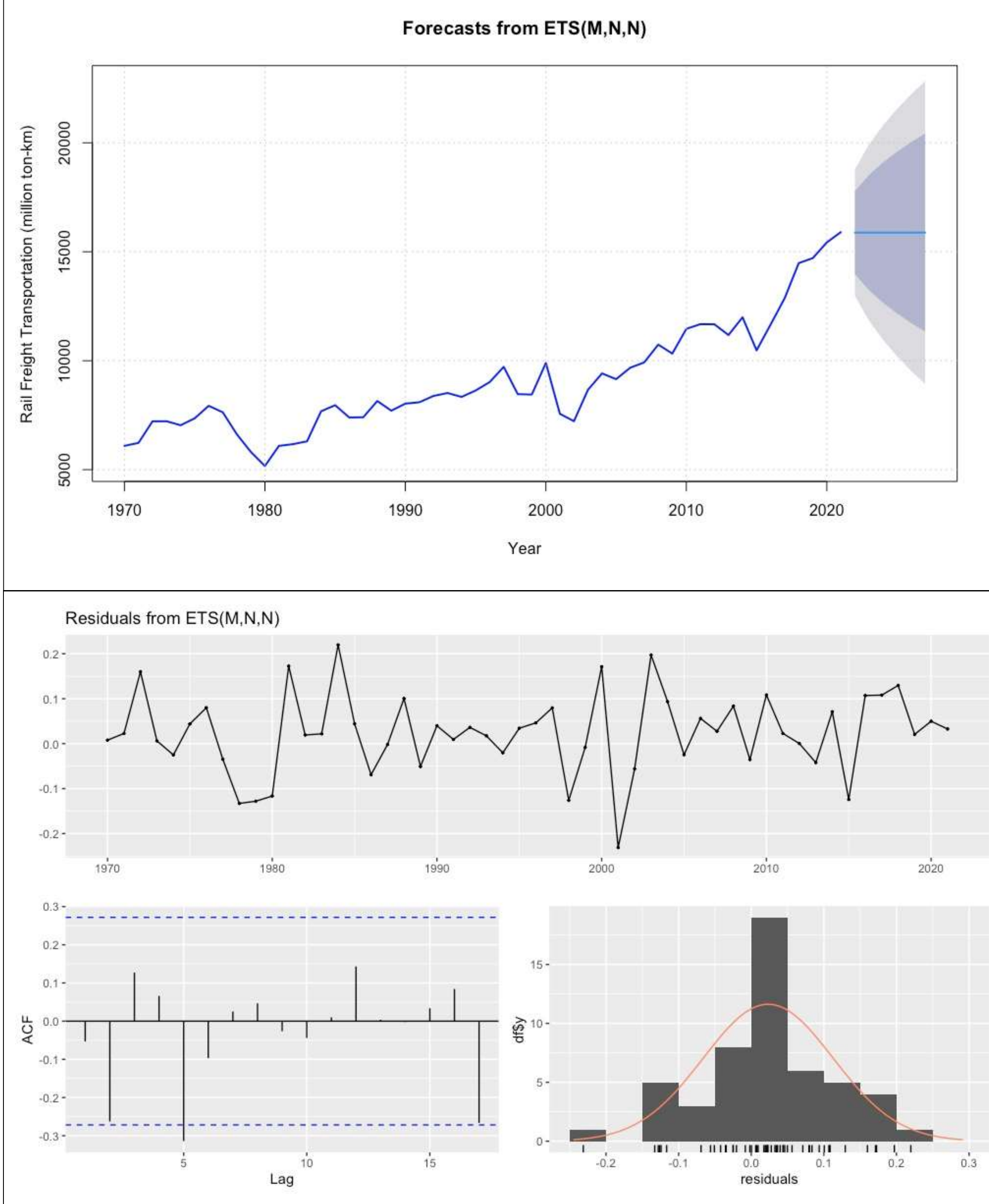
```
ARIMA(0,2,1)

Coefficients:
          ma1
      -0.8922
s.e.   0.0808

sigma^2 = 73037737:  log likelihood = -523.9
AIC=1051.8   AICc=1052.05   BIC=1055.62
> # Check the residuals
> checkresiduals(fit)

        Ljung-Box test

data:  Residuals from ARIMA(0,2,1)
Q* = 5.1933, df = 9, p-value = 0.8171

Model df: 1.   Total lags used: 10

> # forecast for the next 6 years
> forecast_result <- forecast(fit, h=6)
> print(forecast_result)
     Point Forecast    Lo 80    Hi 80    Lo 95    Hi 95
2022       321819.0 310866.6 332771.4 305068.7 338569.3
2023       331820.0 315474.5 348165.6 306821.7 356818.4
2024       341821.1 320738.9 362903.2 309578.7 374063.4
2025       351822.1 326237.3 377406.9 312693.6 390950.6
2026       361823.1 331817.4 391828.8 315933.4 407712.8
2027       371824.1 337407.2 406241.1 319187.9 424460.3
```

The best-fit model of ARIMA on Turkey's road freight transportation is zero autoregressive terms, second order of differencing, and one moving average term ($ARIMA(0,2,1)$). The residual values obtained from the ARIMA forecast for road transportation are distributed equally around 0, indicating that the mean of the residuals is 0. Moreover, the ACF plot of the residuals displays no significant spikes beyond the boundaries. The histogram of the residuals also appears to be normally distributed. The p-value of the Ljung-box test is fairly large, suggesting that the residuals are behaving like white noise, which has been captured by the model.

All these factors satisfy the criteria for a good model and provide confidence in the accuracy of the ARIMA forecast for road transportation.

Additionally, the ARIMA model shows a more robust prediction of an increase in road transportation in the future when compared to the ETS model. Moreover, the confidence interval for the 80% (dark-shaded) and 95% (light-shaded) intervals is narrower than the ETS model, indicating higher precision in the forecasted values. There are no signs of a potential decrease, further reinforcing the model's reliability in predicting an increase in road transportation.

Figure 40: *Rail freight ETS Forecast for Turkey*

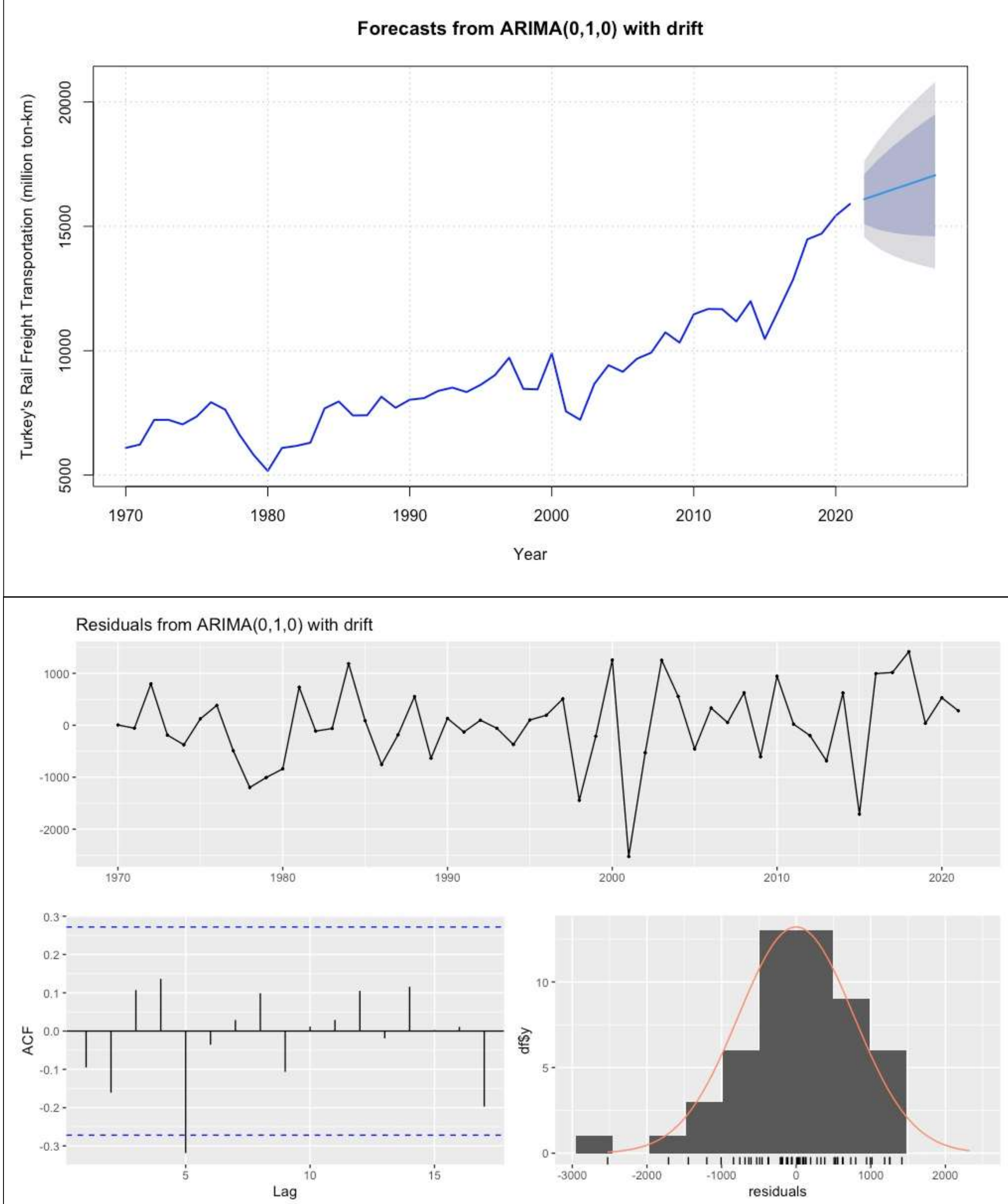


```
ETS(M,N,N)

Call:
 ets(y = timeseries2)

  Smoothing parameters:
    alpha = 0.9594

  Initial states:
    l = 6044.8772

  sigma:  0.0934

     AIC     AICc      BIC
904.8434 905.3434 910.6971
> # Check the residuals
> checkresiduals(fit)

	Ljung-Box test

data:  Residuals from ETS(M,N,N)
Q* = 12.058, df = 10, p-value = 0.2812

Model df: 0.   Total lags used: 10

> # forecast for the next 6 years
> forecast_result <- forecast(fit, h=6)
> print(forecast_result)
     Point Forecast    Lo 80    Hi 80     Lo 95    Hi 95
2022       15879.63 13978.38 17780.89 12971.913 18787.35
2023       15879.63 13239.35 18519.91 11841.675 19917.59
2024       15879.63 12661.82 19097.45 10958.409 20800.86
2025       15879.63 12169.54 19589.72 10205.537 21553.73
2026       15879.63 11732.03 20027.24  9536.418 22222.85
2027       15879.63 11333.42 20425.85  8926.794 22832.47
```

The best-fit model by ETS for rail freight transportation in Turkey differs from the rest, and is an $ETS(M, N, N)$ model, meaning multiplicative error, no trend, and no seasonality with the smoothing parameter $\alpha = 0.9594$, and initial states of $l = 6044.8772$. Because of the ETS(M,N,N) model, the point forecast shows no trend. The 80% prediction interval is showcased by the darker shaded area while the 95% interval is lightly shaded, both very wide, indicating the uncertainty of the forecast. The ranges of the two prediction intervals can be seen in Figure 41 for all the years between 2023 and 2027.

The residuals from the ETS forecast of the rail freight transportation are spread equally around zero, confirming a mean value of zero for all the residuals. However, the ACF plot does have a spike reaching outside of the boundaries, indicating that there might be a significant correlation at lag 5. Although, its only one point that is not far from the boundaries, suggesting that the violation is not a serious cause of concern, but needs to be taken carefully into consideration. The histogram does marginally showcase a normal distribution while the p-value from the Ljung-box test is large, meaning no probable case of autocorrelation for residuals who is not to be distinguishable from a white noise series.

Figure 41: *Rail freight ARIMA Forecast for Turkey*

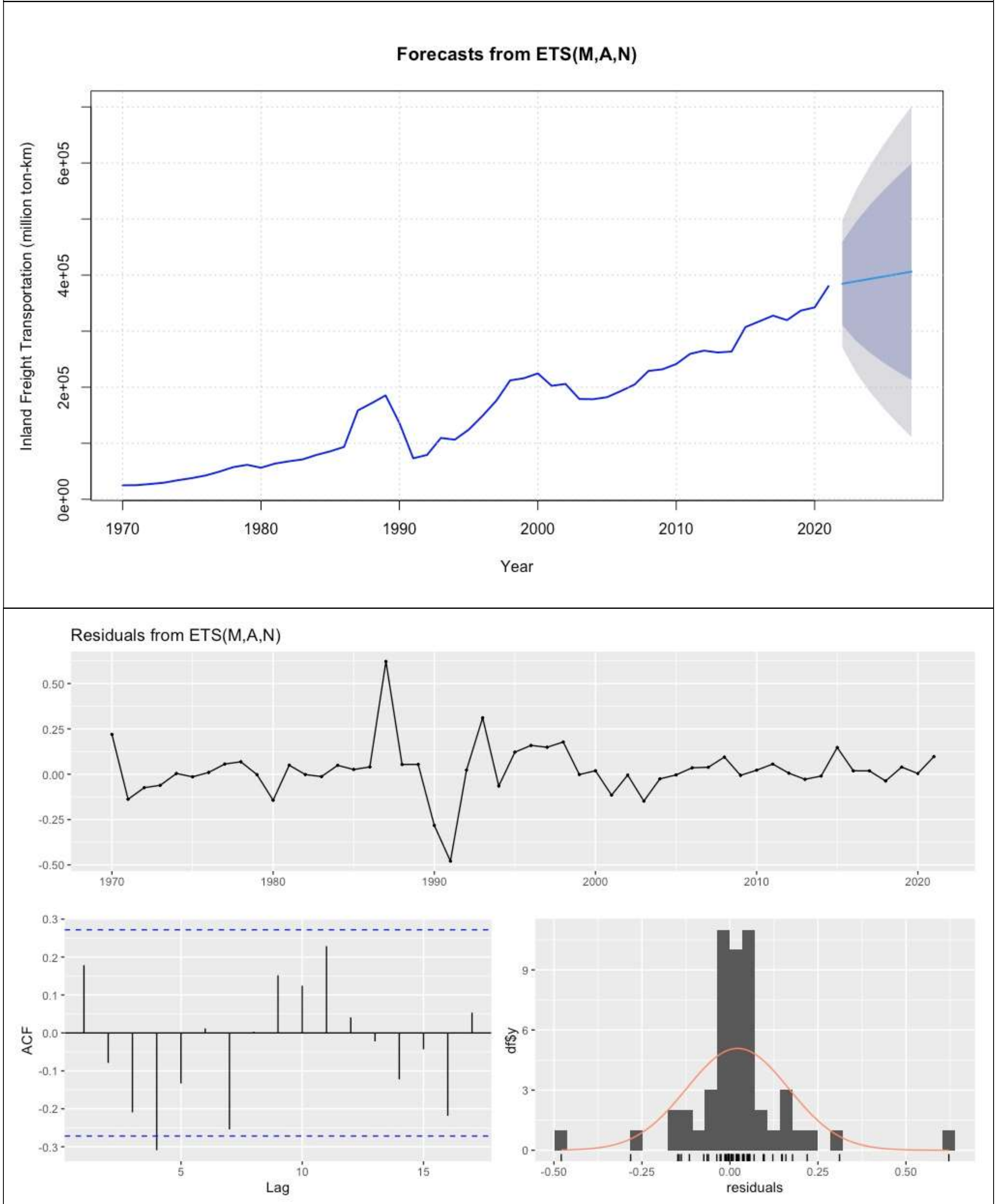


```
ARIMA(0,1,0) with drift

Coefficients:
         drift
      192.3137
s.e.  108.4071

sigma^2 = 611344:  log likelihood = -411.61
AIC=827.22   AICc=827.47   BIC=831.08
> # Check the residuals
> checkresiduals(fit)

        Ljung-Box test

data:  Residuals from ARIMA(0,1,0) with drift
Q* = 11.283, df = 10, p-value = 0.3359

Model df: 0.   Total lags used: 10

> # forecast for the next 6 years
> forecast_result <- forecast(fit, h=6)
> print(forecast_result)
     Point Forecast    Lo 80    Hi 80    Lo 95    Hi 95
2022       16092.31 15090.29 17094.34 14559.85 17624.78
2023       16284.63 14867.55 17701.71 14117.39 18451.86
2024       16476.94 14741.38 18212.50 13822.63 19131.25
2025       16669.25 14665.20 18673.31 13604.32 19734.19
2026       16861.57 14620.97 19102.17 13434.87 20288.27
2027       17053.88 14599.43 19508.33 13300.12 20807.64
```

The best-fit model of ARIMA on Turkey's rail freight transportation is 0 autoregressive terms, first order of differencing, and no moving average terms or $ARIMA(0,1,0)$. The ACF plot of the ARIMA forecast for rail freight transportation also violates the boundaries at lag 5. However, the histogram showcases a more normally distributed outcome. The residuals also have a mean of zero, indicating a good fit of the model. Moreover, The Ljung-box test returned a relatively high p-value, indicating that the residuals in the ARIMA model are likely behaving like white noise. This suggests that the model has effectively captured the patterns present in the residuals.

The ARIMA model gives a narrower prediction interval compared to the ETS, and also indicates an increase through the point prediction. The maximum and minimum values of the dark-shaded 80% and light-shaded 95% prediction intervals between 2023 and 2027 are given in Figure 41, indicating a higher probability of a continuous increase in rail freight transportation.

Figure 42: *Inland freight ETS Forecast for Turkey*



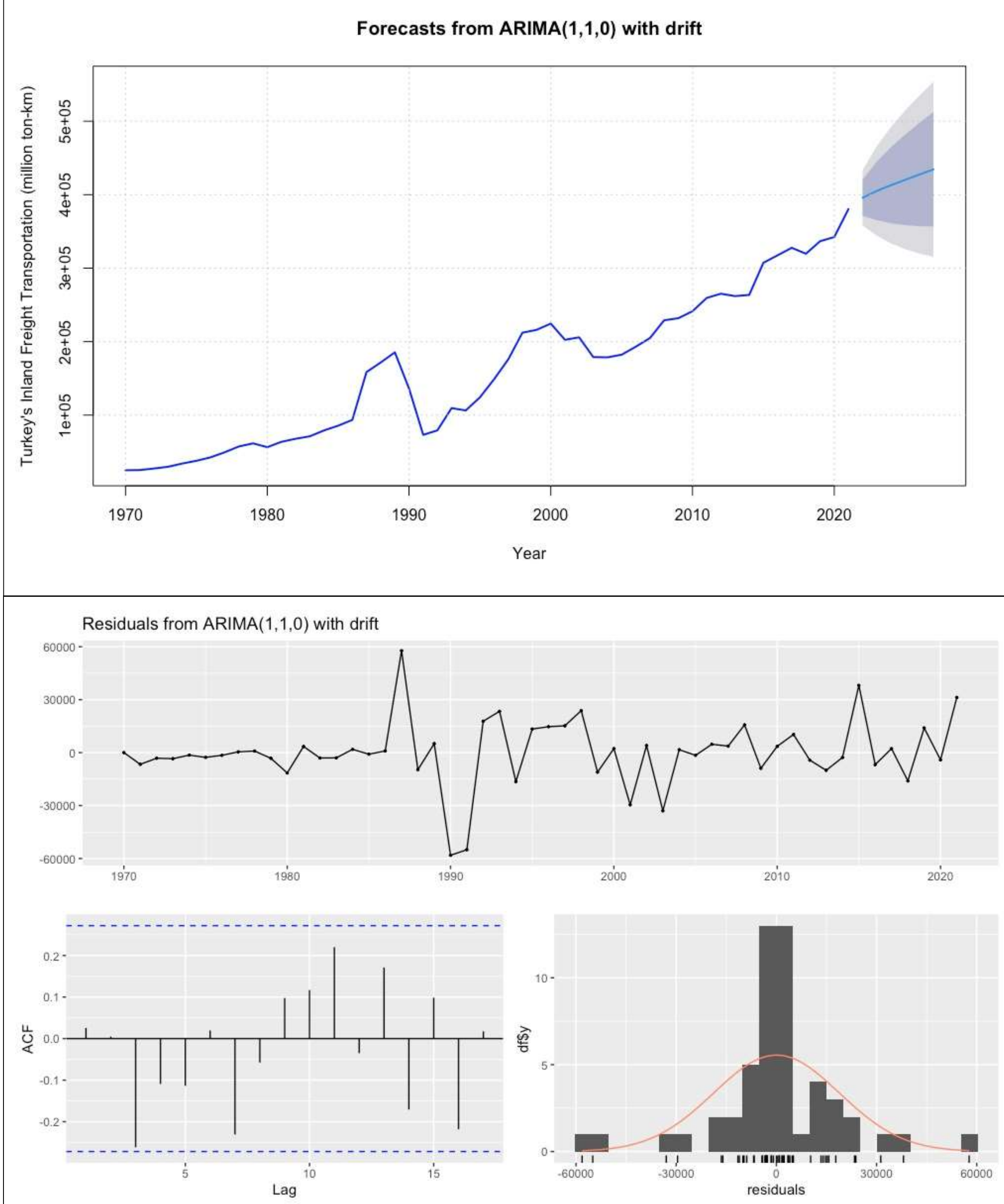
```
ETS(M,A,N)

Call:
 ets(y = timeseries2)

  Smoothing parameters:
    alpha = 0.9999
    beta  = 1e-04

  Initial states:
    l = 16182.625
    b = 4286.1628

  sigma:  0.1512

     AIC     AICc      BIC
1235.568 1236.873 1245.325
> # Check the residuals
> checkresiduals(fit)

        Ljung-Box test

data:  Residuals from ETS(M,A,N)
Q* = 17.87, df = 10, p-value = 0.0572

Model df: 0.   Total lags used: 10

> # forecast for the next 6 years
> forecast_result <- forecast(fit, h=6)
> print(forecast_result)
     Point Forecast    Lo 80    Hi 80    Lo 95    Hi 95
2022       384638.9 310112.5 459165.3 270660.6 498617.2
2023       388939.2 282354.9 495523.5 225932.6 551945.9
2024       393239.5 261230.5 525248.5 191349.2 595129.9
2025       397539.8 243396.9 551682.8 161798.5 633281.1
2026       401840.1 227573.7 576106.6 135322.6 668357.7
2027       406140.4 213111.0 599169.8 110927.4 701353.4
```

Regarding inland freight transportation in Turkey, the $ETS(M, A, N)$ is once again considered the best fit model of the ETS. $ETS(M, A, N)$ means there is multiplicative error, an additive trend, and no seasonality with the following smoothing parameters: $\alpha = 0.9999\ and\ \beta = 10^{-4}$. The initial state of the output gives $l = 16182.625\ and\ b = 4286.1628$.

The future of inland transportation in Turkey has an increasing trend with regards to the point forecast, although a significant possibility of a declining trend considering the prediction interval of 80% (dark-shaded) and 95% (light-shaded). The minimum and maximum values of the different prediction intervals for the year span of 2023 to 2027 are given in Figure 42.

The residual plot depicts that the mean value of the residuals is 0, fulfilling the criteria of a good model fit. However, lag value 4 exceeds the boundary in the ACF plot, the histogram of the residuals does not display a clear sign of normal distribution, and the Ljung-box p-value is very low, all showcasing that we cannot conclude that there are no case of autocorrelation for the residuals. Thus, the model of use might not be suitable for the data.

Figure 43: *Inland freight ARIMA Forecast for Turkey*

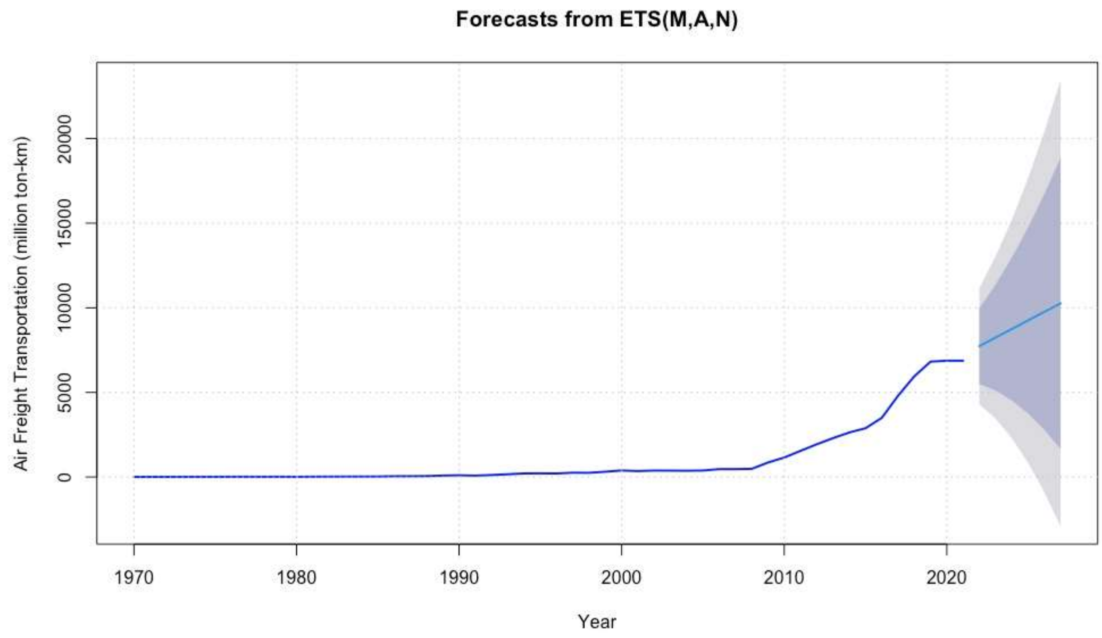


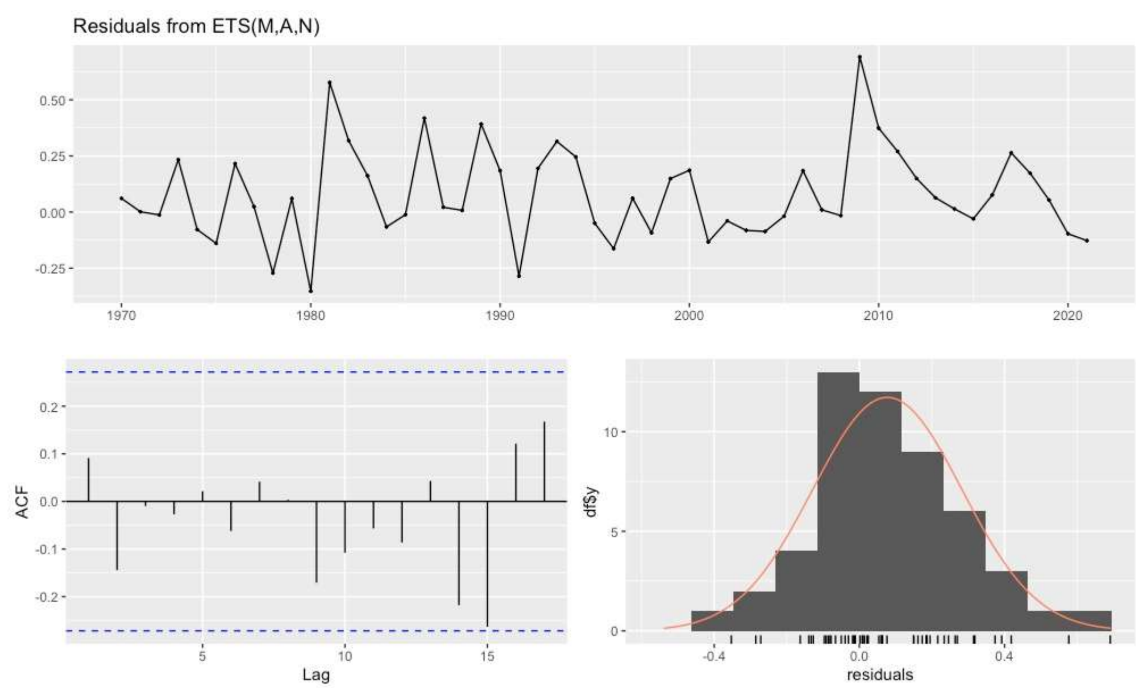
```
Series: timeseries1
ARIMA(1,1,0) with drift

Coefficients:
         ar1     drift
      0.2688  7141.091
s.e.  0.1370  3585.500

sigma^2 = 369706522:  log likelihood = -574.45
AIC=1154.91   AICc=1155.42   BIC=1160.7
> # Check the residuals
> checkresiduals(fit)

        Ljung-Box test

data:  Residuals from ARIMA(1,1,0) with drift
Q* = 10.533, df = 9, p-value = 0.3091

Model df: 1.   Total lags used: 10

> # forecast for the next 6 years
> forecast_result <- forecast(fit, h=6)
> print(forecast_result)
     Point Forecast    Lo 80    Hi 80    Lo 95    Hi 95
2022       395780.8 371139.5 420422.2 358095.1 433466.5
2023       405152.3 365344.2 444960.5 344271.0 466033.7
2024       412892.9 361156.3 464629.6 333768.6 492017.3
2025       420195.2 358546.8 481843.6 325912.1 514478.3
2026       427379.6 357144.2 497615.0 319963.8 534795.4
2027       434532.3 356636.1 512428.6 315400.2 553664.4
```

In Turkey's inland freight transportation, the best fit ARIMA model is identified as $ARIMA(1,1,0)$. The histogram depicts somewhat of a normal distribution, the ACF plot has no spikes violating the boundaries, and the residuals have an approximate mean of zero. Thus, indicating a good fit of the model. Additionally, The ARIMA model appears to have captured the patterns present in the residuals, as evidenced by a relatively high p-value obtained from the Ljung-box test. This suggests that the residuals are behaving like white noise.

In terms of prediction, the ARIMA model provides a narrower prediction interval than the ETS model and indicates an increase through the point interval. Figure 43 shows the maximum and minimum values of the dark-shaded 80% and light-shaded 95% prediction intervals between 2023 and 2027. This suggests a higher probability of a continuous increase in inland freight transportation.

Figure 44: *Air freight ETS Forecast for Turkey*



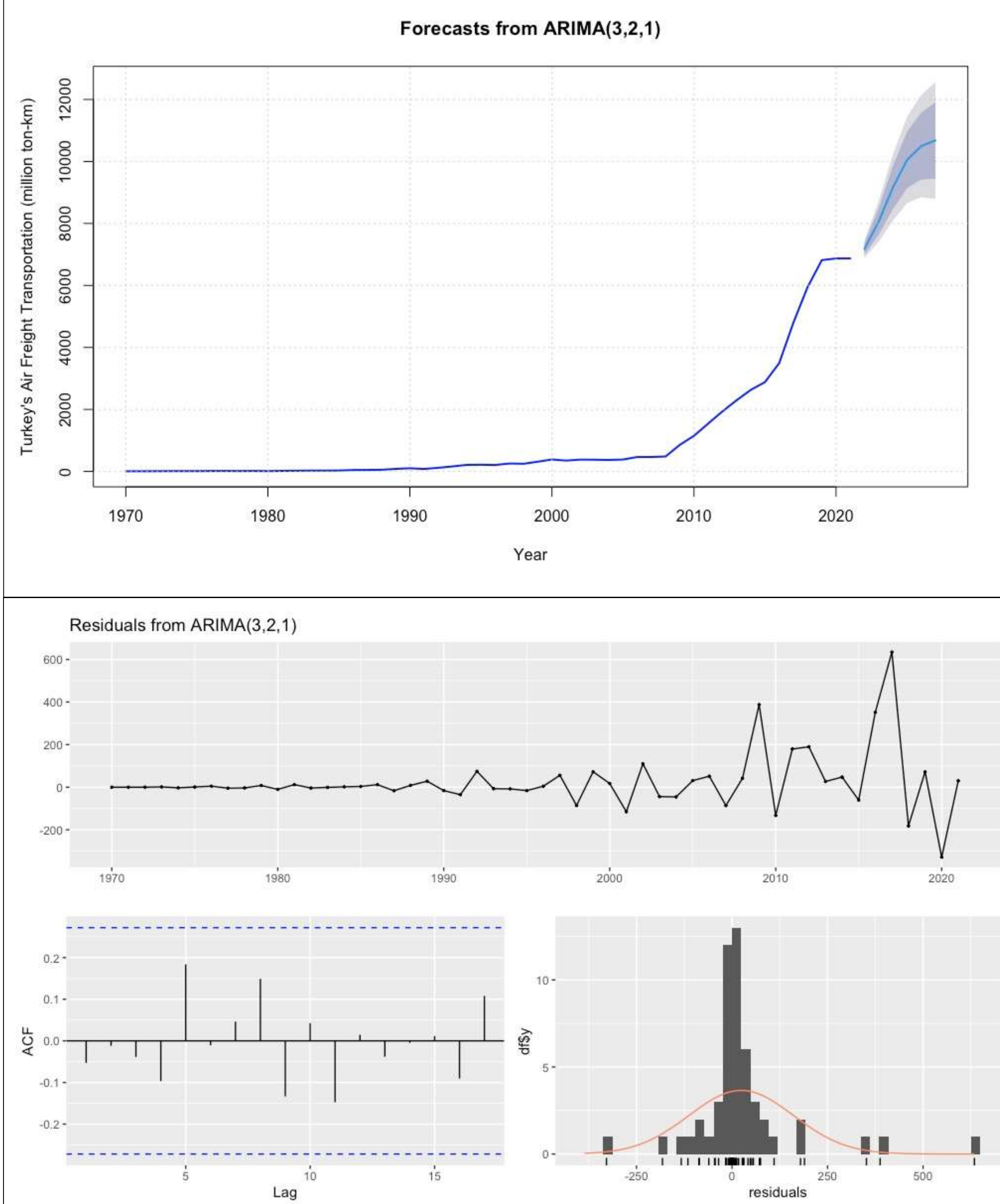


```
ETS(M,A,N)

Call:
 ets(y = timeseries2)

  Smoothing parameters:
    alpha = 0.6456
    beta  = 0.2342

  Initial states:
    l = 3.6879
    b = 1.0249

  sigma:  0.2263

     AIC     AICc      BIC
584.4745 585.7789 594.2308
> # Check the residuals
> checkresiduals(fit)

        Ljung-Box test

data:  Residuals from ETS(M,A,N)
Q* = 4.7228, df = 10, p-value = 0.9089

Model df: 0.   Total lags used: 10

> # forecast for the next 6 years
> forecast_result <- forecast(fit, h=6)
> print(forecast_result)
     Point Forecast     Lo 80     Hi 80      Lo 95    Hi 95
2022       7731.995 5489.184  9974.807  4301.9118 11162.08
2023       8238.172 5107.045 11369.300  3449.5264 13026.82
2024       8744.349 4524.578 12964.121  2290.7658 15197.93
2025       9250.526 3754.041 14747.011   844.3779 17656.67
2026       9756.703 2796.990 16716.416  -887.2595 20400.67
2027      10262.880 1648.335 18877.425 -2911.9296 23437.69
```

When forecasting air freight transportation in Turkey via ETS, $ETS(M, A, N)$ is the model of choice. This stands for multiplicative error, an additive trend, and no seasonality. The best fit model also contains the smoothing factors of $\alpha = 0.6456\ and\ \beta = 0.2342$, while the initial states for $l\ and\ b$ is $3.6879\ and\ 1.0249$ respectively. There is also a rapid increase showcased by the point forecast, however a wide is range is displayed by the prediction intervals of both 80% (dark-shaded) and 90% (light-shaded), making the prediction uncertain. The values of the different intervals are showcased in the data output in Figure 44.

The plot of the residuals displays a mean of zero for the residuals, and the histogram showcases a somewhat normal distribution. The ACF plot displays no signs of spikes outside the boundaries and the Ljung-box test displays a large p-value. Based on the results of the analysis, it can be concluded that the residuals from the ETS forecast model for Turkey's road freight transportation do not show any significant autocorrelation, and they resemble a white noise series. These findings confirm that the ETS model used is suitable and provides a good fit to the data.

Figure 45: *Air freight ARIMA Forecast for Turkey*

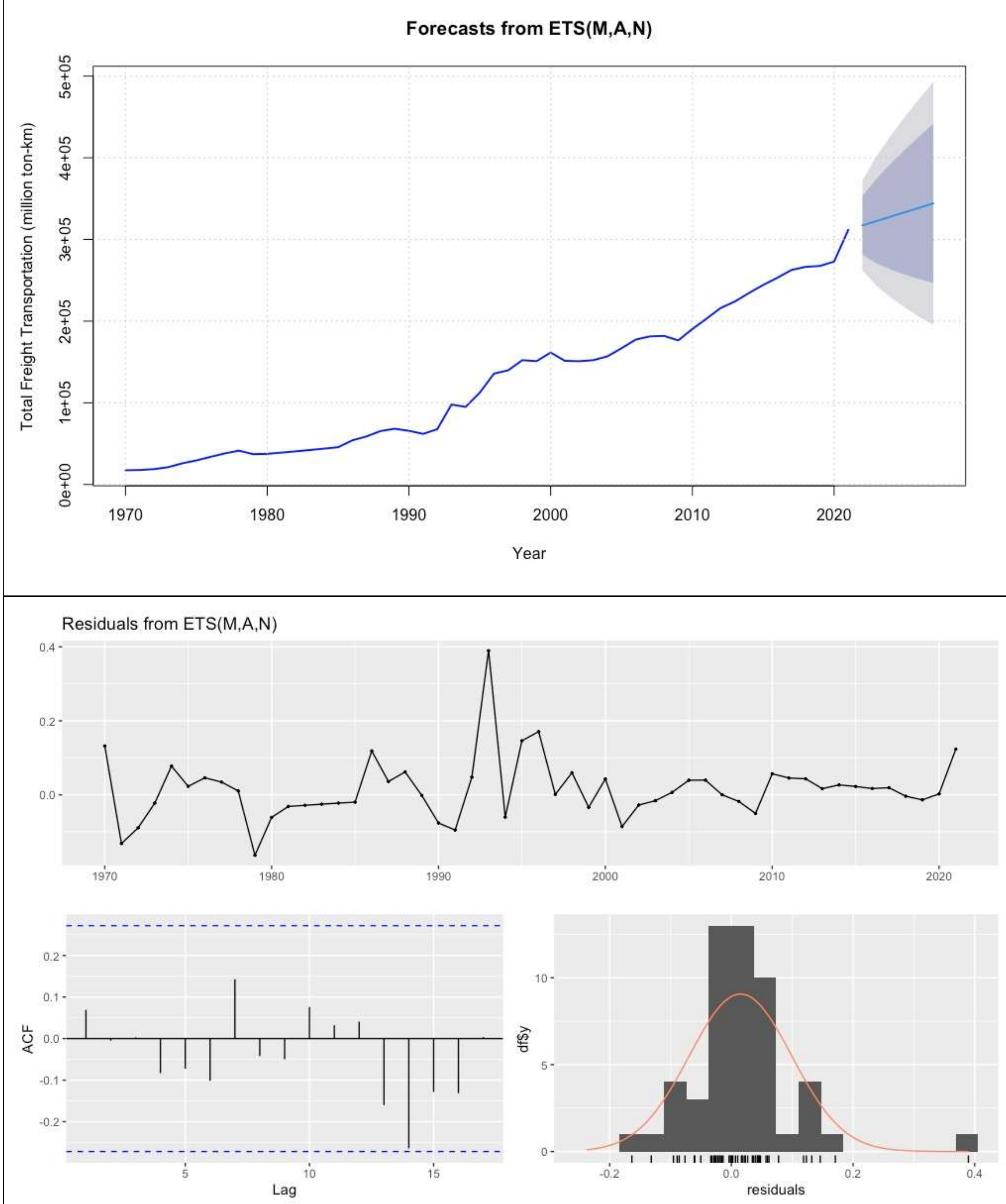


```
Series: timeseries1
ARIMA(3,2,1)

Coefficients:
         ar1      ar2      ar3      ma1
      0.4429  -0.2349  -0.5302  -0.3804
s.e.  0.1854   0.1662   0.1601   0.1881

sigma^2 = 21377:  log likelihood = -319.26
AIC=648.53   AICc=649.89   BIC=658.09
> # Check the residuals
> checkresiduals(fit)

        Ljung-Box test

data:  Residuals from ARIMA(3,2,1)
Q* = 5.6477, df = 6, p-value = 0.4638

Model df: 4.   Total lags used: 10

> # forecast for the next 6 years
> forecast_result <- forecast(fit, h=6)
> print(forecast_result)
     Point Forecast    Lo 80     Hi 80    Lo 95     Hi 95
2022       7175.420 6988.044  7362.796 6888.853  7461.986
2023       8058.108 7628.620  8487.595 7401.264  8714.952
2024       9154.555 8459.319  9849.791 8091.284 10217.827
2025      10048.439 9138.103 10958.775 8656.200 11440.677
2026      10495.799 9417.087 11574.512 8846.051 12145.548
2027      10679.648 9447.114 11912.181 8794.650 12564.645
```

$ARIMA(3,2,1)$ is accordingly the best-fit model by ARIMA when forecasting air freight transportation in Turkey. The histogram exceeds the normal distribution but resembles some likeliness, while the residuals plot and ACF plot follow the criteria of a good fit of model. When the Ljung-Box test returns a large p-value like in this case, the test result suggests that the ARIMA model has captured the patterns present in the residuals, which suggests that the residuals are behaving like white noise.

Regarding the forecast, ARIMA predicts a sharp continuous increase in air freight transportation in Turkey for the upcoming years. Additionally, the prediction intervals are narrow, suggesting a good accuracy of the prediction. The exact numbers of the prediction intervals are given in the output for the years 2023 to 2027 in Figure 45. Interestingly, the forecast suggests a possible slowing of the increase in a few years.

Figure 46: *Total Freight ETS Forecast for Turkey*

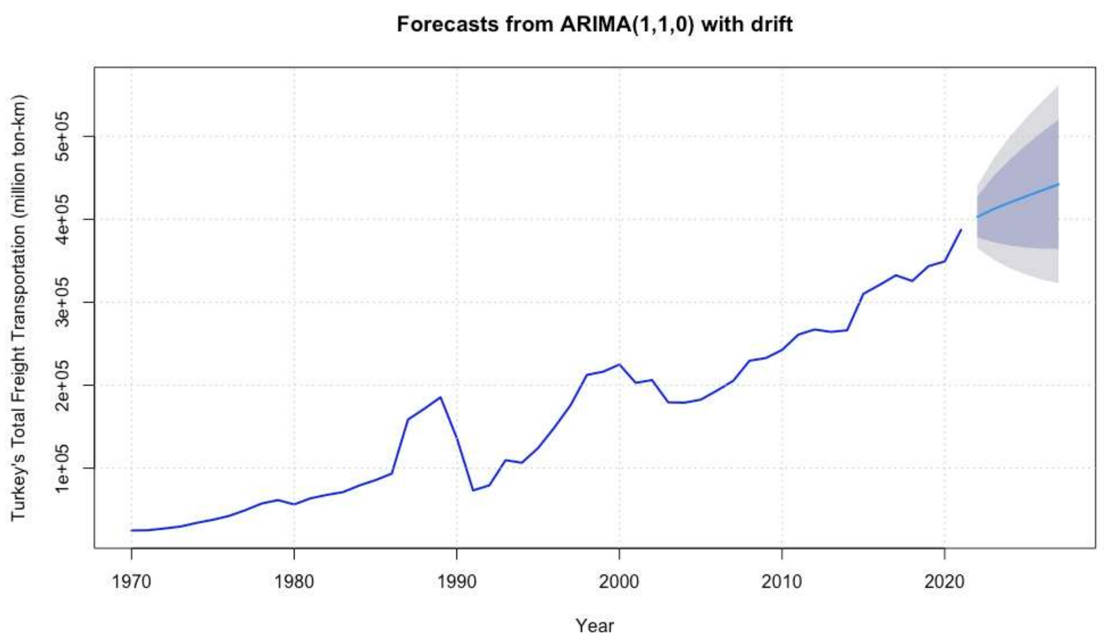


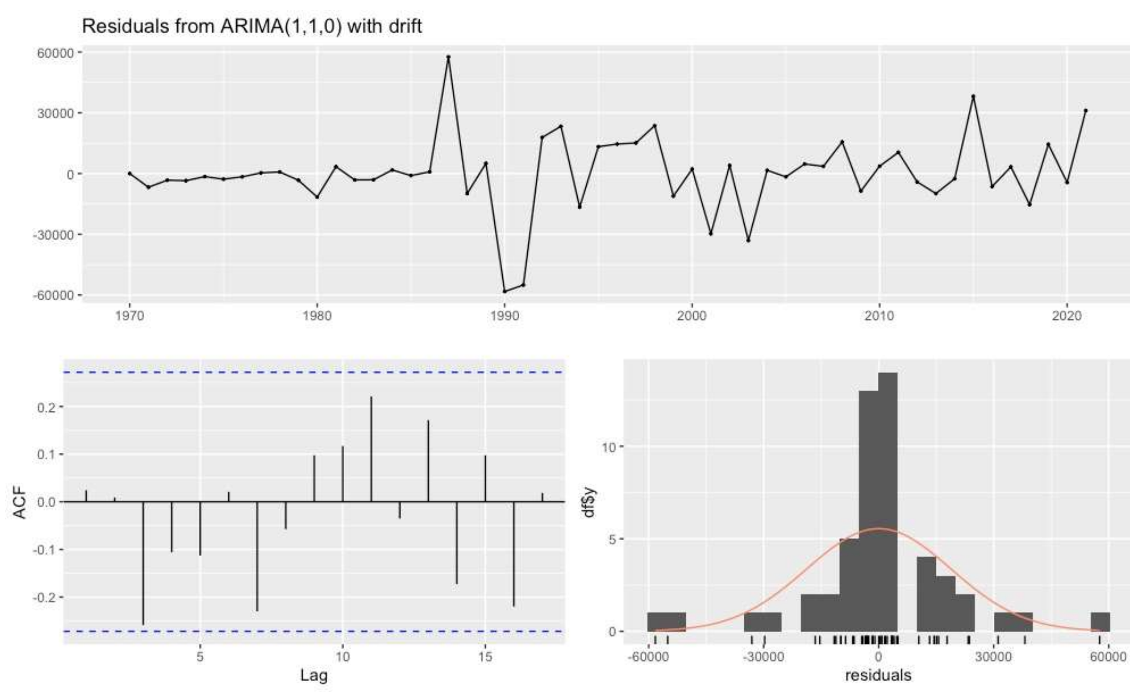
```
ETS(M,A,N)

Call:
 ets(y = timeseries2)

  Smoothing parameters:
    alpha = 0.9999
    beta  = 0.0208

  Initial states:
    l = 12497.3665
    b = 2913.0259

  sigma:  0.0884

     AIC     AICc      BIC
1146.068 1147.372 1155.824
> # Check the residuals
> checkresiduals(fit)

        Ljung-Box test

data:  Residuals from ETS(M,A,N)
Q* = 3.555, df = 10, p-value = 0.9652

Model df: 0.   Total lags used: 10

> # forecast for the next 6 years
> forecast_result <- forecast(fit, h=6)
> print(forecast_result)
     Point Forecast    Lo 80    Hi 80    Lo 95    Hi 95
2022       317202.1 281274.4 353129.7 262255.5 372148.6
2023       322589.5 270721.4 374457.7 243264.1 401915.0
2024       327977.0 263140.2 392813.9 228817.6 427136.5
2025       333364.5 256967.3 409761.7 216525.1 450204.0
2026       338752.0 251609.6 425894.4 205479.2 472024.8
2027       344139.5 246768.3 441510.7 195223.2 493055.8
```

The ETS model with parameters $ETS(M, A, N)$ was applied to model total freight transportation in Turkey, and the estimated smoothing parameters are $\alpha = 0.9999 \; and \; \beta = 0.0208$. The initial states for the model are $l = 312497.3665 \; and \; b = 2913.0259$. The point forecast shows a significant upward trend. The corresponding prediction intervals of 80% and 95%, are showcased by the shaded areas, where the lightly shaded is the 95% and the darker shaded is the 80% prediction interval. The values of the lower and higher values of each forecasted interval is provided in Figure 46.

The analysis of the residuals indicates that they have a mean of approximately zero, and their histogram shows a roughly normal distribution. The ACF plot of residuals does not show any spikes outside the boundaries, and the Ljung-Box test displays a large p-value. These findings suggest that the residuals do not exhibit any significant autocorrelation and resemble a white noise series. Consequently, it can be concluded that the ETS model used for total freight transportation in Turkey is appropriate and fits the data well.

Figure 47: *Total Freight ARIMA Forecast for Turkey*





```
Series: timeseries1
ARIMA(1,1,0) with drift

Coefficients:
         ar1     drift
      0.2702  7275.051
s.e.  0.1369  3592.640

sigma^2 = 369768881:  log likelihood = -574.46
AIC=1154.92   AICc=1155.43   BIC=1160.71
> # Check the residuals
> checkresiduals(fit)

        Ljung-Box test

data:  Residuals from ARIMA(1,1,0) with drift
Q* = 10.361, df = 9, p-value = 0.3221

Model df: 1.   Total lags used: 10

> # forecast for the next 6 years
> forecast_result <- forecast(fit, h=6)
> print(forecast_result)
     Point Forecast    Lo 80    Hi 80    Lo 95    Hi 95
2022       402794.1 378150.7 427437.5 365105.2 440483.0
2023       412313.8 372474.3 452153.2 351384.6 473242.9
2024       420195.4 368398.0 471992.8 340978.1 499412.7
2025       427634.4 365899.5 489369.3 333219.0 522049.7
2026       434953.7 364609.8 505297.7 327372.0 542535.5
2027       442240.7 364216.8 520264.7 322913.4 561568.1
```

Based on the analysis of total freight transportation in Turkey, the best fit ARIMA model is identified as $ARIMA(1,1,0)$. The p-value from the Ljung-box test is large, meaning that the residuals are, in this case, acting like white noise, and there is no autocorrelation in the residuals.

The residuals, ACF, and the histogram plots for the $ARIMA(1,1,0)$ model all meet the criteria for a good fit model. The residuals have a mean of zero, indicating that the model is unbiased. The ACF plot shows that the autocorrelation of the residuals is not significant, which suggests that the model captures the underlying pattern in the data. The histogram of the residuals also shows that the distribution is close to a normal distribution, indicating that the model captures the variability in the data.

In terms of prediction, the forecast using the $ARIMA(1,1,0)$ model has moderately narrow prediction intervals, as seen in Figure 47. The point forecast and the prediction intervals indicate a high probability of an increasing trend for the future.

5. DISCUSSION

This report aims to investigate the logistics sector within an emerging economic country and forecast logistics for a short- and mid-term perspective. This is done by determining the different factors influencing the logistics sector within an emerging market and then using all the available data to create forecasts. The forecast is performed for road-, rail-, inland-, air-, and a total freight transportation. The report is limited to the two emerging economy countries, Mexico, and Turkey. The World Bank database is the source of information for the numerous elements that may affect logistics and freight transportation. Meanwhile, the data we employ to predict freight transportation, categorized by modes, comes from the Organization for Economic Cooperation and Development (OECD) database.

As discussed in the theory section of the report, Mexico and Turkey are developing as emerging economies because of various factors. Two research questions are analyzed in the report and discussed below.

RQ1: What factors can influence the logistics sector in an emerging economy country?

Time series regression model along with the principal component analysis (PCA) method are used to address RQ1. Moreover, the properties of residuals depicted in Figure 25 for Mexico and Figure 27 for Turkey validate our results.

For Mexico, the variables of importance concluded by the analysis are divided into the two principal components of PC1 which is the GDP component, and PC2 which is the mixed demographic and logistical component, as illustrated in Table 5. These are statistically significant in the regression model with a large, adjusted R-squared of 0.963. PC1 includes the variables: *final consumption expenditure (Constant LCU & Constant US$)* and *GDP* (*Constant LCU & Constant US$).* As seen from the variables, PC1 can be considered a GDP factor. Thus, the final consumption expenditure variables are a component of the GDP with the following description: "*Final consumption expenditure is the value of goods and services purchased by households and government for their final use",* suggesting a direct relationship with the GDP.

The mentioned variables in PC1 are highly relevant and illustrate Mexico's logistics growth. With consumption patterns in the population both changing and increasing, as indicated by the final expenditure variables, the current GDP of Mexico illustrates a correlation with the freight transported.

Variables included in PC2 are *changes in inventories (current US$), Final consumption expenditure (% of GDP), Food, beverages and Tobacco (% of value added in manufacturing), Manufactured exports (% of merchandise exports), Merchandise exports to high-income countries (% of total merchandise exports), Population ages 0-14, Real effective exchange rate index (2010=100), Rural population growth (annual %).* The variables of PC2 suggest that it can be interpreted as a mixed factor of significant demographic and logistical variables. As previously discussed, demographic variables, such as population ages and rural population growth, is crucial regarding the growth of an already emerging market. Mapping out these variables through the principal component methods presents a clear answer to the research question of what factors can affect the logistics industry in Mexico, and further strengthens what is written in the theory section.

As seen from the results, *Merchandise exports to high-income economies* significantly impact the growth of freight transportation in Mexico. And once again, this can be contributed greatly to the increase in trade between Mexico and USA. The USA is previously described as the most prominent trading partner for Mexico. Thus, due to the near-shoring trend of U.S.

industries, the incorporation of Mexico into the supply chains is crucial for the positive effect on the development of both logistics and the economy in Mexico.

The young population of Mexico, also previously discussed in section 2.1.1, and its importance in an emerging market discussed in section 2.1**,** substantially impacts freight transportation's growth. The *Rural population growth (annual %)* is another variable that considerably impacts freight transportation as the dependent variable. As discussed in section 2.1, the typical trend of urbanization within an emerging economy is confirmed concerning Mexico, with negative rural population growth, meaning more people are heading towards urban areas. This can, however, also lead to negative consequences as industries are spread around the country, and not only located around large international cities. However, this variable's relevance positively affects freight, hence contradicting that statement.

Financial factors such as the current GDP within Mexico are also revealed as variables affecting freight transportation positively. Merchandise, food, beverages, and tobacco is discovered to be the primary exporting goods regarding the impact on freight transportation from the country. Mexico has many industries of importance that affect the GDP of the country. Industries such as manufacturing, agriculture, and exports are all industries that affect the GDP and, in turn, freight transportation. An increase in economic growth and GDP further affirms the development of the industrial sector, as explained in section 2.1.

For Turkey, we can retrieve three different principal components, as illustrated in Table 6: PC1 which is the logistical component, PC2 which is the economic component, and PC3 which is the travel services component. These are proven statistically significant in the regression model with an adjusted R-squared of 0.8759. A composition of demographic, economic, and logistical factors in PC1 and PC2 affect the future trend of logistics in Turkey. For instance, for PC1, we see the variables *Population ages 16-64, population in urban agglomerations of more than 1 million,* and *population in largest city* as demographic factors, *GDP per capita external debt stocks*, *primary income payments, external debt stocks (public and publicly guaranteed),* and *external debt stocks (total)* as economic factors, and *merchandise imports and exports by the reporting economy, export and import volume index, import and export value index, merchandise imports and exports, industry (value added), imports and exports of goods, services and primary income, service imports, commercial service imports, goods imports and exports, manufacturing value index,* and *commercial service imports and exports* as logistical aspects. For PC2, the significant factors were concluded as the following: *Population ages 0-14* as a demographic factor, *Final consumption expenditure, deposit interest rate, adjusted saving: consumption of fixed capital,* and *net secondary income* as the economic factors, and *Merchandise imports and exports to high-income countries, manufactures imports and exports, agricultural raw materials imports,* and *trade in services* as logistics factors. Finally, for PC3, we can retrieve two variables and determine it to be the travel services factor with the two variables of: *Travel services (% of commercial service exports),* and *Travel services (% of service exports, BoP).*

These variables further strengthen the importance of the population in Turkey, for instance the age distribution of the population. It also depicts the importance of the export market to high-income countries for Turkey, and economic factors such as GDP are also prominent.

The number of factors that affect the logistics trend in Turkey is much more than in Mexico. However, many variables were recognizable and in line with what was retrieved from the principal component analysis of Mexico. Firstly, when comparing the variables between the countries, it is noticeable that Turkey has a better balance between import and export variables. Both service imports and exports and merchandise exports and imports are prevalent. Whereas in Mexico, a larger focus on exports is placed. One aspect that could explain this is Turkey's geographical position and the future strategies set by the country, which are explained in section 2.1.2. With the country being located between the middle east and Europe and the middle rail corridor, which starts in China and goes through Turkey, it further affirms that Turkey's geographical position is vital for its development. Industrial outsourcing to Turkey, which is once again described in section 2.1.2, and the amount of industry and logistics-related variables have a linear relationship where we can see that this does affect the logistics sector in a vastly positive way.

Factors such as GDP and other variables related to the GDP were prominent once again. Moreover, the working generation, namely the *population ages 16-64,* impacts freight transportation. Thus, validating the idea of an increased workforce attributing positively to an emerging market, as discussed in section 2.1. Furthermore, the *population ages 0-14* similarly impacts freight transportation, showcasing the importance of a young population.

Similar to what is proven for Mexico, the trend of urbanization is also a fact in Turkey. Two variables related to a population moving into larger cities are proven impactful on the dependent variable of freight transport. Apparent changes in consumption patterns can also be concluded with factors such as travel services and the import of merchandise from high-income economies, described as typical characteristics of a rising market, and something which revealed as having a very high relevance of the impact on freight transport.

RQ2: How is the future of logistics forecasted for short-term and mid-term in the emerging economy country?

For RQ2, the results of the different forecasts display an increasing trend for all the transportation modes for Mexico and Turkey. Our models for forecasting the data are the ETS and ARIMA models. For Mexico, the forecasts illustrate total growth for all modes of transport. All modes of transportation have a steady growth of forecast every year, except the forecast of air freight transported by ARIMA. Where it is forecasted to increase up until 2025, and then decrease once from 2025 to 2027. However, compared to 2022, it is still an increase in general movement. If we compare this to Turkey, all modes of transport, except for rail freight ETS forecast, also show an increase in freight transported.

One focal point when discussing the forecast models and how freight transport will develop for our time perspective is that the models differ in how they forecast. This is also noticeable in the respective forecast outputs. The prediction intervals for the ETS models are significantly more expansive than that of ARIMA. That is because the models differ in how they model the underlying structure of the data. The ARIMA model is considered more conservative because of how it handles the data. An essential factor of the ARIMA model is to make the data stationary. This directly affects how the forecasts are made because stationarity causes the statistical properties of the data, such as the mean, to be constant and not fluctuate over time. This can result in a more conservative forecast because the model is based on a simplified version of the time series that is less likely to contain extreme or outlying values. In contrast, ETS models do not require the data to be stationary and can capture both short-term and long-term trends and any seasonal patterns that may exist in the data. This means that the statistical properties of the data in the ETS model vary more over time as opposed to the ARIMA model. As a result, ETS models can produce more volatile and less conservative forecasts than ARIMA models.

A common way to decide which forecasting model is the more appropriate for the dataset is through error measurements such as the AIC, AICc, and BIC measurements previously described in section 3.1.4 of the methodology chapter. These measurements are all ways to measure the goodness of fit for a model, and lower values generally indicate a better fit for the dataset.

Starting with viewing the forecasted results for road transportation in Mexico, the values of AIC, AICc, and BIC does not differ much. However, the ARIMA forecast has lower values, indicating a more optimal fit to the data and better forecasts. Continuing with analyzing the two different forecast models for Mexico, it is noticeable that ARIMA is better for forecasting the future for the individual modes of transport. Compared to the ETS forecasts, lower AIC, AICc, and BIC values are prevalent in the ARIMA forecasts. However, when analyzing the complete data set and merging all the modes of transport into a total freight variable, ETS is the better model. One reason for this is that ETS captures the attributes of the data in a better way compared to the ARIMA model.

For Turkey, the results are similar regarding the choice of models and optimal fit. For road transportation in Turkey, the ARIMA model is the better fit-model because of the lower values of AIC, AICc, and BIC. Hence fewer errors in the dataset with this model. For inland and rail for Turkey, ARIMA is once again suited better for forecasting future observations with the help of the dataset. For these two modes of transport, ETS shows a higher error-measurement value than that of ARIMA, hence making ARIMA better suited for these two modes. However, when forecasting air freight for Turkey, ETS is the better-fit model for the data. This could be explained by the variation in the data and how it is structured compared to the other modes of transport. When forecasting the total freight of Turkey, ETS is the optimal model, which is the same as for Mexico. ARIMA might therefore be better when forecasting individual modes. However, for larger datasets such as total freight, ETS may be better when looking at AIC, AICc, and BIC.

The forecasted result for Mexico is in line with what is described in the theory section. Road freight transportation is the most prominent mode of freight transportation in Mexico, and regarding both the forecast models, it is noticeable that this still will be the cause, with road freight transport growing in a linear way from a short- and mid-term perspective.

Furthermore, an interesting observation regarding rail freight transport for Mexico is that the forecasting models differ. The ETS model illustrates a steady growth of forecast for freight transported by rail until 2027, with an increase in volume each year. However, the ARIMA model illustrates growth up until 2024 and then a slight decrease in volume between 2024 and 2027. The increase in freight moved by rail can indicate two aspects. A) Infrastructure for rail freight will be developed, hence enabling a significant increase in freight volume; B) For rail freight transport to grow according to the forecasts, it is crucial that infrastructure investments occur. Enabling further usage of the rail network. The theory section discusses that the rail freight infrastructure is not as clearly developed compared to road transportation, hence illustrating the importance of points A and B.

When looking at the air freight forecasts for Mexico, the models differ again. The best fit ETS model depicts a steady increase in air freight for the entirety of the time series. Whereas the ARIMA model illustrates an increase in freight up until 2025, and from 2026-2027 the volume will decrease. However, the total freight volume is still an increase compared to today's value.

The two models' different structures can explain the reason for the difference in forecasts and how they apply the methods for the data. For instance, ETS takes, as previously explained, patterns and trends more into consideration, so if the data has a specific pattern with an upward trend, this can explain the steady increase in the forecast.

The inland freight forecast shows a similar increase in freight volume for both models. It indicates a strong likelihood that total inland freight will, in fact, increase, and the certainty around might be very high.

The result of the forecasts indicates a clear trend that total freight for Mexico will increase in the short and mid-term perspective. This is also illustrated in the forecast of total freight in Figures 36 and 37. The theory section of this report mentions that it is vital that infrastructure development occurs for freight to develop in an increasing trend. With the increase of e-commerce in the country, and global corporations such as Amazon and Walmart becoming more established in Mexico, the probability of this happening does, in fact, increase. Meaning that the forecasted freight volume has a high likelihood of occurring.

The outcome of the forecasts for Turkey also aligns with what was expected based on the conditions described in the theory section. The emerging market of Turkey is on the rise, regarding all the different modes of transport forecasted. With all new investments in the road segment within the country, described in section 2.2.2, the continuous increase does not come as a surprise. However, there are differences based on the two different forecasting methods. The ETS forecast suggests an increase, although with a broader prediction interval than the ARIMA forecast. With new airports being built and current airports being refreshed, the increase in air freight transportation also aligns with what could be expected. However, regarding rail freight transportation, the ETS forecast depicts neither an increase nor decrease as earlier described, even though the heavy infrastructure investments regarding the segments suppose a substantial increase. The outcome of the forecast can relate to a missing indicator of an increase, supposedly because of the relatively recent changes within the segment, not impacting the available data of the report. Though, the ARIMA forecast for rail freight transportation indicates an increase, suggesting that the aggregating trend will continue, contradicting the ETS forecast. Nonetheless, this outcome may be due to the contrasting forecasting techniques used, as ARIMA typically assigns greater importance to recent data points. An increasing trend is also depicted for inland freight transportation, resulting in an apparent increase in total freight transportation. With heavy investments in infrastructure regarding all segments and to become a more prominent logistic hub, the expected result correlates with the actual results of the forecasts.

6. CONCLUSIONS

This study aims to investigate what factors influence the logistics sector in emerging markets as well as forecast the future of logistics for short and mid-term horizons. The introduction gives a short description of what emerging markets are, and afterward, the theory section maps them out more in detail. What signifies them, what their development is what trends are prevalent, as well as future possibilities are all factors that are included in the section. Furthermore, the two emerging markets of choice, Mexico, and Turkey, are described more in-depth. We clarify why they are chosen, and then proceed to describe them both in a general manner, as well as a deep dive into the logistics sectors of both countries. The choice of quantitative methodology fits the research questions adequately, and results for both questions can be answered.

For first research question of *what factors can affect the logistics sector of the emerging economy country*, we apply various data preparation, reduction, and analysis methods. The data gathered for research question one was retrieved from the World Bank database of world development indicators.

The principal component analysis (PCA) along with time series regression analysis is used to answer the question of determining the influential factors on logistics in emerging markets. It results in some differences between the two countries. However, many factors are related, and a clear conclusion could be drawn. While financial factors such as GDP-related aspects were prominent, demographic factors, such as the age of the population, rural population growth, and population in the larger cities, were also proven to impact the increase in freight transportation significantly. Logistical factors like import and export-related aspects, manufacturing, services, and changes in inventories are also confirmed to have an impact on how logistics function and will further be developed in these two countries.

Regarding Mexico, both the PCA method and time series regression analysis reveal the extraction of two distinct principal components. The first principal component (PC1), which accounts for 0.5991 proportion of the explained variance, encompasses variables associated with the country's logistics sector. These variables include measures of *final consumption expenditure* and *current GDP* in various forms. Therefore, the principal component analysis further supports the theory section's description of an increase in consumption and changing consumption patterns. The second principal component (PC2), explaining 0.08737 proportion of the variance, consists of variables more closely linked to Mexico's industrial sector. Notable variables within PC2 include *Manufactured exports (% of merchandise exports)*, *Merchandise exports to high-income economies (% of total merchandise exports),* and *Food*, *beverages*, and *tobacco (% of value added in manufacturing).* These variables strengthen the assertion that the industrial sector plays a vital role in Mexico. Moreover, the variable *Merchandise exports to high-income economies (% of total merchandise exports)* highlights the significant trade relationship with the United States. Additionally, population-related variables, such as *Population ages 0-14*, total, and *Rural population growth (annual %),* indicate that a young population and increased urbanization are crucial for the development of Mexico's logistics sector.

For Turkey, applying the PCA method and time series regression analysis results in the identification of three principal components. The first principal component (PC1), explaining 0.5524 proportion of the variance, encompasses variables from all three categories: demographic, economic, and logistics-related factors. These variables include *population ages* and the *geographic distribution of the population* within the country (demographic factors), *primary income payments* and *external debt stocks* (economic factors), and *income, export volumes, and values* (logistics-related factors). The second principal component (PC2), accounting for 0.1174 proportion of the variance, also represents all three categories. However, it includes only the demographic variable of *population ages 0-14*. Economic factors, *such as deposit interest rate* and *final consumption expenditure*, are depicted, along with logistics-related factors like *merchandise imports and exports*. Lastly, the third principal component (PC3), explaining 0.08701 proportion of the variance, primarily focuses on travel services. It comprises two variables: *travel services (% of commercial service exports)* and *travel services (% of service export, BoP),* as mentioned previously.

In summary, the application of the PCA method and time series regression analysis enables the identification of distinct principal components for both Mexico and Turkey. These components capture the significant variables associated with various sectors, providing valuable insights into the dynamics of their economies and the key factors driving their logistics sectors' development.

This study also contains short and mid-term forecasts regarding logistics trends within the emerging markets of Mexico and Turkey as the goal of research question two. The forecasted variable used in the project is freight transportation measured in million ton-kilometers for four modes of transport: road, rail, inland, and air. These variables are collected from the Organization of Economic Cooperation and Development (OECD) database. Additionally, the total sum of all transportation modes is considered in the analysis to illustrate the total growth of freight for each country.

For the second research question of *how the future of logistics is forecasted in the emerging economy country for short and mid-term horizons*, we use Exponential Smoothing (ETS) and ARIMA methods. Although the concept of an emerging market implies an expanding market, the logistics sector appears to be keeping pace with this growth. The forecasts for Mexico and Turkey depict an increasing trend regarding all modes of transportation, except for the horizontal no-trend outcome of the rail forecast in Turkey. Hence, the total freight transportation forecasts portray a significant increasing trend for both countries analyzed.

As previously explained, the ETS and ARIMA models differ in how they handle the underlying structure of the data, resulting in a more conservative forecast through ARIMA, and a more fluctuating forecast through ETS. The difference in results between the two models is further discussed in chapter five. Where ARIMA generally gives a more conservative forecast due to making data stationary, ETS is more sensitive to trends, patterns, and seasonality in the dataset. However, both models show, as previously mentioned, an increase in freight volume for both Mexico and Turkey. Furthermore, there is a difference in model fit between the two models. This is determined by the AIC, AICc, and BIC metrics from the regression models. For forecasting individual modes of transport for Mexico the ARIMA model is the better fit. However, for forecasting total freight the ETS model is superior regarding the measurements. For Turkey, the ETS model is the best fit for forecasting total freight and air freight, whereas ARIMA is the better fit for the rest of the modes of transport.

The delimitations set for this report helped with accomplishing the results to the two research questions. When doing a quantitative data analysis such as this, it is important that data for the task exists. Moreover, delimitations regarding not incorporating maritime logistics are also included. The reasoning behind this is that sufficient data for both countries to be able to successfully forecast is not prevalent. Hence not including it in this report. In a nutshell:

1) Variables that influences the logistics trend in the emerging markets are showed in Tables 5 and 6.

2) The forecasted future shows a general increase of freight transport in the short-term and mid-term. With variation in increase of volume differing between the various modes of transportation.

3) Quantitative data analysis including principal component analysis and time series regression model is successful to find factors can influence the logistics sector in in Mexico and Turkey.

4) The forecasting models ETS and ARIMA is successful in forecasting for all modes of transport in Mexico and Turkey.

6.1. Recommendations for further research

Further research directions include but is not limited to:

1) Investigating the impact of technological advancements, such as automation and artificial intelligence, on logistics in emerging economies.
2) Examining the role of government policies and regulations in promoting or hindering logistics development in emerging economies.
3) Exploring the potential for public-private partnerships to improve logistics infrastructure and services in emerging economies.
4) Analyzing the impact of globalization and international trade on logistics in emerging economies.
5) Studying the challenges and opportunities presented by e-commerce and digitalization of logistics in emerging economies.
6) Investigating the impact of environmental sustainability and climate change on logistics in emerging economies, and the potential for sustainable logistics solutions.
7) Comparing and contrasting logistics practices and strategies in different emerging economies, and identify best practices and lessons learned.
8) Analyzing the impact of geopolitical factors, such as trade agreements and political instability, on logistics in emerging economies.
9) Studying the impact of demographic and social changes on logistics in emerging economies, such as urbanization and changing consumer behaviors.
10) Investigating the potential for innovative logistics solutions, such as shared logistics networks and circular supply chains, in emerging economies.
11) Neural network models for forecasting. Neural networks use machine learning techniques to predict future observations. A neural network is a type of artificial intelligence model that is designed to simulate the way the human brain works, with layers of interconnected nodes that process and interpret data.

American Industries. (2022). *Mexico Poised to Become One of the World's Most Important Logistics Areas*. American Industries .

Boske, L., & Harrison, R. (2013). *U.S.-Mexico Transportation and Logistics*.

Callen, T. (n.d.). *Gross Domestic Product: An Economy's All - Back to Basics: Economics Concepts Explained - FINANCE & DEVELOPMENT*.

Davison, A.-C. (2003). *Statistical Models*. Cambridge : Cambridge University Press, 2003.

Deloitte Insights. (2022a). *Mexico Economic Outlook*. Deloitte Insights. https://www2.deloitte.com/uk/en/insights/economy/americas/mexico-economic-outlook.html

Deloitte Insights. (2022b). *Mexico Economic Outlook*. Deloitte Insights.

European Central Bank. (2010). The Growing Importance of Emerging Markets. *Https://Www.Ecb.Europa.Eu/Pub/Pdf/Other/Mb201010_focus01.En.Pdf*.

European Commission. (2022). *Türkiye 2022 Report*.

Fidan, E. Ç. (2020). *Logistics Industry in Turkey*.

G20. (2022). *About G20*. G20.

Hirschinger, M., Spickermann, A., Hartmann, E., von der Gracht, H., & Darkow, I.-L. (2015a). The Future of Logistics in Emerging Markets-Fuzzy Clustering Scenarios Grounded in Institutional and Factor-Market Rivalry Theory. *Journal of Supply Chain Management*, *51*(4), 73–93. https://doi.org/10.1111/jscm.12074

Hirschinger, M., Spickermann, A., Hartmann, E., von der Gracht, H., & Darkow, I.-L. (2015b). The Future of Logistics in Emerging Markets-Fuzzy Clustering Scenarios Grounded in Institutional and Factor-Market Rivalry Theory. *Journal of Supply Chain Management*, *51*(4), 73–93. https://doi.org/10.1111/jscm.12074

Human Rights Watch. (2021). *Turkey - Events of 2021*. Human Rights Watch.

Hymans, Saul. H. (n.d.). *Forecasting and Econometric Models*. Https://Www.Econlib.Org/Library/Enc/ForecastingandEconometricModels.Html.

Hyndman, R. J., & Athanasopoulos, G. (2018). *Forecasting: Principles and Practice* (2nd ed.). Springer.

International Trade Administration. (2022). *Turkey - Country Commercial Guide*. International Trade Administration.

International Trade Administration. (2023). *Transportation Infrastructure Equipment and Services*. https://www.trade.gov/country-commercial-guides/mexico-transportation-infrastructure-equipment-and-services

James, G., Witten, D., Hastie, T., & Tibshirani, R. (2021). *An Introduction to Statistical Learning: with Applications in R (Springer Texts in Statistics)* (2nd ed.). Springer-Verlag .

John Manners-Bell, Thomas Cullen, & Cathy Roberson. (2014). *Logistics and Supply Chains in Emerging Markets*. Kogan Page © 2014.

Johnson, R., & Wichern, D. (2007). *Applied Statistical Multivariate Analysis*. © 2007 Pearson Education, Inc.

Kesgingöz, H., Prof, A., & Dilek, S. (2017). The Impact of G20 On Turkish Foreign Trade 1. In *Afro Eurasian Studies Journal* (Vol. 6).

Kim, Y., & Zangerling, B. (2016a). Setting the Scene. In Y. Kim & B. Zangerling (Eds.), *Mexico Urbanization Review: Managing Spatial Growth for Productive and Livable Cities in Mexico* (pp. 1–14). The World Bank. https://doi.org/10.1596/978-1-4648-0916-3_ch1

Kim, Y., & Zangerling, B. (2016b). Setting the Scene. In Y. Kim & B. Zangerling (Eds.), *Mexico Urbanization Review: Managing Spatial Growth for Productive and Livable Cities in Mexico* (pp. 1–14). The World Bank. https://doi.org/10.1596/978-1-4648-0916-3_ch1

Lopez, E. Y. V., & Lee, S. (2018). The Current State of E-logistics in Mexico. *International Academy of Global Business and Trade*, *14*(2), 31–43. https://doi.org/10.20294/jgbt.2018.14.2.31

Manners-Bell, J., Cullen, T., & Robertson, C. (2014a). *Logistics and Supply Chains in Emerging Markets*. Kogan Page © 2014.

Manners-Bell, J., Cullen, T., & Robertson, C. (2014b). *Logistics and Supply Chains in Emerging Markets*. Kogan Page © 2014.

Manners-Bell, J., Cullen, T., & Robertson, C. (2014c). *Logistics and Supply Chains in Emerging Markets*. Kogan Page © 2014.

Marquis, C., & Raynard, M. (2015). Institutional Strategies in Emerging Markets. *Academy of Management Annals*, *9*(1), 291–335. https://doi.org/10.1080/19416520.2015.1014661

Naz Inci, G., Gonden, T., Avci, M., & Deniz Pala, S. (2022). Turkey Logistics Report 2022. *Cushman & Wakefield*.

OECD. (2021). *Business Insights Emerging Markets*. https://www.oecd.org/dev/Business-Insights-Emerging-Markets-2021.pdf

OECD. (2022). *BUSINESS INSIGHTS ON EMERGING MARKETS 2022*. www.oecd.org/dev.

Petropoulos, F., Apiletti, D., Assimakopoulos, V., Babai, M. Z., Barrow, D. K., Ben Taieb, S., Bergmeir, C., Bessa, R. J., Bijak, J., Boylan, J. E., Browell, J., Carnevale, C., Castle, J. L., Cirillo, P., Clements, M. P., Cordeiro, C., Cyrino Oliveira, F. L., De Baets, S., Dokumentov, A., … Ziel, F. (2022a). Forecasting: theory and practice. In *International Journal of Forecasting* (Vol. 38, Issue 3, pp. 705–871). Elsevier B.V. https://doi.org/10.1016/j.ijforecast.2021.11.001

Petropoulos, F., Apiletti, D., Assimakopoulos, V., Babai, M. Z., Barrow, D. K., ben Taieb, S., Bergmeir, C., Bessa, R. J., Bijak, J., Boylan, J. E., Browell, J., Carnevale, C., Castle, J. L., Cirillo, P., Clements, M. P., Cordeiro, C., Cyrino Oliveira, F. L., de Baets, S., Dokumentov, A., … Ziel, F. (2022b). Forecasting: theory and practice. In *International Journal of Forecasting* (Vol. 38, Issue 3, pp. 705–871). Elsevier B.V. https://doi.org/10.1016/j.ijforecast.2021.11.001

Piotrowicz, W., & Wojciech, C. (2015). *Supply Chain Design and Management for Emerging Markets: Learning from Countries and Regions*.

Rodrigue, J.-P., & Notteboom, T. (2010). Comparative North American and European gateway logistics: the regionalism of freight distribution. *Journal of Transport Geography*, *18*(4), 497–507. https://doi.org/10.1016/j.jtrangeo.2010.03.006

Santander. (2023). *Mexico Foreign Trade in Figures*. Santander. https://santandertrade.com/en/portal/analyse-markets/mexico/foreign-trade-in-figures

Statista. (2022). *Amount of money invested in sea port infrastructure in Turkey from 2006 to 2021*. Statista. https://www.statista.com/statistics/439364/investment-in-sea-port-infrastructure-in-turkey/

The Global Economy. (2022). *Turkey: Political Stability*. The Global Economy. https://www.theglobaleconomy.com/turkey/wb_political_stability/

The World Bank. (2020a, October 10). *Türkiye Overview*. The World Bank.

The World Bank. (2020b, October 10). *Türkiye Overview*. The World Bank. https://www.worldbank.org/en/country/turkey/overview

The World Bank. (2022). *Urban population (% of total population) - Mexico*. The World Bank. https://data.worldbank.org/indicator/SP.URB.TOTL.IN.ZS?locations=MX

The World Bank. (2023a). *Foreign direct investment, net inflows (BoP, current US$) - Turkiye*. The World Bank. https://data.worldbank.org/indicator/BX.KLT.DINV.CD.WD?locations=TR

The World Bank. (2023b). *GDP (current US$) - Turkiye, Mexico*. The World Bank. https://data.worldbank.org/indicator/NY.GDP.MKTP.CD?end=2021&locations=MX-TR&name_desc=false&start=1970

The World Bank. (2023c). *Mexico Overview*. The World Bank. https://www.worldbank.org/en/country/mexico/overview

The World Bank. (2023d). *Urban Population - Turkiye*. The World Bank. https://data.worldbank.org/indicator/SP.URB.TOTL?locations=TR

Trading Economics. (2023a). *GDP from Transport - Turkey*. Trading Economics. https://tradingeconomics.com/turkey/gdp-from-transport

Trading Economics. (2023b). *Mexico GDP From Transport*. Trading Economics.

UNIDO. (n.d.-a). *Country Profile - Mexico*. UNIDO. Retrieved March 20, 2023, from https://stat.unido.org/country-profile/economics/MEX

UNIDO. (n.d.-b). *Country Profile - Mexico*. UNIDO. Retrieved March 20, 2023, from https://stat.unido.org/country-profile/economics/MEX

UNIDO. (n.d.-c). *Country Profile - Turkey*. UNIDO.

World Bank, T., & International, M. F. (2011). *Economic Growth in Developing Countries*. http://ebookcentral.proquest.com/lib/chalmers/detail.action?docID=690506.

World Economics. (n.d.). *Emerging Markets*.

Zotteri, G., Kalchschmidt, M., & Caniato, F. (2005). The impact of aggregation level on forecasting performance. *International Journal of Production Economics*, *93–94*, 479–491. https://doi.org/10.1016/j.ijpe.2004.06.044

APPENDIX A: R CODE

The following R codes were used in the project. The example codes relate to the forecasts for Mexico. For Turkey, all MEX was changed to TUR.

Appendix A.1: R code for data preparation and data reduction by PCA

```
# Read the CSV file into a data frame

MEX_data <- read.csv("MEX-Reg.csv", check.names = FALSE, header = TRUE)

MEX_data

# Transpose the data frame

MEX_data_transposed <- t(MEX_data)

# Write the transposed data frame to a new CSV file

write.csv(MEX_data_transposed, "MEX-Reg-transposed.csv", row.names = TRUE)

# install and call library imputeTS to estimate missing data in time series

install.packages("imputeTS")

library(imputeTS)

# load the data

data_MEX <- read.csv("MEX-Reg-transposedV2.csv")

# loop through each column and remove if it has more than 5 missing values

for (col in colnames(data_MEX)) {

  if (sum(!complete.cases(data_MEX[,col])) > 5) {

    data_MEX[,col] <- NULL

# estimate missing data (NA: Not available) by interpolation

MEX_data_imputed <- na.interpolation(data_MEX)

#remove column which has a the same value for all cells

constant_cols <- which(sapply(MEX_data_imputed, function(x) length(unique(x))) == 1)
#Ask Chat GPT: How can R find a column in a dataset that its all cells have the same value?

MEX_data_imputed_no_constant <- MEX_data_imputed[, -constant_cols]

write.csv(MEX_data_imputed_no_constant, "MEX-Reg-transposedV2-imputed.csv",
row.names = FALSE)

# exclude the first column

MEX_data_imputed_no_constant_WithoutFirstCol <- MEX_data_imputed_no_constant[,-1]

# Standardize the variables

MEX_data_imputed_no_constant_WithoutFirstCol_scaled <-
scale(MEX_data_imputed_no_constant_WithoutFirstCol)

# Perform principal component analysis

pca <- prcomp(MEX_data_imputed_no_constant_WithoutFirstCol_scaled)

# Create scree plot
```

```
plot(pca, type="l", main="Scree Plot")
# View summary of the results
summary(pca)
# Get loadings for each principal component
loadings <- pca$rotation
# View the loadings
print(loadings)
# Convert the row names to a new column and set it as the index
loadings_new <- cbind(Index = row.names(loadings), loadings)
# Remove the row names from the data frame
rownames(loadings_new) <- NULL
# View the new data frame
print(loadings_new)
#non_zero_indexPC1 <- loadings_new$Index[loadings_new$PC1 > 0.08]
#Error in loadings_new$Index : $ operator is invalid for atomic vectors
#convert atomic vectors to data frame
loadings_newDF <- as.data.frame(loadings_new)
# Extract the index names and values for which PC1 cells are more than 0.08
Significant_PC1 <- loadings_newDF[loadings_newDF$PC1 > 0.08, c("Index", "PC1")]
# Print the significant index names and values for PC1
print(Significant_PC1)
# Extract the index names and values for which PC2 cells are more than 0.10
Significant_PC2 <- loadings_newDF[loadings_newDF$PC2 > 0.10, c("Index", "PC2")]
# Print the significant index names and values for PC2
print(Significant_PC2)
# Project original data set onto principal components
data_MEX_pca <- predict(pca, MEX_data_imputed_no_constant_WithoutFirstCol_scaled
data_MEX_pca_1to2 <- data_MEX_pca[,1:2]
# Write the transposed data frame to a new CSV file
write.csv(data_MEX_pca_1to2, "data_MEX_pca_1to2.csv", row.names = TRUE)
```

Appendix A.2: R code for time series regression analysis

```
# Read the CSV file into a data frame
MEX_data <- read.csv("data_MEX_pca_1to2_V2.csv", check.names = FALSE, header =
TRUE)
```

```
# convert the data to a time series object
ts_data <- ts(MEX_data[ ,-1], start = c(1980), frequency = 1)
ts_data
library(forecast)
library(fpp3)
GGally::ggpairs(as.data.frame(ts_data))
fit.consMR_MEX1 <- tslm(Freight.Transportation ~ PC1 + PC2, data=ts_data)
summary(fit.consMR_MEX1)
checkresiduals(fit.consMR_MEX1)
CV(fit.consMR_MEX1)
```

Appendix A.3: R code for ETS

```
data_TURAir <- read.csv('TUR_Air2.csv')
timeseries2 <- ts(data_TURAir, start=c(1970), end=(2021))
# plot the time series
# After running grid(col="gray"), Save the plot
plot(timeseries2, main="Turkey's Air Freight Transportation", xlab="Year", ylab="Million
ton-km", col="blue", lwd=2)
grid(col="gray")
library(forecast)
# ETS() function will find the best exponential smoothing model for the time series data
fit <- ets(timeseries2)
fit
# Check the residuals
checkresiduals(fit)
# forecast for the next 6 years
forecast_result <- forecast(fit, h=6)
print(forecast_result)
# plot the forecasted results
plot(forecast_result, xlab="Year", ylab="Air Freight Transportation (million ton-km)",
col="blue", lwd=2)
grid(col="gray")
```

Appendix A.4: R code for ARIMA

```
data_TURRail <- read.csv('TUR_Rail2.csv')
timeseries1 <- ts(data_TURRail, start=c(1970), end=(2021))
```

```
# plot the time series
# After running grid(col="gray"), Save the plot
plot(timeseries1, main="Turkey's Rail Freight Transportation", xlab="Year", ylab="Million ton-km", col="blue", lwd=2)
grid(col="gray")
library(forecast)
# auto.arima() function will find the best ARIMA model for the time series data
fit <- auto.arima(timeseries1, seasonal=FALSE, stepwise=FALSE, approximation=FALSE)
fit
# Check the residuals
checkresiduals(fit)
# forecast for the next 6 years
forecast_result <- forecast(fit, h=6)
print(forecast_result)
# plot the forecasted results
# After running grid(col="gray"), Save the plot
plot(forecast_result, xlab="Year", ylab="Turkey's Rail Freight Transportation (million ton-km)", col="blue", lwd=2)
grid(col="gray")
```

DEPARTMENT OF TECHNOLOGY MANAGEMENT AND ECONOMICS
DIVISION OF SERVICE MANAGEMENT AND LOGISTICS
CHALMERS UNIVERSITY OF TECHNOLOGY
Gothenburg, Sweden 2023
www.chalmers.se

CHALMERS