**AN OPTIMAL FORECASTING METHOD OF PASSENGER TRAFFIC IN GREEK**

**COASTAL SHIPPING**

**Ioannis Sitzimis**

School of Business Administration and Tourism,

Hellenic Mediterranean University, Heraklion, Greece

**Email:** isitzimis@hmu.gr

**Abstract:** Greek coastal shipping is one of the biggest in Europe serving a large number of passengers and having

a large part of the total shipping fleet. It plays an important role for Greek economy and society, as it connects the

majority of inhabited islands to mainland. There are not a lot of scientific efforts forecasting coastal passenger

traffic in the past. The main goal of this study is to exact an optimal forecasting method by answering the research

question: which is the best model for capturing short-term seasonal components of passenger traffic in Greek

coastal shipping? The analysis results show that in fourteen popular coastal routes Winters’ multiplicative method,

simple seasonal model and decomposition multiplicative trend and seasonal model have the best integration to

the time series data. No coastal line led to better results by seasonal Box-Jenkins ARIMA models.

**Keywords:** Greek coastal shipping, passenger traffic, smoothing forecasting methods, decomposition forecasting

methods, seasonal ARIMA models, measures of forecasting accuracy

**Jel:** R41, M21

**1. Introduction**

GCS is one of the biggest in Europe and performs an important role connecting mainland to

Greek islands1. Its contribution to GDP is € 13.6 billion or 7.4% of total GDP (2019). It employs

approximately 332,000 people and contributes to public revenues with approximately € 3 billion

(IOBE, 2020). It carries over 35 million Greeks and foreigners annually (including ferry lines),

with its fleet accounting for about 7% of the global passenger shipping fleet. It covers about 17%

of total passenger shipping in Europe, with more coastal lines than other countries (due to the

plethora of islands). The sector is characterized by high seasonality with almost half of the

transport traffic taking place in the period June-August (IOBE, 2014).

The listed shipping companies, for the year 2019, employed 2,449 employees, launched 43 ships

and served passenger traffic of 7,543,460 people. Respectively, they transported 1,041,574 cars

and 555,241 trucks. They served about 36% of the total passenger traffic2 and 46% of the total

vehicle traffic. The average age of their fleet is high, with 60% being over 22 years old and 25%

of the total being over 30 years old (XRTC, 2004-2020).

All the above show that GCS is of great importance for Greek economy and society. The volume

of passenger traffic it serves is very high and its forecast is very significant for both business and

government policy. Decisions on the number of routes served by shipping companies, on ships

1

2

It connects about 115 inhabited islands to mainland in Greece.

The remaining 64% for passengers and 54% for vehicles are serviced by smaller companies. For the year 2019, these companies

launched 15 conventional and 21 high-speed vessels on all GCS routes.

1

by coastal line (number and size), on companies' pricing policy, on public service obligations3,

on state port infrastructure policy and for barren lines4 are typical examples.

The main goal of this study is to find an optimal forecasting method of passenger traffic in GCS

by comparing Box-Jenkins ARIMA, smoothing and decomposition methods (Wardono, et al.,

2016; Ahmad & Ahmad, 2013; Trull, et al., 2020). The first methods seem to be effective in

predicting passenger traffic in transport and the other two have not been preferred for research

(Aivazidou, 2015). The basic assumption in all three of the above methods is that the available

observations will continue to behave in the future as in the past (Shim & Siegel, 2001).

In the period 2020-2021, due to the covid-19 pandemic, this did not happen and there was a

sharp drop in passenger traffic. Therefore our forecast will start from the year 2020, ignoring this

decrease so as not to create a problem in the time series forecasts for the coming years (and

especially for the year 2022 when the situation is expected to return to regularity).

**2. Related work**

In relation to the forecasting efforts in the transport sector, about 60% of the publications

concern forecasts for passenger transport (Aivazidou, 2015). These mainly concern air, road and

urban transport. Similar scientific research on maritime passenger transport is absent from the

international literature. The main reason is that in a few countries coastal shipping is a means of

transport. Characteristic is only the research of Ortuzar and Gonzalez (2002), for the coastal line

between the Canary Islands and Tenerife.

In GCS, because it largely connects mainland to Greek islands, the research is more extensive.

Attempts in this direction have been made by various authors, such as Psaraftis (1994) who tried

to systematically analyze possible scenarios for passenger demand after deregulation of the

market. Spathi (2005), attempted to find the function of passenger demand using the dynamic

model with the error correction model mechanism (Ramanathan, 2001). A similar study was

carried out by Tsekeri (2008) who presented an aggregate analysis of substitution and

complementary relationship among all available transport modes for domestic travel in Greece.

He proposed a model based on consumer demand theory.

Simplistic efforts were made to forecast the financial statements and passenger traffic of coastal

companies, with polynomial and hyperbolic functions having the best application (higher R

squared) (Sitzimis, 2012). An important research took place in 2014 (IOBE), which used the

regression method for the estimation of demand elasticity for coastal shipping services with

respect to the price of tickets and household disposable income. Various other forecasting

approaches are performed by XRTC on annual basis (2004-2020).

However, according to Aivazidou (2015) the basic forecasting methods used for other passenger

transport are mainly time series analysis models and less combined time series and regression

analysis models or pure regression analysis models. In fact, the most widely used models are those

based on the Box-Jenkins ARIMA methodology, while very few are based on methods of

3

Although the market is liberalized and a simple declaration to the Ministry of Maritime Affairs and Insular Policy is required to enter

and exit, there are also obligations for shipping companies such as obligatory period of ten-month shipping, prohibition of interruption

and change of routes without approval, mandatory crew compositions of ships.

4 According to the Law 2923/2001, the State characterizes as "barren" those lines for which there is no expression of interest for their

operation from coastal companies.

2

smoothing and time series decomposition. In other words, there is a gap in the relevant literature

that we are going to fill with this research.

**3. Passenger traffic analysis in GCS**

Air, road and urban transport offer useful conclusions about the passenger demand forecasting

to a transportation industry (Sitzimis, 2012; Sabry, et al., 2007; Tamber, et al., 2021; Dingari, et

al., 2019). We could be based on them and reach to the congruent conclusions about GCS.

However, the market conditions differ between those industries. In GCS these assumptions cannot

be unified and undivided (Goulielmos & Sambrakos, 2002). This market consists of several

concentrated submarkets-coastal routes, which should be analyzed individually (Sitzimis, 2012;

Goulielmos & Sitzimis, 2014; Goulielmos & Sitzimis, 2012). There are many studies that make

the mistake of dealing the market GCS as a total (Tsekeris, 2008). We prefered the assiduous

review of the real conditions of GCS, by analyzing it per coastal route.

Coastal lines in Greece are characterized by intense seasonality with the largest percentage of

passenger traffic (about 45%) taking place in the third quarter of each year, with the months of

April to September accounting for about 70% of the total annual number of passengers (Sitzimis,

2012; XRTC, 2004-2020). This fact reflects the strong tourist demand for island destinations

(IOBE, 2014). August, is the month with the greatest traffic, leaving behind July, September and

June. The lowest traffic of passengers traveling within their national borders, mostly appeares

during February, January and November (XRTC, 2004-2020).

In order to analyse the passenger traffic in the shipping itineraries of Greece, we remained in 13

itineraries, which represented the biggest average percentage of total passenger traffic (diagram

2). Those of Argosaronikos (A), Piraeus-Peloponnise (PP), Piraeus-Creta (PC), Piraeus-Creta-

Dodecanese (PCD), Piraeus-Dodecanese (PD), Piraeus-West Cyclades (PWC), Piraeus-East

Cyclades (PEC), Piraeus-Mykonos-Tinos-Samos (PMTS), Piraeus-Chios-Mytilene (PCM),

Patra-Akarnania-Ionian islands (PAII), Rafina-Euboea-Andros-Tinos (REAT), Volos-North

Sporades-Kymi (VNSK) and the rest (L).

The average annual rates of traffic alteration on these lines, between 1970-2000, increased at an

impressive rate. Overall, an average increase of 4.2% was observed (Sitzimis, 2012). This was

due to: (a) the growth of tourism in insular Greece, (b) the increase in GDP per capita of island

inhabitants, (c) the general increase of the permanent population in Greece and (d) the greater

dependence of islands from the mainland (due to the modern tendency for astyphilia) (Spathi,

2005). The lines with the highest traffic were "A", "PEC" and "PC", while the highest growth rate

appeared in the line "PEC" (7.2%), followed by the lines "PC" (7.1%) and "PWC" (6.5%)

(Sitzimis, 2012). It is obvious that the "PWC" and "PEC" lines gathered the largest shipping traffic

in GCS between 1970-2000. This was mainly due to the great growth of tourist arrivals that

occurred in these islands after 1970.

Comparing the years 2001-2019 (table 1 and Figure 1) it is obvious a very large increase of

passenger traffic between the years 2001-2007 (35%), mainly due to the liberalization of the

market (lifting of cabotage privilege), partly in 2002 and fully in 2006 (Law 2932/01, EU

regulation 789/04, Presidential Decree 124/06) (Sitzimis, 2012; Goulielmos & Sitzimis, 2014).

Also, in this increase contributed both the Olympic Games in Athens in 2004 and the increase of

tourist flows to the country. Between 2006-2007 there is stabilization and a small percentage

3

decrease. In the period 2008-2013, passenger transport is significantly reduced due to the global

financial crisis, with the percentage reduction reaching 25%. The decrease was caused by the

descending course of income per capita of Greeks and by the overdraft of the Greek households.

After 2014 and until 2019 the market is recovering but continues to be at lower levels than it was

before the crisis. Despite the sharp increase in tourist traffic5, Greek coastal shipping does not

benefit enough as most foreign tourist arrivals took place by air. Overall, for the years 2001-2019

the average increase was about 10% (table 1).

For 2020, the impact of the Covid-19 pandemic was clear. Greek coastal shipping was subject

to a restriction on passengers’ transportation from March to May 2020 followed by state ceilings

for transported passengers thereafter. If we take into account the big drop in tourism, the passenger

reduction was significant (about 55%) (IOBE, 2020).

The most popular destination and the greatest average traffic (2001-2019) (Figure 2) is

presented in the shipping route "A" (having a part of 16.3% of passengers), followed by Piraeus-

Cyclades ("PEC" and "PWC", having a part of 15,5% of passengers), "PC" (a part of 13,7%) and

"REAT" with 12.9%. This is normal because they constitute very popular touristic destinations.

Especially after the deregulation of the market took place «cream skimming». This means that

most shipping companies preferred the most lucrative coastal markets, mainly in the summer

months. At the same time left non profitable markets in winter (Sitzimis, 2012; Goulielmos &

Sambrakos, 2002). Also, the high levels of traffic are due to the fact that these itineraries are based

on the part of demand with the least seasonality. So the higher levels of fullness and exploitation

of ships are achieved.

**4. Methodology**

The main issue in this research was to predict passenger traffic on the main lines of Greek

coastal shipping. To do this we had to choose between certain quantitative forecasting methods.

Qualitative forecasting methods are based more on human judgment than on the analysis of

existing data (Shim & Siegel, 2001). We had quarterly data on passenger traffic between the years

2004-2019, so we chose the quantitative methodology. In each case we processed our data with

the statistical package SPSS 22. Exception took place for the calculation of time series

decomposition where the minitab 19 software was used. Also the calculation of the augmented

Dickey-Fuller test was done through eviews 11. The reason was that SPSS did not have these

features clearly.

In order to make a prediction for our dependent variable we could use regression analysis

(Petropoulos & Asimakopoulos, 2013). In this way we would recognize the quantitative and

causal relationship between the variables involved in the interpretation of our problem.

Unfortunately, this method is difficult to apply here as the independent variables that affect

passenger traffic are not completely clear, it is difficult to find relevant statistic data and time

series analysis models seem to be better applied in these cases (Sabry, et al., 2007; Wu, et al.,

2013; Tsui, et al., 2014; Rashidi & Ranjitkar, 2015).

5

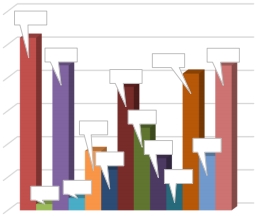
In 2004 there were 13 million tourist arrivals, whereas in 2019 the number was 34 million.

4

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | 20.000.000,00  19.000.000,00  18.000.000,00  17.000.000,00  16.000.000,00  15.000.000,00  14.000.000,00  13.000.000,00  12.000.000,00  2001 2003 2005 2007 2009 2011 2013 2015 2017 2019  Total |  |  | 3.000.000,00             16.31%  2.500.000,00                             13.71%                                                                         13.69%  12.93%  2.000.000,00                                                                11.67%  1.500.000,00                                                                         7.82%  5.62%  1.000.000,00                                                                                  4.93%               5.14%  3.85%  500.000,00                                                                                             2.56%  0.59%      1.16%  0,00  A            PP          PC          PCD       PD         PWC      PEC  PMTS    PCM      PAII       REAT     VNSK     L |  |
| **Figure 1:** Fluctuations of total passenger traffic in GCS  (2001-2019) | | | **Figure 2:** Average passenger traffic per route in GCS  (2001-2019) | | |

|  |  |  |  |
| --- | --- | --- | --- |
|  | |  | |
|  | |

For these reasons we could rely on known smoothing methods or Box-Jenkins ARIMA models,



considering only the existing observations and not the possible relationship with other variables

(Ahmad & Ahmad, 2013; Munarsih & Saluza, 2019; Yonar, et al., 2020).

Starting with smoothing methods, they are easy to apply and have a low degree of computational

difficulty. The basic logic is that we use timeless data, that is, past observations of equal

successive time periods. These time series are not affected by the small amount of available data

and provide satisfactory forecasts in the short term (Petropoulos & Asimakopoulos, 2013). As is

well known, when we do not have a trend and seasonality (stationary time series) for a short

forecast range, the simple moving average and simple exponential smoothing models are best

applied. Respectively, if there is a trend but not seasonality, for a long range of forecasts, the trend

analysis or the exponential smoothing with adaptation to the trend (Holt's method) are suitable.

For a smaller range the double exponential smoothing (Brown's method) or the double moving

average method (double moving average or linear moving average) are more preferable (Chalkos,

2020).

The passenger traffic data available at GCS were quarterly and therefore there were indications

of seasonality and no stationarity. As shown in Figure 3 in all examined lines there is a strong

increase in traffic in the 3rd quarter of the year, with a slightly decreasing or increasing trend over

the years 2004-2019. This means that we could not use prediction techniques such as the above.

Repeated seasonal fluctuations and quarterly available observations made the Winters model

(Winter's triple exponential smoothing) (indicated when we have seasonality rather than a short-

term forecast), the time series decomposition (suitable when we have a trend and seasonality for

a long range of forecasts), the simple seasonal model (suitable when we do not have a trend, but

only a stable seasonal result) and the seasonal ARIMA models suitable for our case (Chalkos,

5

2020; Petropoulos & Asimakopoulos, 2013).

Source: Hellenic Statistical Authority, 2000-2020. Our elaboration.

**Table 1:** Total passenger traffic in GCS (for the 13 main itineraries) (2001-2019).

**%**

**Year** **Passengers** **Average** **alteration**

2001 13,852,000

2002 13,124,000

2003 14,905,000

2004 17,306,000 16,422,862 34.79%

2005 18,257,159

2006 18,844,396

2007 18,671,482

2008 18,068,255

2009 17,442,121

2010 16,587,040 15,729,890 -24.72%

2011 15,071,705

2012 13,608,289

2013 13,601,930

2014 14,463,293

2015 14,323,032

2016 14,542,183 15,598,306 20.40%

2017 15,938,427

2018 16,909,512

2019 17,413,388

**Average (2001-**

**2019)** **15,943,643** **10.16%**

Source: Hellenic Statistical Authority, 2000-2020. Our elaboration.

*4.1Measures of forecasting accuracy*

The basic selection criterion we followed is which method best suited our data, that is, it led to

the smallest values of discrepancies between predicted ( *t* ) and actual values ( *t* ) of the time

series (forecast error). By studying the time behavior of forecast error values, we were able to

arrive at both the evaluation of our forecasting methods and the choice between alternatives

(Agiakloglou & Oikonomou, 2019; Karmaker, et al., 2017).

We used the following precision measures:

a) The mean absolute percentage error (MAPE), which expresses the percentage accuracy.

Defined as:

*MAPE* =



*Yt* −*Y t*

*Yt*

*n*

*x*100

(1)

where n is the number of measurements.

6

b) The mean squared deviation (MSD or MSE) calculated as:

*MSD* =

(*Yt* −*Yt*)2

*n*

(2)

Mean squared error expresses the mean value of the squared deviations and is considered

statistically more reliable, so it is used more often. Because its interpretation is difficult to

understand we mainly used the root mean squared error (RMSE) (Agiakloglou & Oikonomou,

2019).

c) The mean absolute error (MAE), which expresses a measure of the accuracy of the forecast

against the actual values, maintaining the units of measurement of the original time series. It is

set as:

*MAE* =

*Yi* −*Fi* |

*n*

(3)

and its high values show bias of the method.

d) The Bayesian information criterion, developed by Schwarz (1978) (or BIC) selects the model

that minimizes:

2

*BIC* = ln +

ln*n*

*n*

*r*

(4)

where σ2 is the sum of the squares of the residuals, n is the number of observations and r is the

total number of parameters with the constant term.

For all forecasting accuracy measures we considered that the lower the price the better the model

in terms of estimation (Chalkos, 2020). We also accepted that MAPE values below 10% describe

an extremely accurate forecast, below 20% a relatively good forecast and below 30% a marginally

accurate forecast (Dingari, et al., 2019).

*4.2Winter’s triple exponential smoothing*

Winter’s triple exponential smoothing has three smoothing parameters. The parameter α for the

smoothing of time series values (level), the parameter β for the smoothing of trend (slope) and

the parameter γ for seasonality smoothing. These components are either additive or multiplicative.

The multiplication model is selected when the seasonal pattern in the data depends on the size of

the data. In other words, the size of the seasonal pattern increases as the series goes up and

decreases as the series goes down. The additive one is selected when the seasonal pattern in the

data does not depend on the size of the data. In other words, the size of the seasonal pattern does

not change as the time series goes up or down (Hansun, et al., 2019; Dingari, et al., 2019).

The smoothing of the time series values in the additive model is done through the function:

𝛢𝑡 = 𝑎(𝛶𝑡−𝑆𝑡−𝐿) + (1 − 𝑎)(𝛢𝑡−1 + 𝑇𝑡−1)

(5)

whereas in the multiplicative through the function:

(6)

𝑆𝑡−𝐿 𝑡−1 𝑡−1

𝑌𝑡

𝛢𝑡 = 𝑎 + (1 − 𝑎)(𝛢 + 𝑇 )

where α is the smoothing constant (0≤α≤1), Αt the smoothed values of time series, St is the

seasonal factor of the period t and L is the periodicity of the seasonality.

7

In the additive model the smoothing of the trend follows the equation:

𝛵𝑡 = 𝛽(𝛢𝑡 + 𝐴𝑡−1) + (1 − 𝛽)𝛵𝑡−1

(7)

where β is the trend smoothing constant (0≤β≤1) and Τt the smoothed values of trend. The

equation in multiplicative model is transformed as:

𝛵𝑡 = 𝛽(𝛢𝑡/𝐴𝑡−1) + (1 − 𝛽)𝛵𝑡−1

𝛢𝑡 𝑡−𝐿

The seasonality smoothing in additive model follows the equation:

𝑆𝑡 = 𝛾[𝑌𝑡 − 𝐴𝑡] + (1 − 𝛾)𝑆𝑡−𝐿

and in multiplicative the equation:

𝑌𝑡

𝑆𝑡 = 𝛾 + (1 − 𝛾)𝑆

where γ is the smoothing constant of seasonality (0≤γ≤1).

The forecast is:

𝑌̅𝑡+ℎ = (𝐴𝑡 + ℎ𝑇𝑡)𝑆𝑡+ℎ−𝐿

where h=1,2,3…L is the future periods of first year and

𝑌̅𝑡+ℎ = (𝐴𝑡 + ℎ𝑇𝑡)𝑆𝑡+ℎ−2𝐿

(8)

(9)

(10)

(11)

(12)

where h=L+1,L+2,L+3…2L is the future periods of second year etc.

𝑌𝑡

Initialization of the method according to Chatfield (2003) is required. For t=1,2,..,L-1 the values

At are not determined, while for t=L the AL is defined as:

𝑌1 + 𝑌2 + ⋯ + 𝑌𝐿 (13)

𝛢𝐿 =

𝐿

For t=1,2,…,L-1 the values Τt are not determined, for t=L we set ΤL=0 and for t=1,2,…,L the

values of the seasonal coefficients St are calculated as:

(14)

𝑆𝑡 =

𝐴𝐿

The optimal values of α, β and γ were calculated automatically by minimizing the RMSE

criterion in SPSS and for all possible combinations of parameter values in the time series data

(Chalkos, 2020; Dhali, et al., 2019; Ravinder, 2013; Tamber, et al., 2021).

*4.3Τime series decomposition and Simple seasonal exponential smoothing*

The objective of time series decomposition is to identify the mechanism by which time series

values are formed. The decomposition method is used to separate a time series into the trend

component, the seasonality component, the cyclical component and the irregular component to

make predictions (Chalkos, 2020).

It is necessary to choose whether the seasonality works addictively or multiplicatively to the

trend. If it works addictively, the model has the form:

Υt = Tt+St+Ct+It

(15)

where Υt is the real observation in time t, Τt is the trend, St is the seasonality, Ct is the circularity

and It is the randomness. This model is more difficult for further computational analysis and

assumes independence between the factors. This assumption applies to natural phenomena, but

not to business or economic applications, where the trend also affects seasonal fluctuations

8

(Agiakloglou & Oikonomou, 2019). In the case of the multiplicative model the above relation is

transformed as:

Υt = Tt·St⋅Ct⋅It

(16)

It works well when the fluctuations depend on the level of values, which is usually the case

(Kyriakidis, 2018).

The simple seasonal exponential smoothing is suggested for series with no trend and a seasonal

effect that is constant over time. It has two smoothing parameters, level (α) and season (δ). It is

very similar to an ARIMA model with zero orders of autoregression, one order of differencing,

one order of seasonal differencing, and orders 1, p, and p + 1 of moving average, where p is the

number of periods in a seasonal interval (for quartely data, p = 4) (IBM, 2021). The model

equation is:

𝑌𝑡 = 𝜇𝑡 + 𝑆𝑡,𝑝 + 𝑎𝑡

and the smoothing equations are:

𝐿𝑡 = 𝑎(𝑦𝑡 − 𝑆𝑡−𝑃 ) + (1 − 𝑎)𝐿𝑡−1

𝑆𝑡 = 𝛿(𝑦𝑡 − 𝐿𝑡) + (1 − 𝛿)𝑆𝑡−𝑃

The h-stem-ahead equation is:

𝑌̅𝑡+ℎ = 𝐿𝑡 + 𝑆𝑡−𝑃+ℎ

(17)

(18)

(19)

(20)

h=1.2…, where μt is the mean of the observed time series at period t, St-P+h is the seasonal

component, p is the seasonality periodicity, h is the number of periods in forecasting and at is the

forecast error at period t.

*4.4ARIMA: Auto-Regressive Integrated Moving Average*

In relation to the Box-Jenkins ARIMA models, we would say again that in all coastal lines

appear seasonal data that have a distinct pattern which is repeated every year (Figure 3). Our data

are quarterly, so the length of the seasonal period is S = 4. This means that there are observations

that are correlated both within the year and between different years. In these cases, seasonal

ARIMA (SARIMA) models that contain non-seasonal and seasonal autoregressive and moving

average terms are applied (Wardono, et al., 2016; Ma, et al., 2018; Sim, et al., 2019). In fact, in

non-stationary series, a seasonal difference is usually used to completely determine the model.

(Wardono, et al., 2016; Ma, et al., 2018; Sim, et al., 2019).

These models are denoted as ARIMA (p,d,q)(P,D,Q)s where: p are non-seasonal autoregressive

terms, d are regular differences, q are non-seasonal moving average terms, P are seasonal

autoregressive terms, D are seasonal differences, Q are the seasonal terms of moving average and

s is seasonality. Indicatively a SARIMA model is written:

𝛷𝜌(Β)𝛷𝜌(𝐵𝑆)(1 − 𝐵)𝑑(1 − 𝐵𝑠)𝐷𝑦𝑡 = 𝜃𝑞(Β)𝛩𝑄(𝐵𝑆)𝜀𝑡

(21)

where φ and θ, are parameters of autoregressive (AR) and moving average (MA), while Φ and Θ,

are parameters of seasonal autoregressive (SAR) and seasonal moving average (SMA)

respectively. Β is lag operator which defined as BKYt=YT-K (18) (Wang, et al., 2013; Suhartono,

2011).

9

In particular, to implement SARIMA modeling and forecasting in GCS we followed 4 basic

steps (table 2). The first stage was the "recognition" of the model whenever we initially

ascertained the existence or not of stationarity in the time series data. When the series were not

stationary (in the sequence chart the values were not around zero) we applied the method of

differences. In some cases, both first regular differences (ΔΥt = Yt-Yt-1) and seasonal quarterly

differences (ΔΥt = Yt-Yt-4), that is of order S = 4, were needed. After the differences if the

autocorrelation function of the time series were declining rapidly and were zero, we considered

this to be a sign of stationarity. In order to determine whether the time series actually became

stationary we applied the augmented Dickey-Fuller test, which had as null hypothesis that the

data are not stationary (p-value <5% in order to reject the null hypothesis) (Makatjane & Moroke,

2016). We used eviews software for this unit root test.

In the resulting stationary time series, through minitab 19 we checked the importance of the

time series autocorrelation coefficients per lag. We used the t-student distribution, with n-1

degrees of freedom, 95% confidence interval for one-tailed test and null hypothesis that the

coefficients are not autocorrelated (Gujarati & Porter, 2018). The autocorrelation of the

coefficients is desirable, so our goal was to accept the alternative hypothesis (that is approximately

for values t <2). At the same time, we checked the autocorrelation for all Lags, where we used

the Ljung - Box Q statistics through the chi-square distribution, with the same degrees of freedom

as the lags, 95% confidence level, one-tailed test and null hypothesis that the data is random and

without apparent trend (LBQ>chi-squared). That is, not all autocorrelation coefficients are

statistically different from zero (Gujarati & Porter, 2018). The same test was performed through

SPSS where the p-value <5% for the Box-Ljung statistic criterion was required.

The final identification of the appropriate model per coastal line was made by comparing the

ACF and PACF calculated from the data with the theoretical ACF and PACF for the various

ARIMA models. The general logic was that if the sample autocorrelations exponentially drop to

zero and some are interrupted, the model will require autoregressive terms. If the sample

autocorrelations are interrupted and some of them decrease the model will also require moving

average terms (Kyriakidis, 2018). By counting the number of significant sample autocorrelations

and the partial autocorrelations we determined the classes of MA and AR terms.

For instance, in itinerary "A" we got a regular and a seasonal difference, because through the

sequence chart and the ACF diagram of the original series we found no stationarity in the data.

After the differences, the autocorrelations of the time series decreased rapidly and were zero,

which was a sign of stationarity for us. The Augmented Dickey-Fuller test statistic gave a p-

value<5%, which means that the null hypothesis is rejected and indeed the time series was

stationary. Because we wanted our data to have the desired autocorrelation, using the t-statistic

for a significance level of 5%, one-tailed test, and n-1 degrees of freedom, we found that the

autocorrelation coefficients were statistically different from zero. The same conclusions were

emerged by the chi-squared statistic (zero hypothesis rejection), showing that the time series data

as a whole were not random.

So we decided to proceed with the modeling of the seasonal ARIMA model. We definitely had

1 nonseasonal difference (d) and 1 seasonal difference (D). In nonseasonal AR (p) we tested

values 1 and 2 because we had lags which are significantly correlated and in seasonal AR (P) the

value 1 as it is sufficient for most seasonal patterns (IBM, 2021). Considering that sample

10

autocorrelation ceases after the 1st lag and partial autocorrelation decreases, we used values 1 and

2 as nonseasonal MA (q) and obtained value 1 as seasonal MA (Q) as it is sufficient for most

seasonal patterns (IBM, 2021).

The second stage concerned the "assessment" of the model and specifically its parameters.

Based on the previous analysis for line "A" we checked several models, keeping the differences

constant. The model with the lowest RMSE, MAE, MAPE and normalized BIC was the SARIMA

(0,1,1) (1,1,0)4.

In the third stage and before using the models for prediction, we checked them for their

"adequacy". Adequate is the model whose residuals are random and independent (Gujarati &

Porter, 2018). Through minitab 19 we relied on a chi-square test, based on Ljung-Box statistics

with number of lags minus number of parameters degrees of freedom. If p-value>5% for all

individual values (lags), the residual autocorrelations were considered to express consistent and

random errors (white noise). At the same we performed a comprehensive check of the adequacy

of the model, through the chi-square test based on Ljung-Box statistics (SPSS). For line "A" it

appeared that the errors had white noise behavior. Then, the statistical significance of the

parameters of the selected model was checked. In line "A" because p-value<5% the coefficients

were statistically significant and were maintained in the model. Finally, the interpretive power of

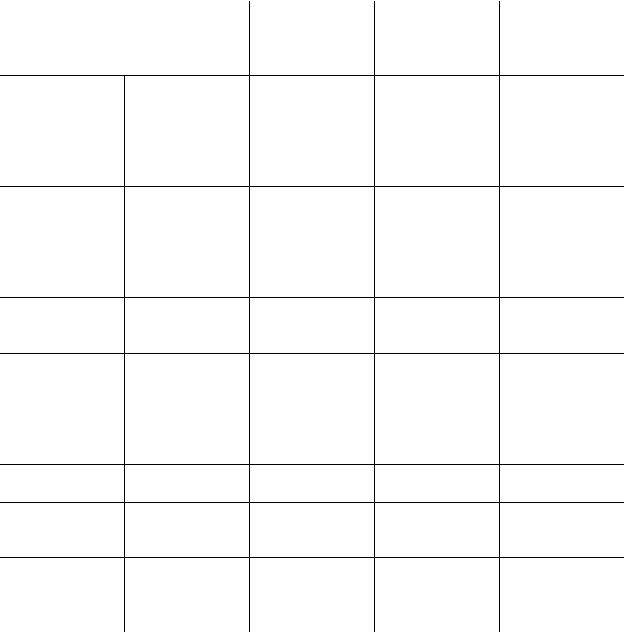
the model was investigated through stationary R squared. Given the adequacy of the models, we

adhered to the principle of parsimony and on every case, we chose the simplest model that

provided an adequate description of the main characteristics of the data (Kyriakidis, 2018).

11

**Table 2:** Basic steps for SARIMA forecasting in GCS (2004-2019).



**STEP 1:**

**Model recognition**

**STEP 2:**

**Model**

**estimation**

**STEP 3:**

**Model**

**adequacy**

**STEP 4:**

**Model**

**forecasting and**

**feedback**

1. Is the best

1. Are data 2. Now the data 1. Which is the fitted model

stationary? are stationary: best fitted model statistically

per route?

adequate?

1. Which are the

quarterly

forecasts  of

passenger traffic

for years 2020-

2022?

a. Check

sequence chart

b. Check ACF,

PACF diagrams

a.   Αre   the

autocorrelation

coefficients

statistically

different   from

zero?

b. Are time series

data as a whole

random?

a. Check which

model has the

lowest result in

MAPE, RMSE,

MAE  and

normalized BIC

a.  Check  the

residuals

whether  are

random  and

independent.

b. Check model

interpretive

power

a.  We  must

forecast

passenger traffic

only  for  year

2022  cause

COVID-19.

b. Compare the

predictive values

to the actuals

c. If there is no

stationarity take

a normal or both

a normal and a

seasonal

difference

c. Which are the

possible  values

of  SARIMA

trend  and

seasonal

components?

d. Check again

sequence chart

e. Check again

ACF, PACF

diagrams

f. Use

augmented

Dickey-Fuller

test

In the last stage and after determining the adequacy of the models, we made forecasts for the

years 2020,2021,2022 per quarter and we compared the predictive values to the actuals.

**5. Analysis and results**

As we said, the quarterly data for most of the coastal lines show a marginally decreasing or

increasing trend and a relatively stable seasonality (repeated) (Figure 1). Using SPSS statistical

software we calculated Winters ’additive and multiplicative model and the additive and

multiplicative model of decomposition in every significant route of GCS. Also we analyzed

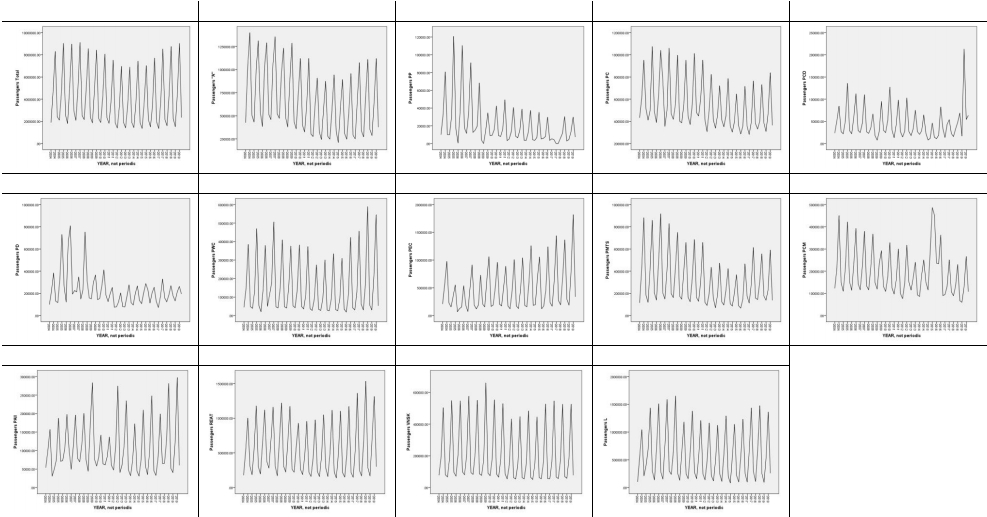
SARIMA models and simple seasonal exponential smoothing models per route. Especially for

decomposition method we calculated both trend and seasonality or only seasonality. So the results

concluded a linear trend and seasonal indices per quarter.

12

T



A

PP

PC

PCD

PD

PWC

PEC

PMTS

PCM

PAII

REAT

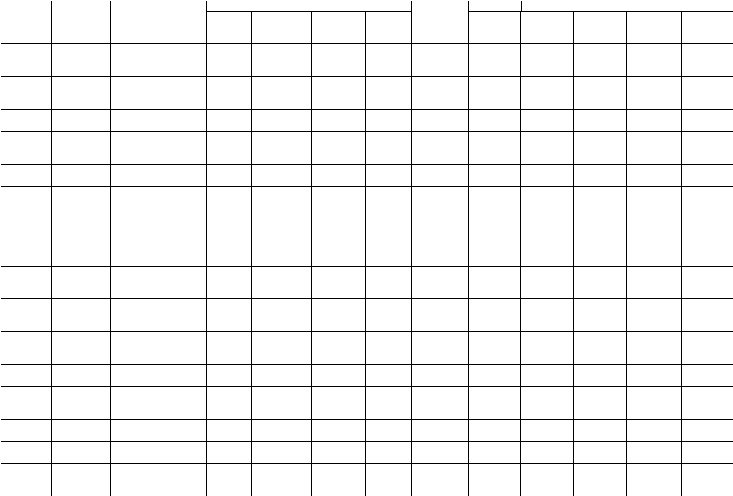
VNSK

L

**Figure 3:** Sequence charts of passenger traffic of the main coastal itineraries in Greece (2004-2019) (data on quarterly basis).

13

The time series on the coastal lines of Greece were examined for the first time, so we considered



it expedient to find the optimal parameters for each method used. That is, those values that

minimize the RMSE criterion (table 3). With SPSS finding the best values of α, β, γ is no longer

the problem (Tamber, et al., 2021). An exception was the SARIMA method where the approach

was done step by step, as described in the methodology. In this case too, however, the exported

model was compared with the excellent one via SPSS (through SPSS modeler). In case of

conflicting results our main selection criteria were RMSE and normalized BIC. Moreover,

through the stationary R squared we performed the interpretation power of the selected model and

with the use of the Ljung Box test we checked its adequacy. Then we made a forecast for the year

2022, which was our final goal, as in theory there will be a return to normality (after the COVID-

19 pandemic).

**Table 3:** Best fitted forecasting methods for all lines in GCS (2004-2019).

OPTIMAL

Decision Criteria

Forecast (2022)

ROUTE

BEST

METHOD

PARAMETERS

OR FORECASTING

EQUATIONS

MAPE

MAE

RMSE

BIC

Stationary

R2

Ljung-

Box (sig)

Q1

Q2

Q3

Q4

α (level)=0.306

T

WM

β (trend)=0.000

6.039

215260.3

294890.3

25.384

0.482

0.913

1540000

4370000

8710000

2220000

γ (seasonal)=0.686

α (level)=0.357

A

WM

β (trend)=0.000

6.749

39243.2

53079.5

21.954

0.538

0.748

259080

623203

1020000

324689

γ (seasonal)=0.686

PP

SS

α (level)=0.110

δ (seasonal)=0.872

82.142

6054.2

10448.6

18.638

0.403

0.840

5822

13977

30044

6825

α (level)=0.285

PC

WM

β (trend)=0.001

8.019

42879.3

61744.5

22.256

0.733

0.412

278050

388919

688388

348976

γ (seasonal)=0.020

PCD

SS

α (level)=0.118

δ (seasonal)=0.596

27.673

12782.6

27146.1

20.548

0.493

0.977

34561

146343

62183

65877

☑ Υt=303722-2287t

☑ Seasonal indices

PD

DMTS

per quarter:

1: 0.60398

29.765

-

108750.6

-

-

-

82617

146142

216239

87566

2: 1.08655

3: 1.63551

4: 0.67396

α (level)=0.152

PWC

WM

β (trend)=0.000

15.982

19917.6

32556.8

20.976

0.646

0.359

31786

195435

552708

55475

γ (seasonal)=0.566

α (level)=0.341

PEC

WM

β (trend)=0.000

23.695

80663.3

115526.3

23.509

0.595

0.769

253116

942605

1960000

349748

γ (seasonal)=0.375

α (level)=0.228

PMTS

WM

β (trend)=0.000

12.544

31153.9

44033.5

21.580

0.568

0.051

111435

205063

478801

121487

γ (seasonal)=0.768

PCM

SS

α (level)=0.800

δ (seasonal)=0.940

20.379

34210.7

50724.4

21.798

0.437

0.966

69395

122896

257577

107133

α (level)=0.800

PAII

WM

β (trend)=0.001

24.531

21330.1

28319.9

20.698

0.630

0.426

51579

128663

281568

60866

γ (seasonal)=0.899

REAT

VNSK

SS

SS

α (level)=0.211

δ (seasonal)=1.000

α (level)=0.299

δ (seasonal)=0.000

10.632

12.430

47950.9

17012.7

70316.0

26667.9

22.451

20.512

0.416

0.573

0.972

0.008

144073

64477

684704

177411

1340000

525451

297672

77157

α (level)=0.209

L

WM

β (trend)=0.000

13.749

58622.3

86110.2

22.922

0.702

0.831

95890

588148

1310000

242620

γ (seasonal)=0.411

Most forecasts gave us MAPE below 20%, so the best fitted methods describe relatively good

forecasts. Particularly, our analysis revealed that surprisingly eight of fourteen itineraries

(including total passenger traffic) integrated better to Winters ’multiplicative method (figure 4

and table 3). It proved the better model for the short - term quarterly seasonality as many

researchers have shown (Makatjane & Moroke, 2016; Dingari, et al., 2019). Other itineraries

14

fitted better to SS model and only in "PD" the best method was DMTS. No line led to better

results through the SARIMA models. The choice of SS and WM methods shows that the

smoothing methods show satisfactory accuracy rates in relation to the SARIMA models and in

general in relation to more complex forecasting methods (Petropoulos & Asimakopoulos, 2013).

This is because they are not affected by the peculiarities of the data patterns or by occasionally

occurring extreme values

What we noticed is that in all the lines selected by the WM method the trend parameter β was

almost zero, which means that the passenger traffic trend does not change over time. The slope

of the trend line was constant during the observation period. In some lines the value of the

parameter α (level) was quite large (eg "PAII") which shows that in this case more weight is given

to the most recent observations and very little weight to the most remote (Agiakloglou &

Oikonomou, 2019; Trull, et al., 2020). In the lines where the value of α was lower, the smoothing

of the time series was more intense, with the respective forecasting models fluctuating around the

initial level and being slow to follow large changes in the historical data. The high weighting

parameter γ for seasonal components showed for the majority of itineraries that seasonal factor

has great influence. This is reasonable because of the observed seasonality in GCS. An exception

is the "PC" line where the low value of γ indicates a stable seasonal effect (Vujko, et al., 2018).

For all coastal lines the Ljung - Box statistical criterion showed that the errors had white noise

behavior and the models were adequate. Also, in all lines the coefficient of determination R

squared was relatively high, which shows the good interpretive power of the models.

The lines that SPSS showed SS as the best model, the conclusions vary. The rule is that the

seasonal factor has a significant influence, except for the "VNSK" line where there is a constant

seasonal effect. The smoothing parameter of level differs per route, considering the importance

of older or newer data. The resulting models, outside the "VNSK" line, are adequate and with

relatively high data adaptability. "PD" line was the only one that gave DMTS as the best model.

The -2287 slope of the linear trend equation shows an average decrease of 2287 passengers per

quarter. The corresponding values of the seasonal indices show that passenger traffic is increased

in the second and third quarters and decreased in the first and fourth. However, the MAPE clarifies

a marginally good forecast.

In conclusion, by comparing predictive values to the actuals, interesting results emerged. For

the first quarter of 2020, when covid-19 pandemic had not fully prevailed, in eight to fourteen

lines the percentage deviation was under 30%. In "PDM" line was -0.72%. Also, in all lines the

actual passenger traffic was inside the upper and low bound of the forecast. Taking into account

firstly that the forecasting methods gave more positive results than the real ones, because of the

long-term upward trend of tourist arrivals and secondly that the Greek government took the first

restrictive decisions for passenger traffic in March of 2020 (a month of the first quarter), we have

to do with a relatively good forecasting result.

**6. Conclusions and discussion**

GCS is one of the biggest in Europe and covers about 17% of total passenger shipping. It plays

an important role for Greek economy and society. There are not a lot of scientific efforts in

forecasting passenger traffic in Greece. In order to fill this gap, the main goal of this study was to

find an optimal forecasting method, by comparing Box-Jenkins ARIMA, smoothing and

15

decomposition methods. As GCS consists of several concentrated submarkets (lines) we remained

in fourteen popular itineraries (including total passenger traffic). Taking into consideration the

high seasonality and no stationarity that characterizes those routes we limited our analysis to

Winter’s triple exponential smoothing, the time series decomposition method, the simple seasonal

model and the seasonal ARIMA models.

Even if we followed a careful step by step approach for SARIMA models (“recognition”,

“assessment”, “adequacy”, “forecasting and feedback”) no coastal line led to better results by this

method. In fact, eight of fourteen itineraries integrated better to WM, five of fourteen to SS and

only one to DTMS. Especially for WM it emerged from the analysis that traffic trend did not

change over time, in some lines the smoothing of the time series was more intense, and the

seasonal factor had great influence. The suggested models were adequate with relatively high

interpretative power. About SS method the smoothing parameter of level differed per route and

seasonality was of great significance. The resulting models presented high data adaptability. In

"PD" line, where DTMS model seemed the best one, the -2287 slope of the linear trend equation

shew an average decrease of 2,287 passengers per quarter.

In general, most forecasts gave as MAPE below 20%, so the best fitted methods described

relatively good forecasts. Of course, the results should be treated with caution since COVID-19

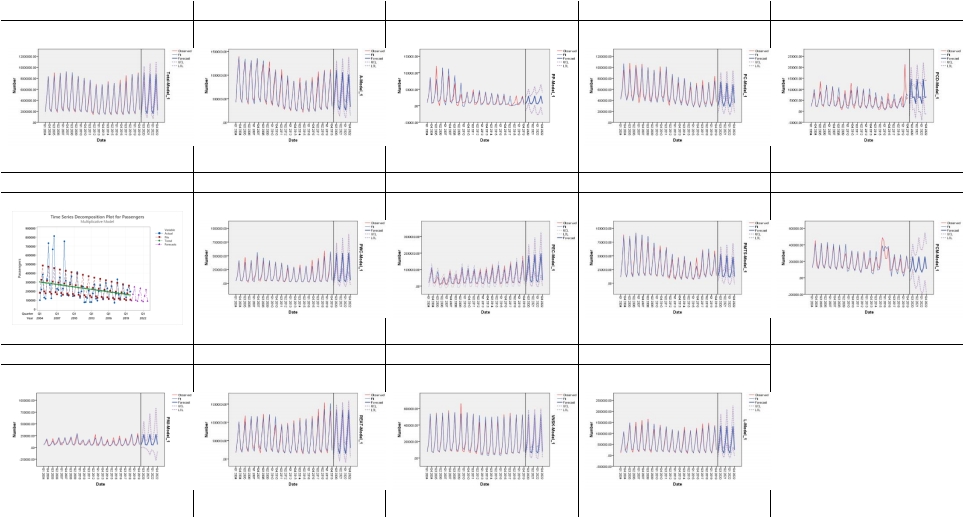
pandemic does not allow safe conclusions for the forecasting period 2020-2022 in GCS. However,

the forecasting of the first quarter of 2020, when pandemic had not fully prevailed, gave

encouraging results with little deviations between predicted and actual values.

16

T (method: WM)



A (method: WM)

PP (method: SS)

PC (method: WM)

PCD (method: SS)

PD (method: DMTS)

PWC (method: WM)

PEC (method: WM)

PMTS (method: WM)

PCM (method: SS)

PAII (method: WM)

REAT (method: SS)

VNSK (method: SS)

L (method: WM)

**Figure 4:** Best fitted method and forecasting results for the main coastal itineraries in Greece (2020-2022) (data on quarterly basis).

17

*Nomenclature*

GCS Greek Coastal Shipping

WM Winters’ Multiplicative method

DTMS Decomposition Multiplicative Trend and Seasonal method

SS Simple Seasonal method

SARIMA Seasonal ARIMA models

ACF Autocorrelation

PACF Partial autocorrelation

GDP Gross domestic product

*References*

Agiakloglou, C. N. & Oikonomou, G. S., 2019.*Forecasting methods and decision analysis.* Athens: Benou.

Ahmad, W. & Ahmad, S., 2013.*Arima model and exponential smoothing method: a comparison.* s.l.,

Department of Mathematics, Faculty of Science and technology.

Aivazidou, E., 2015.*Development of time series and regression models for the assessment of the effects of the*

*economic crisis on maritime passenger and freight traffic in Greece,* Thessaloniki: UTH.

Chalkos, G. E., 2020.*Statistics: Theory and Practice.* Thessaloniki: Disigma.

Chatfield, C., 2003.*The analysis of time series: An introduction.* New York: Chapman and Hall.

Dhali, N., Barman, N. & Hasan, B., 2019. Determination of optimal smoothing constants for Holt-Winter's

multiplicative method.*Dhaka Univ. J.*, July, 2(67), pp. 99-104.

Dingari, M., Reddy, M. & Sumalatha, V., 2019. Air traffic forecasting using time series models.*International*

*journal of recent technology and engineering*, November, 8(4).

Goulielmos, A. & Sambrakos, E., 2002.*Ferry and short-sea shipping.* Piraeus: Stamoulis.

Goulielmos, A. & Sitzimis, I., 2012. Measuring market concentration in the Aegean Ferry System.*Spoudai*

*Journal*, 62(1-2), pp. 7-27.

Goulielmos, A. & Sitzimis, I., 2014. The Liberalization process of the Ferry System in Greece, 2001-2009:

What have been the benefits to users of Aegean Sea Transportation?.*Spoudai Journal*, 64(4), pp. 39-66..

Gujarati, D. & Porter, D., 2018.*Econometrics: Principles and applications.* Thessaloniki: Tziolas.

Hansun, S., Vincent, C. & Subanar, C., 2019. Revitising the Holt-Winters' additive method for better

forecasting.*International journal of enterprise information systems*, 2(15), pp. 43-57.

IBM, 2021. *IBM* *SPSS* *Forecasting:* *Build* *expert* *forecasts* *in* *a* *flash.* [Online]

Available at: https://www.ibm.com/downloads/cas/OP3RLVLR

IOBE, 2014.*The contribution of passenger shipping to the Greek economy: performance and prospects,*

Athens: Foundation for economical and industrial research.

IOBE, 2020.*The Greek coastal shipping during period 2016-2020: Performance, contribution in economy*

*and prospects.* Athens: Foundation for economic and industrial research.

Karmaker, C., Halder, P. & Sarker, E., 2017. A study of time series model for predicting Jute Yarn demand:

Case study.*Hindawi, Journal of Industrial Engineering*.

Kyriakidis, M., 2018.*Τelecommunications market analysis and forecasting techniques.* Athens: National and

Kapodistrian university of Athens.

Makatjane, K. & Moroke, N., 2016. Comparative study of Holt-Winters triple exponential smoothing and

seasonal ARIMA: Forecasting short term seasonal car sales in South Africa.*Risk governance: financial*

*markets & institutions*, 6(1).

Ma, L., Hu, C. & Han, Y., 2018.*ARIMA model forecast based on eviews software.* s.l., IOP conf. series: earth

and environmental science 208.

Munarsih, E. & Saluza, I., 2019.*Comparison of exponential smoothing method and autoregressive integrated*

*moving average (ARIMA) method in predicting dengue fever cases in the city of Palembang.* s.l., Journal

of Physics: conference series.

Ortuzar, D. & Gonzalez, M., 2002. Inter-island demand response with discrete choice models.*Journal of*

*transport economic and policy*, January, 46(1).

Petropoulos, F. & Asimakopoulos, V., 2013.*Business forecasting.* Athens: Symmetria.

18

Psaraftis, C., 1994.*Greek coastal shipping system: impact of market deregulation and new technologies on*

*modal split.* Capri, Italy, s.n.

Ramanathan, R., 2001. The long-run behavior of transport performance in India: a cointegration approach.

*Transportation research*, 35(A).

Rashidi, S. & Ranjitkar, P., 2015. Estimation of bus dwell time using univariate time series models.*Journal*

*of advanced transportation*, Issue 49, pp. 139-152.

Ravinder, H., 2013. Determing the optimal values of exponential smoothing constants - does solver really

work?.*American journal of business education*, May/June, 6(3), pp. 347-360.

Sabry, M., Abd-El-Latif, H. & Badra, N., 2007. Comparison between regression and arima models in

forecasting traffic volume.*Australian journal of basic and applied sciences*, 126-136, 2(1).

Schwarz, G., 1978. Estimating the dimension of a model.*The annals of statistics*, 2(6), pp. 461-464.

Shim, J. K. & Siegel, J. G., 2001.*Managerial economics.* Athens: Kleidarithmos.

Sim, S., Tay, K., Huong, A. & Tiong, W., 2019. Forecasting electricity consumption using SARIMA method

in IBM SPSS software.*Universal journal of electrical and electronic engineering*, 6(5B), pp. 103-114.

Sitzimis, I., 2012.*Aegean coastal market: the consequences of cabotage removal by microeconomic tools.*

Thesis. Peiraus: University of Peiraus.

Spathi, S., 2005.*A comparison between air and coastal services in internal routes: An econometric estimation*

*of demand,* Athens: Center of programming and economical research.

Suhartono, 2011. Time series forecasting by using seasonal autoregressive integrated moving average: Subset,

multiplicative or additive model.*Journal of mathematics and statistics*, 1(7), pp. 20-27.

Tamber, Jighjigh, A., Michael, O. & Ojowu, O., 2021. The Holt-Winters multiplicative model of passengers'

traffic forecast of the Nigeria airports.*Internationa journal of engineering in computer science*, 3(1), pp.

35-40.

Trull, O., Garcia-Diaz, J. & Troncoso, A., 2020. Stability of multiple seasonal Holt-Winters models applied

to hourly electricity demand in Spain.*Applied Sciences*, February, Volume 10.

Tsekeris, T., 2008. Consumer demand analysis of complementarity-substitution relationships among

passenger transport modes in Greece.*International journal of transport economics*, 35(3), pp. 415-449.

Tsui, W., Balli, H., Gilbrey, A. & Gow, H., 2014. Forecasting of Hong Kong airport's passenger throughput.

*Tourism management*, Issue 42, pp. 62-76.

Vujko, A., Papic-Blagojevic, N. & Gajic, T., 2018. Applying the exponential smoothing model for forecasting

tourists' arrivals - example of Novisad, Belgrade and Nis.*Ekonomika Poljoprivrede*, July, pp. 757-473.

Wang, S., Feng, J. & Liu, G., 2013. Application of seasonal time series model in the precipitation forecast.

*Mathematical and computer modelling*, Issue 58, pp. 677-683.

Wardono, Arief, A. & Siti, R., 2016. Arima method with the software minitab and eviews to forecast inflation

in semarang indonesia.*Journal of theoritical and applied information technology*, December, 94(1).

Wu, J., Zhong, L., Li, L. & Lu, A., 2013.*A prediction model bases on time series data in intelligent*

*transportation system.* Berlin: Springer.

XRTC, B. c., 2004-2020.*Annual survey on Greek coastal shipping.* Athens: s.n.

Yonar, H., Yonar, A., Tekindal, M. & Tekindal, M., 2020. Modeling forecasting for the number of cases of

the COVID-19 pandemic with the curve estimation models, the Box-Jenkins and exponential smoothing

method.*EJMO*, 4(2), pp. 160-165.

19