ΠΑΝΕΠΕΙΣΤΗΜΙΟ ΠΕΙΡΑΙΩΣ

Τμήμα πληροφορικής

ΠΡΟΓΡΑΜΜΑ ΜΕΤΑΠΤΥΧΙΑΚΩΝ ΣΠΟΥΔΩΝ
ΣΤΑ ΠΡΟΗΓΜΕΝΑ ΣΥΣΤΗΜΑΤΑ ΠΛΗΡΟΦΟΡΙΚΗΣ

ΠΡΟΒΛΕΨΕΙΣ ΧΡΟΝΟΣΕΙΡΩΝ ΚΑΙ ΕΦΑΡΜΟΓΗ ΣΤΟΝ ΤΟΜΕΑ ΤΩΝ LOOGISTICS

ΜΠΕΚΙΟΣ ΔΗΜΗΤΡΙΟΣ

Διπλωματική Εργασία υποβληθείσα στο Τμήμα Πληροφορικής Πανεπιστημίου Πειραιώς ως μέρους των απαιτήσεων για την απόκτηση Μεταπτυχιακού Διπλώματος Προηγμένα Συστήματα Πληροφορικής

UNIVERSITY OF PIRAEUS

Department of informatics

MASTER PROGRAMM IN

ADVANCED INFORMATION SYSTEMS

TIME SERIES PREDICTIONS AND APPLICATION IN LOGISTICS

BY

BEKIOS DIMITRIOS

Master Thesis submitted to the Department of Ινφορματιψσ of the University of Piraeus in partial fulfillment of the requirements for the degree of Master of Arts in Advanced Information Systems

ΠΡΟΒΛΕΨΕΙΣ ΧΡΟΝΟΣΕΙΡΩΝ ΚΑΙ ΕΦΑΡΜΟΓΗ ΣΤΟΝ ΤΟΜΕΑ ΤΩΝ LOOGISTICS

Περίληψη

Η παρούσα διπλωματική εργασία μελετά την πρόβλεψη χρονοσειρών με τεχνικές μεθόδους πρόβλεψης και με την χρήση του στατιστικού εργαλείου R, με στόχο την εφαρμογή των διαφόρων μοντέλων πρόβλεψης στον τομέα των logistics.

Με βάση όλα τα παραπάνω, μελετώνται αρχικά οι βασικές σημασιολογικές έννοιες που αφορούν τη διαδικασία της πρόβλεψης και το πόσο σημαντική είναι στις μέρες μας.

Επιπλέον μελετώνται οι διαφορετικές μέθοδοι πρόβλεψης ανάλογα με τον αριθμό και την ποιότητα των διαθέσιμων παρατηρήσεων με συγκεκριμένα παραδείγματα.

Παράλληλα δίνεται ιδιαίτερη βαρύτητα στην πιο σημαντική μέθοδο πρόβλεψης που είναι η ανάλυση των χρονοσειρών με την βοήθεια του στατιστικού εργαλείου R.Αναλύονται και παρουσιάζονται πλήρως όλα τα μοντέλα ανάλυσης χρονοσειρών,καθώς επίσης και τα αποτελέσματα απο την εφαρμογή τους σε πραγματικά παραδείγματα.Επίσης γίνεται γραφική απεικόνιση των αποτελεμάτων και έλεγχος σφαλμάτων πρόβλεψης

Τέλος, γίνεται μια προσπάθεια διεξαγωγής συμπερασμάτων ως προς την καταλληλότητα και την αξιοπιστία του κάθε μοντέλου πρόβλεψης σύμφωνα με τις κατηγορίες χρονοσειρών.

TIME SERIES PREDICTIONS AND

APPLICATION IN LOGISTICS

Abstract

This dissertation studies the forecast time series with time series techniques and methods using the statistical tool R, in order to implement the varius forecasting models in logistics.

Based on the above, initially studied the basic concepts concerning the process of forecasting and the importance nowadays.

Moreover studied different forecasting methods depending on the number and quality of available data with specific examples.

While special attention is given to the most important method of forecasting is the analysis of time series with the help of statistical tool R. Analysed and fully presented all time series analysis model ,as well as the results of their application to real examples. In addition we have graphical representation of the results and control forecasting errors.

Table of Contents

Περίληψη .... Error! Bookmark not defined.

Κατάλογος Εικόνων 10

ΚΕΦΑΛΑΙΟ 1 12

Εισαγωγή 12

1.1 Περιγραφή του αντικειμένου μελέτη της εργασίας 12

1.2 Κύριος Σκοπός της εργασίας 13

1.3 Περιγραφή του τρόπου και των μεθόδων διρεύνησης του αντικειμένου της εργασίας 13

1.4 Σύνοψη 14

ΚΕΦΑΛΑΙΟ 2 16

Χρονοσειρές 16

2.1 Βασικά χαρακτηριστικά χρονοσειράς 20

2.1.2 Στασιμότητα, τάση και περιοδικότητα 20

2.2 Απαλοιφή τάσης και περιοδικότητας 21

2.2.1 Λευκός θόρυβος 24

2.2.2 Τυχαίος περίπατος 24

2.2.4 Παράδειγμα με δεδομένα με σταθερή διακύμανση και χωρις εποχικότητα στο εργαλείο R 25

2.3 Παράδειγμα με τάση και εποχικότητα στο εργαλείο R 28

2.3.1 Παράδειγμα χωρίς τάση και χωρίς εποχικότητα 31

2.4 Αποσύνθεση Χρονοσειρών 34

2.4.1 Αποσύνθεση δεδομένων χωρίς εποχικότητα 35

2.4.2 Αποσύνθεση δεδομένων με εποχικότητα 38

2.5 Τροποποίηση Εποχικότητας 41

Κεφάλαιο 3 43

Προβλέψεις 43

3.1 Απλή Εκθετική Εξομάλυνσης σε δεδομένα με τάση και χωρίς εποχικότητα 43

3.1.1 Παράδειγμα πρόβλεψης με απλή εκθετική εξομάλυνση 44

3.1.2 Χρήση λειτουργίας forecast.HoltWinters 49

3.1.3 Σφάλματα Πρόβλεψης 52

3.1.4 Εγκυρότητα και αξιοπιστία πρόβλεψης 52

3.2 Εκθετική εξομάλυνση Holt 58

3.2.1 Προβλέψεις με δεδομένα χωρίς τάση και εποχικότητα 58

3.2.2 Έλεγχος αξιοπιστίας αποτελεσμάτων πρόβλεψης 65

3.3 Προβλέψεις σε δεδομένα χωρίς τάση και με εποχικότητα ... 69
3.3.1 Έλεγχος αξιοπιστίας αποτελεσμάτων πρόβλεψης ... 74
Κεφάλαιο 4 ... 77
4.1 Μοντέλα ARIMA ... 77
4.1.1 Η Μεθοδολογία των Box & Jenkins ... 78
4.1.2 Πρώτο Στάδιο: Ταυτοποίηση ... 78
4.1.3 Δεύτερο Στάδιο: Εκτίμηση ... 79
4.1.4 Διαγνωστικός έλεγχος ... 80
4.1.5 Τέταρτο Στάδιο: Μεταδιάγνωση (Πρόβλεψη) ... 81
4.2 Επιλογή κατάλληλου μοντέλου ARIMA ... 82
4.2.1 Διαφοροποίηση χρονοσειράς ... 84
4.2.2 Απεικόνιση διαφοροποιημένης χρονοσειράς ... 84
4.2.3 Διαδικασία Επιλογής Μοντέλου ARIMA για παράδειγμα με καθυστερήσεις πτήσεων ... 86
4.2.4 Συμπέρασμα Επιλογής Μοντέλου ARIMA για παράδειγμα με καθυστερήσεις πτήσεων . 89
4.3 Πρόβλεψη με την χρήση Μοντέλου ARIMA(0,1,1) ... 90
4.3.1 Απεικόνιση πρόβλεψης μοντέλου ARIMA(0,1,1) ... 91
4.3.2 Έλεγχος αξιοπιστίας αποτελέσματος πρόβλεψης με την δοκιμή Ljung-Box ... 92
Κεφάλαιο 5 ... 96
Αποτελέσματα ... 96
5.1 Εισαγωγή ... 96
5.2 Αποτελέσματα απλής εκθετικής εξομάλυνσης ... 96
5.3 Αποτελέσματα εκθετικής εξομάλυνσης HoltWinters σε χρονοσειρά χωρίς τάση και εποχικότητα ... 99
5.4 Αποτελέσματα εκθετικής εξομάλυνσης HoltWinters σε χρονοσειρές χωρίς τάση αλλά με εποχικότητα ... 101
5.5 Αποτελέσματα ARIMA(0,1,1) ... 102
Κεφάλαιο 6 ... 104
Συμπεράσματα ... 104
6.1 Εισαγωγή ... 104
6.2 Συμπεράσματα ... 104

Κατάλογος Πινάκων

Πίνακας 2 1 :Πίνακας με καθυστερήσεις 42 πτήσεων 25

Πίνακας 2. 2:Προιόντα που δεν αφίχθησαν έγκαιρα στους παραλήπες 29

Πίνακας 2. 3:Καθυστερήσεις σε μεταφορά εμπορευμάτων με φορτηγά πλοία 31

Πίνακας 3 1:Προβλέψεις για τις επόμενες 8 πτήσεις 50

Πίνακας 3. 2:Δεδομένα με χρόνους μεταφοράς προιόντων σε διάστημα 46 ημερών 59

Πίνακας 3. 3:Πρόβλεψη για τις επόμενες 19 ημέρες 63

Πίνακας 4. 1:Αποτελέσματα καθυστέρησης για τις επόμενες 5 πτήσεις 90

Πίνακας 5. 1:Πίνακας με τις τιμές πρόβλεψης της απλής εκθετικής εξομάλυνσης 98

Πίνακας 5. 2:Πίνακας με τις προβλέψεις σε χρονοσειρά χωρίς τάση και εποχικότητα 100

Κατάλογος Διαγραμμάτων

Διάγραμμα 2. 1:Απεικόνιση με τις καθυστερήσεις 42 πτήσεων 27

Διάγραμμα 2. 2:Διαγραμματική απεικόνιση προιόντων που αφίχθησαν αργοπορημένα 30

Διάγραμμα 2. 3 Διαγραμματική απεικόνιση προιόντων που καθυστέρησαν κατα την μεταφορά τους με φορτηγά πλοία 32

Διάγραμμα 2. 5 Διαγραμματική απεικόνιση με καθυστερήσεις 42 πτήσεων 36

Διάγραμμα 2. 6 Διαγραμματική απεικόνιση με καθυστερήσεις 42 πτήσεων 37

Διάγραμμα 2. 7 Διαγραμματική απεικόνιση προιόντων πρόσθετης χρονοσειράς 40

Διάγραμμα 2. 8 Εποχική προσαρμογή χρονοσειράς 42

Διάγραμμα 3. 1 Απεικόνιση με καθυστερήσεις 100 πτήσεων 46

Διάγραμμα 3. 2 Απεικόνιση αρχικής χρονοσειράς με πρόβλεψη 48

Διάγραμμα 3. 4 Απεικόνιση πρόβλεψης για τις επόμενες 8 πτήσεις 51

Διάγραμμα 3. 5 Απεικόνιση στο κορελόγραμμα 53

Διάγραμμα 3. 6 Απεικόνιση σφαλμάτων πρόβλεψης 55

Διάγραμμα 3. 7 Γραφική απεικόνιση Ιστογράμματος σφαλμάτων πρόβλεψης 56

Διάγραμμα 3. 8 Απεικόνιση χρόνων μεταφοράς σε διάστημα 42 ημερών 60

Διάγραμμα 3. 9 Προσαρμογή μοντέλου πρόβλεψης HoltWinters 62

Διάγραμμα 3. 10 Γραφική απεικόνιση με την μέθοδο HoltWinters εκθετικής εξομάλυνσης . 64

Διάγραμμα 3. 11 Απεικόνιση στο κορελόγραμμα 66

Διάγραμμα 3. 12 Απεικόνιση σφαλμάτων πρόβλεψης συναρτήση του χρόνου 67

Διάγραμμα 3. 13 Απεικόνιση Ιστογράμματος... 68

Διάγραμμα 3. 14 Απεικόνιση πρόσθετου μοντέλου πρόβλεψης επι της αρχικής χρονοσειράς72

Διάγραμμα 3. 15 Απεικόνιση πρόβλεψης για τους επόμενους 48 μήνες 73

Διάγραμμα 3. 16 Απεικόνιση κορελογράμματος 74

Διάγραμμα 3. 17 Απεικόνιση σφαλμάτω πρόβλεψης 75

Διάγραμμα 3. 18 Απεικόνιση Ιστογράμματος... 76

Διάγραμμα 4. 1 Απεικόνιση πρώτης διαφοροποίησης της αρχικής χρονοσειράς 85

Διάγραμμα 4. 2 Απεικόνιση κορελογράμματος 86

Διάγραμμα 4. 3 Απεικόνιση μερικού κορελογράμματος 87

Διάγραμμα 4. 4 Απεικόνιση μοντέλου πρόβλεψης ARIMA(0,1,1) 91

Διάγραμμα 4. 5 Απεικόνιση κορελογράμματος για έλεγχο μη μηδενικών αυτοσυσχετισμών 93

Διάγραμμα 4. 6 Απεικόνιση σφαλμάτων πρόβλεψης 94

Διάγραμμα 4. 7 Απεικόνιση Ιστογράματος σφαλμάτων πρόβλεψης 95

Διάγραμμα 5. 1 Μέθοδος απλής εκθετικής εξομάλυνσης σε δεδομένα με τάση και χωρίς εποχικότητα a=0.024 και b=1 97

Διάγραμμα 5. 2 Holt Winters εκθετική εξομάλυνση σε δεδομένα χωρίς τάση και εποχικότητα 99

Διάγραμμα 5. 3 Πρόβλεψη HoltWinters σε χρονοσειρά χωρίς τάση και με εποχικότητα με a=0.83 b=0 και gamma=1 101

Διάγραμμα 5. 4 Απεικόνιση πρόβλεψης μοντέλου ARIMA... 102

Διάγραμμα 5. 5 Απεικόνιση αυτοσυσχετίσεων σφαλμάτων πρόβλεψης 103

Κατάλογος Εικόνων

Εικόνα 2. 1 Κώδικας σε γλώσσα R 26

Εικόνα 2. 2 Κώδικας σε γλώσσα R για την απεικόνιση της χρονοσειράς 32

Εικόνα 2. 3 Κώδικας σε γλώσσα R για την απεικόνιση εποχικότητας 38

Εικόνα 2. 4 Εκτίμηση των τιμών της εποχικότητας,τάσης και των ακανόνιστών στοιχείων .. 39

Εικόνα 2. 5 Κώδικας για αποσύνθεση χρονοσειράς 41

Εικόνα 2. 8 Αποθηκευμένες τιμές εξόδου και προβλέψεων HoltWInters εντολής 47

Εικόνα 3. 2 Κώδικας σε γλώσσα R για χρήση της εντολής HoltWinters 44

Εικόνα 3. 3 Συνέχεια Κώδικας σε γλώσσα R για χρήση της εντολής HoltWinters 45

Εικόνα 3. 4 Κώδκας σε γλώσσα R για χρήση εντολής Box.Tset() 54

Εικόνα 3. 5 Κώδικας σε γλώσσα R για εισαγωγή δεδομένων και μοντέλου HoltWinters 61

Εικόνα 3. 6 Κώδικας σε γλώσσα R για έλεγχο αξιοπιστίας πρόβλεψης 65

Εικόνα 3. 7 Τιμές μοντέλου πρόβλεψης 70

ΚΕΦΑΛΑΙΟ 1

Εισαγωγή

1.1 Περιγραφή του αντικειμένου μελέτη της εργασίας

Η παρούσα διπλωματική εργασία επικεντρώνεται στην μελέτη της πρόβλεψης χρονοσειρών με μεθόδους πρόβλεψης και του στατιστικού εργαλείου R, με στόχο την εφαρμογή των διαφόρων μοντέλων στην πρόβλεψη μελλοντικών τιμών των καθυστερήσεων στην κατηγορία των logistics.

Η πρόβλεψη είναι μια από τις σημαντικότερες λειτουργίες μέσα σε μια επιχείρηση και εν γένει σε έναν οργανισμό, για τη λήψη κάθε κρίσιμης απόφασης: ο έλεγχος του κόστους, ο σχεδιασμός νέων προϊόντων, η πρόσληψη προσωπικού, ο όγκος της παραγωγής, το ύψος των αποθεμάτων, όλα καθορίζονται από την πρόβλεψη. Χωρίς αυτήν, κάθε απόφαση θα λαμβανόταν στην τύχη. (Waddell and Sohal, 1994 ,Newbold and Bos ,1994,Klassen and Flores Benito 2001).

Η ανάλυση των χρονοσειρών και οι μέθοδοι που θα χρησιμοποιήσουμε είναι δυο επιστημονικά πεδία που θα μας βοηθήσουν, έτσι ώστε να έχουμε σχετικά ακριβείς προβλέψεις με βάση στατιστικά και ιστορικά στοιχεία του παρελθόντος.

Απο την άλλη πλευρά όσο αφορά τον τομέα των logistics, οι παράγοντες που επηρεάζουν και αποτελούν τον κύριο λόγο για καθυστερήσεις στα προιόντα, στα δρομολόγια, στις πτήσεις αεροπλάνων κ.α, είναι πάρα πολλοί.Τέτοια παράγοντες μπορεί να είναι ο καιρός, η εποχή, οι άνθρωποι ,ακόμη και οι ίδιες οι χώρες.Γιαυτό τον λόγο και στην εργασία αυτή θα προσπαθήσουμε να κατηγοριοποιήσουμε τα στατιστικά και τα στοιχεία του παρελθόντος.

Στο πλαίσιο αυτο, με την συσκεκριμένη εργασία θα επικεντρωθούμε στον τρόπο με τον οποίο μπορούμε να συνδυάσουμε όλα τα παραπάνω, δηλαδή στον τρόπο με τον οποίο μπορεί η ανάλυση των χρονοσειρών και το εργαλείο R να χρησιμοποιηθούν κατάλληλα, με σκοπό να επιτεύξουν αξιόπιστες και αποτελεσματικές προβλέψεις.

Όσο αφορά την αξιοπιστία των προβλέψεων κρίνεται ιδιαίτερα σημαντικό, καθώς το να προβεί κάποιος σε προβλέψεις είναι σχετικά εύκολο, αλλά το να προβεί σε αξιόπιστες προβλέψεις είναι ένας διαφορετικός τομέας, τον οποίο θα αναλύσουμε παρακάτω.

Στις επόμενες ενότητες παρουσιάζονται οι κύριοι στόχοι της εργασίας αυτης καθώς επίσης και οι μέθοδοι ανάπτυξής της.

1.2 Κύριος Σκοπός της εργασίας

Οι κύριοι στόχοι της παρούσας εργασίας παρουσιάζονται παρακάτω:

- Η ανάλυση των εννοιών της πρόβλεψης και των χρονοσειρών.
- Η ανάλυση και αποσύνθεση των χρονοσειρών μέσα απο το στατιστικό εργαλείο R,καθώς και τα μοντέλα ανάλυσης τους.
- Η εφαρμογή των διαφόρων μοντέλων χρονοσειρών και πρόβλεψης με παραδείγματα,αναλύοντας τις μεθόδους πρόβλεψης, τις κατηγορίες που θα χωρίσουμε τις χρονοσειρές,καθώς και τον έλεγχο αξιοπιστίας των αποτελεσμάτων μέσω του εργαλείου R.

1.3 Περιγραφή του τρόπου και των μεθόδων διερεύνησης του αντικειμένου της εργασίας

Τα βασικά βήματα που θα ακολουθήσουμε για την προσέγγιση του θέματος της εργασίας παρουσιάζονται παρακάτω:

- Αρχικά αυτο που θα επιχειρήσουμε είναι να επιτύχουμε να χωρίσουμε τα δεδομένα σε κατηγορίες που εμφανίζουν κάποια ιδιαίτερα χαρακτηριστικά σε συγκεκριμένες περιόδους του έτους, όπως καλοκαίρι ή χειμώνας, σε κατηγορίες που εμφανίζουν πιο ομαλές διακυμάνσεις στα ιστορικά στοιχεία τους, αλλά και σε κατηγορίες όπου τα δεδομένα τους παρουσιάζουν διακυμάνσεις χώρις τάση, εποχικότητα ή σταθερότητα.
- Βέβαια όλα αυτά για να επιτευχθούν και να αναλυθούν σωστά είναι απαραίτητο να χρησιμοποιήσουμε τις σωστές μεθόδους πρόβλέψεων.Στην δική μας περίπτωση θα γίνει η εφαρμογή των μεθόδων πρόβλεψης που βασίζονται σε μαθηματικά μοντέλα όπως η απλή εκθετική εξομάλυνση,η εκθετική εξομάλυνση HaltWinters, τα μοντέλα Arima και η χρήση του στατιστικού εργαλείου R που θα μας βοηθήσει καθοριστικά στην εφαρμογή αυτή.

- Ο έλεγχος της αξιοπιστίας των προβλέψεων μπορεί να επιτευχθεί με αρκετούς τρόπους.Στην δική μας περίπτωση ο έλεγχος θα γίνει διαγραμματικά.Πιο συγκεκριμένα με τον όρο διαγραμματικά εννοούμε ότι θα ελέγχουμε τα αποτελέσματα τω προβλέψεων μας σύμφωνα με τη διαγραμματική απεικόνιση των σφαλμάτων πρόβλεψης και ελέγχοντας την κατανομή τους και τον μέσο όρο των σφαλμάτων.Σε αυτό βέβαια καθοριστικό ρόλο θα έχει το εργαλείο R, όπου με τον κατάλληλο χειρισμό του θα απλοποιήσει αρκετά τις μεθόδους που αναφέραμε παραπάνω

.

1.4 Σύνοψη

Στο πρώτο αυτο κεφάλαιο της παρούσας εργασίας, παρουσιάστηκαν το αντικείμενο και οι στόχοι της διπλωματικής αυτης εργασίας ,οι λόγοι που κάνουν το θέμα της εργασίας να είναι σημαντικό καθώς και οι μεθοδολογίες που θα χρησιμοποιηθούν για την προσέγγιση του θέματος που πραγματεύεται η εργασία.

ΚΕΦΑΛΑΙΟ 2

Χρονοσειρές

Όταν παρατηρούμε το μέγεθος στο χρόνο με σταθερό χρονικό βήμα, δηλαδή με σταθερό χρόνο δειγματοληψίας (sampling time). Ένα σύνολο τέτοιων παρατηρήσεων λέγεται χρονική σειρά ή χρονοσειρά (time series). Σε κάποια προβλήματα ο χρόνος δειγματοληψίας μπορεί να μην είναι σταθερός και τότε χρειάζεται ειδικότερη επεξεργασία της χρονοσειράς για να γίνει η ανάλυση. Για παράδειγμα οι ημερήσιες τιμές ενός χρηματιστηριακού δείκτη (π.χ. του Χρηματιστηρίου Αξιών Αθηνών) συνιστούν μια χρονοσειρά με μεταβλητό φυσικό χρόνο δειγματοληψίας, αφού μεσολαβούν Σαββατοκύριακα και γιορτές που είναι κλειστό το χρηματιστήριο. Η πιο απλή προσέγγιση σε αυτήν την περίπτωση είναι να ορίσουμε ως χρόνο αναφοράς όχι το φυσικό χρόνο αλλά τον οικονομικό χρόνο συναλλαγών και να θεωρήσουμε σταθερό χρονικό βήμα μιας (οικονομικής)ημέρας ακόμα και από Παρασκευή σε Δευτέρα.(Λέκκας Φραγκίσκος,Πανεπιστήμιο Αιγαίου). Στη συνέχεια θα θεωρούμε το χρόνο δειγματοληψίας σταθερό που είναι και η τυπική περίπτωση στην ανάλυση μιας χρονοσειράς.Το πρόβλημα στην ανάλυση χρονοσειρών είναι να εκτιμήσουμε το σύστημα που παράγει τη χρονοσειρά και ενδεχομένως να κάνουμε προβλέψεις μελλοντικών τιμών του μεγέθους που παρατηρούμε. Η πρώτη υπόθεση που θα πρέπει να απορρίψουμε για να έχει νόημα η ανάλυση της χρονοσειράς είναι ότι η μεταβολή των τιμών του μεγέθους που παρατηρούμε είναι εντελώς τυχαία, δηλαδή το σύστημα που παρατηρούμε είναι λευκός θόρυβος.

Άν οι παρατηρήσεις της χρονοσειράς δεν είναι ανεξάρτητες, η πληροφορία που υπάρχει στη χρονοσειρά μπορεί να δίνεται με διαφορετικές μορφές και τα κυριότερα χαρακτηριστικά που θα πρέπει να μελετήσουμε πριν προχωρήσουμε να προσαρμόσουμε κάποιο μοντέλο στη χρονοσειρά είναι :

1.Στασιμότητα (Stationarity): Απλά αυτό σημαίνει ότι οι διακυμάνσεις των τιμών της χρονοσειράς δε διαφοροποιούνται με το χρόνο. Μια μη στάσιμη χρονοσειρά μπορεί να έχει τάσεις (trends), δηλαδή (αργές) αλλαγές στη μέση τιμή της με το χρόνο, π.χ. η τιμή βενζίνης μπορεί να έχει διακυμάνσεις λόγω της διεθνούς αγοράς αλλά και να παρουσιάζει μια αυξητική τάση σε βάθος χρόνου λόγω πληθωρισμού. Μια μη-στάσιμη χρονοσειρά μπορεί επίσης να παρουσιάζει περιοδικότητα (periodicity), που όταν αναφέρεται σε συγκεκριμένες περιόδους που σχετίζονται με φυσικές εποχές του έτους (μήνα, τρίμηνο, τετράμηνο)
λέγεται και εποχικότητα (seasonality), π.χ. η τιμή του όζοντος στην ατμόσφαιρα υπόκειται σε εποχικές διακυμάνσεις πέρα από τις διακυμάνσεις που μπορεί να οφείλονται στην εξέλιξη του οικοσυστήματος.(Univeristy of Aegean doc)

2. Αιτιοκρατία (determinism) και στοχαστικότητα (stochasticity): Όλες οι χρονοσειρές από πραγματικά μεγέθη περιέχουν θόρυβο και με αυτήν την έννοια όλες οι πραγματικές χρονοσειρές είναι στοχαστικές. Η μεγαλύτερη πρόκληση στην ανάλυση πραγματικών χρονοσειρών είναι η διερεύνηση και ταύτιση ή εντοπισμός του αιτιοκρατικού μέρους του συστήματος που παράγει τη χρονοσειρά. Όταν αυτό είναι κρυμμένο μέσα στο θόρυβο ή γενικότερα δεν κυριαρχεί στην εξέλιξη της χρονοσειράς,τότε θεωρούμε πως το σύστημα είναι στοχαστικό και περιοριζόμαστε σε στατιστική περιγραφή του συστήματος, όπως κάναμε και για τις τυχαίες μεταβλητές στην στατική περίπτωση (ανεξάρτητες μετρήσεις ενός μεγέθους)
Αν για κάποιο λόγο μπορούμε να υποθέσουμε ότι το σύστημα που παράγει τη χρονοσειρά είναι κυρίως αιτιοκρατικό με κάποιες στοχαστικές διαταραχές που όμως δεν κυριαρχούν στην εξέλιξη του συστήματος (και της χρονοσειράς που μελετάμε), τότε μπορούμε να χρησιμοποιήσου με διαφορετικές προσεγγίσεις που είναι κατάλληλες για αιτιοκρατικά
δυναμικά συστήματα, π.χ. ανίχνευση κύριων περιόδων αν το σύστημα φαίνεται να είναι περιοδικό ή διερεύνηση της μη-γραμμικής δυναμικής αν το σύστημα φαίνεται να είναι χαοτικό.

3. Γραμμικότητα (linearity) και μη-γραμμικότητα (nonlinearity): Σύμφωνα με τα παραπάνω φαίνεται αυτές οι δύο έννοιες να σχετίζονται με την αιτιοκρατία και στοχαστικότητα αλλά γενικά μπορούν να ορισθούν ανεξάρτητα από αυτές. Η γραμμικότητα του συστήματος σημαίνει πως οι μεταβλητές του συστήματος (που μπορεί να μην έχουμε τη δυνατότητα να τις παρατηρήσουμε) αλληλο-επιδρούν γραμμικά, δηλαδή αν θα εκφράζαμε το σύστημα με αναλυτική μορφή όλοι οι όροι θα ήταν γραμμικοί ως προς τις μεταβλητές του συστήματος. Σε αντίθετη περίπτωση το σύστημα είναι μη-γραμμικό. Για τη χρονοσειρά αυτό σημαίνει πως για ένα γραμμικό σύστημα ορίζουμε την εξέλιξη της χρονοσειράς ως γραμμικό συνδυασμό των προηγούμενων παρατηρήσεων της χρονοσειράς, ενώ για ένα μη-γραμμικό σύστημα μπορούμε να ορίσουμε την εξέλιξη της χρονοσειράς με μεγαλύτερη ακρίβεια αν θεωρήσουμε και τη συνδυασμένη επίδραση των προηγούμενων παρατηρήσεων σε διαφορετικές χρονικές στιγμές ή τις ίδιες.΄Αρα λοιπόν ένα στοχαστικό σύστημα μπορεί να είναι γραμμικό ή μη γραμμικό και το ίδιο ισχύει για ένα αιτιοκρατικό σύστημα. Βέβαια ένα αιτιοκρατικό γραμμικό σύστημα δεν παρουσιάζει ιδιαίτερο ενδιαφέρον γιατί τα γραμμικά αιτιοκρατικά δυναμικά συστήματα έχουν απλές λύσεις που στην απουσία θορύβου μπορούμε εύκολα να εντοπίσουμε (σταθερό σημείο, περιοδικά σημεία ή τροχιές). Εδώ σημειώνεται ότι κάποια δυσκολία μπορεί να παρουσιαστεί αν το σύστημα είναι πολλών διαστάσεων, υπάρχει κάποια τυχαία διαταραχή και το πλήθος των παρατηρήσεων είναι σχετικά μικρό. Από την άλλη μεριά, είναι ιδιαίτερα δύσκολο να εντοπίσουμε μη-γραμμικότητα σε ένα στοχαστικό σύστημα (ή διαδικασία όπως συνήθως λέγεται) αφού ο θόρυβος στο σύστημα δεν επιτρέπει τον εντοπισμό πολύπλοκων μη-γραμμικών σχέσεων. Σε μια τέτοια περίπτωση θα πρέπει να έχουμε ορίσει μια συγκεκριμένη μη γραμμική μορφή που θέλουμε να διερευνήσουμε. Συνήθως λοιπόν οι δύο κυρίαρχες κλάσεις συστημάτων που υποθέτουμε για στάσιμες χρονοσειρές είναι η γραμμική στοχαστική διαδικασία (linear stochasticprocess) και το μη γραμμικό δυναμικό (πιθανώς χαοτικό) σύστημα .

Στη συνέχεια θα αναλύσουμε τη χρονοσειρά επικεντρώνοντας σε κάθε μια από τις παραπάνω μορφές χρονοσειρών. Γενικά για κάθε χρονική στιγμή t θεωρούμε την τιμή xt της μεταβλητής X που παρατηρούμε. Το σύνολο των παρατηρήσεων xt για κάποια χρονική περίοδο n (σε μονάδες του χρόνου δειγματοληψίας), δηλαδή για χρονικές στιγμές t = 1, . . . , n, αποτελεί τη χρονοσειρά {xt }nt=1 = {x1, . . . , xn}. ΄Οταν έχουμε τη δυνατότητα ταυτόχρονης παρατήρησης πολλών μεγεθών για το ίδιο σύστημα, όπως π.χ. καταγραφές σεισμικών κυμάτων από διαφορετικούς σταθμούς ή καταγραφή θερμοκρασίας και πίεσης, έχουμε πολλαπλές ταυτόχρονες χρονοσειρές ή αλλιώς έχουμε μια πολυ-διάστατη χρονοσειρά (multivariate timeseries). Συνήθως όμως η μελέτη περιορίζεται στη συλλογή τιμών ενός μεγέθους, δηλαδή μιας μονοδιάστατης χρονοσειράς, και εδώ δε θα επεκταθούμε σε πολυ-διάστατες χρονοσειρές.Η μελέτη θα αρχίσει δίνοντας τα πρώτα στάδια της ανάλυσης χρονοσειρών και περιγράφοντας βασικά χαρακτηριστικά,όπως στασιμότητα, τάση, περιοδικότητα και αυτοσυσχέτιση. Στη συνέχεια θα μελετήσουμε γραμμικά μοντέλα για την πρόβλεψη χρονοσειρών και ειδικότερα τη διαδικασία **Box-Jenkins**. Η μελέτη θα ολοκληρωθεί με την παρουσίαση κάποιων μη γραμμικών μεθόδων ανάλυσης χρονοσειρών καθώς και κάποιων απλών μη-γραμμικών μοντέλων πρόβλεψης.

2.1 Βασικά χαρακτηριστικά χρονοσειράς

Θα μελετήσουμε πρώτα κάποια βασικά χαρακτηριστικά των χρονοσειρών μέσα από πραγματικά παραδείγματα.

2.1.2 Στασιμότητα, τάση και περιοδικότητα

Η εμφάνιση τάσης ή περιοδικότητας σε μια χρονοσειρά υποδηλώνει ότι τα στατιστικά χαρακτηριστικά του συστήματος που παράγει η χρονοσειρά αλλάζουν με το χρόνο και η χρονοσειρά δεν είναι στάσιμη. Η αυστηρή στασιμότητα (strict stationarity) ορίζεται μαθηματικά ως η διατήρηση στο χρόνο t της κοινής κατανομής των {xt , xt+1, . . . , xt+τ} για κάποιο αυθαίρετο παράθυρο υστερήσεων τ. Η συνθήκη στασιμότητας περιορίζεται συνήθως στη διατήρηση της μέσης τιμής και αυτοδιασποράς (θα ορίσουμε αυτόν τον όρο στη συνέχεια) και αναφέρεται ως ασθενής στασιμότητα (weak stationarity). Μάλιστα πρακτικά αντί της αυτοδιασποράς εξετάζουμε μόνο τη σταθερότητα της διασποράς με το χρόνο.
Η μη-στασιμότητα αποτελεί σοβαρό πρόβλημα στην ανάλυση χρονοσειρών και ιδιαίτερα όταν προσπαθούμε να κάνουμε προβλέψεις.Σε χρονοσειρές με έντονη περιοδικότητα ή εποχικότητα θα θέλαμε πρώτα να ουδετεροποιήσουμε την επίδραση της περιοδικής ή εποχικής συνιστώσας πριν αναλύσουμε τη χρονοσειρά. Υπάρχουν συγκεκριμένες τεχνικές καθώς και στατιστικοί έλεγχοι για να διερευνήσουμε τη στασιμότητα σε μια χρονοσειρά αλλά δε θα μας απασχολήσουν εδώ. Θα δούμε όμως κάποιους βασικούς μετασχηματισμούς παρακάτω που θα εφαρμόζουμε σε μια φανερά μη-στάσιμη χρονοσειρά για να την κάνουμε στάσιμη και να προχωρήσουμε με την ανάλυση.(Γ.κοκκολάκης,σημειώσεις,Τομέας μαθηματικών,Αναπληρωτής καθηγητής Σχολή εφαρμοσμένων μαθηματικών)

2.2 Απαλοιφή τάσης και περιοδικότητας

Γενικά μια χρονοσειρά μπορούμε να τη χωρίσουμε σε τρεις συνιστώσες ως εξής:

- **xt = μt + st + yt ,**

  όπου μt είναι ή συνιστώσα της τάσης,

  st η συνιστώσα της περιοδικότητας για κάποια περίοδο d (st−d = st)

  yt είναι η χρονοσειρά των υπολοίπων αν αφαιρέσουμε από την παρατηρούμενη χρονοσειρά την τάση και την περιοδικότητα.

Η τάση και η περιοδικότητα είναι και οι δυό συναρτήσεις του χρόνου και δεν περιέχουν πληροφορία για τη δυναμική του συστήματος, δηλαδή την εξάρτηση της παρατήρησης xt από τις προηγούμενες παρατηρήσεις. Σε κάποια προβλήματα όλη η πληροφορία που μας ενδιαφέρει μπορεί να εντοπίζεται στη συνιστώσα της τάσης και της περιοδικότητας, οπότε το πρόβλημα περιορίζεται στην εκτίμηση των μt και st . Σε άλλες περιπτώσεις θέλουμε να εξαλείψουμε την τάση και την περιοδικότητα για να διερευνήσουμε τη δυναμική του συστήματος και θα πρέπει αφού εκτιμήσουμε τις συναρτήσεις μt και st να τις αφαιρέσουμε από της xt για να πάρουμε τη χρονοσειρά των υπολοίπων yt.(Δημήτρης Κουγιουμντζής users.auth.timeseries)Παρακάτω θα τα δούμε όλα σε πραγματικά παραδείγματα.

2.2.1 Λευκός θόρυβος

Θεωρώντας διαδοχικά στοιχεία της χρονοσειράς ως τυχαίες μεταβλητές,η χρονοσειρά λέγεται ότι αποτελείται από **ανεξάρτητες τυχαίες μεταβλητές με ίδια κατανομή** (independent and identically distributed, iid) όταν οι xt , xt+1, . . . , xt+τ τυχαίες μεταβλητές για τ > 1 έχουν την ίδια κατανομή και είναι ανεξάρτητες μεταξύ τους .Μια iid χρονοσειρά είναι εντελώς τυχαία και δεν περιέχει αυτοσυσχετίσεις (γραμμικές ή μη-γραμμικές), δηλαδή συσχετίσεις μεταξύ στοιχείων της χρονοσειράς. Μια iid χρονοσειρά λέγεται και **λευκός θόρυβος** (white noise) και θα συμβολίζουμε την κατανομή της **ως WN(0, σ2 þ),** με μέση τιμή 0 και διασπορά **σ2þ** . Αν επιπλέον τα στοιχεία της χρονοσειράς λευκού θορύβου ακολουθούν κανονική (Γκαουσιανή) κατανομή, τότε η χρονοσειρά λέγεται Γκαουσιανός λευκός θόρυβος **(Gaussian white noise).** (Γ.παύλος ,Βιβλίο 2)

2.2.2 Τυχαίος περίπατος

Ο τυχαίος περίπατος (random walk) είναι μια μη-στάσιμη χρονοσειρά,όπου το κάθε στοιχείο της προκύπτει από το προηγούμενο με την πρόσθεση μιας τυχαίας τιμής, δηλαδή η χρονοσειρά είναι τυχαίος περίπατος αν

xt = xt−1 + þt **,**

όπου {þt } είναι χρονοσειρά λευκού θορύβου, þt ∼ WN(0, σ2þ). Το όνομα υποδηλώνει ακριβώς ότι η χρονοσειρά παράγεται από την κίνηση κάποιου πάνω σε μια ευθεία γραμμή (στο IR), που σε κάθε χρονική στιγμή κάνει ένα τυχαίο βήμα μπρος ή πίσω (þt) από το ημείο που βρίσκεται (xt−1) στο επόμενο (xt).Σημειώνεται ότι αρχίζοντας από κάποια τιμή μ (για t=0) και αντικαθιστώντας επαναληπτικά τον ορισμό (6.3) του τυχαίου περιπάτου για χρόνους ως t ο ορισμός του τυχαίου περιπάτου μπορεί να γραφεί ως

$$x_t = \mu + \Sigma_{tj=1} þ_j,$$

δηλαδή ως άθροισμα όλων των τυχαίων βημάτων ως τη στιγμή t. Οι φαινομενικές συσχετίσεις που φαίνονται από ένα διάγραμμα ιστορίας τυχαίου περιπάτου οφείλονται στο ότι το τυχαίο βήμα σε κάθε χρονική στιγμή, έχει γνωστό σημείο εκκίνησης. Αυτό δημιουργεί στοχαστικές τάσεις για αυτό και η χρονοσειρά του τυχαίου περιπάτου είναι μη-στάσιμη.

Παίρνοντας τις πρώτες διαφορές προκύπτει η στάσιμη χρονοσειρά του λευκού θορύβου.(Πολυτεχνείο Κρήτης,Σημειώσεις στοχαστικών Ανελίξεων)

2.2.4 Παράδειγμα με δεδομένα με σταθερή διακύμανση και χωρις εποχικότητα στο εργαλείο R

Με τον όρο **εποχικότητα** αναφερόμαστε στα εποχικά πρότυπα που παρατηρούνται σε δεδομένα που έχουν ταξινομηθεί ανά τρίμηνο, μήνα ή εβδομάδα. Λαμβάνουν χώρα ή επανακυκλώνονται μέσα στην περίοδο ενός χρόνου, επαναλαμβάνονται από τον ένα χρόνο στον άλλο και είναι προβλέψιμα. Οι εποχικές διακυμάνσεις, επειδή παρουσιάζονται συνήθως με συστηματικό τρόπο, μπορούν σχετικά εύκολα να αναλυθούν και να προσδιοριστούν και κατά συνέπεια να χρησιμοποιηθούν για πρόβλεψη, και που συμβαίνει άλλωστε και με την τάση.

Θα παρουσιάσουμε έναν πίνακα δεδομένων που αναφέρεται στις καθυστερήσεις 42 πτήσεων γνωστής αεροπορικής εταιρείας και θα εξηγήσουμε τι ακριβώς θέλουμε να πούμε με την φράση δεδομένα χωρίς εποχικότητα και σταθερή διακύμανση.(Bureau transotation stat)

Ακολουθεί ο πίνακας με τα δεδομένα:

Πίνακας 2. 1

Πίνακας με καθυστερήσεις 42 πτήσεων

Πτήση	Καθυστέρηση	Πτήση	Καθυστέρηση	Πτήση	Καθυστέρηση	Πτήση	Καθυστέρηση
1	60	11	65	21	16	34	67
2	43	12	34	22	43	35	81
3	67	13	47	23	69	36	67
4	50	14	34	24	59	37	71
5	56	15	49	25	48	38	81
6	42	16	41	26	59	39	68
7	50	17	13	27	86	40	70
8	65	18	35	28	55	41	77
9	68	19	53	29	68	42	56
10	43	20	56	30	51	42	56
				33	67		

Αφού εισάγουμε τα δεδομένα της χρονοσειράς μας στο εργαλείο R, το επόμενο βήμα είναι να αποθηκεύσουμε τα δεδομένα μας σε ένα αντικείμενο της χρονοσειράς, έτσι ώστε να μπορούμε να χρησιμοποιήσουμε πολλές απο τις λειτουργίες του R για την ανάλυση των δεδομένων της χρονοσειράς. Για να επιτύχουμε κάτι τέτοιο χρησιμοποιούμε την συνάρτηση ts(). Στην δική μας περίπτωση για να αποθηκεύσουμε τα δεδομένα της χρονοσειράς στην μεταβλητή data ως αντικείμενο χρονοσειρών σε γλώσσα R , γράφουμε:

Εικόνα 2. 1

Κώδικας σε γλώσσα R

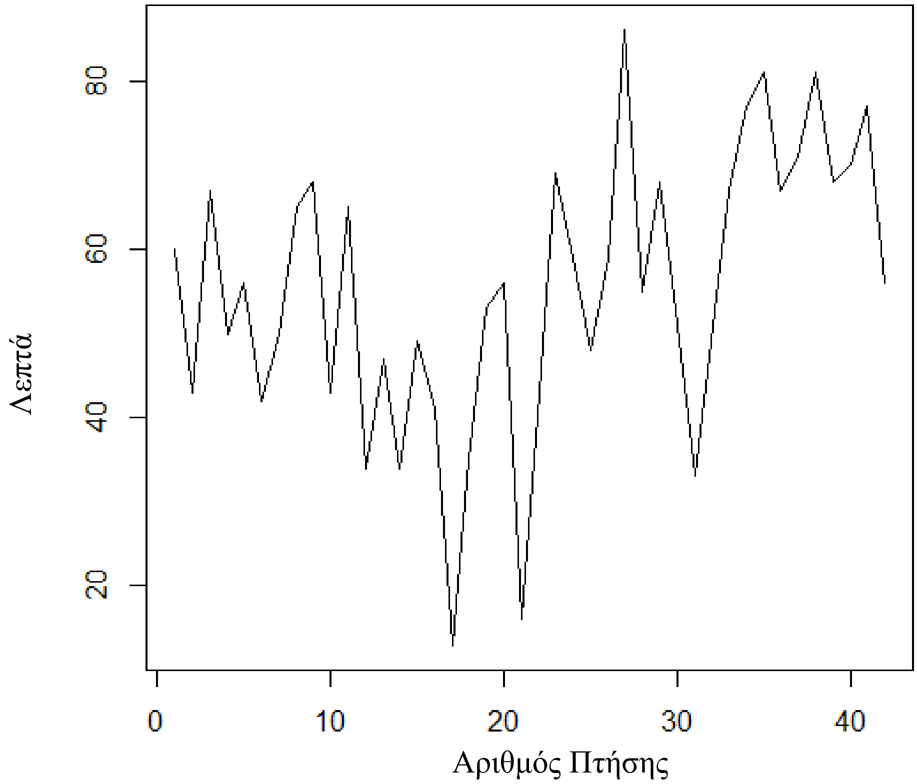
```
> data <- read.csv(C:\\Users\\dimitris\\Desktop\\ptyxiaki\\data.csv”)
Read 42 items
> data
 [1] 60 43 67 50 56 42 50 65 68 43 65 34 47 34 49 41 13 35 53 56 16 43 69 59 48
[26] 59 86 55 68 51 33 49 67 77 81 67 71 81 68 70 77 56
> dataseries<-ts (data)
> dataseries
Time Series:
Start = 1
End = 42
Frequency = 1
 [1] 60 43 67 50 56 42 50 65 68 43 65 34 47 34 49 41 13 35 53 56 16 43 69 59 48
[26] 59 86 55 68 51 33 49 67 77 81 67 71 81 68 70 77 56
```

Μετά την εισαγωγή των δεδομένων της χρονοσειράς ακολουθεί η γραφική απεικόνιση της χρονοσειράς, έτσι ώστε να είμαστε σε θέση να παρατηρήσουμε καλύτερα τα δεδομένα μας.

Εισάγουμε πρώτα την εξής εντολή στο εργαλείο R:

➤ Plot.ts(dataseries)

Έτσι έχουμε την εξής γραφική απεικόνιση:



Διάγραμμα 2. 1

Απεικόνιση με τις καθυστερήσεις 42 πτήσεων

Στο Διάγραμμα 2.1 παρατηρούμε ότι η απεικόνιση των δεδομένων της χρονοσειράς θα μπορούσε να περιγραφεί με την χρήση ενός επιπλέον μοντέλου, αφού οι διακυμάνσεις στο διάγραμμα είναι σχεδόν σταθερές συναρτήση των πτήσεων. Πιο συγκεκριμένα,αυτό που μπορούμε να παρατηρήσουμε με μία πρώτη ματιά είναι ότι οι διακυμάνσεις είναι κοντά ως προς τις μεταβολλές και οι τιμές μεταβάλλονται σχετικά σταθερά

Θα μπορούσαμε να πούμε ότι η χρονοσειρά μας δεν εμφανίζει κάποια τάση. Σε μία πτήση μπορεί να δούμε 10 ή ακόμη και 90 λεπτά καθυστέρηση.

2.3 Παράδειγμα με τάση και εποχικότητα στο εργαλείο R

Μερικές φορές τα δεδομένα των χρονοσειρών μας που έχουμε συλλέξει μπορεί να είναι σε διάστημα μικρότερου του ενός έτους,όπως για παράδειγμα να είναι μηνιαία ή ακόμη και ανα τετράμηνο. Σε αυτή την περίπτωση μπορούμε να συγκεκρινοποιήσουμε το χρονικό διάστημα που συλλέξαμε τα δεδομένα μας χρησιμοποιώντας την παράμετρο της συχνότητας (frequency) στην εντολή ts(). Για δεδομένα που αναφέονται μηνιαιά θέτουμε συχνότητα (frequency)=12.Μπορούμε επίσης να καθορίσουμε την περίοδο αλλά και το πρώτο έτος που συλλέχθηκαν τα δεδομένα χρησιμοποιώντας την παράμετρο start() στην συνάρτηση ts().

Αυτό θα μας βοηθήσει στο δεύτερο παραδειγμά μας όπου έχουμε μια βάση δεδομένων που αναφέρεται σε προιόντα που καθυστέρησαν να παραληφούν απο την Κίνα την περίοδο του πρώτου εξαμήνου του 2001 έως και 2013. Θα πρέπει επίσης να γνωρίζουμε ότι το εργαλείο R μας παρέχει πολλές δυνατότητες όσο αφορά του τι δεδομένα χρονοσειρών μπορούμε να εισάγουμε κάθε φορά.

Μπορούμε λοιπόν να εισάγουμε απο την βάση δεδομένων που έχουμε μόνο τα δεδομένα που αφορούν απο εκείνη την περίοδο πληκτρολογώντας την εντολή: start=c(2001,1).

Συνοψίζοντας λοιπόν όλα τα παραπάνω εισάγουμε τις εντολές στο εργαλείο R:

- products <- read("<- read.csv(C:\\Users\\dimitris\\Desktop\\ptyxiaki\\products.csv ")
- productstimeseries <- ts(products, frequency=12, start=c(2001,1))
- Productstimeseries

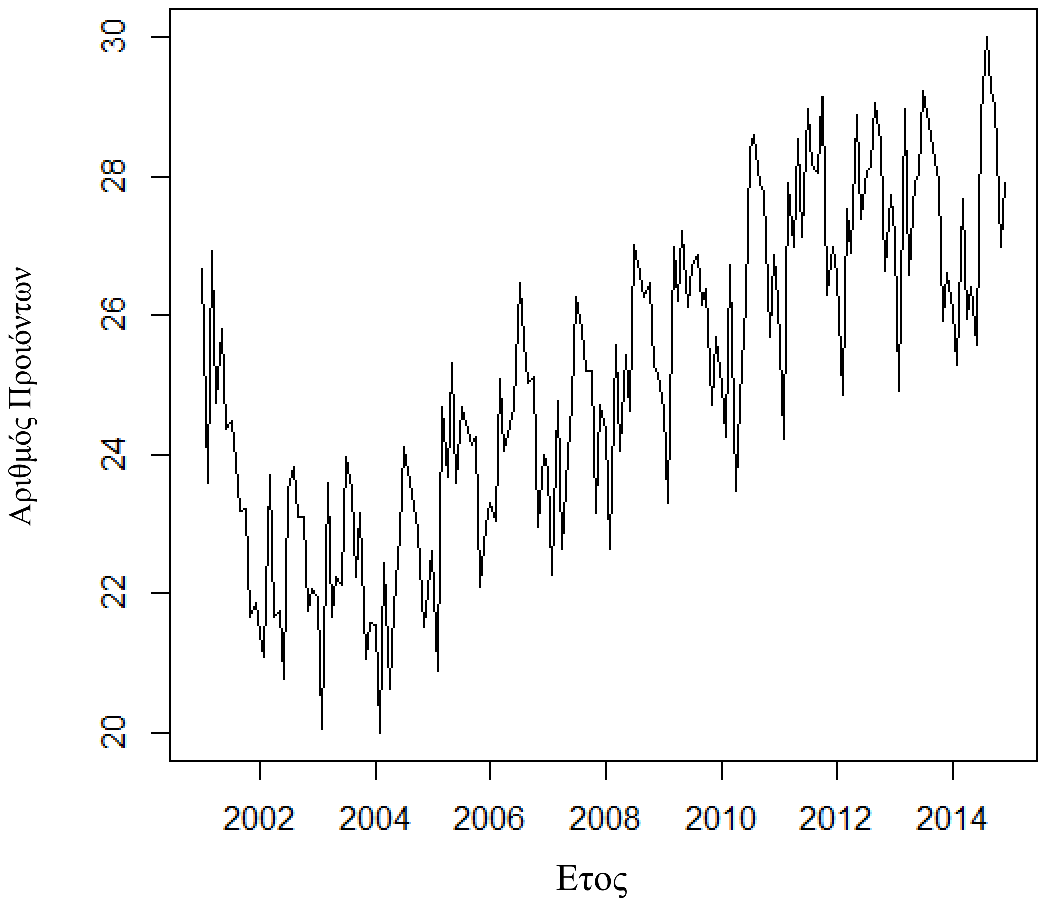
Μετά την εισαγωγή των δεδομένων στο εργαλείο R ακολουθεί ο πίνακας με τα δεδομένα:

Η μορφή των δεδομένων μας έχουν την εξής μορφή:

Πίνακας 2. 2:Προιόντα που δεν αφίχθησαν έγκαιρα στους παραλήπες

Ετος	Ιανουάριος	Φεβρουάριος	Μάρτιος	Απρίλιος	Μάιος	Ιούνιος
2001	26.663	23.598	26.931	24.740	25.806	24.364
2002	21.439	21.089	23.709	21.669	21.752	20.761
2003	21.937	20.035	23.590	21.672	22.222	22.123
2004	22.604	20.894	24.677	23.673	25.320	23.583
2005	23.287	23.049	25.076	24.037	24.430	24.667
2006	23.798	22.270	24.775	22.646	23.988	24.737
2007	24.364	22.644	25.565	24.062	25.431	24.635
2008	24.657	23.304	26.982	26.199	27.210	26.122
2009	24.990	24.239	26.721	23.475	24.767	26.219
2010	26.217	24.218	27.914	26.975	28.527	27.139
2011	26.589	24.848	27.543	26.896	28.878	27.390
2012	27.132	24.924	28.963	26.589	27.931	28.009
2013	26.076	25.286	27.660	25.951	26.398	25.565
	Ιούλιος	**Αύγουστος**	**Σεπτέμβριος**	**Οκτόβριος**	**Νοέμβριος**	**Δεκέμβριος**
2001	24.477	23.901	23.175	23.227	21.672	21.870
2002	23.479	23.824	23.105	23.110	21.759	22.073
2003	23.950	23.504	22.238	23.142	21.059	21.573
2004	24.671	24.454	24.122	24.252	22.084	22.991
2005	26.451	25.618	25.014	25.110	22.964	23.981
2006	26.276	25.816	25.210	25.199	23.162	24.707
2007	27.009	26.606	26.268	26.462	25.246	25.180
2008	26.706	26.878	26.152	26.379	24.712	25.688
2009	28.361	28.599	27.914	27.784	25.693	26.881
2010	28.982	28.169	28.056	29.136	26.291	26.987
2011	28.065	28.141	29.048	28.484	26.634	27.735
2012	29.229	28.759	28.405	27.945	25.912	26.619
2013	28.865	30.000	29.261	29.012	26.992	27.897

Επειτα ακολουθεί η γραφική απεικόνιση:



Διάγραμμα 2. 2:Διαγραμματική απεικόνιση προιόντων που αφίχθησαν αργοπορημένα

Απο τα δεδομένα της χρονοσειράς του δεύτερου παραδείγματος και σε συνδυασμό με το Διάγραμμα 2.2 ,παρατηρούμε ότι υπάρχει τους καλοκαιρινούς μήνες μία ανοδική τάση καθυστερήσεων στα προιόντα ,σε αντίθεση με τους χειμερινούς μήνες που υπάρχει μία αισθητή πτώση.

Όπως και στο προηγούμενο παράδειγμα θα μπορούσαμε να πούμε ότι μπορεί η χρονοσειρά μας να περιγραφεί απο ένα πρόσθετο μοντέλο,καθώς οι **εποχιακές** διακυμάνσεις είναι σχετικά σταθερές συναρτήση του χρόνου και δεν φαίνονται να εξαρτώνται απο το επίπεδο της χρονοσειράς.

2.3.1 Παράδειγμα χωρίς τάση και χωρίς εποχικότητα

Παρομοίως για το τρίτο παραδειγμά μας έχουμε μία χρονοσειρά η οποία αναφέρεται σε δρομολόγια φορτηγών πλοίων. Το συγκεκριμένο παράδειγμα είναι μία ιδιαίτερη περίπτωση ,καθώς υπάρχουν πολλοί παράγοντες που είναι απρόβλεπτοι ,οι οποίοι μπορούν να προκαλέσουν μεγάλες καθυστερήσεις σε μία μεταφορά προιόντων,όπως είναι ο καιρός ,οι καθυστερήσεις σε ξένα λιμάνια, κρίσεις που ξεσπούν σε ξένες χώρες κ.α

Γιαυτό τον λόγο στο παρακάτω παράδειγμα έχουμε καθυστερήσεις σε μεταφορές προιόντων εκφρασμένες σε ώρες .Τα δεδομένα μας αναφέρονται στην περίοδο 2003 έως 2009 και έχουν την εξής μορφή:

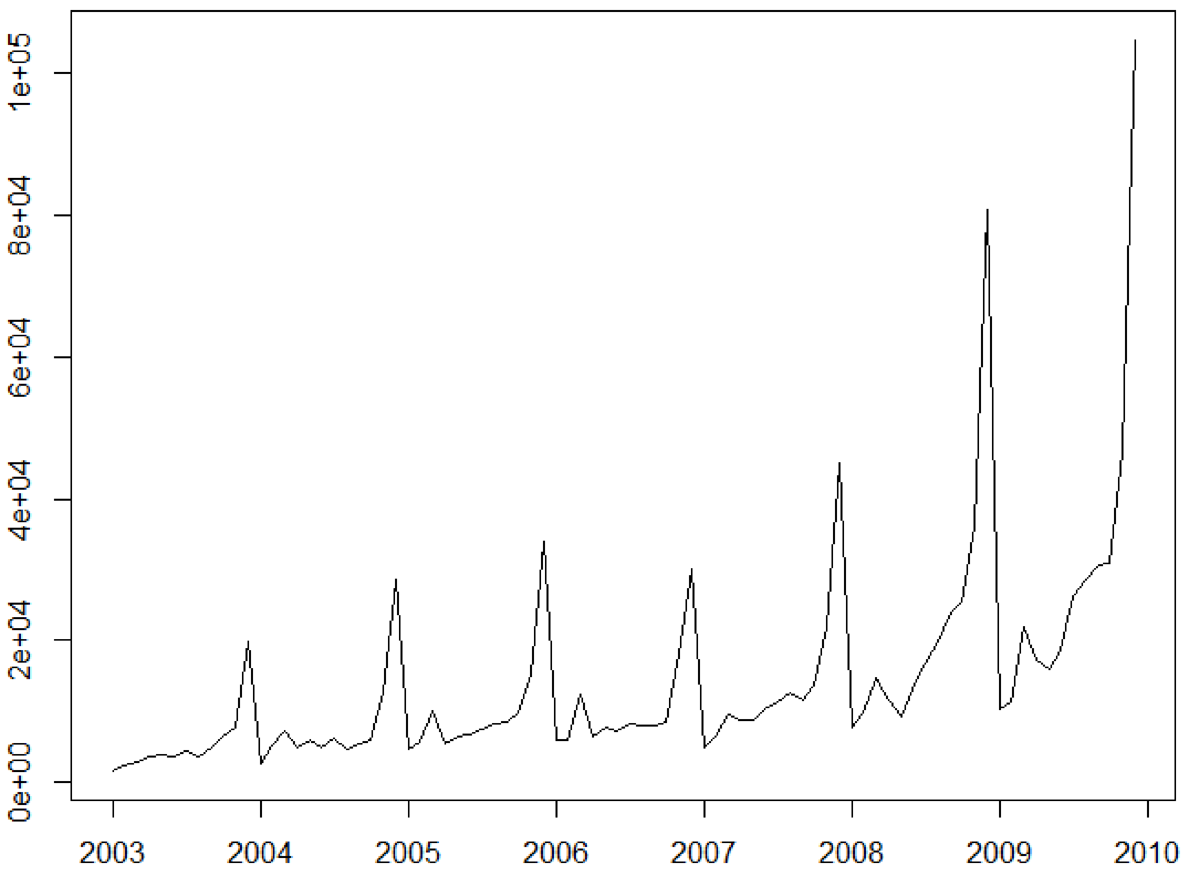
Πίνακας 2. 3:Καθυστερήσεις σε μεταφορά εμπορευμάτων με φορτηγά πλοία

Έτος	Ιανουάριος	Φεβρουάριος	Μάρτιος	Απρίλιος	Μάιος	
2003	1664.81	2397.53	2840.71	3547.29	3752.96	
2004	2499.81	5198.24	7225.14	4806.03	5900.88	
2005	4717.02	5702.63	9957.58	5304.78	6492.43	
2006	5921.10	5814.58	12421.25	6369.77	7609.12	
2007	4826.64	6470.23	9638.77	8821.17	8722.37	
2008	7615.03	9849.69	14558.40	11587.33	9332.56	
2009	10243.24	11266.88	21826.84	17357.33	15997.79	
Ιούνιος	**Ιούλιος**	**Αύγουστος**	**Σεπτέμβριος**	**Οκτώβριος**	**Νοέμβριος**	**Δεκέμβριος**
3714.74	4349.61	3566.34	5021.82	6423.48	7600.60	19756.21
4951.34	6179.12	4752.15	5496.43	5835.10	12600.08	28541.72
6630.80	7349.62	8176.62	8573.17	9690.50	15151.84	34061.01
7224.75	8121.22	7979.25	8093.06	8476.70	17914.66	30114.41
10209.48	11276.55	12552.22	11637.39	13606.89	21822.11	45060.69
13082.09	16732.78	19888.61	23933.38	25391.35	36024.80	80721.71
18601.53	26155.15	28586.52	30505.41	30821.33	46634.38	104660.67

Ακολουθούν οι αντίστοιχες εντολές στο εργαλείο R και η απεικόνιση μέσω του εργαλείου R:

Εικόνα 2. 2 Κώδικας σε γλώσσα R για την απεικόνιση της χρονοσειράς

```
> ship <- read.csv(C:\\Users\\dimitris\\Desktop\\ptyxiaki\\ship.csv «)»)

Read 84 items

> shiptimeseries <- ts(ship, frequency=12, start=c(2003,1))

> shiptimeseries
```



Διάγραμμα 2. 3 Διαγραμματική απεικόνιση προιόντων που καθυστέρησαν κατα την μεταφορά τους με φορτηγά πλοία

Στην συγκεκριμένη περίπτωση παρατηρούμε απο την απεικόνιση ότι δεν μπορεί η χρονοσειρά μας να περιγραφεί με την χρήση ενός πρόσθετου μοντέλου,αφού το μέγεθος των εποχικών διακυμάνσεων φαίνεται να αυξάνεται στην πάροδο του χρόνου.

.

2.4 Αποσύνθεση Χρονοσειρών

Οι χρονοσειρές αναλύονται για να κατανοήσουμε το παρελθόν και να προβλέψουμε το μέλλον, επιτρέποντας στους διαχειριστές ή τους φορείς χάραξης πολιτικής να κάνουν σωστά ενημερωμένες αποφάσεις. Μια ανάλυση χρονοσειρών ποσοτικοποιεί τα κύρια χαρακτηριστικά των δεδομένων και την τυχαία παραλλαγή.

Οι λόγοι αυτοί, σε συνδυασμό με τη βελτιωμένη υπολογιστική ισχύ, μας επιτρέπουν να εφαρμόζουμε τις χρονοσειρές ευρέως απο την βιομηχανία και τα logistics μέχρι και πολλές άλλες εφαρμογές

2.4.1 Αποσύνθεση δεδομένων χωρίς εποχικότητα

Μια χρονοσειρά της οποίας τα δεδομένα δεν παρουσιάζουν καμία εποχικότητα ,συνήθως αποτελούνται απο μία συνιστώσα τάσης και δεδομένα τα οποία είναι ακανόνιστα. Η διαδικασία αποσύνθεσης περιέχει την προσπάθεια να διαχωρίσουμε την χρονοσειρά μας σε αυτά τα στοιχεία,δηλαδή την εκτίμηση των στοιχείων που παρουσιάζουν κάποια τάση στα δεδομένα μας και τα ακανόνιστα στοιχεία

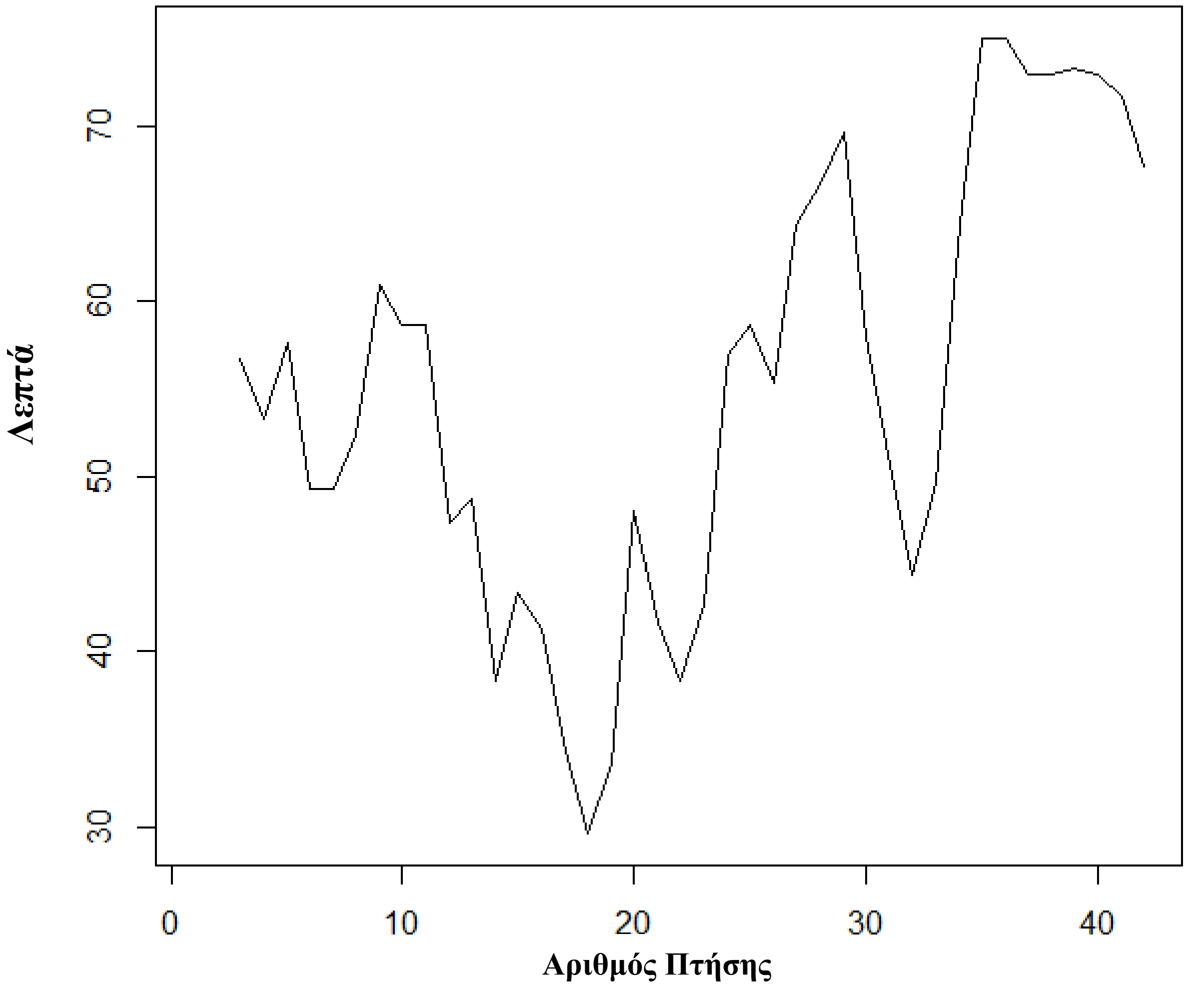
Για αυτό τον λόγο χρησιμοποιούμε μία μέθοδο εξομάλυνσης των δεδομένων και της γραφικής μας απεικόνισης. Μία τέτοια μέθοδος είναι να χρησιμοποιήσουμε τον **‘’κινητό μέσο όρο´´** των μεταβλητών των δεδομένων μας.

Η λειτουργεία της εντολής SMA() στην βιβλιοθήκη TTR του εργαλείου R χρησιμοποιείται για να εξομαλύνουμε την χρονοσειρά μας.

Για να χρησιμοποιήσουμε την εντολή SMA() πρέπει να προσδιορίσουμε τον τον ’κινητό μέσο όρο´´ χρησιμοποιώντας την μεταβλητή n.Η μεταβλητη n στην πραγματικότητα αυτό που εφαρμόζει είναι να ´´ανοίγει´´ στην ουσιά την γραφική μας παράσταση ‘η ακόμη καλύτερα θα μπορούσαμε να πούμε ότι απλοποιεί την γραφική μας απεικόνιση. Για παράδειγμα για να υπολογίσουμε έναν ‘’κινητό μέσο όρο´´ με άνοιγμα 5 θέτουμε n=5.

Στο πρώτο παράδειγμά μας που αναφέρεται στις καθυστερήσεις πτήσεων σε διάστημα 42 πτήσεων, όπως αναφέραμε τα δεδομένα μας είναι μη εποχιακά. Ετσι θα προσπαθήσουμε να εξομαλύνουμε την χρονοσειρά μας ,δηλαδή τα δεδομένα μας με την μέθοδο που περιγράψαμε.Θα προσπαθήσουμε δηλαδή να εξομαλύνουμε την φραφική απεικόνιση της χρονοσειράς μας με την τεχνική του τεχνικού μέσου όρου.Ακολουθεί ο κώδικας που εισάγουμε το εργαλείο R για να προκύψει η γραφική μας απεικόνιση.

- datatimeseriesSMA3 <- SMA(datatimeseries,n=3)
- plot.ts(datatimeseriesSMA3)

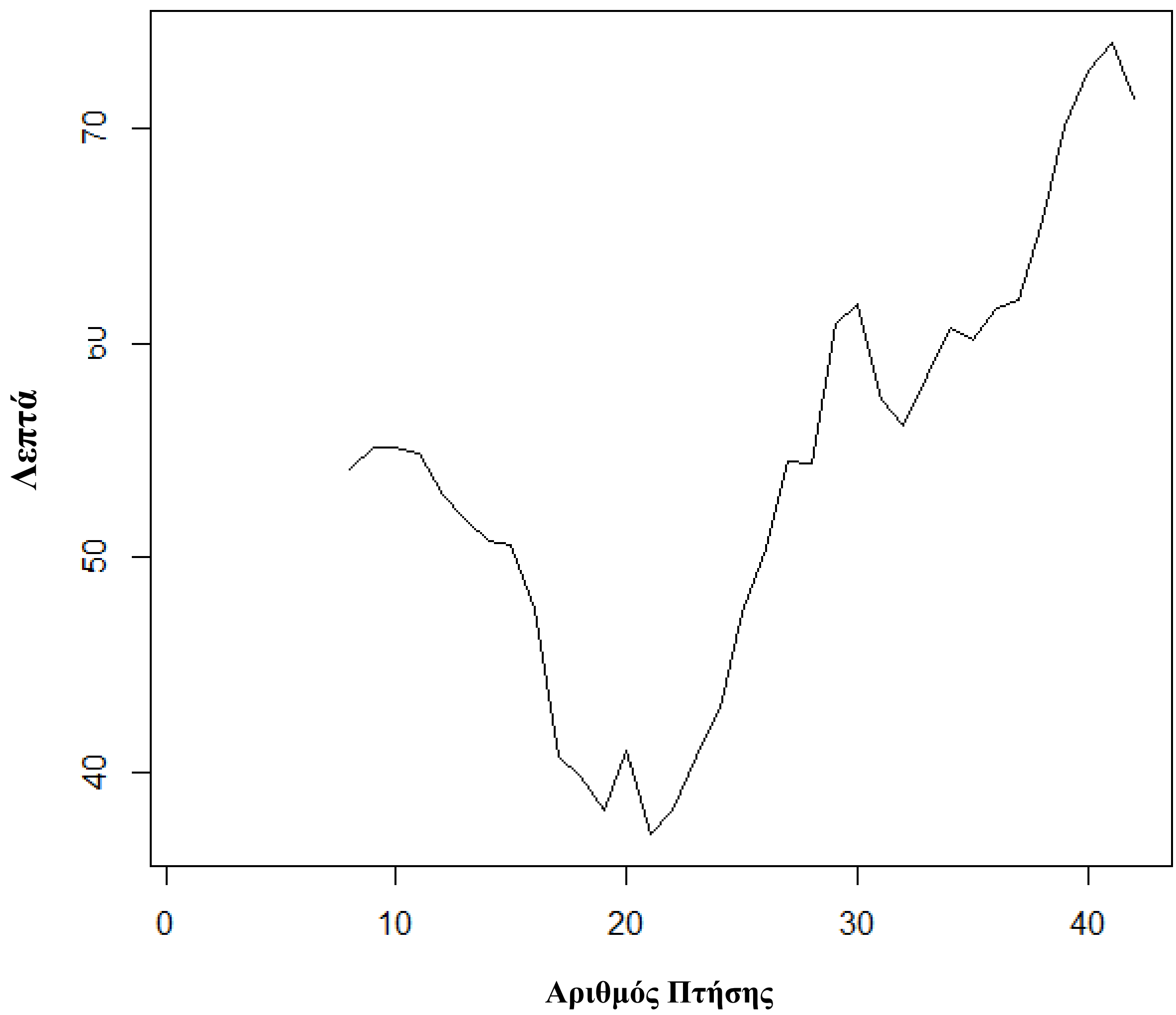


Διάγραμμα 2. 4 Διαγραμματική απεικόνιση με καθυστερήσεις 42 πτήσεων

Απο το αποτέλεσμα της απεικόνισης παρατηρούμε ότι υπάρχει αισθητή βελτίωση σε σχέση με την αρχική μας απεικόνιση. Ωστόσο δεν θα μπορούσαμε να πούμε ότι βρισκόμαστε στο επιθυμητο αποτέλεσμα.Υπενθυμίζουμε ότι θέλουμε να φτάσουμε σε μία μορφή η οποία να εμφανίζει πιο σταθερές διακυμάνσεις. Γιαυτό τον λόγο θα δοκιμάσουμε εκ νέου μια απεικόνιση ,αλλα αυτή την φορά με άνοιγμα n=8.

Επομένως ακολούθεί μια ένα νεό διάγραμμα παρακάτω με τον αντίστοιχο νέο κώδικά:

```
datatimeseriesSMA8 <- SMA(datatimeseries,n=8)
plot.ts(datatimeseriesSMA8)
```



Διάγραμμα 2. 5 Διαγραμματική απεικόνιση με καθυστερήσεις 42 πτήσεων

Αυτή την φορά έχουμε μία ξεκάθαρη εικόνα απο την εξομάλυνση που κάναμε. Παρατηρούμε πλέον ότι υπάρχει μία τάση στην γραφική μας απεικόνιση.Πιο συγκεκριμένα φαίνεται ότι στις πρώτες 20 πτήσεις η καθυστέρηση μειώνεται απο τα 50 λεπτά στα 38 και έπειτα αυξάνεται απο τα 40 λεπτά στα 73.

2.4.2 Αποσύνθεση δεδομένων με εποχικότητα

Μια εποχιακή χρονοσειρά αποτελείται από μια συνιστώσα τάσης, μία εποχική συνιστώσα και ένα ακανόνιστο στοιχείο. Η διακικασία της αποσύνθεσης σε αυτές τις χρονοσειρές περιλαμβάνει τον διαχωρίσμό των χρονοσειρών σε αυτές τις τρεις συνιστώσες: δηλαδή, την εκτίμηση αυτών των τρίων στοιχείων.

Για να εκτιμήσουμε την συνιστώσα τάσης και την εποχική συνιστώσα μιας χρονοσειράς που μπορεί να περιγραφεί απο ένα πρόσθετο μοντέλο,μπορούμε να χρησιμοποιήσουμε την λειτουργία της εντολής decompose() η οποία επιστρέφει ως αποτέλεσμα μια λίστα αντικειμένων ,όπου οι εκτιμήσεις των εποχικών δεδομένων ,της συνιστώσας τάσης και του ακανόνιστου στοιχείου αποθηκεύονται σε στοιχεία του εν λόγω καταλόγου αντικειμένων με την ονομασία ´´εποχικά´´.´´τάσης´´ και ´´τυχαίων´´ αντίστοιχα.

Πιο συγκεκριμένα, όπως αναφερθήκαμε παραπάνω, στο **παράδειγμα 2** που αναφέρεται στις καθυστερημένες παραλαβές προιόντων, υπάρχει μία εποχική τάση με την εμφάνιση των μεγαλύτερων καθυστερήσεων να γίνονται την περίοδο του καλοκαιριού και την μικρότερη τους χειμερινούς μήνες.

Επομένως αυτό που θα επιχειρήσουμε παρακάτω είναι ο διαχωρισμός των στοιχείων που αναφέραμε μέσω της γραφικής απεικόνισης στο δικό μας παράδειγμα.Ακολουθεί παρακάτω ο κώδικας σε γλώσσα R και η αντίστοιχη απεικόνιση.

Εικόνα 2. 3 Κώδικας σε γλώσσα R για την απεικόνιση εποχικότητας

```
products <- read(“<- read.csv(C: \\Users\\dimitris\\Desktop\\ptyxiaki\\products.csv “)

Read 168 items

> productstimeseries <- ts(products, frequency=12, start=c(2001,1))

> productstimeseries

>productstimeseriescomponents <- decompose(productstimeseries)

> birthstimeseriescomponents$seasonal

>plot(producttimeseriescomponents)
```

Με τις εντολές που εισάγαμε στο εργαλείο R η εκτίμηση των τιμών της εποχικότητας, τάσης και των ακανόνιστων στοιχείων, είναι πλέον αποθηκευμένες σε μεταβλητές με την ονομασία **productstimeseriescomponents$seasonal,productstimeseriescomponents$trend and productstimeseriescomponents$random**.

Πιο αναλυτικά, μπορούμε να εμφανίσουμε τις παραπάνω τιμές:

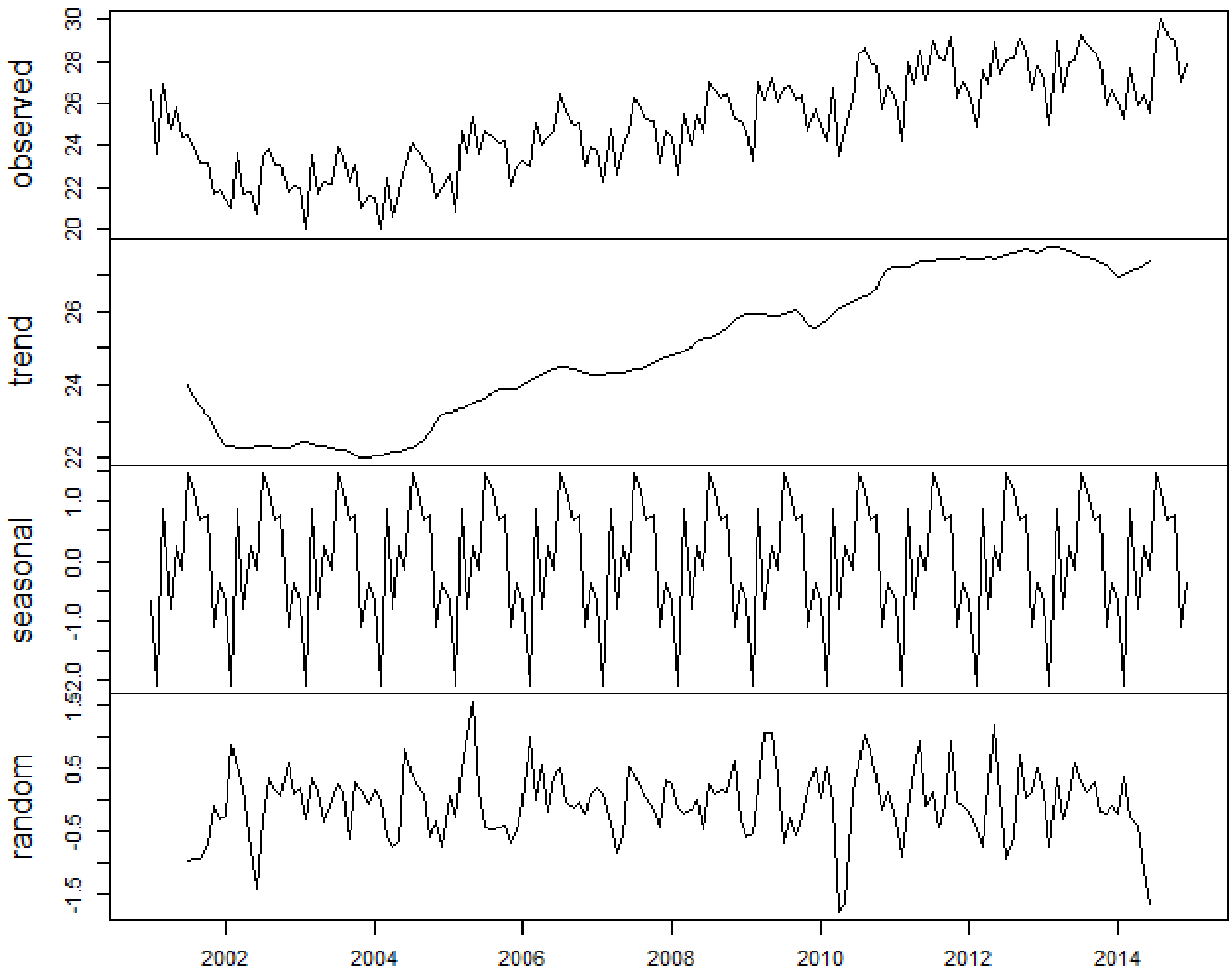
Εικόνα 2. 4 Εκτίμηση των τιμών της εποχικότητας,τάσης και των ακανόνιστών στοιχείων

	Jan	Feb	Mar	Apr	May	Jun
2001	-0.6771947	-2.0829607	0.8625232	-0.8016787	0.2516514	-0.1532556
2002	-0.6771947	-2.0829607	0.8625232	-0.8016787	0.2516514	-0.1532556
2003	-0.6771947	-2.0829607	0.8625232	-0.8016787	0.2516514	-0.1532556
2004	-0.6771947	-2.0829607	0.8625232	-0.8016787	0.2516514	-0.1532556
2005	-0.6771947	-2.0829607	0.8625232	-0.8016787	0.2516514	-0.1532556
2006	-0.6771947	-2.0829607	0.8625232	-0.8016787	0.2516514	-0.1532556
2007	-0.6771947	-2.0829607	0.8625232	-0.8016787	0.2516514	-0.1532556
2008	-0.6771947	-2.0829607	0.8625232	-0.8016787	0.2516514	-0.1532556
2009	-0.6771947	-2.0829607	0.8625232	-0.8016787	0.2516514	-0.1532556
2010	-0.6771947	-2.0829607	0.8625232	-0.8016787	0.2516514	-0.1532556
2011	-0.6771947	-2.0829607	0.8625232	-0.8016787	0.2516514	-0.1532556
2012	-0.6771947	-2.0829607	0.8625232	-0.8016787	0.2516514	-0.1532556
2013	-0.6771947	-2.0829607	0.8625232	-0.8016787	0.2516514	-0.1532556
2014	-0.6771947	-2.0829607	0.8625232	-0.8016787	0.2516514	-0.1532556

	Jul	Aug	Sep	Oct	Nov	Dec
2001	1.4560457	1.1645938	0.6916162	0.7752444	-1.1097652	-0.3768197
2002	1.4560457	1.1645938	0.6916162	0.7752444	-1.1097652	-0.3768197
2003	1.4560457	1.1645938	0.6916162	0.7752444	-1.1097652	-0.3768197
2004	1.4560457	1.1645938	0.6916162	0.7752444	-1.1097652	-0.3768197
2005	1.4560457	1.1645938	0.6916162	0.7752444	-1.1097652	-0.3768197
2006	1.4560457	1.1645938	0.6916162	0.7752444	-1.1097652	-0.3768197
2007	1.4560457	1.1645938	0.6916162	0.7752444	-1.1097652	-0.3768197
2008	1.4560457	1.1645938	0.6916162	0.7752444	-1.1097652	-0.3768197
2009	1.4560457	1.1645938	0.6916162	0.7752444	-1.1097652	-0.3768197
2010	1.4560457	1.1645938	0.6916162	0.7752444	-1.1097652	-0.3768197
2011	1.4560457	1.1645938	0.6916162	0.7752444	-1.1097652	-0.3768197
2012	1.4560457	1.1645938	0.6916162	0.7752444	-1.1097652	-0.3768197
2013	1.4560457	1.1645938	0.6916162	0.7752444	-1.1097652	-0.3768197

Οι εκτιμήσεις των εποχικών παραγόντων που δόθηκαν παραπάνω είναι απο τον Ιανουάριο έως τον Δεκέμβριο και είναι ίδιες για κάθε έτος. Ο μεγαλύτερος εποχικός παράγοντας είναι τον μήνα Ιούλιο με τιμή περίπου 1,46 και ο χαμηλότερος είναι τον μήνα Φεβρουάριο με τιμή περίπου -2,08, υποδεικνύοντας έτσι ότι φαίνεται να υπάρχει μία μέγιστη τιμή των προιόντων που καθυστέρησαν τον μήνα Ιούλιο και μία ελάχιστη τον μήνα Φεβρουάριο.

Ακολουθεί η γραφική απεικόνιση:



Διάγραμμα 2. 6 Διαγραμματική απεικόνιση προιόντων πρόσθετης χρονοσειράς

Η παραπάνω απεικόνιση στο πρώτο κομμάτι μας δείχνει την αρχική χρονοσειρά πριν περιγραφεί με το πρόσθετο μοντέλο.Στο δεύτερο κομμάτι μας δείχνει την εκτιμώμενη συνιστώσα τάσης.Στο τρίτο κομμάτι έχουμε την εκτίμηση των ακανόνιστων στοιχείων.

Επίσης παρατηρούμε ότι η εκτίμηση της συνιστώσας τάσης φαίνεται να έχει μία μείωση απο τα 24 το 2001 στα 22 το 2002,με μία μεγάλη άνοδο στα 27 μέχρι το 2013.

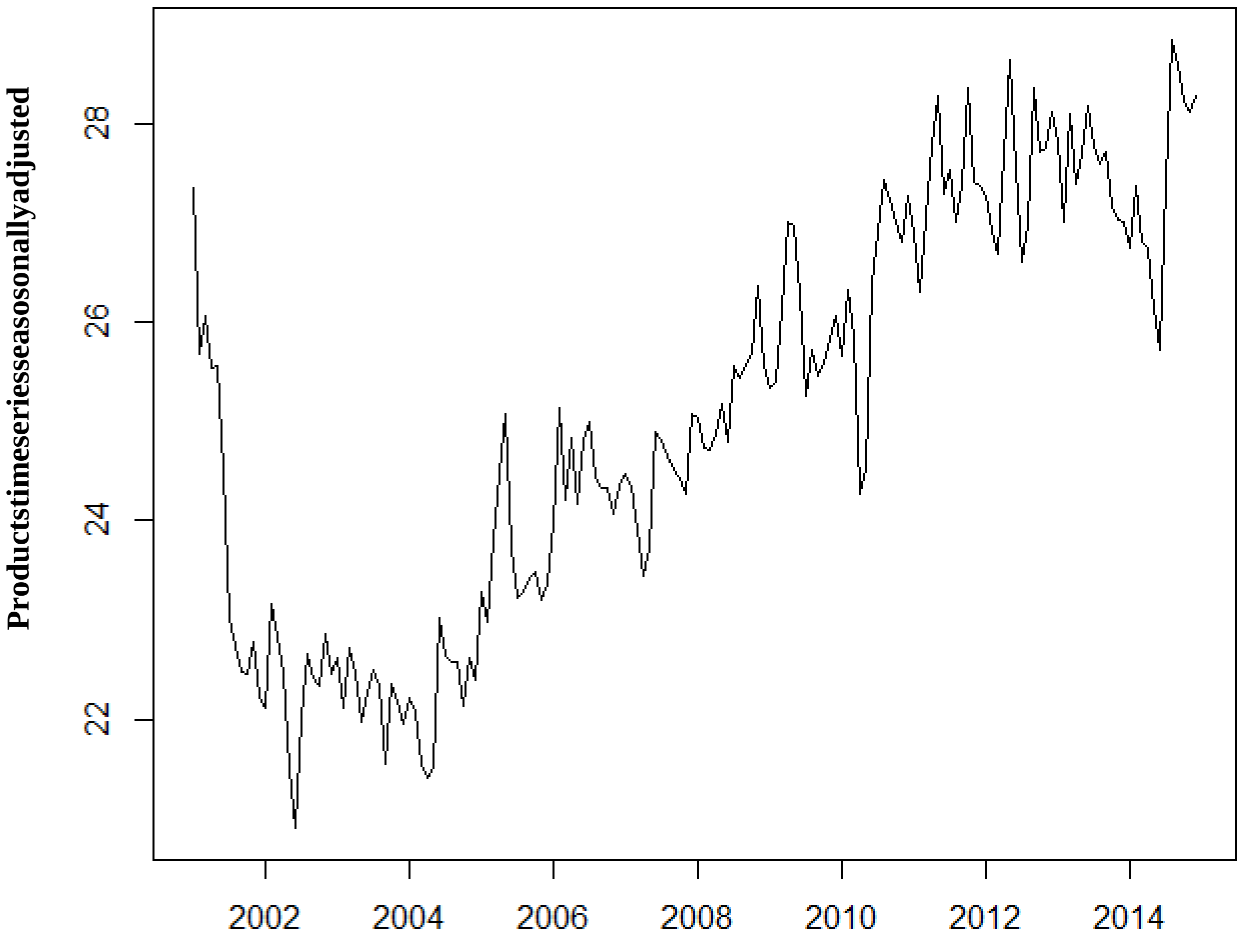
2.5 Τροποποίηση Εποχικότητας

Όταν έχουμε εποχικές χρονοσειρές οι οποίες μπορούν να περιγραφούν, απο ένα πρόσθετο μοντέλο, τότε μπορούμε να προσαρμόσουμε εποχικά την χρονοσειρά αυτη εκτιμώντας το εποχιακό στοιχείο και αφαιρώντας το απο την αρχική μορφή της χρονοσειράς.Με την βοήθεια του εργαλείου R αυτό επιτυγχάνεται εύκολα με την εντολή decompose().

Ακολουθεί ο κώδικας σε γλώσσα R και η αντίστοιχη γραφική απεικόνιση.

Εικόνα 2. 5 Κώδικας για αποσύνθεση χρονοσειράς

```
>productstimeseriescomponents <- decompose(productstimeseries)

>productstimeseriesseasonallyadjusted <- productstimeseries

 >productsstimeseriescomponents$seasonal

>plot(productstimeseriesseasonallyadjusted)
```



Διάγραμμα 2. 7Εποχική προσαρμογή χρονοσειράς

Μπορούμε να δούμε απο την απεικόνιση ότι η εποχική διακύμανση έχει αφαιρεθεί απο την τροποποιημένη εποχικά διακύμανση.Η τροποποιημένη εποχικά χρονοσειρά περιλαμβάνει πλέον μόνο την **συνιστώσα τάσης** και το **ακανόνιστο στοιχείο**.

Κεφάλαιο 3

Προβλέψεις

3.1 Απλή Εκθετική Εξομάλυνσης σε δεδομένα με τάση και χωρίς εποχικότητα

Στην περίπτωση που έχουμε μία χρονοσειρά η οποία δεν παρουσιάζει κάποια εποχικότητα στα δεδομένα και χαρακτηρίζεται απο ένα σχετικά σταθερό επίπεδο διακυμάνσεων, τότε μπορούμε να εφαρμόσουμε την χρήση της εκθετικής εξομάλυνσης για να έχουμε βραχυπρόθεσμες **προβλέψεις**.

Η μέθοδος της απλής εκθετικής εξομάλυνσης παρέχει έναν τρόπο για την εκτίμηση του επιπέδου κατα την τρέχουσα χρονική στιγμή.Η εξομάλυνση ελέγχεται απο την παράμετρο alpha για την εκτίμηση του επιπέδου κατα την τρέχουσα χρονική στιγμή.Η τιμή της παραμέτρου **alpha** κυμαίνεται μεταξύ των τιμών 0 και 1.Οι τιμές που είναι κοντά στο 0 σημαίνει ότι δίνεται μικρή βαρύτητα στις πιο πρόσφατες παρατηρήσεις-δεδομένα κατα την διαδικασία της πρόβλεψης.

Στο πρώτο παράδειγμα που αναφέρονται στις καθυστερήσεις των πτήσεων όπως είχαμε αναφέρει οι διακυμάνσεις είναι σχεδόν σταθερές ,επομένως μπορούμε να κάνουμε χρήση ενός πρόσθετου μοντέλου .Έτσι μπορούμε να εφαρμόσουμε την μέθοδο της εκθετικής εξομάλυσης.

Ωστόσο θα εφαρμόσουμε πρόβλεψη με άλλη μέθοδο όπως θα δουμε παρακάτω αν και μπορούμε και με την απλή εκθετική εξομάλυνση.Θα χρησιμοποιήσουμε ένα παρόμοιο παράδειγμα 100 πτήσεων μόνο που αυτή την φορά οι διακυμάνσεις είναι λίγο πιο ομαλές,αυτό θα έχει σαν αποτέλσμα την ευκολότερη κατανόηση της μεθόδου.

Για να προβούμε σε προβλέψεις με την χρήση της εκθετικής εξομάλυνσης στο εργαλείο R, μπορούμε να προσαρμόσουμε ένα εκθετικό μοντέλο εξομάλυνσης χρησιμοποιώντας την λειτουργία της εντολής **HoltWinters().**

Στην απλή εκθετική εξομάλυνση θέτουμε τις παραμέτρους **beta, gamma** ίσο με FALSE.Παρακάτω θα εξηγήσουμε αναλυτικά τον λόγο για τον οποίο οι παράμετροι αυτοί χρησιμοποιούνται μόνο στην περίπτωση της HoltWinters εκθετικής εξομάλυνσης και όχι στην απλή εκεθτική εξομάλυνση.(RobHyndman,forecasting principles and practice)

3.1.1 Παράδειγμα πρόβλεψης με απλή εκθετική εξομάλυνση

Όπως αναφέραμε θα εισάγουμε στο εργαλείο R μια βάση δεδομένων που αναφέρεται σε 100 πτήσεις και τις αντίστοιχες καθυστερήσεις τους.

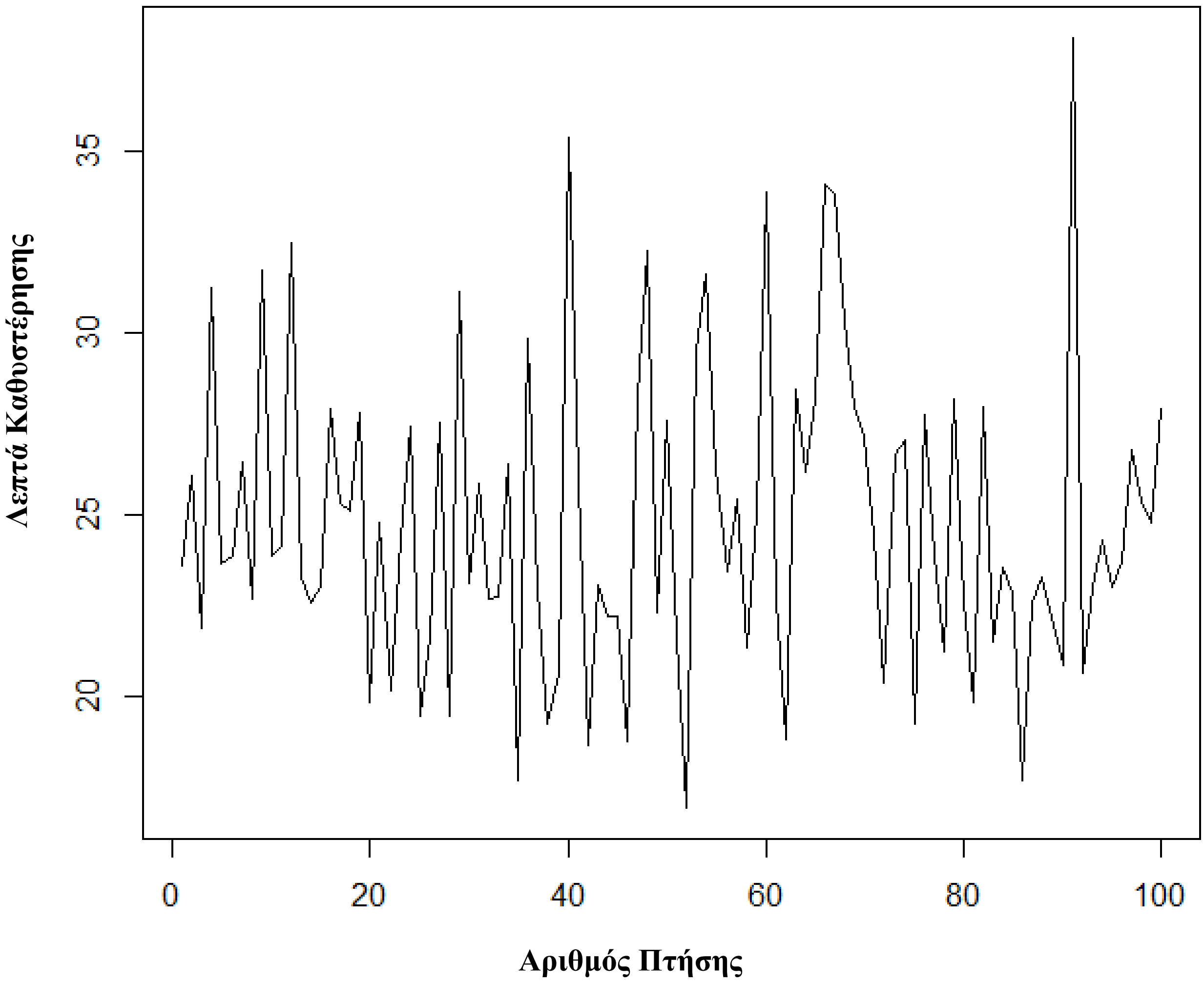
Η εντολή HoltWinters επιστρέφει μία λίστα μεταβλητών η οποία περέχει αρκετά στοιχεία με όνοματα.Παρακάτω ακολουθεί ο πίνακας ο κώδικας σε γλώσσα R και η γραφική απεικόνιση της πρόβλεψης.

Εικόνα 3. 1 Κώδικας σε γλώσσα R για χρήση της εντολής HoltWinters

```
> data <- read.csv(C:\\Users\\dimitris\\Desktop\\ptyxiaki\\data.csv")

Read 100 items

Start = 1

End = 100

Frequency = 1

> data

>dataseries <- ts(data,start=c(1))

>plot.ts(dataseries)

>rainseriesforecasts <- HoltWinters(rainseries, beta=FALSE, gamma=FALSE)
> rainseriesforecasts
Smoothing parameters: alpha: 0.02412151
 beta : FALSE
```

Εικόνα 3. 2 Συνέχεια Κώδικας σε γλώσσα R για χρήση της εντολής HoltWinters

```
gamma: FALSE
Coefficients: [,1] a 24.67819
> datseriesforecasts$fitted
Time Series:
 Start = 1
 End = 100
Frequency = 1
> plot(dataseriesforecasts)
> dataseriesforecasts$SSE
 [1] 1828.855
> HoltWinters(dataseries, beta=FALSE, gamma=FALSE, l.start=23.56)
> HoltWinters(rainseries, beta=FALSE, gamma=FALSE, l.start=23.56)
> library("forecast")
> rainseriesforecasts2 <- forecast.HoltWinters(rainseriesforecasts, h=8) > rainseriesforecasts2
```



Διάγραμμα 3. 1 Απεικόνιση με καθυστερήσεις 100 πτήσεων

Μπορούμε να καταλάβουμε απο την απεικόνιση ότι υπάρχει ένα σταθερό επίπεδο γύρω απο το οποίο κυμαίνονται οι διακυμάνσεις, στα 25 λεπτά.Οι τυχαίες διακυμάνσεις της χρονοσειράς τείνουν να είναι σταθερές συναρτήση των πτήσεων ,επομένως τηρούμε ολές τις προυποθέσεις που αναφέραμε πρίν για την χρήση της εκθέτικής εξομάλυνσης.

Η έξοδος της εντολής HoltWinters στο εργαλείο R μας δίνει την τιμή της παραμέτρου alpha ίση με 0,024,όπως φαίνεται και παραπάνω. Αυτή η τιμή είναι πολυ κοντά στο 0 κάτι που σημαίνει ότι οι προβλέψεις βασίζονται πιο πολύ στις τιμές των παρατηρήσεων-δεδομένων του παρελθόντος και λιγότερο στις πρόφατες.

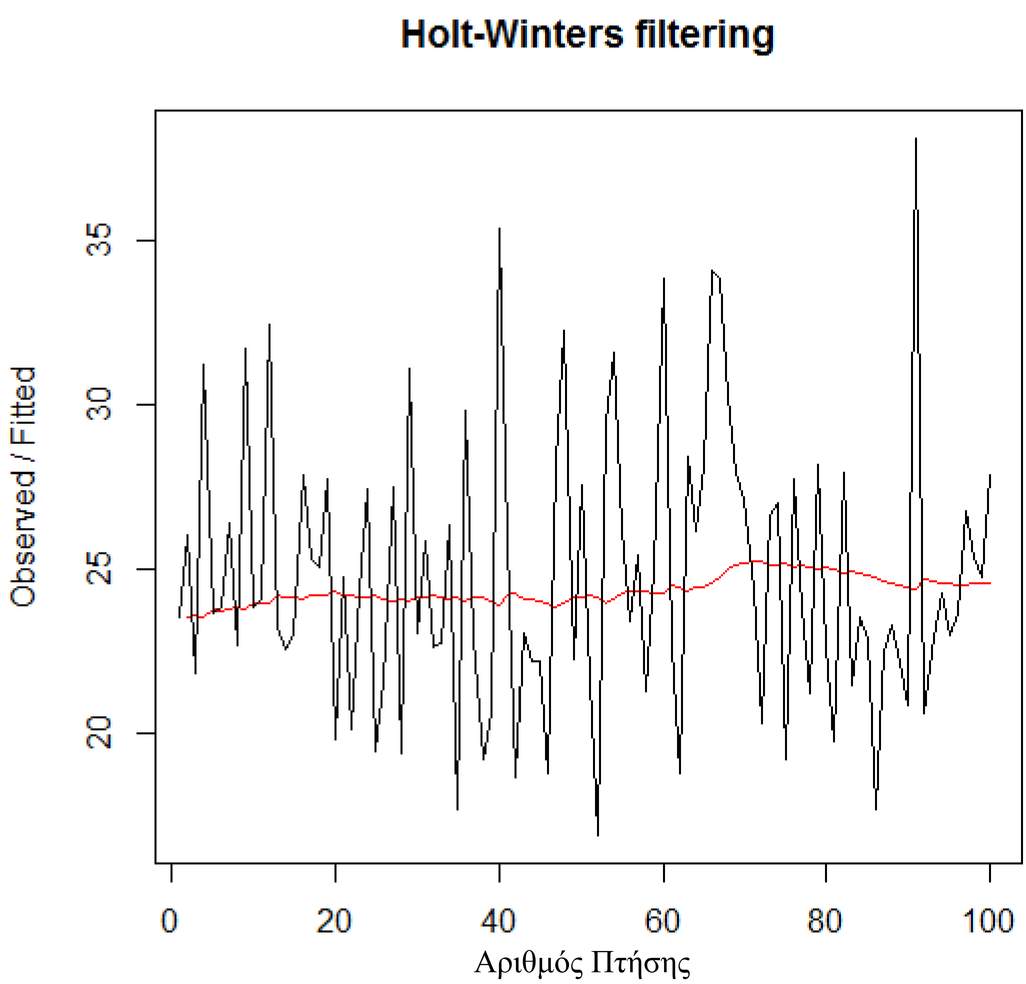
Από προεπιλογή, η εντολή Holt Winters **() πραγματοποιεί τις προβλέψεις ακριβώς για την ίδια χρονική περίοδο που καλύπτεται από την αρχική χρονοσειρά μας**.Στην δική μας περίπτωση οι αρχικές μας χρονοσειρές αποτελούνταν απο καθυστερήσεις 100 πτήσεων,επομένως οι προβλέψεις είναι επίσης για τις 100 πρώτες πτήσεις.

Στο κώδικα μας έχουμε αποθηκεύσει την έξοδο της εντολής HoltWinters() σε μία λίστα μεταβλητών με το όνομα ´´dataseriesforecasts´´.Οι προβλέψεις που πραγματοποιήθηκαν έχουν αποθηκευτεί σε ένα στοιχείο με το όνομα´´fitted´´ ,έτσι ώστε να μπορούμε να τις λάβουμε πληκτρολογώντας ‘dataseriesforecastsfitted’ όπως βλέπουμε και στον κώδικα.Έτσι προκύπτουν οι λίστες ,κάτι που αρχικά θα μας φανεί παράδοξο αφου οι τιμές και στις δύο στήλες θα είναι ίδιες.Υπενθυμίζουμε όμως ότι αυτό είναι λογικό, αφού απο προεπιλογή με αυτό τον τρόπο λειτουργεί η εντολή HoltWinters.

Εικόνα 2. 6 Αποθηκευμένες τιμές εξόδου και προβλέψεων HoltWInters εντολής

	xhat	level
2	23.56000	23.56000
3	23.62054	23.62054
4	23.57808	23.57808
5	23.76290	23.76290
6	23.76017	23.76017
7	23.76306	23.76306
8	23.82691	23.82691
9	23.79900	23.79900
10	23.98935	23.98935
11	23.98623	23.98623
12	23.98921	23.98921
13	24.19282	24.19282
14	24.17032	24.17032
15	24.13171	24.13171
16	24.10442	24.10442
17	24.19549	24.19549
18	24.22261	24.22261
19	24.24329	24.24329
20	24.32812	24.32812
21	24.21938	24.21938
22	24.23290	24.23290
23	24.13369	24.13369
24	24.13867	24.13867
25	24.21782	24.21782

Για να καταλάβουμε ακόμη καλύτερα μπορούμε να απεικονίσουμε τις αρχικές μας χρονοσειρές σε σύγκριση με τις πρόβλέψεις.



Διάγραμμα 3. 2 Απεικόνιση αρχικής χρονοσειράς με πρόβλεψη

Παρατηρούμε απο την γραφική απεικόνιση ότι με την μαύρη γραμμή έχουμε την αρχική μας χρονοσειρά και με την κόκκινη φραμμή τις προβλέψεις με την λειτουργια της HoltWinters εντολής.

Είναι προφανές ότι υπάρχει μεγάλο σφάλμα στο αποτέλεσμα των προβλέψεων,αλλά ήταν κάτι που το περιμέναμε.Μπορούμε να υπολογίσουμε το σύνολο των λάθος προβλέψεων σε σχέση με την αρχική μας χρονοσειρά και αυτό γιατί οι τιμές αυτές αποθηκεύτηκαν αυτόματα σε ένα στοιχείο της λίστας ‘dataseriesforecasts’ με την ονομασία ‘’SSE’’ ,όπως βλέπουμε και απο τον κώδικα που έχουμε ήδη παραπάνω.Οι τιμές λοιπόν αυτές είναι 1828.855

Είναι συνηθισμένο σε μία απλή εκεθετική εξομάλυνση να χρησιμοποιούμε την πρώτη τιμή της χρονοσειράς μας ως αρχική τιμή.Για παράδειγμα στην δική μας περίπτωση η πρώτη καθυστέρηση για την πρώτη πτήση είναι 23,56.Μπορούμε να θέσουμε στην εντολή

HoltWinters ως αρχική τιμή την 23,56 μέσω της παραμέτρου ‘l.start’ ,όπως έχουμε ήδη πράξει παραπάνω στον κώδικα που παραθέσαμε.

Σε αυτό το σημείο θα πρέπει να υπενθυμίσουμε ότι απο προεπιλογή η λειτουργίας της εντολής HoltWinters πραγματοποιεί απλά προβλέψεις για την ίδια χρονική περίοδο κάτι που ίσως με μία πρώτη σκέψη θα μπορούσε να αναρωτηθεί τι νόημα μπορεί να έχει κατι τέτοιο.

Ωστόσο αυτό τελικά έχει νόημα αφού έχουμε καταφέρει στην ουσία να εκπαιδεύσουμε και να προσαρμόσουμε το μοντέλο της πρόβλεψης μας στην αρχική μας χρονοσειρά,έτσι ώστε στην συνέχεια να μπορέσουμε να επιτύγχουμε **σωστές και αξιόπιστες προβλέψεις.**

3.1.2 Χρήση λειτουργίας forecast.HoltWinters

Θα εφαρμόσουμε λοιπόν προβλέψεις τώρα όχι για την ίδια χρονική περίοδο ,αλλά για περαιτέρω χρονικά σημεία.Για να επιτύχουμε κάτι τέτοιο θα πρέπει να κάνουμε χρήση της εντολής ‘forecast.HoltWinters()’ στο πακέτο βιβλιοθήκης ‘forecast’.Κάνοντας λοιπόν εγκατάσταση της βιβλιοθήκης στο εργαλείο R έπειτα την καλούμε με την εντολή που πληκτρολογήσαμε παραπάνω στον κώδικα.

Με την εφαρμογή της λειτουργίας αυτής σαν πρώτο στοιχείο εισόδου θέτουμε το μοντέλο πρόβλεψης που έχουμε ήδη προσαρμόσει πιο πριν με την χρήση της λειτουργίας HoltWnters().Πιο συγκεκριμένα υπενθυμίζουμε ότι στο δικό μας παράδειγμα αποθηκεύσαμε το μοντέλο της πρόβλεψης στο στοιχείο ‘dataseriesforecasts’’

Αυτο που έχει απομείνει πλέον είναι να θέσουμε στην παράμετρο ‘h’ της εντολής ‘foreast.HoltWinters()’ τον αριθμό των περαιτέρω χρονικά προβλέψεων που επιθυμούμε να κάνουμε. Έστω λοιπόν ότι εμείς θέουμε να κάνουμε πρόβλεψη για τις επόμενες 8 πτήσεις,θα πληκτρολογήσουμε h=8 όπως φαίνεται και στην εικόνα 2.8.

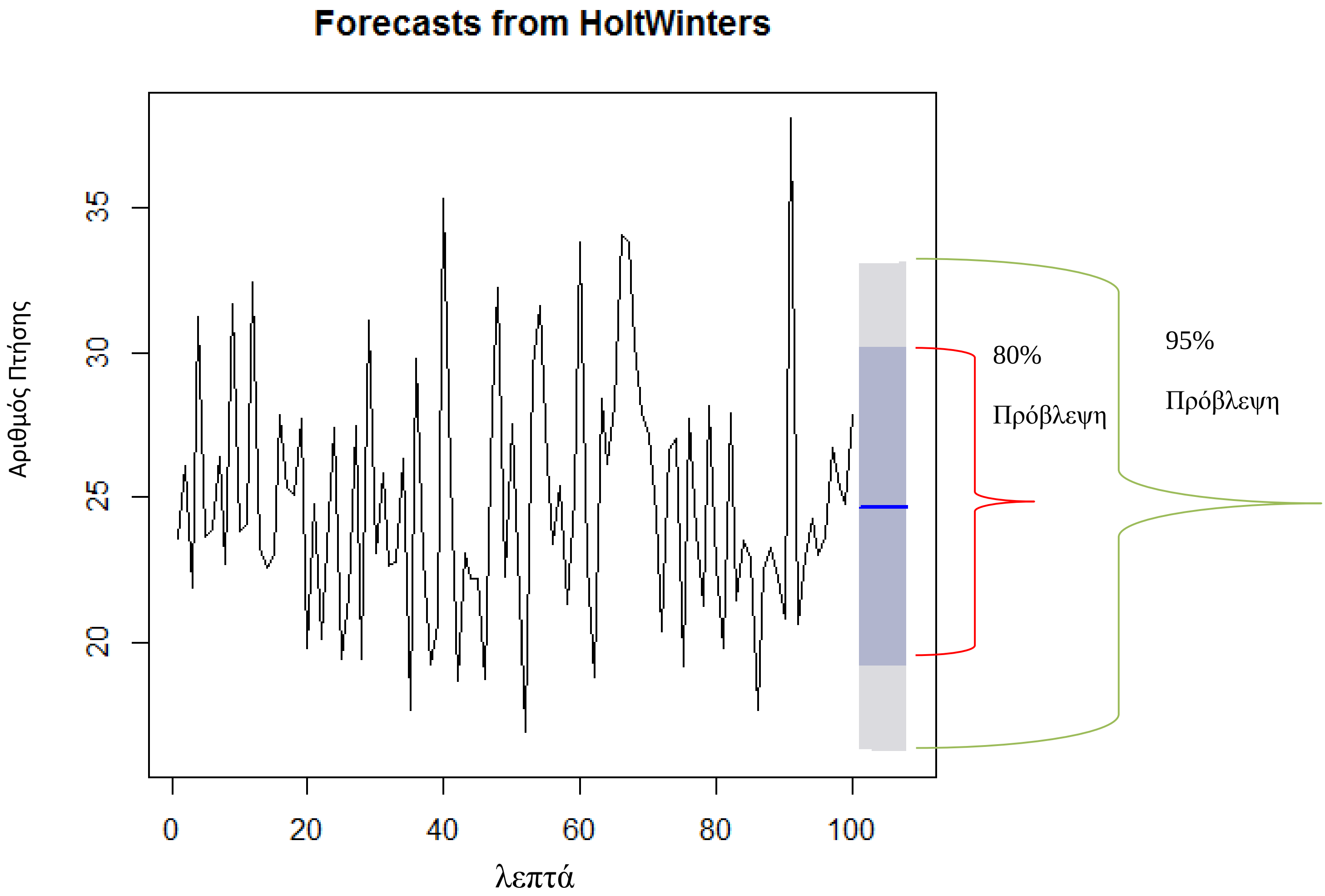
Φτάσαμε στον επιθυμητό πίνακα προβλέψεων που προκύπτει παρακάτω:

Πίνακας 3 1:Προβλέψεις για τις επόμενες 8 πτήσεις

flight	Forecast	Lo 80	Hi 80	Lo 95	Hi
101	24.67819	19.17493	30.18145	16.26169	33.09470
102	24.67819	19.17333	30.18305	16.25924	33.09715
103	24.67819	19.17173	30.18465	16.25679	33.09960
104	24.67819	19.17013	30.18625	16.25434	33.10204
105	24.67819	19.16853	30.18785	16.25190	33.10449
106	24.67819	19.16694	30.18945	16.24945	33.10694
107	24.67819	19.16534	30.19105	16.24701	33.10938
108	24.67819	19.16374	30.19265	16.24456	33.11182

Ο πίνακας που προκύπτει μας εμφανίζει τις προβλέψεις για τις καθυστερήσεις των επόμενων 8 πτήσεων.Πιο συγκεκριμένα για την πτήση 101 μας δείχνει ότι η πρόβλεψη είναι 24 λεπτά με 95% η μικρότερη καθυστέρηση να είναι 16 λεπτά και με 95% η μεγαλύτερη καθυστέρηση να είναι 33 λεπτά.Αντίστοιχα 80% η μικρότερη καθυστέρηση 19 λεπτά και 80% η μεγαλύτερη καθυστέρηση να είναι 30.Φτάνοτας στην τελική τιμή πρόβλεψης που είναι **24 λεπτά**.

Μπορούμε επίσης να απεικονίσουμε τα αποτελέσματα αυτά για να έχουμε και μία καλύτερη εικόνα.Στο τελευτάιο κομμάτι του κώδικα στην εικόνα 2.8 είναι οι εντολές για την γραφική απεικόνιση του αποτελέσματος των προβλέψεων που εφαρμόσαμεΑκολουθεί η γραφική απεικόνιση:



Διάγραμμα 3. 3 Απεικόνιση πρόβλεψης για τις επόμενες 8 πτήσεις

Απο το διάγραμμα 2.9 παρατηρούμε ότι με την μπλέ γραμμή είναι η τελική τιμή πρόβλεψης που αντιστοιχεί στην τιμή των 24 λεπτών.Με την κόκκινη παρένθεση απεικόνίζουμε τις τιμές που αντιστοιχούν στο 80% της μέγιστης και της ελάχιστης τιμής πρόβλεψης και αντίστοιχα με την πράσινη παρένθεση το 95% της μέγιστης και της ελάχιστης τιμής πρόβλεψης.

3.1.3 Σφάλματα Πρόβλεψης

Τα σφάλματα της πρόβλεψης που πραγματοποιήσαμε μπορούν να υπολογισθούν αφαιρόντaις τις **προβλεπόμενες τιμές** μείον **τις παρατηρούμενες τιμές .**Φυσικά μπορούμε μόνο τα σφάλματα πρόβλεψης της αρχικής μας χρονοσειράς να προβλέψουμε τα οποία αναφέρονται στις 100 πρώτες πτήσεις.

Όπως αναφέραμε παραπάνω, ένα μέτρο της ακρίβειας του μοντέλου πρόβλεψης είναι το άθροισμα των σφαλμάτων (SSE) στην πρόβλεψη της αρχικής χρονοσειρά που εφαρμόσαμε.

Τα σφάλματα αυτά έχουν αποθηκευτεί σε ένα στοιχείο της λίστα των μεταβλητών με την ονομασία 'residuals' τα οποία τα έχει επιστρέψει η εντολή forecast.HoltWinters().

Εφόσον το μοντέλο της πρόβλεψης δεν μπορεί να βελτιωθεί αυτό θα σημαίνει ότι δεν θα πρέπει να υπάρχουν συσχετισμοί ανάμεσα στα σφάλματα της πρόβλεψης για διαδοχικές προβλέψεις.

Με άλλα λόγια, αν υπάρχουν συσχετισμοί ανάμεσα στα σφάλματα της πρόβλεψης για διαδοχικές προβλέψεις αυτό θα σημαίνει ότι η τεχνική της απλής εκθετικής εξομάλυνσης που επιλέξαμε δεν ήταν κατάλληλη και θα μπορούσαμε να εφαρμόσουμε μια άλλη μέθοδο πρόβλεψης με πολύ πιο αγιόπιστες προβλέψεις.

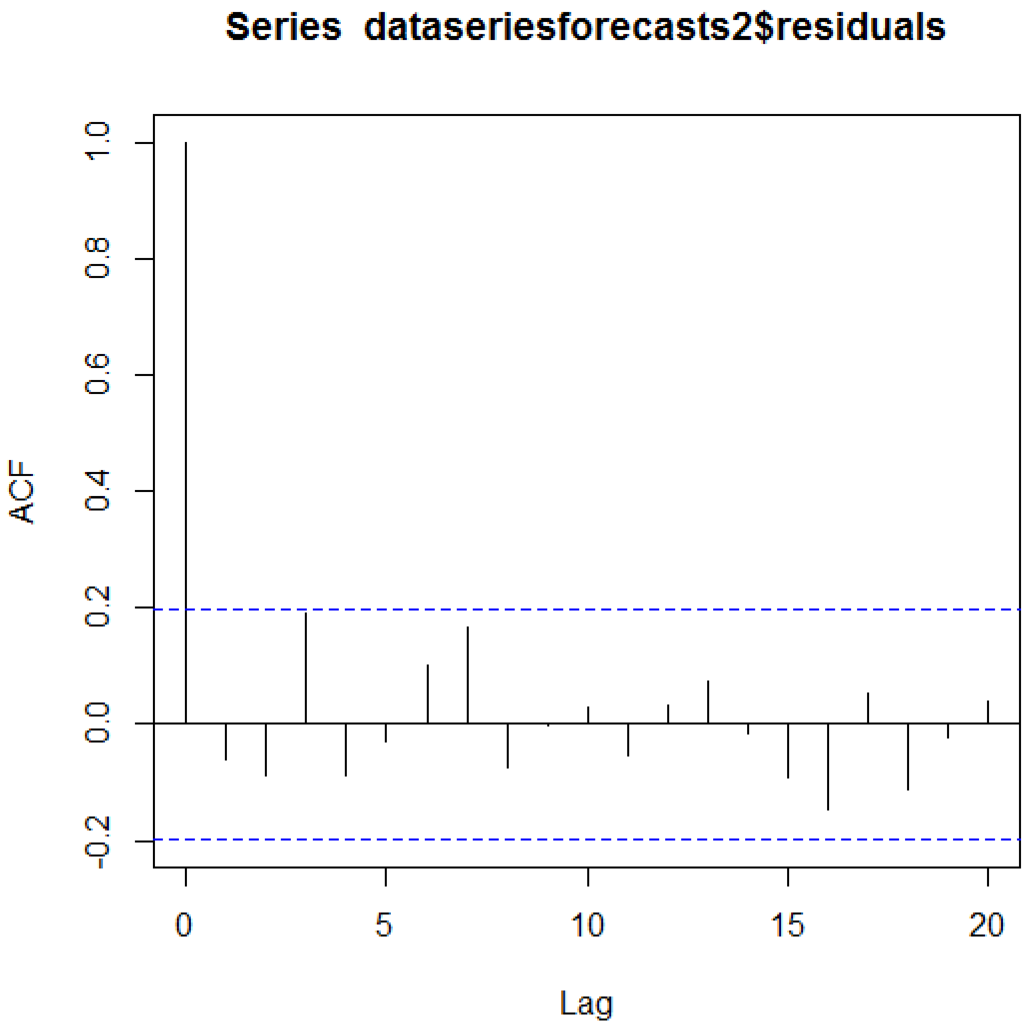
Υπενθυμίζουμε σε αυτο το σημείο ότι για να έχουμε μία επιτυχη και αξιόπιστη πρόβλεψη με την μέθοδο της εκθετικής εξομάλυνσης θα πρέπει οι διακυμάνσεις της χρονοσειράς μας να είναι σχεδόν σταθερές συναρτήση του χρόνου.Θα μπορούσαμε να εφαρμόσουμε στο παράδειγμα 1 την μέθοδο αυτή,γιατί και εκείνη τηρούσε τα κριτήριο αυτά, ωστόσο αυτο το παράδειγμα είχε πιο σταθερές διακυμάνσεις και άλλωστε θα καταλάβουμε παρακάτω ότι ταριάζει μια καλύτερη μέθοδος πρόβλεψης για το παράδειγμα 1.

3.1.4 Εγκυρότητα και αξιοπιστία πρόβλεψης

Υπάρχει τρόπος να είμαστε σίγουροι ότι η μέθοδος που επιλέξαμε για να προβούμε σε ορθή και αξιόπιστη πρόβλεψη στο δικό μας παράδειγμα είναι σωστή. Ο τρόπος αυτός είναι να απεικονίσουμε το κορελόγραμμα της αρχικής πρόβλεψης της χρονοσειράς για υστερήσεις 1 έως 20.Μπορούμε λοιπόν να φτιάξουμε το κορελόγραμμα αυτό πληκτρολογώτας στο εργαλείο R την εντολή ACF() και σε συνέχεια του παραπάνω κώδικα:

- > acf(dataseriesforecasts2$residuals, lag.max=20)

Ακολουθεί το κορελόγραμμα:



Διάγραμμα 3. 4 Απεικόνιση στο κορελόγραμμα

Μπορούμε να δούμε απο το Διάγραμμα 2.10 ότι η αυτοσυσχέτιση στην υστέρηση 3 αγγίζει τα όρια της ΄΄σημαντικότητας΄΄.Ωστόσο το αποτέλεσμα αυτο είναι πολυ ενθαρρυντικό αφού σε κανένα σημείο η αυτοσυσχέτιση δεν ξεπερνά τα όρια,κάτι που σημαίνει ότι η μεθοδος που έχουμε επιλέξει είναι σαν πρώτη εκτίμηση σωστή.

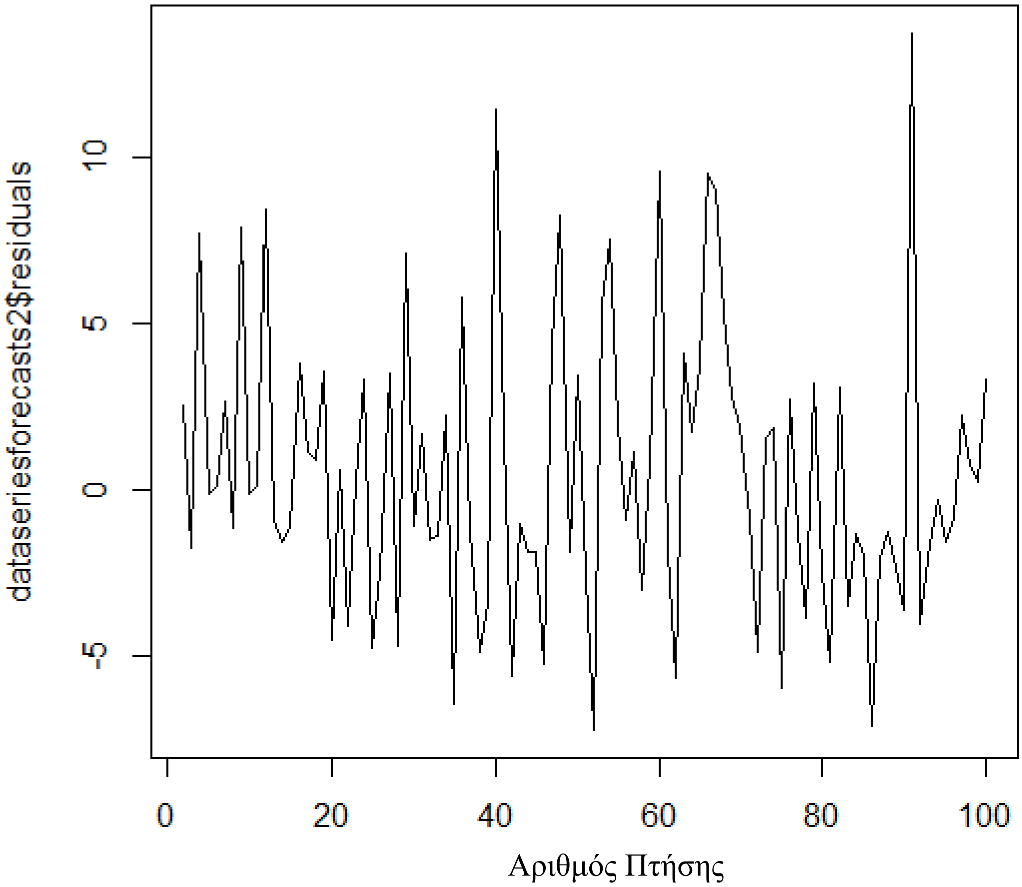
Ωστόσο θα πρέπει να εμβαθύνουμε για να σιγουρευτούμε ότι δεν υπάρχουν σημαντικές ενδείξεις αυτοσυσχέτισης για υστερήσεις 1 έως 20.Μπορούμε να κάνουμε χρήση της Ljungbox δοκιμής.Για να επιτύχουμε κάτι τέτοιο στο εργαλείο R καλουμε την εντολή ‘Box.test()’ θέτοντας στην παράμετο υστέρηση 1 έως 20.Ακολουθεί ο κώδικας σε γλώσσα R και η γραφική απεικόνιση:

Εικόνα 3. 3 Κώδκας σε γλώσσα R για χρήση εντολής Box.Tset()

```
> Box.test(rainseriesforecasts2$residuals, lag=20, type=”Ljung-Box”)
 Box-Ljung test
 data: rainseriesforecasts2$residuals
 X-squared = 17.4008, df = 20, p-value = 0.6268
> plot.ts(rainseriesforecasts2$residuals)
> plotForecastErrors <- function(forecasterrors)
(
# make a histogram of the forecast errors:
 mybinsize <- IQR(forecasterrors)/4
 mysd <- sd(forecasterrors)
mymin <- min(forecasterrors) – mysd*5
 mymax <- max(forecasterrors) + mysd*3
 # generate normally distributed data with mean 0 and standard deviation mysd
 mynorm <- rnorm(10000, mean=0, sd=mysd)
 mymin2 <- min(mynorm)
 mymax2 <- max(mynorm)
 if (mymin2 < mymin) { mymin <- mymin2 }
if (mymax2 > mymax) { mymax <- mymax2 }
 # make a red histogram of the forecast errors, with the normally distributed data overlaid:
mybins <- seq(mymin, mymax, mybinsize)
 hist(forecasterrors, col=”red”, freq=FALSE, breaks=mybins)
  # freq=FALSE ensures the area under the histogram = 1
 # generate normally distributed data with mean 0 and standard deviation mysd
 myhist <- hist(mynorm, plot=FALSE, breaks=mybins)
# plot the normal curve as a blue line on top of the histogram of forecast errors:
points(myhist$mids, myhist$density, type=”l”, col=”blue”, lwd=2)
}
```

Προκύπτει απο το εργαλείο R ότι η τιμή της δοκιμής του Ljung-box είναι 17,4 η p-value τιμή είναι 0,62, έτσι υπάρχουν λίγα στοιχεία για την ύπαρξη μη μηδενικών αυτοσυσχετισμών για υστερήσεις 1-20. Για να σιγουρευτούμε ότι το μοτέλο πρόβλεψης δεν βελτιώνεται άλλο και ότι η μέθοδος της εκθετικής εξομάλυνσης που επιλέξαμε είναι σίγουρα σωστή ,μπορούμε να ελένξουμε τα σφάλματα της πρόβλεψης διανέμονται κανονικά με μέση τιμή 0 και σταθερή διακύμανση.

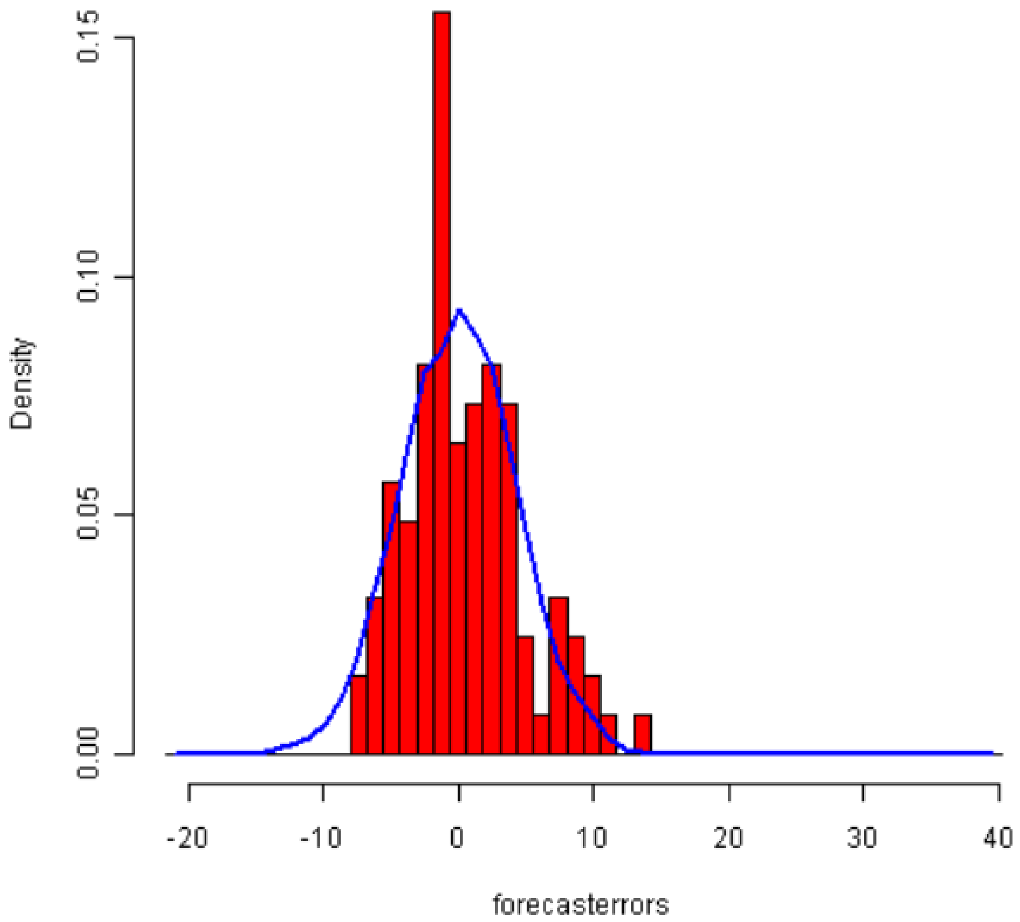
Επομένως για να ελέγξουμε αν τα σφάλματα πρόβλεψης έχουν σταθερή διακύμανση μπορούμε να κάνουμε μία απεικόνιση των σφαλμάτων πρόβλεψης στην αρχική μας χρονοσειρά. Η οποία είναι η εξής:



Διάγραμμα 3. 5 Απεικόνιση σφαλμάτων πρόβλεψης

Πράγματι φαίνεται απο το Διάγραμμα 2.11 ότι οι διακυμάνσεις είναι σχετικά σταθερές αν και το μέγεθος τω διακυμάνσεων φαίνεται να είναι λίγο μικρότερο στις πτήσεις 20 – 30 απο μεταγενέστερες πτήσεις όπως 40-50.

Για να ελέγξουμε αν τα σφάλματα της πρόβλεψης είναι κανονικά κατανεμημένα με μέση τιμή 0, μπορούμε να απεικονίσουμε ένα ιστόγραμμα των προβλέψεων με τα σφάλματα, με μία υπέρθεση της της κανονικής καμπύλης που θα εκφράζει την τιμή 0 και την ίδια τυπικη απόκλιση, όπως η διανομή των σφαλμάτων πρόβλεψης.Για επιτύγχουμε κάτι τέτοιο μπορούμε να ορίσαμε μια συνάρτηση στο εργαλείο R ‘plotforecasterrors()’.Η οποία έχει την εξής μορφή:



Διάγραμμα 3. 6 Γραφική απεικόνιση Ιστογράμματος σφαλμάτων πρόβλεψης

Το Διάγραμμα 2.11 μας δείχνει ότι η κατανομή των σφαλμάτων πρόβλεψης είναι περίπου στο κέντρο,δηλαδή στο 0 και είναι σχεδόν κανονικά κατανεμημένη,αν και υπάρχει μια μικρή μεγαλύτερη απόκλιση δεξιά.Ωστόσο είναι πολύ μικρή αυτή η διαφορά και έτσι εύλογα μπορούμε να πούμε ότι έχουμε μια κανονικά κατανεμημένη τυπικη απόκλιση ίση με 0 .

Επιπλέον,η δοκιμή Ljung-Box μας έδειξε ότι υπάρχουν ελάχιστα στοιχεία μη μηδενικών αυτοσυσχετίσεων .Φτανουμε λοιπόν στο συμπέρασμα ότι η μέθοδος της εκθετικής εξομάλυνσης για αυτό το συγκεκριμένο παράδειγμα είναι πράγματι επιτυχής.

3.2 Εκθετική εξομάλυνση Holt

3.2.1 Προβλέψεις με δεδομένα χωρίς τάση και εποχικότητα

Στην περίπτωση που έχουμε χρονοσειρές οι οποίες εμφανίζουν απότομες διακυμάνσεις και δεν εμφανίζουν κάποια εποχικότητα στα δεδομένα τους, τότε μπορούμε να εφαρμόσουμε το μοντέλο πρόβλεψης της εκθετικής εξομάλυνσης Holt.

Η συγκεκριμένη μέθοδος κάνει μία εκτίμηση του επιπέδου των διακυμάνσεων και της κλίσης την τρέχουσα χρονική στιγμή.Η εξομάλυνση ελέγχεται απο τρείς παραμέτρους:

- **Alpha**
- **Betta**

Όπου η παράμετρος alpha εκτιμά το επίπεδο των διακυμάνσεων την τρέχουσα χρονική στιγμή και η παράμετρος betta εκτιμά την κλίση της τάσης που εμφανίζουν τα δεδομένα την τρέχουσα χρονική στιγμή.

Όπως και στην απλή εκθετική εξομάλυνση οι παράμετροι alpha και betta κυμαίνονται απο 0 έως 1 και όσο πιο κοντά στην τιμή 0 βρίσκεται η κάθε παράμετρος τόσο πιο μικρή βαρύτητα θέτουμε στις πιο πρόσφατες παρατηρήσεις-δεδομένα.

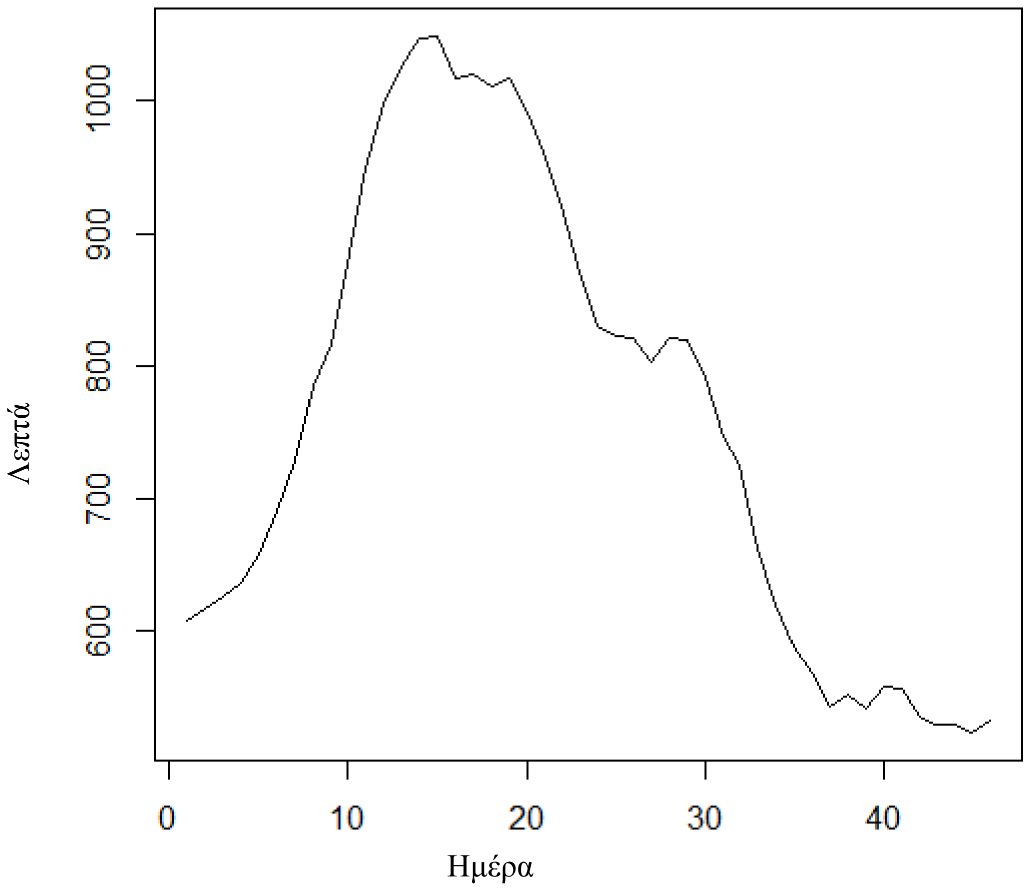
Το εργαλείο R κατα την διάρκεια της συγγραφής κώδικα για την εισαγωγή της χρονοσειράς μας παρέχει αυτόματα τις τιμές των παραμέτρων alpha και betta, ωστόσο μας παρέχει και την δυνατότητα να πειραματιστούμε μόνοι μας,κάτι το οποίο την συγκεκριμένη στιγμή θα αποφύγουμε.

Θα χρησιμοποιήσουμε ένα παράδειγμα απο εταιρεία logistics που αναφέρεται στους χρόνους παράδοσης προιόντων εκφρασμένους σε λεπτά με φορτηγά οχήματα.σε διάστημα 46 ημερών .Ακολουθούν τα δεδομένα σε μορφή πίνακα

Πίνακας 3. 2:Δεδομένα με χρόνους μεταφοράς προιόντων σε διάστημα 46 ημερών

Χρόνος σε λεπτά	**Ημέρα**	**Χρόνος σε λεπτά**	**Ημέρα**
608.	1	829.	24
617.	2	822.	25
625.	3	820.	26
636.	4	802.	27
657.	5	821.	28
691.	6	819.	29
728.	7	791.	30
784.	8	746.	31
816.	9	726.	32
876.	10	661.	33
949.	11	620.	34
997.	12	588.	35
1027.	13	568.	36
1047.	14	542.	37
1049.	15	551.	38
1018.	16	541.	39
1021.	17	557.	40
1012.	18	556.	41
1018.	19	534.	42
991.	20	528.	43
962.	21	529.	44
921.	22	523.	45
871.	23	531	46

Παρατηρούμε απο την εικόνα 2.4 ότι τα δεδομένα πράγματι δεν είναι καθόλου σταθερά κάτι που μας βοηθά να εφαρμόσουμε τη νέα μας μέθοδο σε αυτό το παράδειγμα.Αυτό βέβαια φαίνεται ακόμη καλύτερα παρακάτω στηην γραφική απεικόνιση.



Διάγραμμα 3. 7 Απεικόνιση χρόνων μεταφοράς σε διάστημα 42 ημερών

Μπορούμε να καταλάβουμε απο το Διάγραμμα 2.12 ότι υπάρχει μία αύξηση απο τα 600 λεπτά την πρώτη ημέρα στα 1050 την 10^{η} μέρα και έπειτα μία πτώση μέχρι την 42^{η} ημέρα.

Στην συνέχεια για να εφαρμόσουμε προβλέψεις θα προσαρμόσουμε το μοντέλο στο δικό μας παράδειγμα με τον κώδικα που παραθέτουμε παρακάτω.

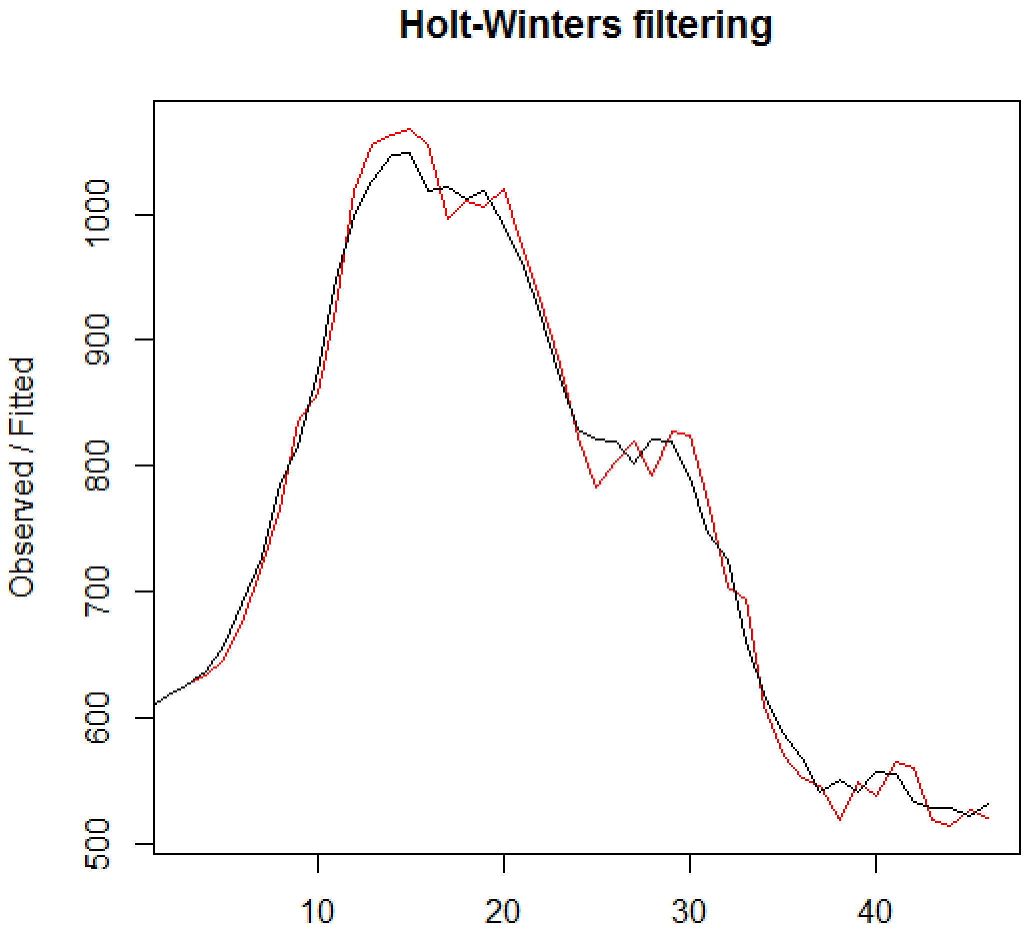
Εικόνα 3. 4 Κώδικας σε γλώσσα R για εισαγωγή δεδομένων και μοντέλου HoltWinters

```
truckseries <- ts(truck,start=c(1))

> plot.ts(truckseries)

> truckseriesforecasts <- HoltWinters(truckseries, gamma=FALSE)

> truckseriesforecasts

Holt-Winters exponential smoothing with trend and without seasonal component.

Call:

HoltWinters(x = truckseries, gamma = FALSE)

Smoothing parameters:

 alpha: 0.8383481

 beta : 1

 gamma: FALSE

Coefficients:       [,1]

a 529.308585

b   5.690464

> truckseriesforecasts$SSE

[1] 16954.18

> plot(truckseriesforecasts)

>> HoltWinters(skirtsseries, gamma=FALSE, l.start=608, b.start=9)  plot(truckseriesforecasts)

> skirtsseriesforecasts2 <- forecast.HoltWinters(skirtsseriesforecasts, h=19)

> plot.forecast(skirtsseriesforecasts2)
```

Παρατηρούμε απο την εικόνα 2.12 ότι το εργαλείο R επέστρεψε για την παράμετρο alpha την τιμή 0,84 και betta 1,00.Οι τιμές αυτές είναι πολύ υψηλές για τις δύο παραμέτρους κάτι που σημαίνει ότι το μοντέλο βασίζεται κυρίως με πολυ μεγάλη βαρύτητα στις πιο πρόσφατες παρατηρήσεις-δεδομένα της χρονοσειράς.

Αυτό το γεγονός βέβαια θα μπορούσαμε να πούμε ότι είναι λογικό,αφού κοιτάζοντας την γραφική απεικόνιση φαίνεται ότι πράγματι όσο στις πιο πρόσφατες ημέρες φτάνει τόσο μεταβάλεται η χρονοσειρά.

Ας δούμε τώρα λοιπόν την προσαρμογή του μοντέλου μας στην αρχική μας χρονοσειρά.



Διάγραμμα 3. 8 Προσαρμογή μοντέλου πρόβλεψης HoltWinters

Φαίνεται ξεκάθαρα απο το Διάγραμμα 2.13 ότι το μοντέλο πρόβλεψης HoltWinters με την κόκκινη γραμμή ταιριάζει πολύ καλά πάνω στην αρχική μας χρονοσειρά.Ετσι μπορούμε να πούμε με βεβαιότητα πλέον ότι είμαστε σε θέση να εφαρμόσουμε μία πρόβλεψη πέρα απο την αρχική μας χρονοσειρά.

Για παράδειγμα ας επιχειρήσουμε να προβλέψουμε για τις επόμενες 19 ημέρες τον χρόνο παραδόσεων.Αυτό που θα πρέπει να κάνουμε είναι να χρησιμοποιήσουμε την λειτουργία της εντολής forecast.HoltWinters με την εντολή

- > truckseriesforecasts2 <- forecast.HoltWinters(truckseriesforecasts, h=19)
- > plot.forecast(truckseriesforecasts2)

Ετσι έχουμε τις εξής τιμές πρόβλεψης:

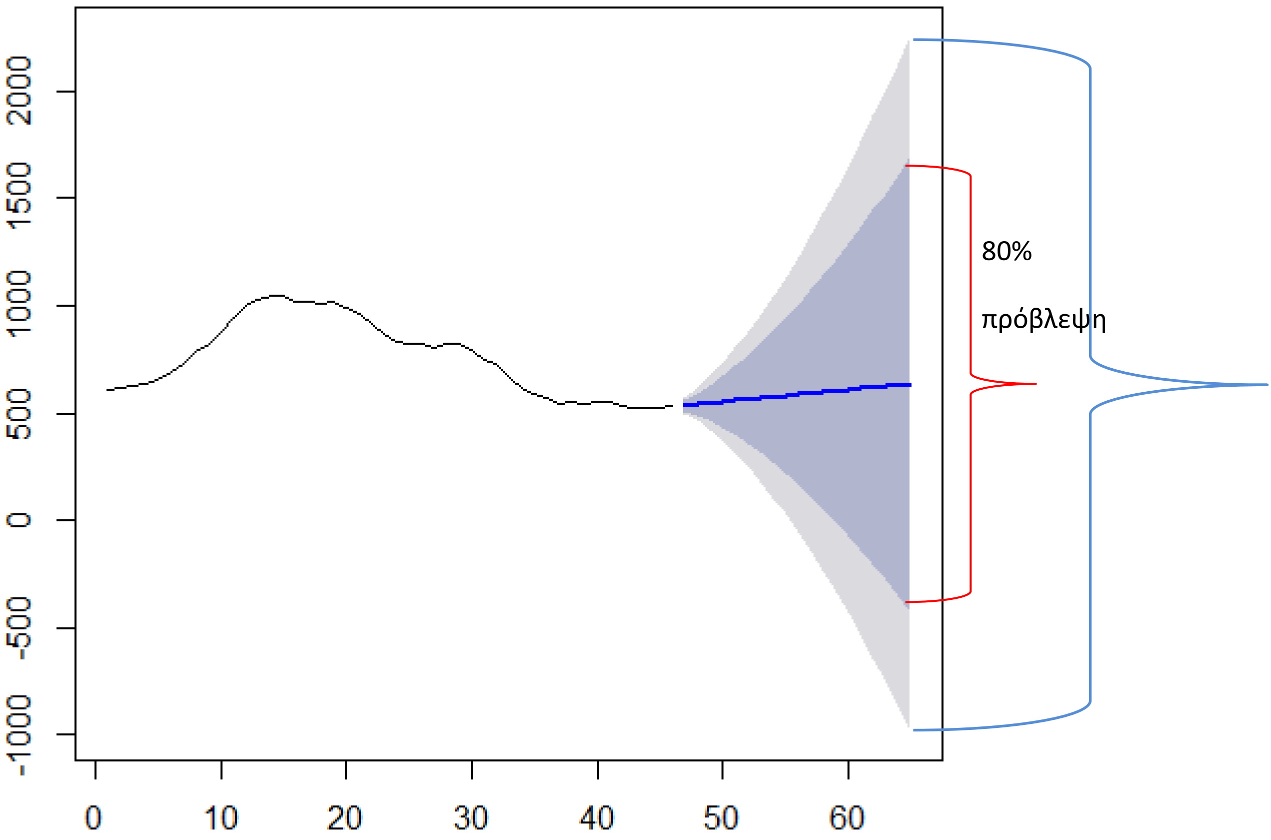
Πίνακας 3. 3:Πρόβλεψη για τις επόμενες 19 ημέρες

Point	Forecast	Lo 80	Hi 80	Lo 95	Hi 95
47	534.9990	509.55210	560.4460	496.08130	573.9168
48	540.6895	491.01052	590.3685	464.71204	616.6670
49	546.3800	465.36129	627.3987	422.47258	670.2874
50	552.0704	434.40205	669.7388	372.11216	732.0287
51	557.7609	398.94120	716.5806	314.86713	800.6547
52	563.4514	359.47147	767.4313	251.49103	875.4117
53	569.1418	316.34076	821.9429	182.51596	955.7677
54	574.8323	269.81480	879.8498	108.34829	1041.3163
55	580.5228	220.10648	940.9391	29.31362	1131.7319
56	586.2132	167.39191	1005.0345	-54.31870	1226.7452
57	591.9037	111.82029	1071.9871	-142.32052	1326.1279
58	597.5942	53.52019	1141.6681	-234.49517	1429.6835
59	603.2846	-7.39593	1213.9652	-330.67069	1537.2399
60	608.9751	-70.82861	1288.7788	-430.69495	1648.6451
61	614.6655	-136.68903	1366.0201	-534.43211	1763.7632
62	620.3560	-204.89720	1445.6092	-641.75986	1882.4719
63	626.0465	-275.38060	1527.4736	-752.56728	2004.6602
64	631.7369	-348.07309	1611.5470	-866.75319	2130.2271

Στην πρώτη στήλη του πίνακα αναγράφονται οι τελικές τιμές πρόβλεψης για την κάθε ημέρα.Στην δεύτερη και την τέταρτη στήλη αναγράφεται η πρόβλεψη για το 80% της μικρότερης και μεγαλύτερης τιμής.Αντίστοιχα στην Τρίτη και την Πέμπτη στήλη αναγράφεται η πρόβλεψη για το 95% της μικρότερης και μεγαλύτερης τιμής.

Ωστόσο οι τελικές τιμές πρόβλεψεις έιναι οι τιμές στην στήλη με την ονομασία **forecast.** Θα πρέπει στο σημείο αυτό να πούμε ότι ο Πίνακας 2.5 περιέχει και αρνητικές τιμές,κάτι που μας προβληματίζει ως προς την ορθότητα των αποτελεσμάτων.Πιο συγκεκριμένα,θα μπορούσε να αναρωτηθεί κάποιος αν η επιλογής της μεθόδου HoltWinters εκθετικής εξομάλυνσης είναι η κατάλληλη.Θα εξετάσουμε λοιπόν παρακάτω την ορθότητα των αποτελεσμάτων.

Ας δούμε πρώτα όμως την γραφική απεικόνιση των προβλέψεων



Διάγραμμα 3. 9 Γραφική απεικόνιση με την μέθοδο HoltWinters εκθετικής εξομάλυνσης

Στο Διάγραμμα 2.14 με την μπλέ γραμμή απεικονίζονται οι προβλέψεις για τον χρόνο μεταφοράς προιόντων με φορτηγά οχήματα για την επόμενες 19 ημέρες. Με την κόκκινη αγκύλη στην πραγματικότητα είναι δεύτερη και η τέταρτη στήλη που αναλύσαμε παραπάνω και αντίστοιχα η μπλέ αγκύλη η Τρίτη και η Πέμπτη στήλη.

Όπως και στο παράδειγμα της απλής εκθετικής εξομάλυνσης ,μπορούμε να ελέγξουμε αν το μοντέλο πρόβλεψης μπορεί να βελτιωθεί περαιτέρω.Αυτο γίνεται με τον έλεγχο του αθροίσματος των σφαλμάτων επι της αρχικής χρονοσειράς.

Πιο συγκεκριμένα θα πρέπει να εξετάσουμε αν υπάρχουν μη μηδενικοί συσχετισμοί σε ανάμεσα σε διαδοχικές προβλέψεις σε διάστημα διακύμανσης 1 έως 20 μέσω του κορελογράματος.

3.2.2 Έλεγχος αξιοπιστίας αποτελεσμάτων πρόβλεψης

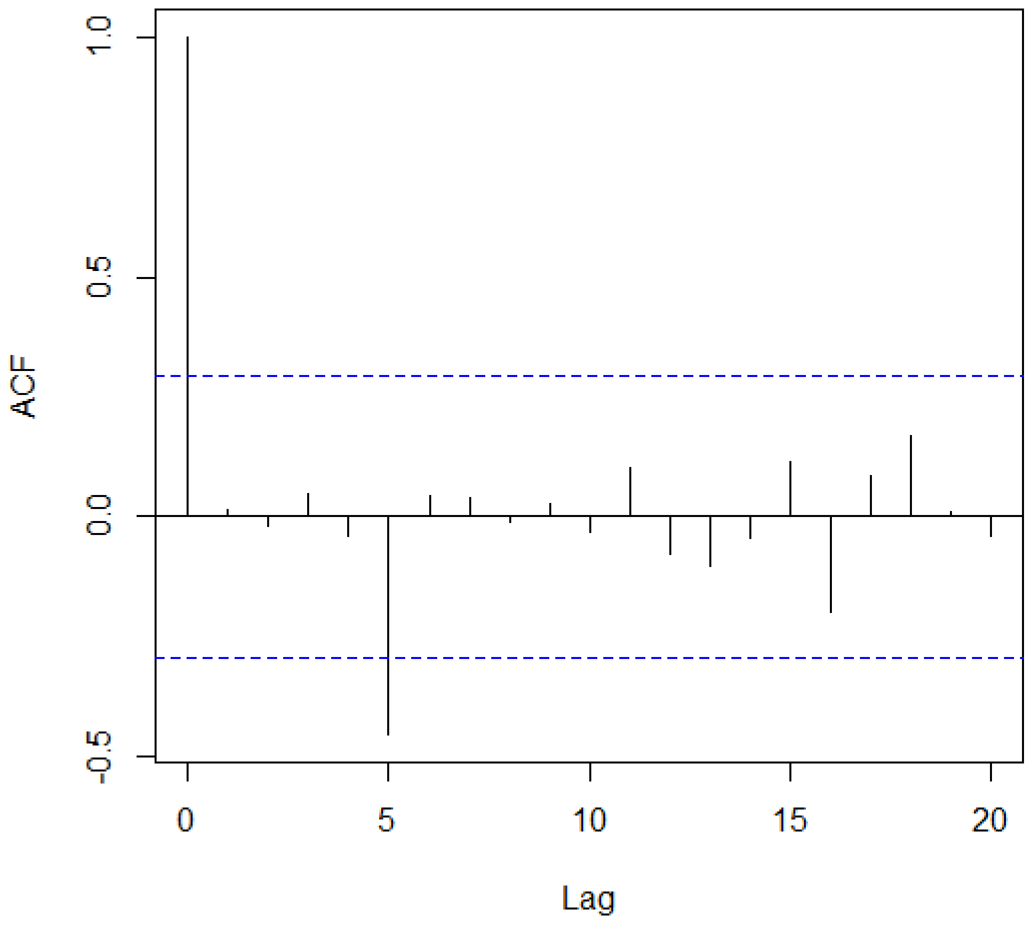
Όπως ήδη γνωρίζουμε για να είμαστε αξιόπιστοι ως προς τα αποτελέσματα των προβλέψεων κάνουμε την δοκιμή του **Ljung-Box.**

Επομένως πληκτρολογούμε στο εργαλείο R τον παρακάτω και κώδικα:

Εικόνα 3. 5 Κώδικας σε γλώσσα R για έλεγχο αξιοπιστίας πρόβλεψης

```
> acf(skirtsseriesforecasts2$residuals, lag.max=20)

> Box.test(skirtsseriesforecasts2$residuals, lag=20, type=”Ljung-Box”)

 Box-Ljung test

data: skirtsseriesforecasts2$residuals

 X-squared = 19.7312, df = 20, p-value = 0.4749
```

Ακολουθεί η απεικόνιση στο κορελόγραμμα:



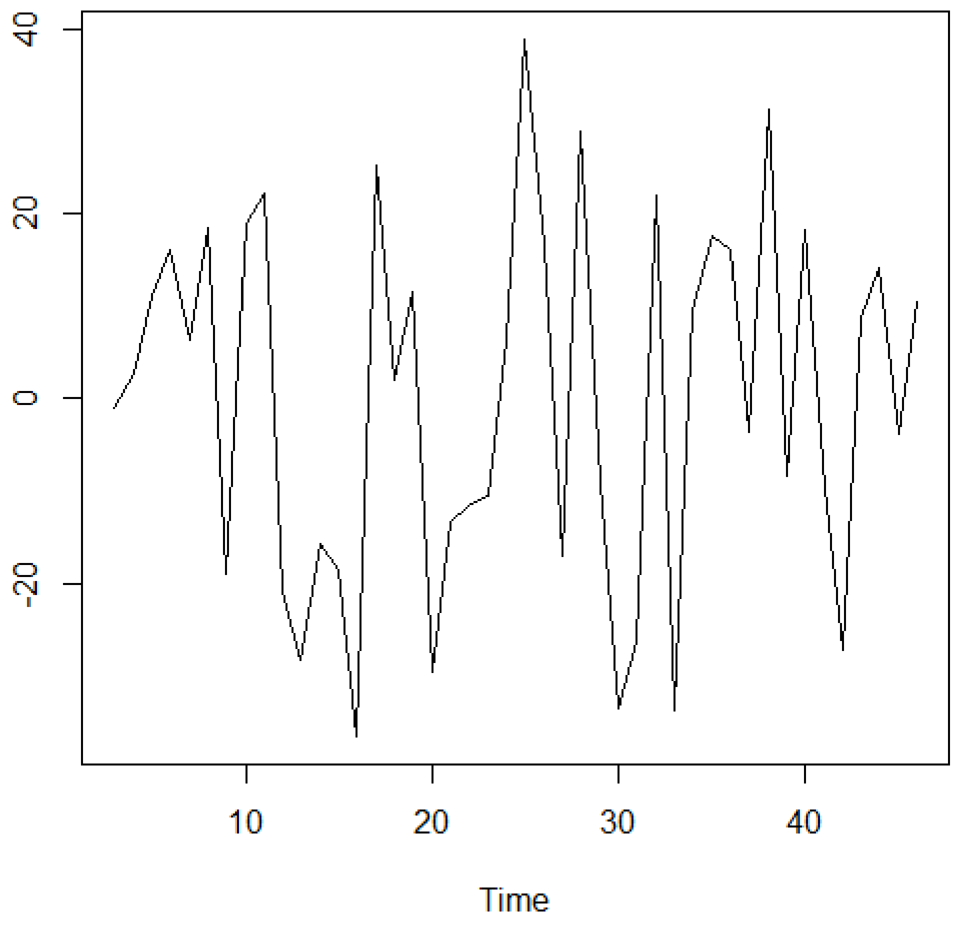
Διάγραμμα 3. 10 Απεικόνιση στο κορελόγραμμα

Απο το κορελόγραμμα παρατηρούμε ότι η αυτοσυσχέτιση του δείγματος στη διακύμανση 5 ξεπερνά τα επιτρεπτά όρια.Ωστόσο σε ένα δείγμα 20 διακυμάνσεων ,ίσως και να περιμέναμε να υπάρχει μια τέτοια αυτόσυσχέτιση να υπερβαίνει το 95% απο τύχη και μόνο.

Επιπλέον, απο τις τιμές της δοκιμής Ljung-Box, η τιμή της p-value είναι 0,47,κάτι που σημάινει ότι υπάρχουν ελάχιστα σημεία των μη μηδενικών αυτοσυσχετίσεων στα σφάλματα του δείγματος προβλέψεων σε υστερήσεις 1 έως 20.

Τέλος όπως και στην απλή εκθετική εξομάλυνση για να είμαστε απολύτως σίγουροι για την καταλληλότητα της επιλογής του μοντέλου πρόβλεψης που κάναμε ,μπορούμε να απεικονίσουμε τα σφάλματα πρόβλεψης καθώς και την τυπική απόκλιση.

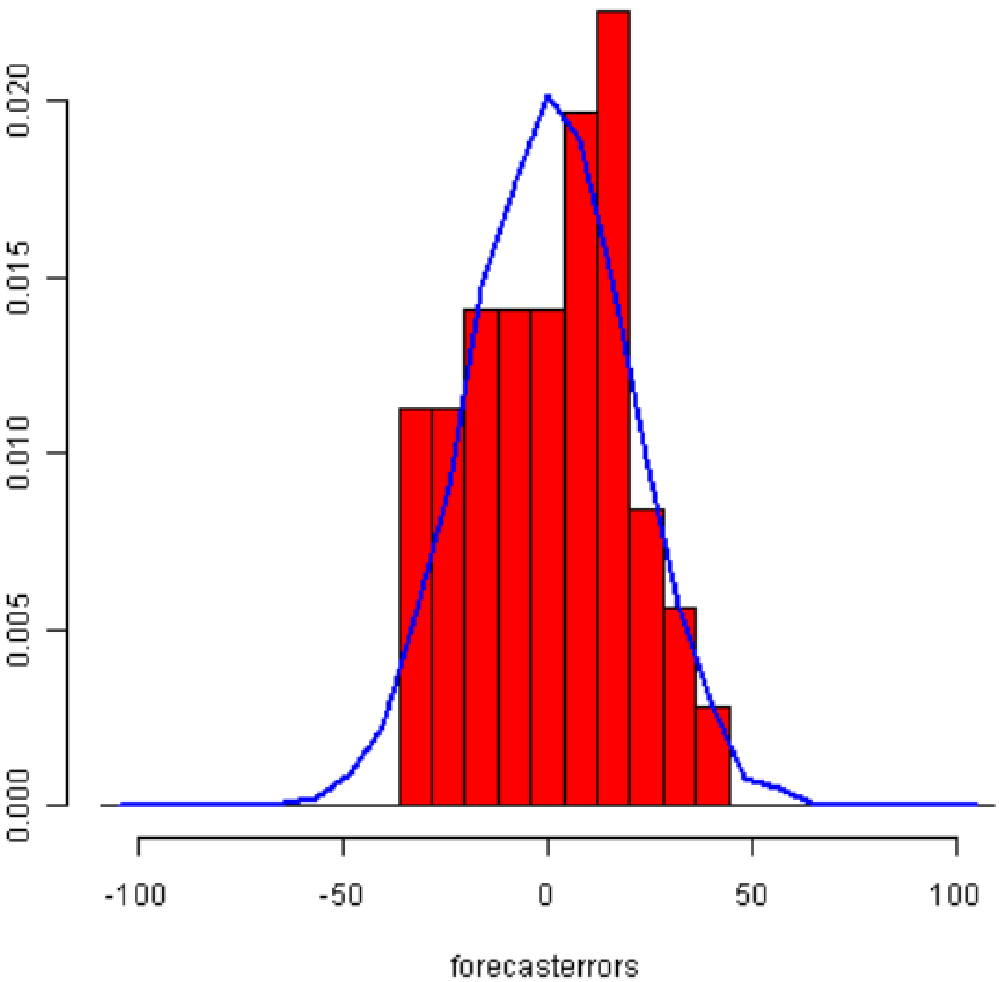
Οταν τα σφάλματα πρόβλεψης είναι σταθερά στην διάρκεια του χρόνου και η τυπική απόκλιση διανέμεται με μέση τιμή 0,τότε μπορούμε να έχουμε έναν καλό λόγο να πιστεύουμε ότι η μέθοδος μας είναι σωστή.



Διάγραμμα 3. 11 Απεικόνιση σφαλμάτων πρόβλεψης συναρτήση του χρόνου

Το αποτέσμα της απεικόνισης των σφαλμάτων πρόβλεψης είναι πολυ ικανοποιητική, αφού πράγματι τα σφάλματα πρόβλεψης έχουν σταθερή διακύμανση.Ετσι μια απο τις δύο προυποθέσεις για την εγκυρότητα του μοντέλου έχει καλυφθεί,Αυτο που απομένει τώρα είναι η απεικόνιση του ιστογράμματος,ώστε να δούμε την αν η τυπική απόκλιση κυμαίνεται με μέση τιμή 0.

Ακολουθεί το ιστόγραμμα:



Διάγραμμα 3. 12 Απεικόνιση Ιστογράμματος

Το ιστόγραμμα μας δείχνει ότι τα σφάλματα πρόβλεψης είναι κανονικά κατανεμημένα με μέση τιμή 0 και σταθερή δικαύμανση,επομένων είμαστε σε θέση να πουμε ότι πράγματι η πρόβλεψη και το μοντέλο αντίστοιχα που επιλέξαμε είναι σωστά.

3.3 Προβλέψεις σε δεδομένα χωρίς τάση και με εποχικότητα

Μεχρι τώρα εφαρμόσαμε την μέθοδο της εκθετικής εξομάλυνσης που αναφερόταν σε παράδειγμα με σταθερές διακυμάνσεις συναρτήση του χρόνου και χωρίς να εμφανίζουν τα δεδομένα κάποια εποχικότητα, δηλαδή κάποια ιδιαίτερη τάση συγκεκριμένες εποχές.Αυτο που θα εξετάσουμε τώρα είναι για την περίπτωση που έχουμε δεδομένα με απότομες διακυμάνσεις και με εμφάνιση κάποιας εποχικότητας στα δεδομένα,κάτι που θα αναλύσουμε καλύτερα παρακάτω.

Σε αυτήν την περίπτωση μπορούμε να χρησιμοποιήσουμε την μέθοδο της εκθετικής εξομάλυνσης HoltWinters για να προβούμε σε προβλέψεις.Αυτή η μέθοδο κάνει μία εκτίμηση στο επίπεδο των διακυμάνσεων, στην κλίση και στην εποχιακή συνιστώσα κατα την τρέχουσα χρονική στιγμή.

Η εξομάλυνση ελέγχεται απο τρείς παραμέτρους:

- **Alpha**
- **Betta**
- **Gamma**

Όπου **alpha** και **betta** κυμαίνονται ανάμεσα στις τιμές 0 έως 1.Όποια παράμετρος είναι πιο κοντά στην τιμή 0 επιλέγουμε και την αντίστοιχα βαρύτητα που θα του θέσουμε.Ο παράγοντας gamma αναφέρονται στην εποχικότητα των δεδομένων και κυμαίνεται επίσης απο 0 έως 1,θέτοντας την αντίστοιχη βαρύτητα κάθε φορά.

Όσο αφορά την νέα παράμετρο **gamma**, πρόκειται για την παράμετρο που αναφέρεται στην επαναληψιμότητα των δεδομένων που εμφανίζουν εποχικότητα.Αντίστοιχα και εδώ όσο η τιμή είναι πιο κοντά στο 0 ,τόσο πιο μικρή βαρύητητα θέτουμε στις πιο πρόσφατες παρατηρήσεις-δεδομένα.

Το παράδειγμα που θα χρησιμοποιήσουμε σε αυτή την μέθοδο είναι η περίπτωση που συζητήσαμε παραπάνω τον χρόνο μεταφοράς προιόντων απο φορτηγά πλοία.Όπως είχαμε αναλύσει το παράδειγμα μας ταιριάζει στην μέθοδο που θα εφαρμόσουμε αφού οι διακυμάνσεις δεν είναι σταθερές και παρουσίαζαν τα δεδομένα μία μεγιστοποίηση των προιόντων που καθυστέρησαν στην διάρκεια των καλοκαιρινών μηνών και μία αισθητή πτώση τους χειμερινούς μήνες.

Όπως είναι ήδη γνωστό για να εφαρμόσουμε μια πρόβλεψη HoltWinters εκθετικής εξομάλυνσης αρκεί να προσαρμόσουμε το μοντέλο αυτό στα δεδομένα της χρονοσειράς.Το επιτυγχάνουμε με τις παρακάτω εντολές στο εργαλείο R .

- logshiptimeseries <- log(shiptimeseries)
- shiptimeseriesforecasts <- HoltWinters(logshiptimeseries)
- shiptimeseriesforecasts

Οι τιμές που μας επιστρέφει το εργαλείο R είναι οι εξής:

Εικόνα 3. 6 Τιμές μοντέλου πρόβλεψης

```
Holt-Winters exponential smoothing with trend and additive seasonal component.
Smoothing parameters:
alpha: 0.413418
beta : 0
gamma: 0.9561275
Coefficients:
 [,1]
 a 10.37661961
b 0.02996319
 s1 -0.80952063
s2 -0.60576477
 s3 0.01103238
s4 -0.24160551
 s5 -0.35933517
 s6 -0.18076683
 s7 0.07788605
s8 0.10147055
```

Οι εκτιμώμενες τιμές των τιμών alpha,betta και gamma είναι 0,40 , 0 , και 0,96 αντίστοιχα.

Η τιμή της παραμέτρου alpha είναι σχετικά χαμηλή, υποδεικνύοντας ότι η εκτίμηση του επιπέδου κατα την τρέχουσα χρονική στιγμή βασίζεται τόσο στις πρόσφατες παρατηρήσεις όσο και σε κάποιες πιο παρελθοντικές παρατηρήσεις.

Η τιμή της παραμέτρου betta είναι 0, υποδεικνύοντας ότι η εκτίμηση της κλίσης του στοιχείου τάσης δεν ενημερώνεται ουσιαστικά επι της χρονοσειράς ,αντιθέτως θέτει την παράμετρο betta ίση με την αρχική τιμή του.Αυτο είναι λογικό αφού το επίπεδο αλλάζει πολύ λίγο στην διάρκεια της χρονοσειράς με αποτέλεσμα η παράμετρος της κλίσης της τάσης Betta να είναι η ίδια,δηλαδή 0.

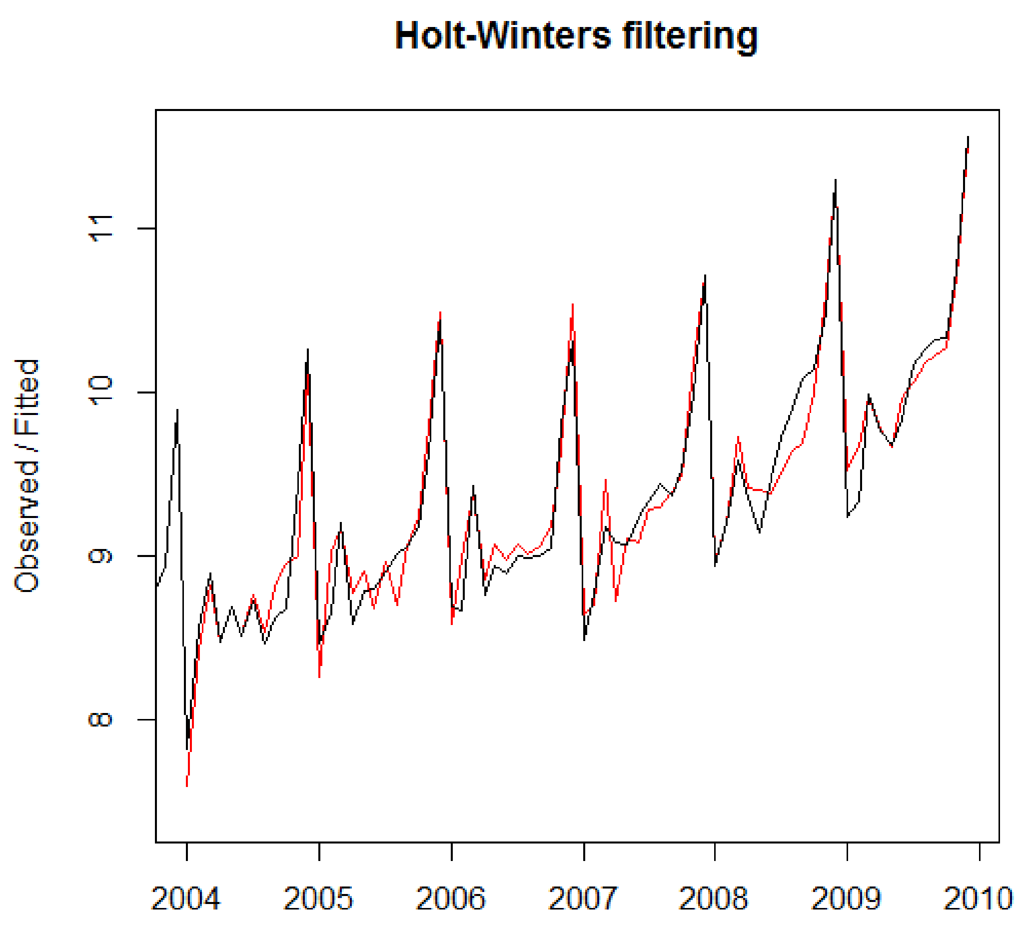
Σε αντίθεση με την παράμετρο gamma η τιμή είναι πολυ υψηλή, υποδεικνύοντας έτσι ότι η εκτίμηση της εποχικότητας βασίζεται πολύ στις πρόσφατες παρατηρήσεις.

Όπως και την απλή εκθετική εξομάλυνση μπορούμε να απεικονίσουμε την αρχική χρονοσειρά με μια μαύρη γραμμή και το πρόσθετο μοντέλο πρόβλεψης με μία κόκκινη γραμμή πάνω σε αυτή.

Επομένως εισάγουμε στο εργαλείο R την εντολή:

- plot(shiprtimeseriesforecasts)

Επομένως έχουμε την παρακάτω απεικόνιση:

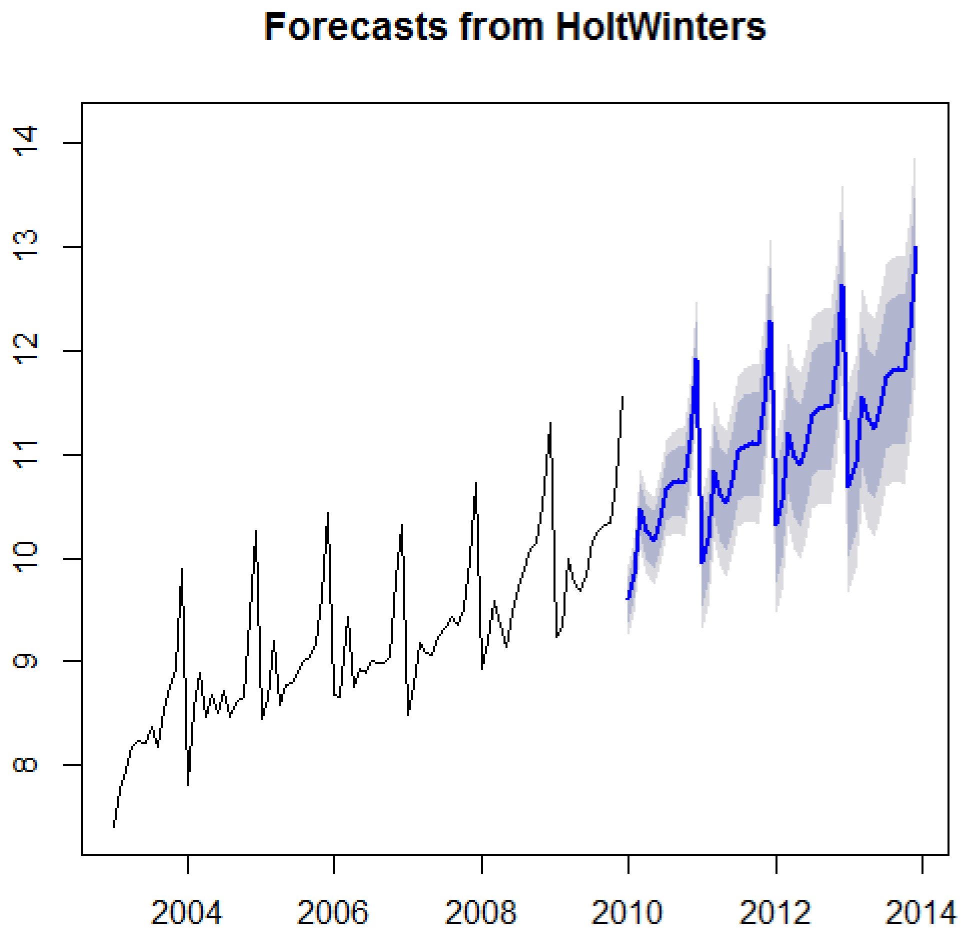


Διάγραμμα 3. 13 Απεικόνιση πρόσθετου μοντέλου πρόβλεψης επι της αρχικής χρονοσειράς

Παρατηρούμε ότι η μέθοδος του HoltWinters είναι πολύ επιτυχής όσον αφορά την πρόβλεψη των εποχιακών αιχμών, που συμβαίνουν περίπου τον Νοέμβριο κάθε έτους.Για να προβούμε σε μελλοντικές προβλέψεις που δεν περιλαμβάνονται στην αρχική χρονοσειρά όπως ήδη γνωρίζουμε, θα χρησιμοποιήσουμε την λειτουργία forecast.HoltWinters().

Για παράδειγμα τα αρχικά δεδομένα αναφέρονται στην περίοδο 2003 έως 2009, στην περίπτωση που εμείς θέλουμε να προβλέψουμε για τα έτη 2010 έως 2014 (48 μήνες) και στην συνέχεια να απεικονίσουμε τις προβλέψεις,πληκρολογούμε στο εργαλείο R τις εξής εντολές:

- shiptimeseriesforecasts2 <- forecast.HoltWinters(shiptimeseriesforecasts, h=48)
- plot.forecast(shiptimeseriesforecasts2)



Διάγραμμα 3. 14 Απεικόνιση πρόβλεψης για τους επόμενους 48 μήνες

Οι προβλέψεις απεικονίζονται με την μπλε γραμμή ,ενώ οι σκιαγραφημένες περιοχές με το έντονο και αχνό γκρι απεικονίζουν το 80 % και 90% των προβλέψεων ,αντίστοιχα.

3.3.1 Έλεγχος αξιοπιστίας αποτελεσμάτων πρόβλεψης

Μπορούμε να διερευνήσουμε αν το προγνωστικό μοντέλο μπορεί να βελτιωθεί ελέγχοντας αν σε ένα δείγμα προβλέψεων ,τα σφάλματα προβλέψεων δείχνουν μη μηδενική αυτοσυσχέτιση σε υστερήσεις 1-20.Αυτό επιτυγχάνεται όπως είναι ήδη γνωστό με την απεικόνιση του κορελογράμματος και της δοκιμής του Ljung-Box.

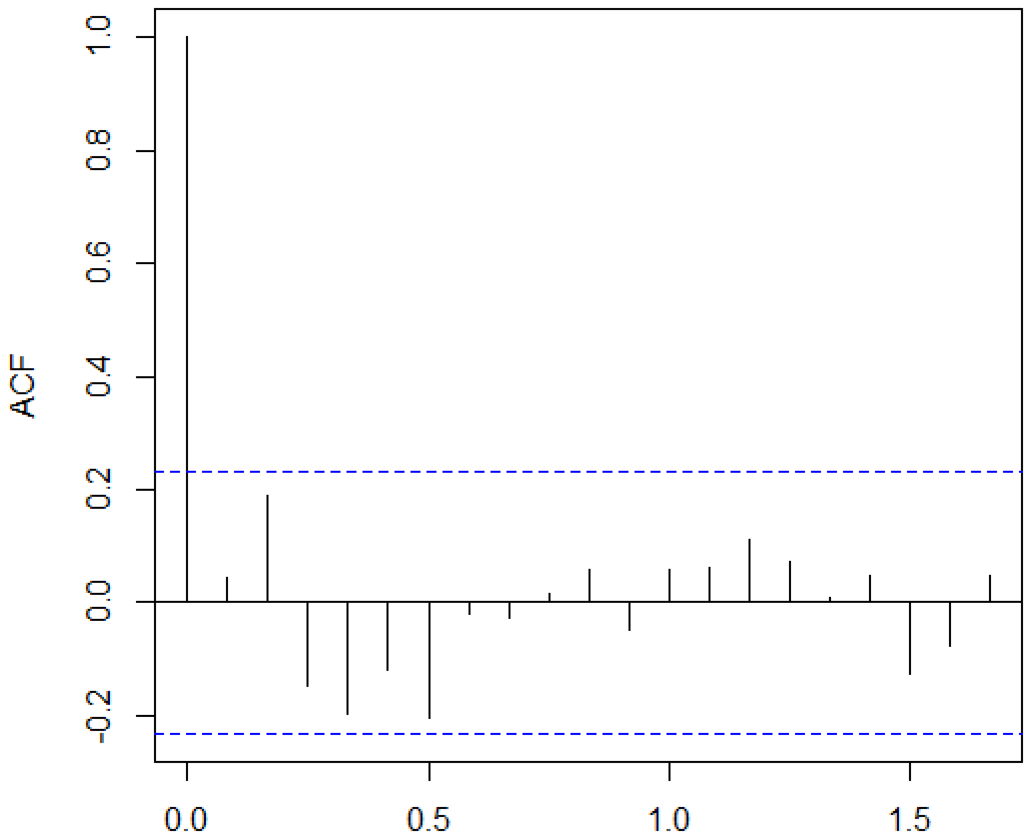
Επομένως πληκτρολογούμε στο εργαλείο R τις εξής εντολές:

- acf(shiptimeseriesforecasts2$residuals, lag.max=20)
- Box.test(shiptimeseriesforecasts2$residuals, lag=20, type="Ljung-Box")

- Box-Ljung test

  data: souvenirtimeseriesforecasts2$residuals

  X-squared = 17.5304, df = 20, p-value = 0.6183



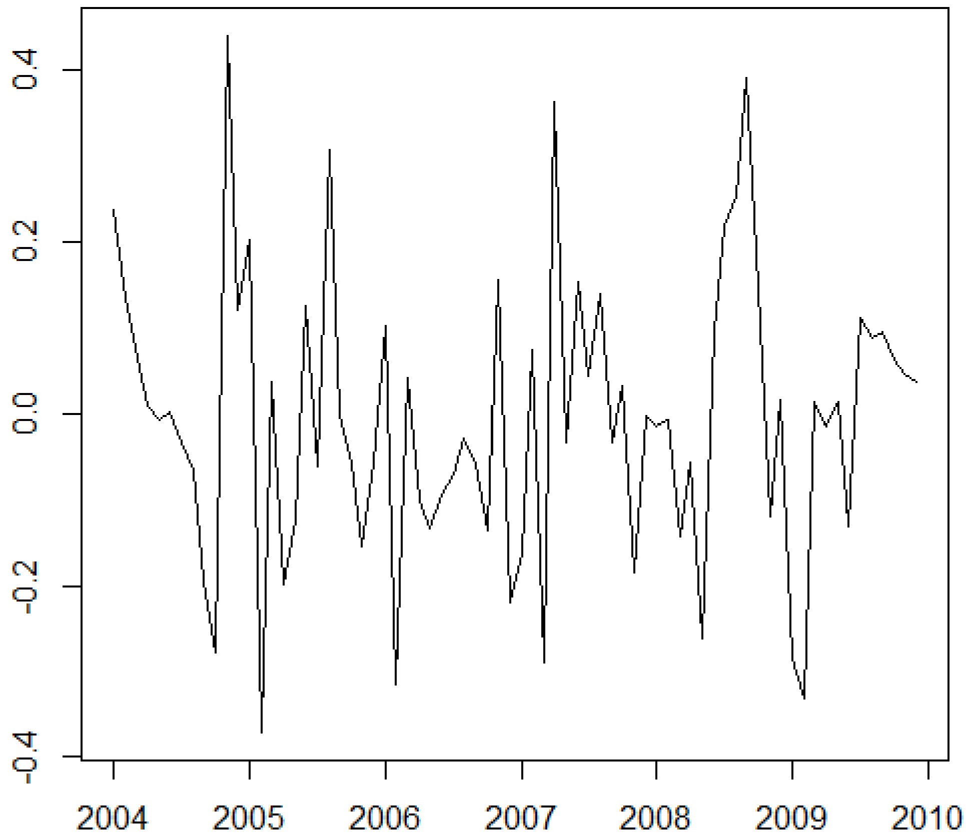
Διάγραμμα 3. 15 Απεικόνιση κορελογράμματος

Παρατηρούμε απο το κορελόγραμμα ότι τα σφάλματα πρόβλεψης του δείγματος δεν υπερβαίνουν πουθενά τα σημαντικά όρια.Επιπλέον η τιμή p του Ljung-Box είναι 0,6 αποδεικνύοντας ότι υπάρχουν ελάχιστα στοιχεία μη μηδενικών αυτοσυσχετίσεων στο διάστημα χρονικής υστέρησης 1 έως 20.

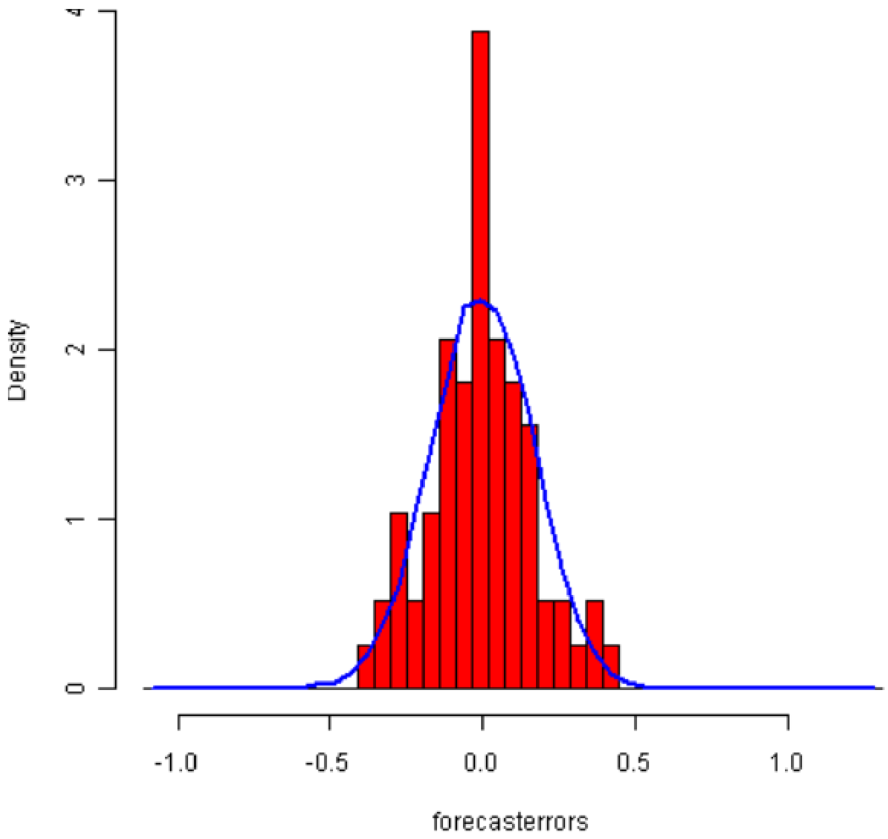
Αυτο που απομένει για να είμαστε σίγουροι για την επιτυχία της πρόβλεψης, είναι να ελένξουμε αν τα σφάλματα πρόβλεψης έχουν σταθερή διακύμανση στην πάροδο του χρόνου και αν είναι κανονικά κατανεμημένα με μέση τιμή 0 .

Γιαυτό τον λόγο παρακάτω θα απεικονίσουμε τα σφάλματα πρόβλεψης και το ιστόγραμμα με τις εντολές:

- plot.ts(shiptimeseriesforecasts2$residuals)
- plotForecastErrors(shiptimeseriesforecasts2$residuals)



Διάγραμμα 3. 16 Απεικόνιση σφαλμάτω πρόβλεψης



Διάγραμμα 3. 17 Απεικόνιση Ιστογράμματος

Πράγματι απο το διάγραμα 2.21 φαίνεται ότι τα σφάλματα πρόβλεψης έχουν σταθερή διακύμανση στην πάροδο του χρόνου.Απο την απεικόνιση του ιστογράμματος φαίνεται ότι τα σφάλματα πρόβλεψης έχουν είναι κανονικά κατανεμημένα με μέση τιμή 0.Αυτα τα στοιχεία λοιπόν δηλώνουν ότι η μέθοδος HoltWinters() παρέχει ένα επαρκές μοντέλο πρόβλεψης των χρόνων παράδοσης για τα προιόντα που μεταφέρονται με φορτηγά πλοία.

Κεφάλαιο 4

4.1 Μοντέλα ARIMA

Τα αυτοπαλινδρομικά μοντέλα κινητού μέσου όρου είναι στοχαστικά μαθηματικά με τα οποία προσπαθούμε να περιγράψουμε τη διαχρονική εξέλιξη κάποιου φυσικού μεγέθους. Δεδομένου ότι για την πλειοψηφία των φυσικών μεγεθών είναι αδύνατη η πλήρης γνώση και καταγραφή όλων των παραγόντων που επηρεάζουν την εξέλιξη τους στο χρόνο, είναι πολύ δύσκολη η διαχρονική περιγραφή του μεγέθους από ένα ντετερμινιστικό μοντέλο. Από την άλλη μεριά, η εξάρτηση τέτοιων μεγεθών από μη ντετερμινιστικούς παράγοντες (π.χ. καιρός, τυχαία γεγονότα) καθιστά δυνατή την περιγραφή της διαχρονικής τους εξέλιξης από ένα στοχαστικό μοντέλο, με το οποίο μπορεί να υπολογιστεί η πιθανότητα με την οποία η τιμή του μεγέθους βρίσκεται σε κάποιο διάστημα.

Τα στοχαστικά μοντέλα περιέχουν τον τυχαίο παράγοντα (τυχαίο σφάλμα ή σφάλμα πρόβλεψης), τις τιμές του μεγέθους οι οποίες εμφανίστηκαν σε προηγούμενες χρονικές στιγμές και ίσως κάποιους άλλους στοχαστικούς παράγοντες. Το μοντέλο που προκύπτει είναι ένας γραμμικός συνδυασμός των παραπάνω ποσοτήτων.

Γενικά, έχουν αναπτυχθεί πολλά και ποικίλα τέτοια μοντέλα για την περιγραφή των διακυμάνσεων κάποιου μεγέθους μέσα στο χρόνο. Τα μοντέλα ARIMA χρησιμοποιούνται ευρύτατα γιατί βρίσκουν εφαρμογή στη μελέτη πολλών μεγεθών και φαίνεται να δίνουν μια «καλή» εικόνα της διαχρονικής τους συμπεριφοράς, καθώς και ικανοποιητικά αποτελέσματα στην πρόβλεψη των μελλοντικών τιμών του μεγέθους.

Τα μοντέλα ARIMA έχουν μελετηθεί εκτεταμένα από τους Box και Jenkins, σε βαθμό που τα ονόματα των παραπάνω να είναι σχεδόν συνώνυμα με τις ARIMA διαδικασίες και τις εφαρμογές τους στην ανάλυση και την πρόβλεψη χρονοσειρών.

Οι Box-Jenkins πρότειναν μια οικογένεια αλγεβρικών μοντέλων πρόβλεψης, από τα οποία μπορεί κάποιος να διαλέξει το «καταλληλότερο» για την πρόβλεψη μιας δεδομένης χρονοσειράς. Στα μοντέλα αυτά οι προβλέψεις βασίζονται αποκλειστικά στις παρελθούσες τιμές και τα εμφανισθέντα πρότυπα συμπεριφοράς της χρονοσειράς που εξετάζεται.

4.1.1 Η Μεθοδολογία των Box & Jenkins

Η ανάπτυξη και η κατασκευή υποδειγμάτων ARIMA ως εργαλεία πρόβλεψης των τιμών οικονομικών μεταβλητών είναι γνωστή ως μεθοδολογία Box-Jenkins. Πρόκειται στην πραγματικότητα για γραμμικά στατιστικά μοντέλα που μπορούν να περιγράψουν ικανοποιητικά τις διάφορες συνιστώσες της χρονολογικής σειράς.

Μεθοδολογία Box-Jenkins Η προσέγγιση των Box-Jenkins στην ανάλυση χρονοσειρών είναι μια μέθοδος εύρεσης ενός στατιστικού υποδείγματος ARIMA που να παριστάνει ικανοποιητικά τη στοχαστική διαδικασία από την οποία προήλθαν τα δεδομένα, δηλαδή το δείγμα μας. Η μέθοδος αυτή εφαρμόζεται όταν η χρονοσειρά δεν είναι στάσιμη και περιλαμβάνει τέσσερα στάδια, την ταυτοποίηση (identification), την εκτίμηση (estimation), και το διαγνωστικό έλεγχο (diagnostic checking) και την πρόβλεψη(forecasting)και τα οποία θα αναλύσουμε στη συνέχεια.

4.1.2 Πρώτο Στάδιο: Ταυτοποίηση

Λέγοντας ταυτοποίηση του υποδείγματος εννοούμε ότι θα πρέπει να προσδιορισθούν: α) η τάξη της μη στασιμότητας β) η τάξη των AR και/ή πολυωνύμων Αυτό επιτυγχάνεται με σύγκριση της μορφής των δειγματικών συναρτήσεων αυτοσυσχετίσεως και μερικής αυτοσυσχετίσεως με τη μορφή θεωρητικών συναρτήσεων αυτοσυσχέτισης και μερικής αυτοσυσχέτισης που αντιστοιχούν σε διαδικασίες με άπειρο πλήθος όρων.

Πιο αναλυτικά, σε αυτό το στάδιο γίνεται η εξειδίκευση ενός ARIMA υποδείγματος με βάση τις πληροφορίες που παίρνουμε από το δείγμα. Αυτό σημαίνει ότι καθορίζονται οι τιμές των d,p και q. Δηλαδή, καθορίζεται ο αριθμός d των διαφορών που απαιτούνται για να μετατραπεί η σειρά σε στάσιμη, από τη στιγμή βέβαια που δεν είναι, και στη συνέχεια καθορίζεται η τάξη p της αυτοπαλίνδρομης διαδικασίας και η τάξη της q διαδικασίας κινητού μέσου.

Για να διαπιστωθεί αν η σειρά είναι στάσιμη ή όχι, θα εξεταστεί η συμπεριφορά της δειγματικής συνάρτησης αυτοσυσχέτισης. Αν οι αυτοσυσχετίσεις συγκλίνουν ταχύτατα προς το μηδέν σημαίνει ότι η σειρά μάλλον είναι στάσιμη. Αντίθετα, αν οι αυτοσυσχετίσεις φθίνουν με αργό ρυθμό, είναι σοβαρή ένδειξη ότι η σειρά είναι μη στάσιμη, οπότε πρέπει να γίνει στάσιμη. Σε αυτή την περίπτωση θα χρησιμοποιήσουμε τις πρώτες ή τις δεύτερες ή κ.τ.λ. διαφορές για να μετατραπεί η σειρά σε στάσιμη. Αφού η σειρά έχει γίνει στάσιμη, προσδιορίζεται στη συνέχεια η τάξη του υποδείγματος ARIMA, δηλαδή προσδιορίζονται οι τιμές του p και q του. Ο προσδιορισμός τους βασίζεται στις δειγματικές απλές και μερικές, αυτοσυσχετίσεις.

4.1.3 Δεύτερο Στάδιο: Εκτίμηση

Αφού καθορίσουμε το υπόδειγμα ARIMA(p,d,q) στο προηγούμενο στάδιο της αναγνώρισης, προχωρούμε στην εκτίμηση των παραμέτρων του με την μέθοδο των ελαχίστων τετραγώνων (OLS). Εδώ είναι αναγκαίο να σημειωθεί ότι, όταν στο μοντέλο συμπεριλαμβάνονται MA όροι ($q > 0$), είναι δύσκολη η εφαρμογή της OLS. Στην περίπτωση αυτή, μπορεί να χρησιμοποιηθεί μια επαναληπτική μέθοδος. Αφού επιλεχθούν αρχικές τιμές για τις παραμέτρους, στη συνέχεια βελτιώνονται μέσω μιας επαναληπτικής διαδικασίας μέχρι να ελαχιστοποιηθεί το άθροισμα των τετραγώνων των σφαλμάτων. Μια άλλη επαναληπτική μέθοδος η οποία χρησιμοποιείται συχνά είναι η μέθοδος μεγίστης πιθανοφάνειας. Η πιθανοφάνεια ενός συνόλου δεδομένων συμβολίζεται με L και είναι ανάλογη με την πιθανότητα να παραχθούν τα πραγματικά δεδομένα από το μοντέλο. Η μέθοδος μεγίστης πιθανοφάνειας χρησιμοποιείται για την εκτίμηση των τιμών των παραμέτρων που μεγιστοποιούν την πιθανότητα L. Κάθε εκτιμούμενος συντελεστής έχει ένα τυπικό σφάλμα γιατί είναι μια στατιστική τιμή που βασίζεται σε πληροφορία από ένα μόνο δείγμα. Ένα διαφορετικό δείγμα πιθανόν να έδινε διαφορετικές εκτιμήσεις για τους συντελεστές του μοντέλου. Η σημαντικότητα των συντελεστών ελέγχεται μέσω του στατιστικού δείκτη t. Στην πράξη απορρίπτουμε κάθε εκτιμούμενη τιμή συντελεστή με απόλυτη t-τιμή μικρότερη του 2.0. Κάθε συντελεστής με απόλυτη τιμή $t < 2$ δεν είναι σημαντικά διάφορος του μηδενός για επίπεδο σημαντικότητας 5% και οδηγεί στη δημιουργία μη φειδωλών μοντέλων και άρα σε λιγότερο ακριβείς προβλέψεις. Τα περισσότερα στατιστικά υπολογιστικά πακέτα προσαρμόζουν αυτόματα ένα μοντέλο ARIMA σε μια χρονοσειρά, εκτελούν όλους τους αναγκαίους στατιστικούς ελέγχους και παράγουν εκθέσεις με τις τιμές των παραμέτρων, το τυπικό τους σφάλμα, την t-τιμή τους, το επίπεδο σημαντικότητας, την διακύμανση των υπολοίπων και γενικότερα τις τιμές όλων των στατιστικών δεικτών που χρησιμοποιούνται για τον έλεγχο της καταλληλότητας του μοντέλου

4.1.4 Διαγνωστικός έλεγχος

Αν και το επιλεγμένο μοντέλο θεωρείται το καλύτερο από αυτά που εξετάσθηκαν, είναι επίσης αναγκαίο να επιβεβαιωθεί η επάρκεια του μοντέλου. Οι διαγνωστικοί έλεγχου προσδιορίζουν αν το μοντέλο είναι επαρκές ή απαιτείται ανασκευή του. Εδώ εξετάζουμε δύο κυρίως θέματα.

4. Αν το μοντέλο είναι υπέρ-επαρκές (Overfitting), αν δηλαδή έχει γίνει εφαρμογή μεγαλύτερου μοντέλου από αυτού που χρειάζεται για να απεικονιστεί επαρκώς η δυναμική συμπεριφορά των δεδομένων όπως αυτή παρουσιάζεται στο βήμα 1. Έτσι, αν το μοντέλο είναι επαρκές, κάθε επιπλέον προστιθέμενος όρος θα πρέπει να είναι στατιστικά μη σημαντικός.

II. Διάγνωση των καταλοίπων: Συνήθως ελέγχουμε αν υπάρχει γραμμική εξάρτιση (αυτοσυσχέτιση) μεταξύ των καταλοίπων η οποία δηλώνει ότι το μοντέλο είναι ανεπαρκές για να αναπαράγει με ακρίβεια τη δυναμική συμπεριφορά των δεδομένων. Τα κατάλοιπα ενός καλού μοντέλου πρόβλεψης πρέπει να είναι «λευκός θόρυβος» και συνεπώς οι ACF και PACF των υπολοίπων δεν πρέπει να παρουσιάζουν στατιστικά σημαντικές αυτοσυσχετίσεις και μερικές αυτοσυσχετίσεις αντίστοιχα. Εάν τα κατάλοιπα δεν είναι λευκός θόρυβος τότε το μοντέλο είναι ανεπαρκές και πρέπει να εξετασθούν άλλα μοντέλα ARIMA. Το πρότυπο που ακολουθούν οι στατιστικά σημαντικοί συντελεστές αυτοσυσχέτισης και μερικής αυτοσυσχέτισης των καταλοίπων, υποδεικνύουν τον τρόπο βελτίωσης του μοντέλου. Για παράδειγμα σημαντικές τιμές για εποχιακές καθυστερήσεις υποδεικνύουν την προσθήκη μιας εποχιακής συνιστώσας ή σημαντικές τιμές για μικρές καθυστερήσεις υποδεικνύουν την αύξηση των μη εποχιακών AR ή MA συνιστωσών του μοντέλου

Ο σκοπός μας γενικά είναι να δημιουργήσουμε ένα απέριττο (parsimonious) μοντέλο δηλαδή ένα μοντέλο που να περιγράφει όλα τα χαρακτηριστικά των δεδομένων χρησιμοποιώντας όσο το δυνατό λιγότερες παραμέτρους (δηλαδή όσο το δυνατό πιο απλό μοντέλο γίνεται). Υπάρχουν αρκετοί λόγοι γι’ αυτό, οι σημαντικότεροι εκ των οποίων αναφέρονται παρακάτω:

4) Η διακύμανση των εκτιμητών είναι αντιστρόφως ανάλογη των βαθμών ελευθερίας. Ένα μοντέλο που περιλαμβάνει άσχετες υστερήσεις των μεταβλητών ή των λαθών (και επομένως μη αναγκαίες παραμέτρους) θα οδηγήσει σε αύξηση των τυπικών λαθών των συντελεστών με συνέπειες να είναι δύσκολο να βρεθούν στατιστικά σημαντικές σχέσεις στα δεδομένα. Β) Είναι πολύ πιθανό ένα μεγάλο (με πολλές παραμέτρους) μοντέλο να προσαρμόζεται πολύ καλά στα δεδομένα με υψηλό R2 αλλά να δίνει πολύ ανακριβείς προβλέψεις. Δηλαδή, όπως και στη φυσική, θα πρέπει να γίνει διάκριση μεταξύ «σήματος» (signal) και «θορύβου» (noise). Αυτό σημαίνει ότι θα πρέπει να προσαρμόζουμε ένα υπόδειγμα (μοντέλο) στα δεδομένα, που θα παρουσιάζει το «σήμα» (τα σημαντικά χαρακτηριστικά των 96 δεδομένων ή την υποκείμενη τάση τους) και όχι τον «θόρυβο» (τα εντελώς τυχαία φαινόμενα της χρονολογικής σειράς).

4.1.5 Τέταρτο Στάδιο: Μεταδιάγνωση (Πρόβλεψη)

Το γεγονός ότι ένα δοκιμαστικό υπόδειγμα δεν απορρίφθηκε από το διαγνωστικό έλεγχο δε σημαίνει ότι μπορεί αυτόματα να γίνει αποδεκτό, καθώς είναι πιθανό να υπάρχουν και άλλα υποδείγματα που να ανταποκρίνονται στις απαιτήσεις των σταδίων 2 και 3. Έτσι δυνατόν να έχουμε περισσότερα του ενός αποδεκτά κατ’ αρχήν υποδείγματα. Στο στάδιο της μεταδιάγνωσης επιλέγεται τελικά εκείνο το υπόδειγμα το οποίο εμφανίζει την καλύτερη προσαρμογή, ή/και την καλύτερη προβλεπτική ικανότητα.

Έλεγχος της Τάξης του υποδείγματος

Η καταλληλότητα του εκτιμόμενου υποδείγματος ελέγχεται συγκρίνοντας το με ένα άλλο υπόδειγμα μεγαλύτερης τάξης. Δηλαδή, το εκτιμόμενο υπόδειγμα ARMA (p, q) συγκρίνεται με τα υποδείγματα ARMA (p+1, q) και ARMA (p, q+1). Αν το υπόδειγμα που εκτιμήθηκε περιγράφει τη διαδικασία που παρήγαγε τα δεδομένα, οι επιπλέον συντελεστές στα μεγαλύτερα υποδείγματα δεν θα πρέπει να είναι στατιστικά διαφορετικοί από το μηδέν. Η 3παραπάνω διαδικασία ελέγχου ονομαζεται υπερπροσαρμογη.

4.2 Επιλογή κατάλληλου μοντέλου ARIMA

Μετά την εκτίμηση των παραμέτρων ενός μοντέλου ARIMA είναι αναγκαία η εκ νέου αναγνώριση προκειμένου να διαπιστωθεί εάν το επιλεγμένο μοντέλο μπορεί να βελτιωθεί. Συγκεκριμένα σε αυτό το στάδιο της διαδικασίας μοντελοποίησης πρέπει:

- Εάν προκύψουν συντελεστές στατιστικά μη σημαντικοί, οι αντίστοιχοι όροι πρέπει να αφαιρεθούν από το μοντέλο.
- Οι ACF και PACF παρέχουν κάποια καθοδήγηση στην επιλογή ενός απλού AR ή MA μοντέλου. Εάν το καταλληλότερο μοντέλο είναι ένα σύνθετο ARMA μοντέλο, αυτό είναι πολύ δύσκολο να αναγνωρισθεί από τις ACF και PACF. Μετά την επιλογή ενός απλού μοντέλου πρέπει να μελετηθεί εάν αυτό μπορεί να επεκταθεί σε ένα σύνθετο ARMA μοντέλο.
- Εάν έχουν εκτιμηθεί περισσότερα από ένα «καλά» μοντέλα, πρέπει να εφαρμοσθεί μια μέθοδος επιλογής του καλύτερου από αυτά.

Επειδή πολλές φορές τα δεδομένα είναι ακατάστατα, αυτό έχει σαν αποτέλεσμα η γραφική τεχνική με τη χρήση των ACF και PACF να μην μπορεί να δώσει ακριβή πληροφόρηση για τη δημιουργία του σωστού μοντέλου ARIMA. Για την αποφυγή της υποκειμενικότητας που υπάρχει με τις ACF και PACF χρησιμοποιούνται τα κριτήρια πληροφορίας (information criteria). Τα κριτήρια πληροφορίας ενσωματώνουν δύο παράγοντες: έναν όρο που είναι συνάρτηση του αθροίσματος των τετραγώνων των καταλοίπων και μια ποινή για τη μείωση των βαθμών ελευθερίας λόγω της πρόσθεσης επιπλέον παραμέτρων. Έτσι η πρόσθεση νέας μεταβλητής ή νέας χρονικής υστέρησης στο μοντέλο θα έχει δύο αντίθετα αποτελέσματα στα κριτήρια πληροφορίας: η τιμή των αθροισμάτων των τετραγώνων των καταλοίπων θα μειωθεί αλλά θα αυξηθεί η τιμή της ποινής λόγω απώλειας των βαθμών ελευθερίας. Επομένως ο αντικειμενικός σκοπός είναι να επιλέξουμε των αριθμό των παραμέτρων ο οποίος μεγιστοποιεί την τιμή των κριτηρίων πληροφορίας.

Τα πιο δημοφιλή κριτήρια πληροφορίας, που έχουν αναφερθεί και προηγουμένως, είναι: Akaike΄s information criterion (AIC), Schwarz΄s Bayesian information criterion, και Hannan-Quinn criterion (HQ).

Τα κριτήρια πληροφορίας ουσιαστικά μεγιστοποιούνται υπό τον περιορισμό $p \leq p(\varpi)$, $q \leq q(\varpi)$, δηλαδή ένα ανώτερο όριο ορίζεται για τον αριθμό του MA(q) και του AR(p) μέρους του μοντέλου ARIMA.

Όσον αφορά την αυστηρότητα της ποινής για την μείωση των βαθμών ελευθερίας το SBIC έρχεται πρώτο μετά το HQ και τέλος το AIC. Θα πρέπει να σημειωθεί ότι το R μπορεί να θεωρηθεί ως κριτήριο πληροφορίας αν και είναι το πιο επιεικές από όλα τα προηγούμενα κριτήρια και επιλέγει το μεγαλύτερο μοντέλο από όλα (δηλαδή τον μεγαλύτερο αριθμό χρονικών υστερήσεων).

4.2.1 Διαφοροποίηση χρονοσειράς

Τα μοντέλα ARIMA ορίζονται για χρονοσειρές που είναι στάσιμες, ωστόσο στην περίπτωση που έχουμε μη στάσιμη χρονοσειρά μπορούμε να διαφοροποιήσουμε την χρονοσειρά μας μέχρι να έχουμε μία σταθερή χρονοσειρά.Αν πρέπει να διαφοροποιήσουμε d φορές την χρονοσειρά μας για να έχουμε μία στάσιμη χρονοσειρά τότε έχουμε ένα μοντέλο ARIMA (p,d,q) όπου η μεταβλητή d είναι η σειρά διαφοροποίησης που χρησιμοποιούμε.

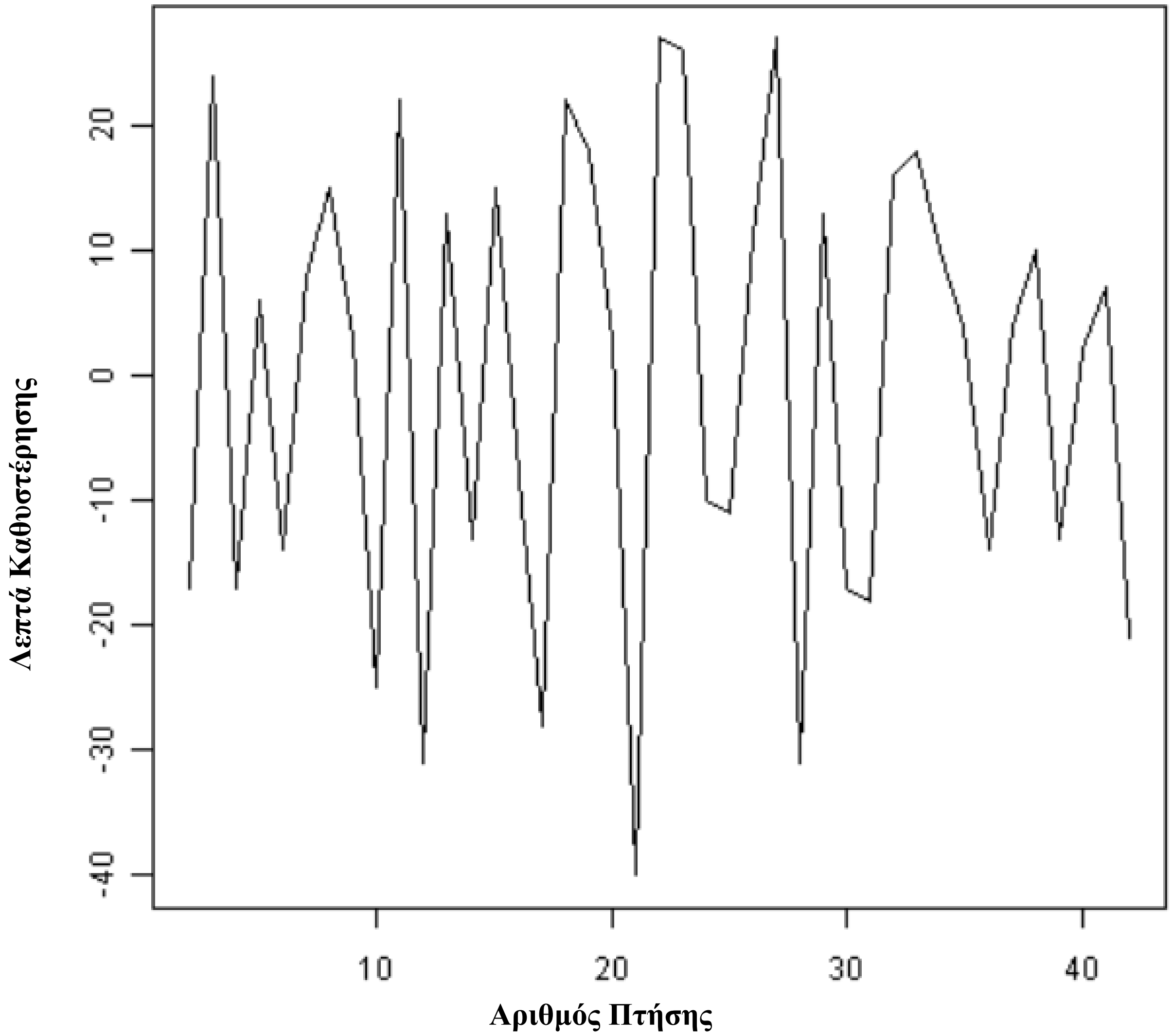
Στο εργαλείο R για να διαφοροποιήσουμε την χρονοσειρά χρησιμοποιούμε την εντολή **diff().**Πιο συγκεκριμένα θα αναφερθούμε στο παράδειγμα που αναφέρεται στις καθυστερήσεις 42 πτήσεων.Στο συγκεκριμένο παράδειγμα είχαμε προβεί σε πρόβλεψη με την μέθοδο της εκθετικής εξομάλυνσης HoltWinters για τον λόγο ότι τα δεδομένα της χρονοσειράς ήταν μη στάσιμα.Αυτό μάλιστα φαίνεται ξεκάθαρα αν κοιτάξουμε το διάγραμμα 2.1.

Θα επιχειρήσουμε λοιπόν να διαφοροποιήσουμε την χρονοσειρά σε στάσιμη ,έτσι ώστε να εφαρμόσουμε την μέθοδο ARIMA.Εισάγουμε στο εργαλείο R τις εξής εντολές:

- datatimeseriesdiff1 <- diff(datatimeseries, differences=1)
- plot.ts(datatimeseriesdiff1)

4.2.2 Απεικόνιση διαφοροποιημένης χρονοσειράς

Αφού εισάγουμε στο εργαλείο R την εντολή ώστε να μετατρέψουμε σε στάσιμη την χρονοσειρά, το επόμενο βημα είναι να απεικονίσουμε την διαφοποιημένη χρονοσειρά για να διαπιστώσουμε αν τελικά πράγματι έχουμε μια στάσιμη χρονοσειρά.Ακολουθεί η απεικόνιση:



Διάγραμμα 4. 1 Απεικόνιση πρώτης διαφοροποίησης της αρχικής χρονοσειράς

Πράγματι φαίνεται να προκύπτει ότι απο την πρώτη διαφοροποίηση η χρονοσειρά μας έχει μετατραπεί σε μία στάσιμη χρονοσειρά σε μέση τιμή και διασπορά. Θα μπορούσαμε να πούμε ότι το μοντέλο ARIMA (p,1,q) είναι το κατάλληλο για την δική μας περίπτωση. Τώρα αυτό που μένει είναι να προσπαθήσουμε να βρούμε αν υπάρχουν συσχετισμοί μεταξύ τως διαδοχικών όρων και του ακανόνιστου στοιχείου.

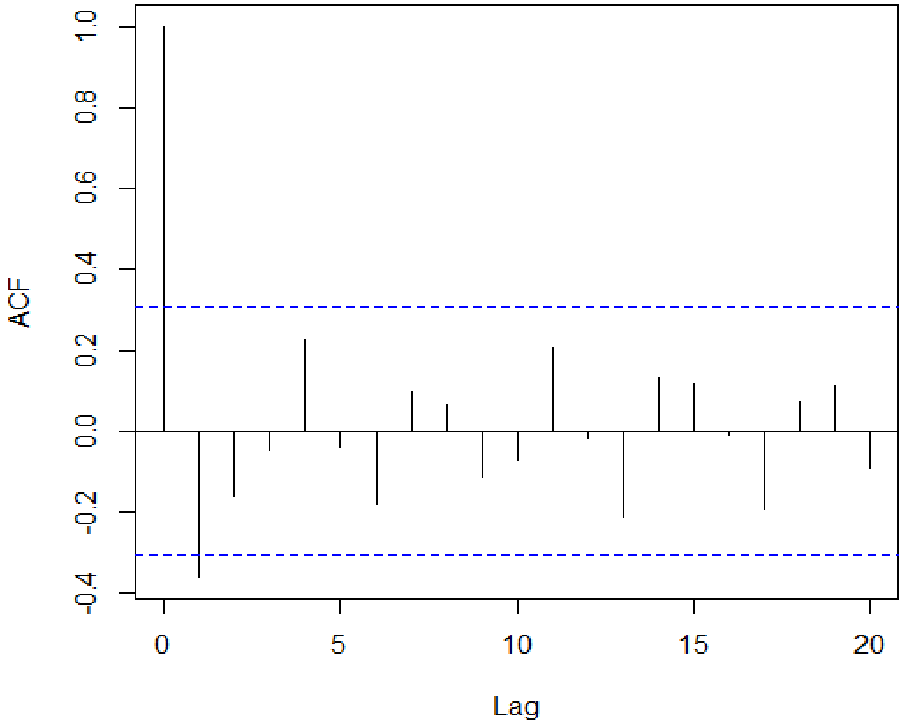
4.2.3 Διαδικασία Επιλογής Μοντέλου ARIMA για παράδειγμα με καθυστερήσεις πτήσεων

Γνωρίζουμε ότι όταν η χρονοσειρά μας είναι σταθερή ή αν την έχουμε διαφοροποιήσει σε σταθερή ,το επόμενο βήμα είναι να επιλέξουμε το κατάλληλο μοντέλο ARIMA. Αυτό συνεπάγεται την βέλτιστη εύρεση τιμών των παραμέτρων p, q του μοντέλου.Για να επιτευχθεί κάτι τέτοιο θα πρέπει να απεικονίσουμε το **κορελόγραμμα** και το **μερικό κορελόγραμμα** της στάσιμης χρονολογικής σειράς.

Για να απεικονίσουμε το κορελόγραμμα και το μερικό κορελόγραμμα χρησιμοποιούμε τις εντολές **'acf' ,'pacf'** αντίστοιχα στο εργαλείο R. Πρώτα για το κορελόγραμμα πληκτρολογούμε στο εργαλείο R τις εξής εντολές:

- acf(datatimeseriesdiff1, lag.max=20) # plot a correlogram
- acf(datatimeseriesdiff1, lag.max=20, plot=FALSE) # get the autocorrelation values
- Autocorrelations of series 'kingtimeseriesdiff1', by lag

Έπειτα έχουμε την απεικόνιση του κορελογράμματος:



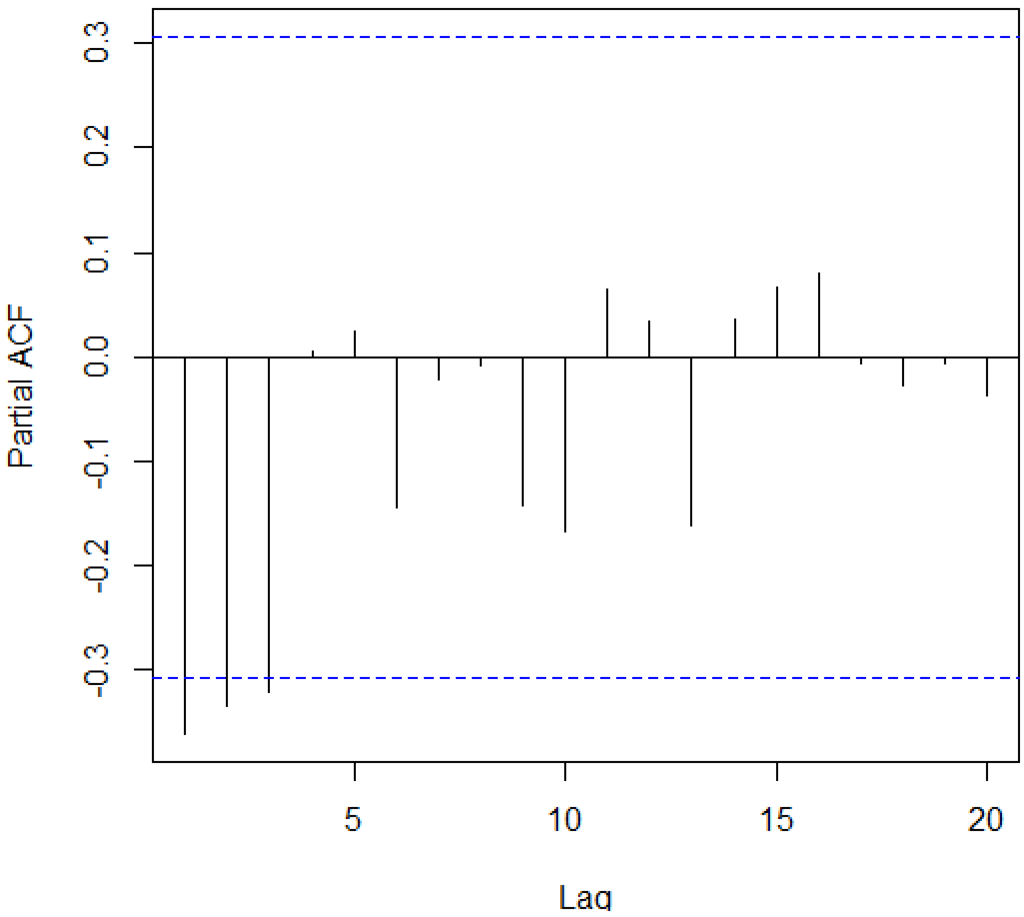
Διάγραμμα 4. 2 Απεικόνιση κορελογράμματος

Παρατηρούμε απο το κορελόγραμμα ότι η αυτοσυσχέτιση στην διακύμανση 1 είναι -0,36 και ξεπερνάει τα επιτρεπτά όρια,αλλά όλοι οι υπόλοιποι αυτοσυσχετισμοί δεν υπερβαίνουν τα επιτρεπτά όρια.

Για να απεικονίσουμε το μερικό κορελόγραμμα εισάγουμε στο εργαλείο R τις εξής εντολές:

- pacf(datatimeseriesdiff1, lag.max=20) # plot a partial correlogram
- pacf(datatimeseriesdiff1, lag.max=20, plot=FALSE) # get the partial autocorrelation values

Έχουμε την απεικόνιση του μερικού κορελογράμματος:



Διάγραμμα 4. 3 Απεικόνιση μερικού κορελογράμματος

Το μερικό κορελόγραμμα παρουσιάζει στις καθυστερήσεις 1,2 και 3 τους αυτοσυσχετισμους να υπερβαίνουν τα επιτρεπτά όρια .Πιο συγκεκριμένα στα σημεία αυτα οι αυτοσυσχετισμοί είναι αρνητικοί ,και μειώνονται αργά όσο μεγαλώνουν οι διακυμάνσεις.

Συνοψίζοντας τα στοιχεία του κορελογράμματος και του μερικού κορελογράμματος γνωρίζουμε ότι:

1. Στο κορελόγραμμα απο την διακύμανση 1 και μετά οι αυτοσυσχετισμοί τείνουν να μηδενίστούν.
2. Στο μερικό κορελόγραμμα απο την διακύμανση 3 και έπειτα οι αυτοσυσχετισμοί τείνουν να μηδενιστούν.

Αυτό σημαίνει ότι τα παρακάτω ARMA μοντέλα είναι πιθανά για την χρονοσειρά μας.

- Ένα ARMA (3,0) μοντέλο.Δηλαδή ένα μοντέλο αυτοπαλινδρόμησης p=3,δεδομένου ότι στο μερικό κορελόγραμμα οι αυτοσυσχτίσεις είναι 0 μετά την καθυστέρηση 3 και στο κορελόγραμμα έχουν τάση μηδενισμού όσο μεγαλώνει η καθυστέρηση.
- Ένα ARMA (0,1) μοντέλο ,δηλαδή ένα μοντέλο κινητού μέσου όρου της ταξης q=1,το κορελόγραμμα είναι 0 μετα την καθυστέρηση 1 και στο μερικό κορελόγραμμα οι αυτοσυσχτετισμοί τείνουν να μηδενιστούν όσο μεγαλώνει η καθυστέρηση.
- Ενα ARMA (p,q) μοντέλο δηλαδή ένα συνδυαστικό μοντέλο με παραμέτρους p,q μεγαλύτερους απο το 0 ,δεδομένου ότι και στα δύο κορελογράμμματα οι αυτοσυσχετισμοί τείνουν να μηδενιστούν όσο μεγαλώνει η καθυστέρηση.

4.2.4 Συμπέρασμα Επιλογής Μοντέλου ARIMA για παράδειγμα με καθυστερήσεις πτήσεων

Θα μπορούσαμε να πούμε ότι θα χρησιμοποιήσουμε την αρχή της φειδούς για να αποφασίσουμε ποιο μοντέλο είναι το καλύτερο.Πιο συγκεκριμένα θα επιλέξουμε το μοντέλο με τις λιγότερες παραμέτρους .Το ARMA (3,0) μοντέλο έχει 3 παραμέτρους ,το ARMA (0,1) μοντέλο έχει 1 παράμετρο και τέλος το ARMA (p,q) έχει τουλάχιστον 2 παραμέτρους.

Επομένως φτάνουμε στο συμπέρασμα ότι η πιο κατάλληλη επιλογή είναι το μοντέλο **ARMA (0.1)**.

Ένα μοντέλο ARMA (0,1) είναι ένα μοντέλο κινητού μέσου όρου της τάξης του 1, ή MA (1) μοντέλο.Αυτό το μοντέλο μπορεί να περιγραφεί ως:

- : X_t – mu = Z_t – (Θ * Z_t-1)

Όπου: :

- **X_t** είναι η στάσιμη χρονολογική σειρά που μελετάμε,δηλαδή η πρώτη διαφοροποιημένη χρονολογική σειρά με τις καθυστερήσεις των πτήσεων.
- **Mu** είναι η μέση τιμή της χρονοσειράς **X_t**
- **Z_t** είναι λευκός θόρυβος με μέση τιμή μηδέν και σταθερή διακύμανση.
- **Θ** είναι μία παράμετρος που μπορεί να εκτιμηθεί.

Ενα μοντέλο κινητού μέσου όρου συνήθως χρησιμοποιείται για να διαμορφώσει μια χρονοσειρά που δείχνει βραχυπρόθεσμη εξαρτήση μεταξύ διαδοχικών παρατηρήσεων.

Εφόσον φτάσαμε στην επιλογή του **ARMA(0,1)** μοντέλου της πρώτης διαφοροποιημένης χρονοσειράς, συνεπάγεται ότι στην αρχική μας χρονοσειρά δηλαδή πρωτού την διαφοροποιήσουμε ,μπορούμε να την μοντελοποιήσουμε με το μοντέλο **ARIMA(0,1,1)** (με p=0,d=1,q=1) όπου d είναι ο αριθμός των διαφοροποιήσεων.

4.3 Πρόβλεψη με την χρήση Μοντέλου ARIMA(0,1,1)

Μετά την επιλογή του κατάλληλου μοντέλου ARIMA μπορούμε να υπολογίσουμε τις παραμέτρους p,d,q του μοντέλου και να το χρησιμοποιήσουμε για ένα μοντέλο πρόβλεψης.Εφόσον έχουμε υπολογίσει πιο πάνω τις παραμέτρους αυτές φτιάχνουμε στο εργαλείο R το μοντέλο πρόβλεψης.

Επομένως πληκτρολογούμε στο εργαλείο R τις εξής εντολές:

- datatimeseriesarima <- arima(datatimeseries, order=c(0,1,1)) # fit an ARIMA(0,1,1) model
- datatimeseriesarima

  ARIMA(0,1,1)

Έπειτα για να προβούμε σε πρόβλέψεις όπως για παράδειγμα ας πούμε τις 5 επομενες πτήσεις πληκτρολογούμε στο εργαλείο R:

- library("forecast") # load the "forecast" R library
- datatimeseriesforecasts <- forecast.Arima(datatimeseriesarima, h=5)
- datatimeseriesforecasts

Έτσι έχουμε τις εξής προβλέψεις:

Πίνακας 4. 1:Αποτελέσματα καθυστέρησης για τις επόμενες 5 πτήσεις

	Point Forecast	Lo 80	Hi 80	Lo 95	Hi 95
43	67.75063	48.29647	87.20479	37.99806	97.50319
44	67.75063	47.55748	87.94377	36.86788	98.63338
45	67.75063	46.84460	88.65665	35.77762	99.72363
46	67.75063	46.15524	89.34601	34.72333	100.77792
47	67.75063	45.48722	90.01404	33.70168	101.79958

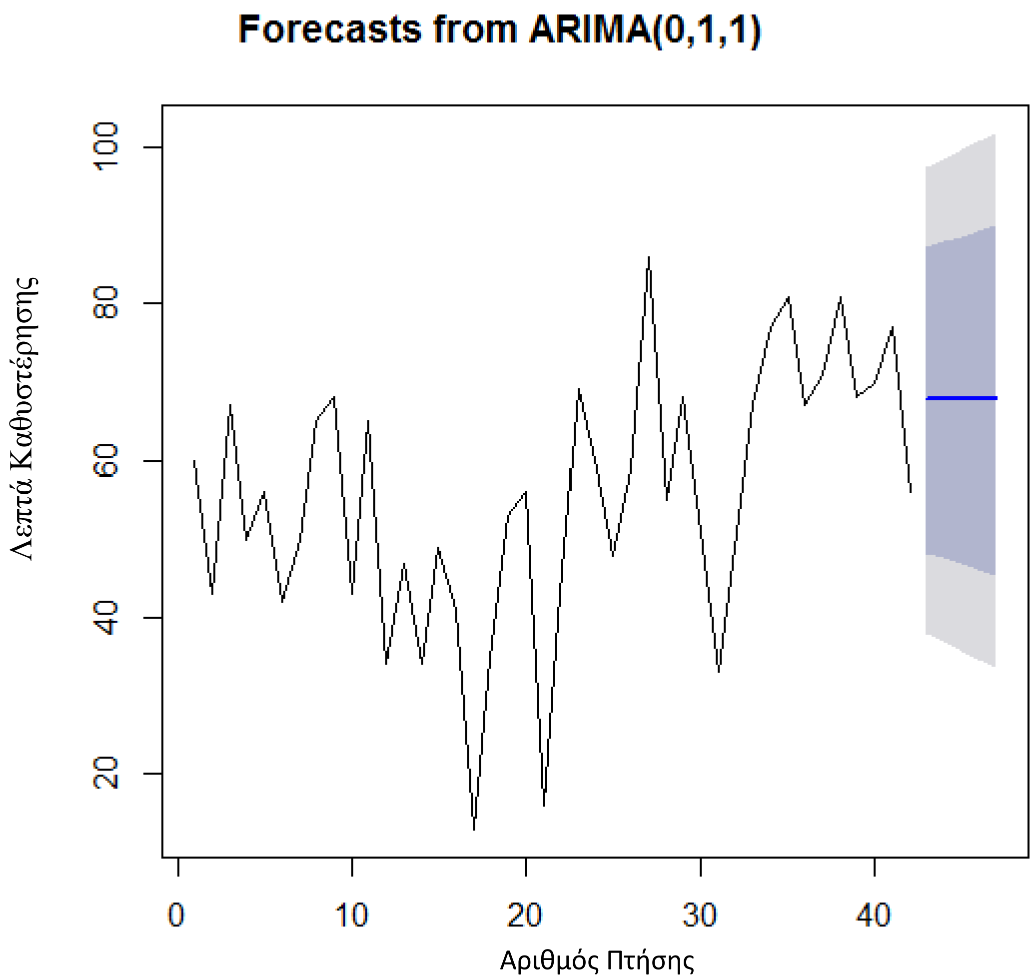
Η τιμή καθυστέρησης της τελευταίας πτήσης ήταν 56 λεπτά και η πρόβλεψη για τις επόμενες 5 πτήσεις είναι 67 λεπτά.

4.3.1 Απεικόνιση πρόβλεψης μοντέλου ARIMA(0,1,1)

Μπορούμε να απεικονίσουμε την αρχική χρονοσειρά των πτήσεων και τις αντίστοιχες προβλέψεις που κάναμε παραπάνω ,πληκτρολογώντας στο εργαλείο R την εξής εντολή:

- plot.forecast(datatimeseriesforecasts)

Έτσι προκύπτει η παρακάτω γραφή απεικόνιση:



Διάγραμμα 4. 4 Απεικόνιση μοντέλου πρόβλεψης ARIMA(0,1,1)

Με την μπλε γραμμή απεικονίζεται η πρόβλεψη του μοντέλου ARIMA (O,1,1,) για τις επόμενες 5 πτήσεις και αντιστοιχή στα 67 λεπτά.Όπως είχαμε αναλύσει και παραπάνω η έντονα σκιαγραφημένη περιοχή αντιστοιχή στο 80% της ελάχιστης και της μέγιστης τιμής πρόβλεψης αντίχτοιχα.Ενω η πιο ανοιχτή σκιαγραφημένη περιοχή αντιστοιχεί στο 95% της μέγιστης και ελάχιστης τιμής πρόβλεψης αντίστοιχα.

4.3.2 Έλεγχος αξιοπιστίας αποτελέσματος πρόβλεψης με την δοκιμή Ljung-Box

Όπως είναι ήδη γνωστό στην περίπτωση που έχουμε προβεί σε μια πρόβλεψη ανεξάρτητα απο το μοντέλο και την μέθοδο πρόβλεψης που έχουμε επιλέξει, για να είμαστε αξιόπιστοι με τα αποτελέσματα μας μπορούμε να κάνουμε την δοκιμή **Ljung-Box** και έπειτα να ελένξουμε αν τα σφάλματα των προβλέψεων είναι **κανονικά κατανεμημένα με μέση τιμή 0**.

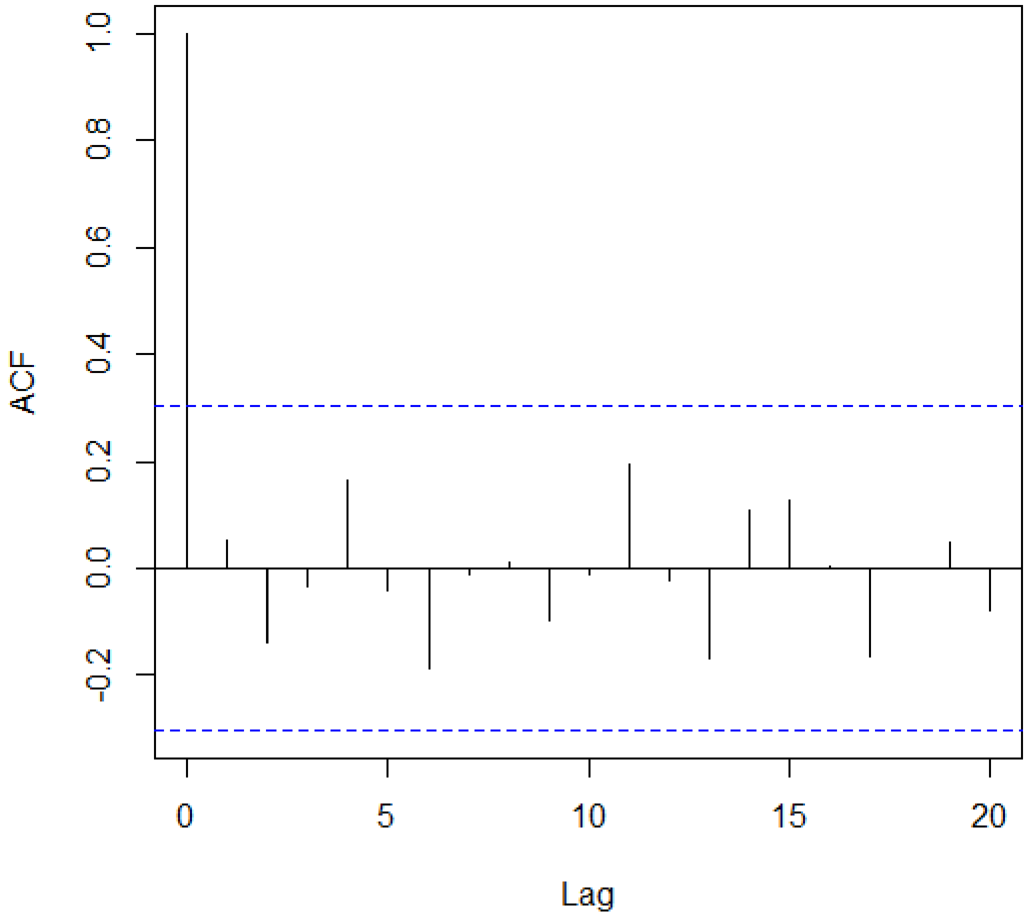
Η δοκιμή Ljung-Box θα μας δώσει σαν αποτέλεσμα στοιχεία για το αν υπάρχουν **μη μηδενικοί αυτοσυσχετισμοί**.Στην περίπτωση που υπάρχουν θα πρέπει να επαναεξετάσουμε την μέθοδο η το μοντέλο πρόβλεψης που έχουμε επιλέξει,γιατί δεν πρέπει να υπάρχουν συσχετισμοί. Για το αν υπάρχουν αυτοσυσχετισμοί είναι εύκολο να το εξετάσουμε με την απεικόνιση των σφαλμάτων πρόβλεψης σε ένα κορελόγραμμα.

Προχωράμε λοιπόν στις εντολές που εισάγουμε στο εργαλείο R για την απεικόνιση του κορελογράμματος:

- acf(datatimeseriesforecasts$residuals, lag.max=20)
- Box.test(datatimeseriesforecasts$residuals, lag=20, type="Ljung-Box")
- Box-Ljung test data: kingstimeseriesforecasts$residuals
  X-squared = 13.5844, df = 20, p-value = 0.851

Αφού πληκτρολογήσουμε τις εντολές στο εργαλείο R παρατηρούμε ότι μας επιστρέφει το πρόγραμμα την τιμή p-value=0.851 που είναι τα στοιχεία μη μηδενικών αυτοσυσχετίσεων ,κάτι που είναι ενθαρυντικό ,γιατί η τιμή είναι πολυ μικρή.

Ας δούμε βέβαια και την απεικόνιση που παίρνουμε απο το πρόγραμμα R.



Διάγραμμα 4. 5 Απεικόνιση κορελογράμματος για έλεγχο μη μηδενικών αυτοσυσχετισμών

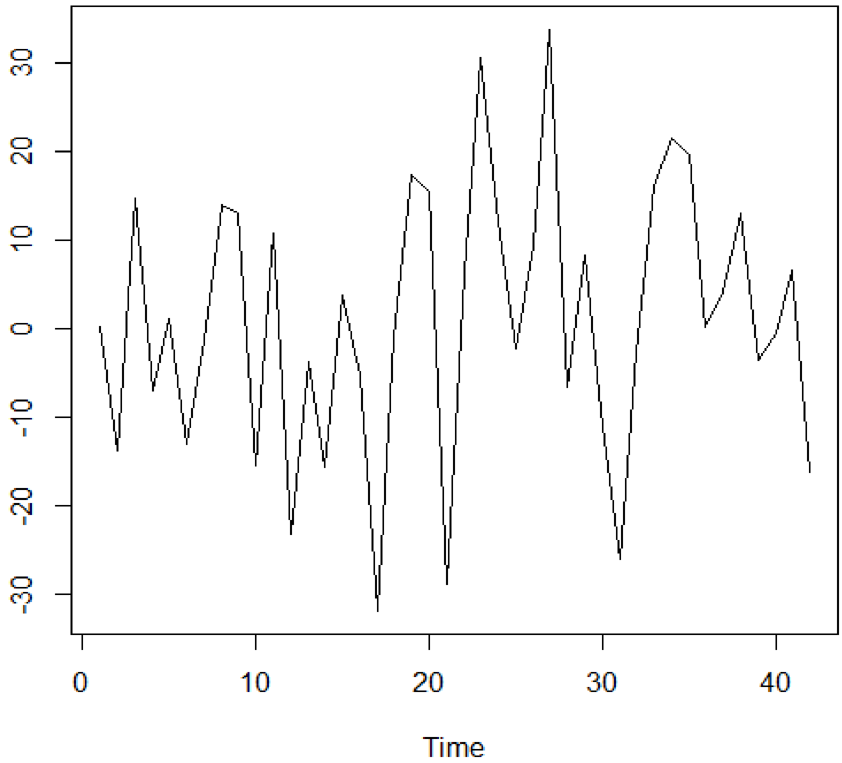
Πράγματι παρατηρούμε απο το Διάγραμμα 2.27 ότι σε κανένα σημείο του δείγματος των αυτοσυσχετίσεων για καθυστέρηση 1 – 20 δεν υπερβαίνει τα επιτρεπτά όρια.Επομένως δεν έχουμε μη μηδενικούς αυτοσυσχετισμούς κάτι που σημαίνει ότι μέχρι αυτό το σημείο οι προβλέψεις μας είναι αξιόπιστες.

Σαν πρώτο βήμα για την αξιοπιστία των προβλέψεων ελένξαμε αν τα σφάλματα των προβλέψεων εμφάνιζαν μη μηδενικούς αυτοσυσχετισμούς και καταλήξαμε πράγματι ότι δεν υπάρχουν.Για να είμαστε ακόμη πιο σίγουροι για την αξιοπιστία των αποτελεσμάτων της πρόβλεψης αυτό που απομένει να κάνουμε είναι να απεικονίσουμε τα σφάλματα πρόβλεψης και ένα ιστόγραμμα.Αυτο που θα θέλαμε ιδανικά να δούμε θα ήταν μια **κανονικά κατανεμημένη απεικόνιση** με τα σφάλματα πρόβλεψης και ένα ιστόγραμμα των σφαλμάτων πρόβλεψης **με μέση τιμή κοντά στο 0.**

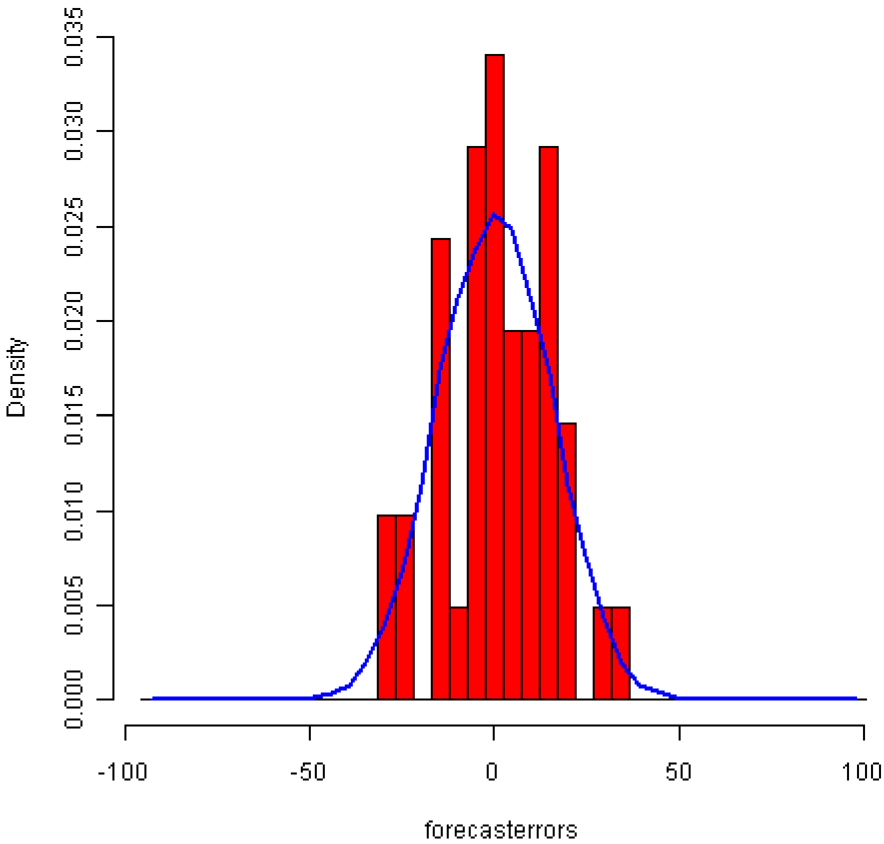
Προχωράμε με τιε εντολές που εισάγουμε στο πρόγραμμα R:

- plot.ts(datatimeseriesforecasts$residuals) # make time plot of forecast errors
- plotForecastErrors(datatimeseriesforecasts$residuals) # make a histogram

Έτσι έχουμε την απεικόνιση των σφαλμάτων πρόβλεψης και το ιστόγραμμα:



Διάγραμμα 4. 6 Απεικόνιση σφαλμάτων πρόβλεψης



Διάγραμμα 4. 7 Απεικόνιση Ιστογράματος σφαλμάτων πρόβλεψης

Παρατηρούμε απο το Διάγραμμα 2.28 ότι πράγματι τα σφάλματα της πρόβλεψης είναι **κανονικά κατανεμημένα** και απο το Διάγραμμα 2.29 ότι η **μέση τιμή** των σφαλμάτων πρόβλεψης είναι **0**.

Επομένως είμαστε πλέον σε θέση να μπορούμε να πούμε ότι έχουμε κάνει μια σωστή και αξιόπιστη πρόβλεψη για τις επόμενες 5 πτήσεις του παραδείγματός μας.

Κεφάλαιο 5

Αποτελέσματα

5.1 Εισαγωγή

Στο Κεφάλαιο αυτό παρουσιάζονται αναλυτικά όλοι οι πίνακες και τα διαγράμματα που παρουσιάστηκαν στα παραδείγματα μας καθώς και τα αποτελέσματα που προέκυψαν απο την εφαρμογή τόσο των μεθόδων πρόβλεψης όσο και του εργαλείου R, στις προβλέψεις του κάθε παραδείγματος που πραγματοποιήσαμε.

Όπως είναι ήδη γνωστό για την εξαγωγή των πινάκων και των διαγραμμάτων χρισιμοποιήθηκε το στατιστικό εργαλείο R.

5.2 Αποτελέσματα απλής εκθετικής εξομάλυνσης

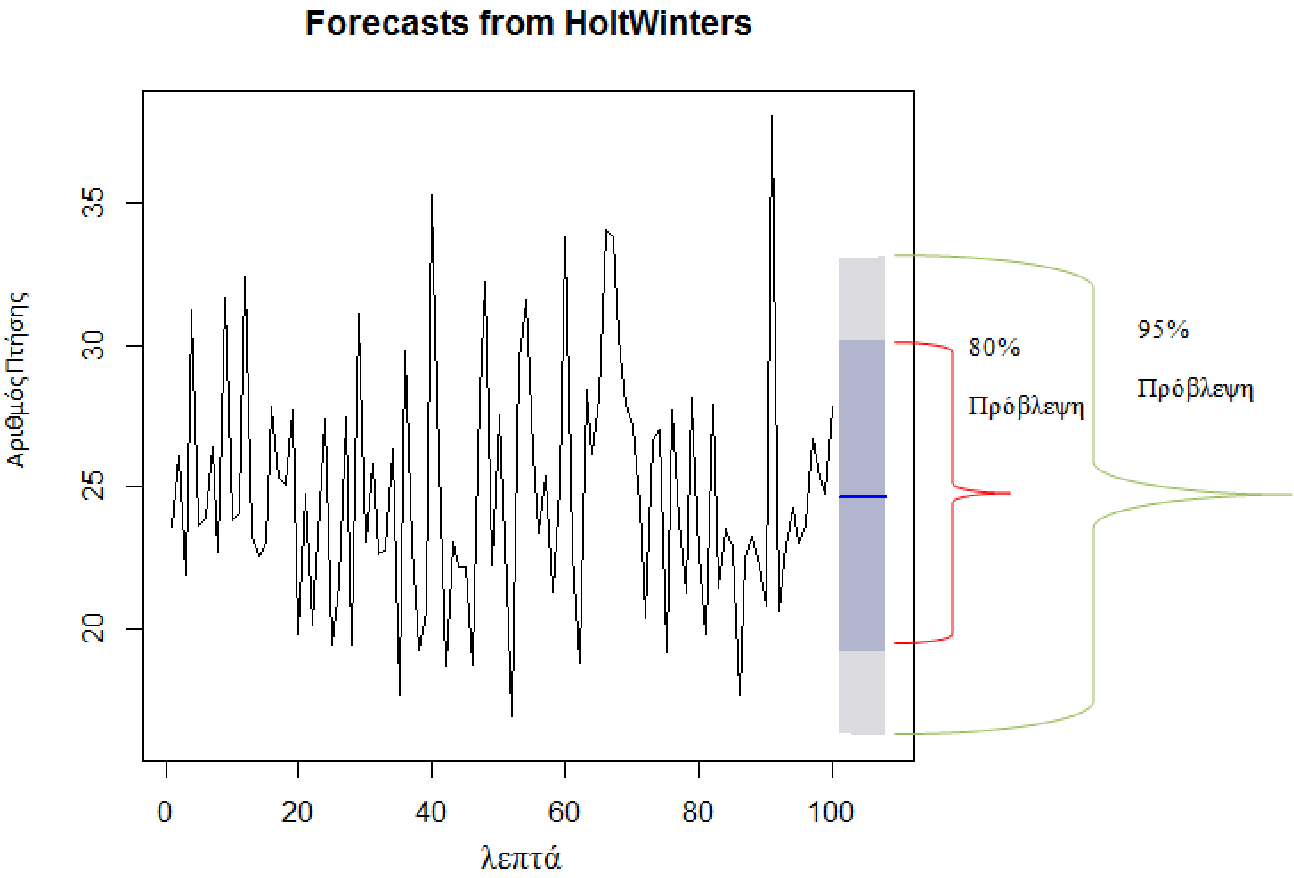
Τα μοντέλα που περιγράψαμε στην παράγραφο των μεθόδων εξομάλυνσης ήταν το μοντέλο της απλή εκθετικής εξομάλυνσης και το μοντέλο του HoltWinters εκθετικής εξομάλυνσης.Όπως είναι ήδη γνωστό οι χρονοσειρές μας είναι χωρισμένες στις εξής κατηγορίες:

1. χρονοσειρές που παρουσιάζουν τάση και όχι εποχικότητα

2. χρονοσειρές που δεν παρουσιάζουν τάση ούτε εποχικότητα

3. χρονοσειρές που δεν παρουσιάζουν τάση αλλά παρουσιάζουν εποχικότητα

Ας ξεκινήσουμε αρχικά με το μοντέλο της απλης εκθετικής εξομάλυνσης που εφαρμόσαμε στο παράδειγμα της κατηγορίας 1.



Διάγραμμα 5. 1 Μέθοδος απλής εκθετικής εξομάλυνσης σε δεδομένα με τάση και χωρίς εποχικότητα a=0.024 και b=1

Απο το Διάγραμμα 2.30 παρατηρούμε ότι με την μπλέ γραμμή είναι η τελική τιμή πρόβλεψης που αντιστοιχεί στην τιμή των 24 λεπτών.Με την κόκκινη παρένθεση απεικόνίζουμε τις τιμές που αντιστοιχούν στο 80% της μέγιστης και της ελάχιστης τιμής πρόβλεψης και αντίστοιχα με την πράσινη παρένθεση το 95% της μέγιστης και της ελάχιστης τιμής πρόβλεψης. Η έξοδος της εντολής HoltWinters στο εργαλείο R μας δίνει την τιμή της παραμέτρου alpha ίση με 0,024. Αυτή η τιμή είναι πολυ κοντά στο 0 κάτι που σημαίνει ότι οι προβλέψεις βασίζονται πιο πολύ στις τιμές των παρατηρήσεων-δεδομένων του παρελθόντος και λιγότερο στις πρόφατες.

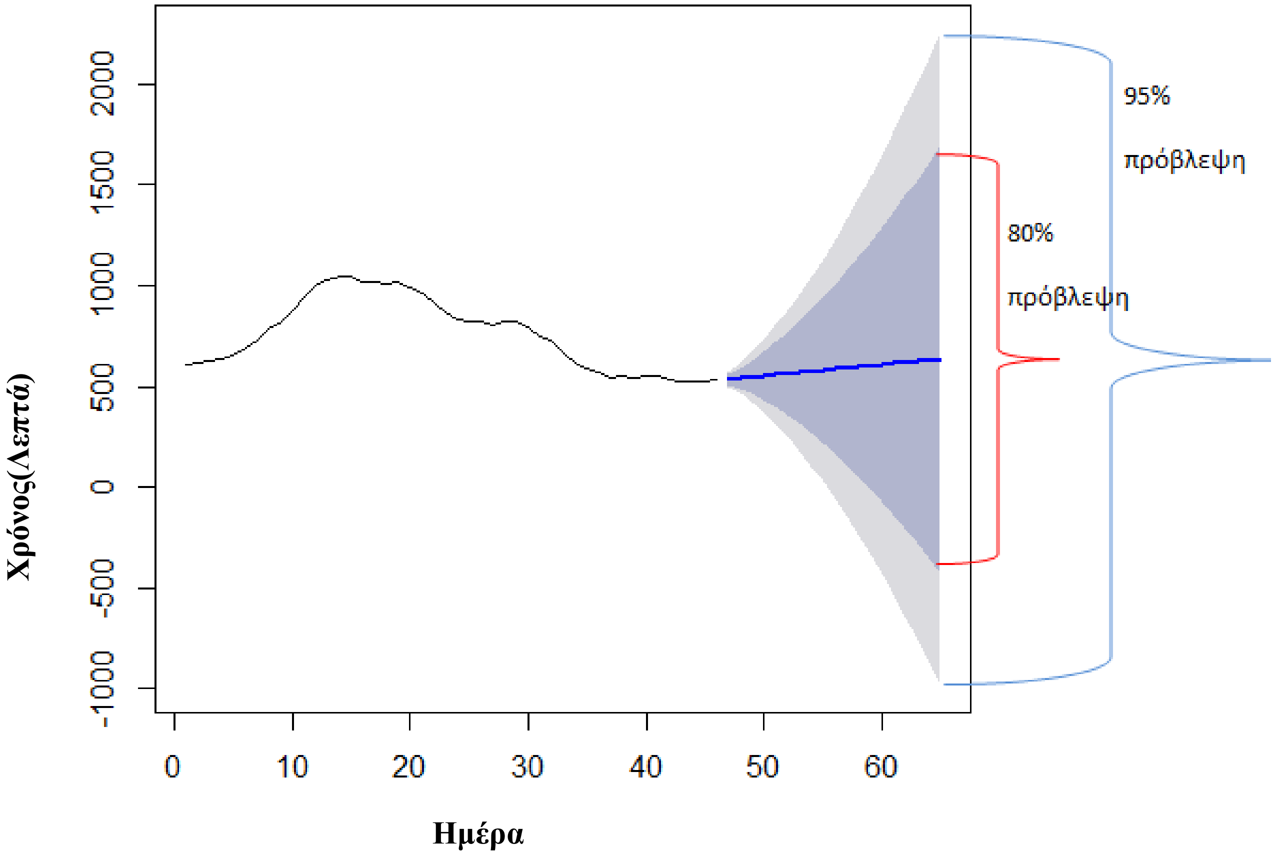
Ακολουθεί ο πίνακας με τις τελικές τιμές πρόβλεψης καθώς και τα ποσοστά της μέγιστης και της ελάχιστης τιμής πρόβλεψης.

Πίνακας 5. 1:Πίνακας με τις τιμές πρόβλεψης της απλής εκθετικής εξομάλυνσης

flight	Forecast	Lo 80	Hi 80	Lo 95	Hi
101	24.67819	19.17493	30.18145	16.26169	33.09470
102	24.67819	19.17333	30.18305	16.25924	33.09715
103	24.67819	19.17173	30.18465	16.25679	33.09960
104	24.67819	19.17013	30.18625	16.25434	33.10204
105	24.67819	19.16853	30.18785	16.25190	33.10449
106	24.67819	19.16694	30.18945	16.24945	33.10694
107	24.67819	19.16534	30.19105	16.24701	33.10938
108	24.67819	19.16374	30.19265	16.24456	33.11182

5.3 Αποτελέσματα εκθετικής εξομάλυνσης HoltWinters σε χρονοσειρά χωρίς τάση και εποχικότητα.

Με την χρήση του εργαλείου R εφαρμόσαμε προβλέψεις για το παράδειγμα της μεταφοράς προιόντων με φορτηγό όχημα στις καθυστερήσεις παράδοσης.Η τιμή a και b που μας πρότεινε το εργαλείο R είναι 0.083 και 1 αντίστοιχα.Ακολουθεί το αποτέλεσμα της πρόβλεψης:



Διάγραμμα 5. 2 Holt Winters εκθετική εξομάλυνση σε δεδομένα χωρίς τάση και εποχικότητα

Ακολουθεί και ο πίνακας με τις αντίστοιχες προβλέψεις:

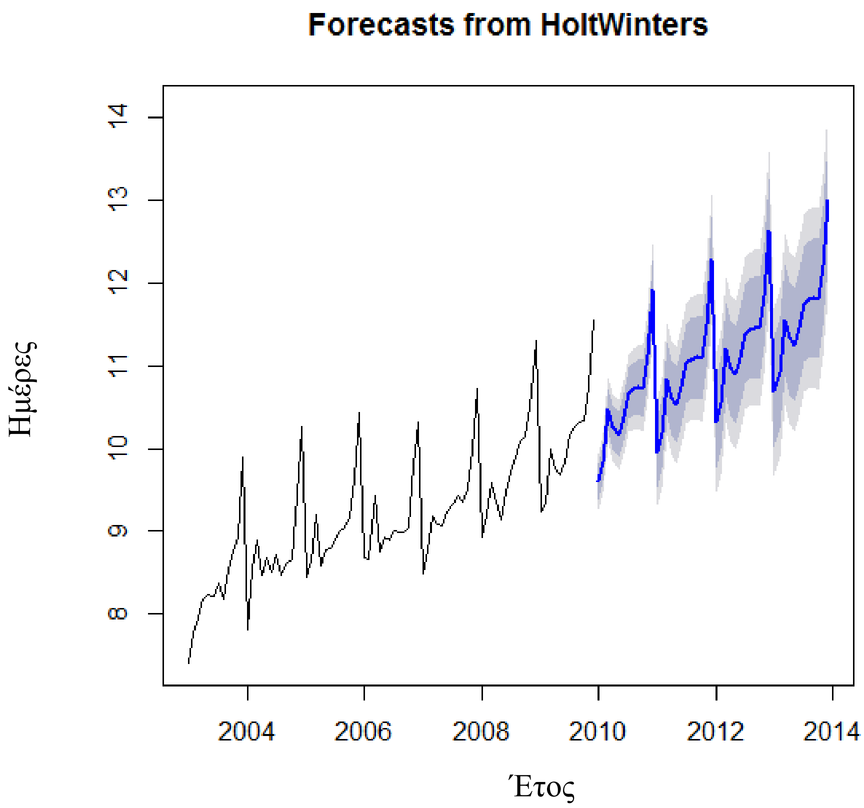
Πίνακας 5. 2:Πίνακας με τις προβλέψεις σε χρονοσειρά χωρίς τάση και εποχικότητα

Point	Forecast	Lo 80	Hi 80	Lo 95	Hi 95
47	534.9990	509.55210	560.4460	496.08130	573.9168
48	540.6895	491.01052	590.3685	464.71204	616.6670
49	546.3800	465.36129	627.3987	422.47258	670.2874
50	552.0704	434.40205	669.7388	372.11216	732.0287
51	557.7609	398.94120	716.5806	314.86713	800.6547
52	563.4514	359.47147	767.4313	251.49103	875.4117
53	569.1418	316.34076	821.9429	182.51596	955.7677
54	574.8323	269.81480	879.8498	108.34829	1041.3163
55	580.5228	220.10648	940.9391	29.31362	1131.7319
56	586.2132	167.39191	1005.0345	-54.31870	1226.7452
57	591.9037	111.82029	1071.9871	-142.32052	1326.1279
58	597.5942	53.52019	1141.6681	-234.49517	1429.6835
59	603.2846	-7.39593	1213.9652	-330.67069	1537.2399
60	608.9751	-70.82861	1288.7788	-430.69495	1648.6451
61	614.6655	-136.68903	1366.0201	-534.43211	1763.7632
62	620.3560	-204.89720	1445.6092	-641.75986	1882.4719
63	626.0465	-275.38060	1527.4736	-752.56728	2004.6602
64	631.7369	-348.07309	1611.5470	-866.75319	2130.2271

5.4 Αποτελέσματα εκθετικής εξομάλυνσης HoltWinters σε χρονοσειρές χωρίς τάση αλλά με εποχικότητα

Το παράδειγμα που χρησιμοποιήσαμε σε αυτή την μέθοδο είναι η περίπτωση που συζητήσαμε παραπάνω τον χρόνο μεταφοράς προιόντων απο φορτηγά πλοία.Όπως είχαμε αναλύσει το παράδειγμα μας ταιριάζει στην μέθοδο που εφαρμόσαμε αφού οι διακυμάνσεις δεν είναι σταθερές και παρουσίαζαν τα δεδομένα μία μεγιστοποίηση των προιόντων που καθυστέρησαν στην διάρκεια των καλοκαιρινών μηνών και μία αισθητή πτώση τους χειμερινούς μήνες. Όσο αφορά την νέα παράμετρο **gamma**, πρόκειται για την παράμετρο που αναφέρεται στην επαναληψιμότητα των δεδομένων που εμφανίζουν εποχικότητα.Αντίστοιχα και εδώ όσο η τιμή είναι πιο κοντά στο 0 ,τόσο πιο μικρή βαρύτητα θέτουμε στις πιο πρόσφατες παρατηρήσεις-δεδομένα.

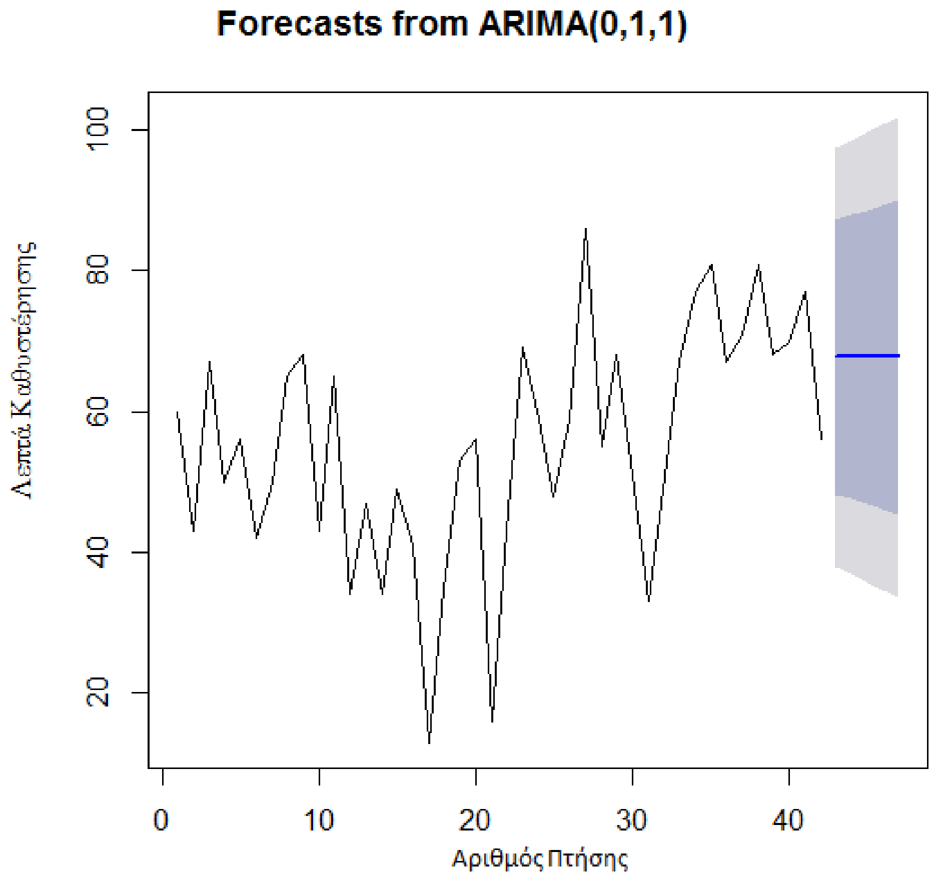
Οι προβλέψεις που μας έδωσε το εργαλείο R είναι:



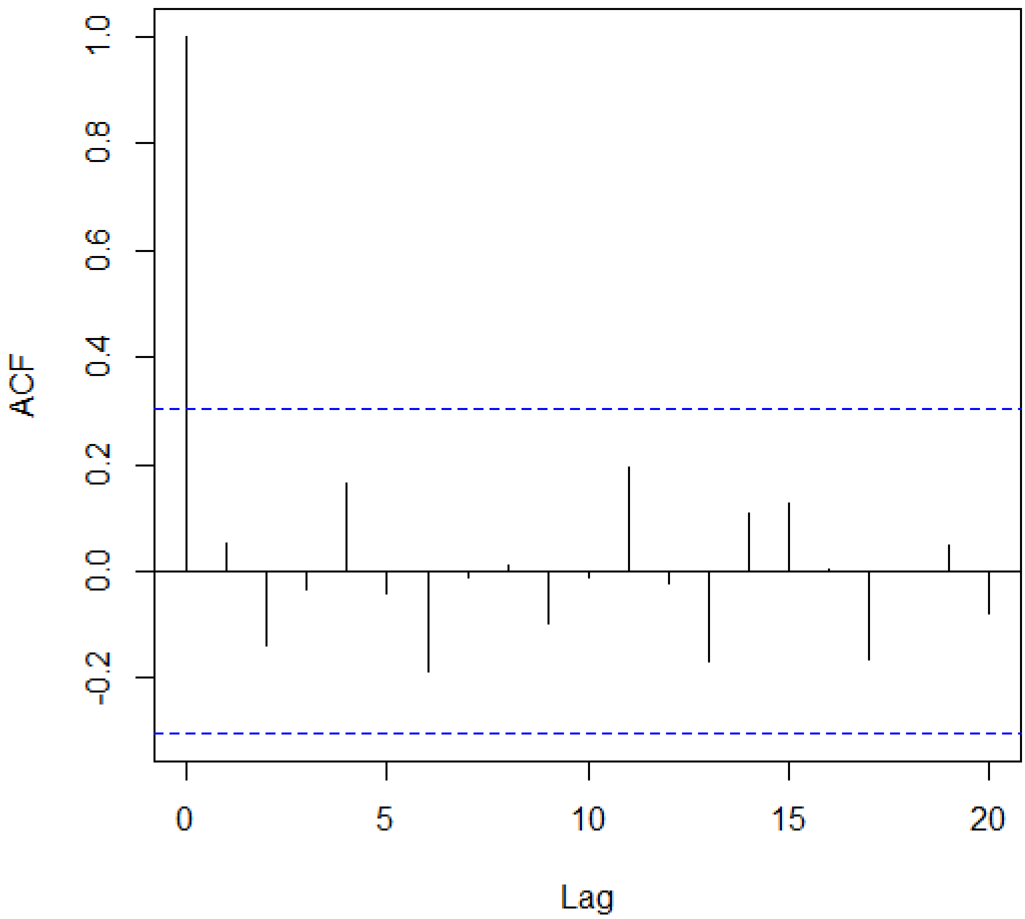
Διάγραμμα 5. 3 Πρόβλεψη HoltWinters σε χρονοσειρά χωρίς τάση και με εποχικότητα με a=0.83 b=0 και gamma=1

5.5 Αποτελέσματα ARIMA(0,1,1)

Αν και στην περίπτωση των αποτελεσμάτων του μοντέλου ARIMA για το παράδειγμα των πτήσεων με τις καθυστερήσεις έχει γίνει πλήρη ανάλυση των διαδικασιών και του αποτελέσματος ,παρακάτω θα απεικονίσουμε συνοπτικά τα αποτελέσματα του μοντέλου ARIMA(0,1,1) που επιλέξαμε για τους λόγους που είχαμε αναλύσει στο 3^{ο} κεφάλαιο.Θα απεικονίσουμε την πρόβλεψη μας καθώς και το διάγραμμα αυτοσυσχετίσεων σφαλμάτων πρόβλεψης.



Διάγραμμα 5. 4 Απεικόνιση πρόβλεψης μοντέλου ARIMA



Διάγραμμα 5. 5 Απεικόνιση αυτοσυσχετίσεων σφαλμάτων πρόβλεψης

Κεφάλαιο 6

Συμπεράσματα

6.1 Εισαγωγή

Στην παρούσα εργασία το αντικείμενο που μελετάμε είναι η καταλληλότητα των μοντέλων πρόβλεψης σε εφαρμογές logistics.Συνήθως οι προβλέψεις αυτές δεν είναι μακροπρόθεσμες και αφορούν προβλέψεις στο πρόσφατο μέλλον,δηλαδή προβλέψεις που μπορεί να αναφέρονται σε διάστημα 10 ημερών ή ενός μήνα.Αυτο βέβαια είναι και το μεγαλύτερο εμπόδιο για να επιτεύξουμε σύντομες προβλέψεις.

Μία πρόβλεψη για να είναι αξιόπιστη αποτελείται απο πολλά δεδομένα σε ένα μεγάλο χρονικό διάστημα,διότι όσο περισσότερα είναι τα δεδομένα τόσο πιο εύκολα γίνεται η ανάλυση,η διαφοροποίηση και η κατηγοριοποίηση μιας χρονοσειράς.

Πιο συγκεκριμένα ,ένα χαρακτηριστικό παράδειγμα που αναφερθήκαμε παραπάνω ήταν οι χρονοσειρές που εμφάνιζαν κάποια εποχικότητα στα δεδομένα τους.Όπως καταλαβαίνουμε σε ένα δείγμα σχετικά μικρό δεν μπορούμε να αντιληφθούμε μέσα απο την ανάλυση των χρονοσειρών μια τέτοια τάση, κάτι το οποίο κάνει την πρόβλεψή μας πιο δύσκολη ως προς την επιλογή του μοντέλου πρόβλεψης αλλά και την αξιοπιστία.

6.2 Συμπεράσματα

Στα κεφάλαια 2 και 3 ασχοληθήκαμε με την αποσύνθεση των χρονοσειρών και τις μεθόδους πρόβλεψης.Αυτο που είναι ευρέως γνωστό είναι ότι η πρόβλεψη μιας μελλοντικής τιμής μιας μεταβλητής προέρχεται απο την εκθετική εξομάλυνση την διάσπαση και την αποσύνθεση των χρονοσειρών και την ανάλυση ARIMA.

Πιο συγκεκριμένα αρχικά αυτο που αρχικά εφαρμόσαμε ήταν η διάσπαση των χρονοσειρών .Η μέθοδος αυτη είναι απαραίτητη να γίνεται προτου εφαρμόσουμε κάποια πρόβλεψη.Η διάσπαση των χρονοσειρών μας εξασφαλίσει βελτίωση του σφάλματος σε μία πρόβλεψη που σίγουρα θα έχουμε.Αυτο συμβαίνει διότι μέσω της διάσπασης της αρχικής χρονοσειράς σε πολλές υπο-χρονοσειρές,μπορούμε να αντιληφθούμε την τάση ή την εποχικότητα και να κάνουμε πιο ακριβείς προβλέψεις

Στην συνέχεια εφαρμόσαμε ορισμένες απο τις πλέον γνωστές μεθόδους εξομάλυνσης και με την βοήθεια του εργαλείου R.Το πλεονέκτημα με αυτές τις μεθόδους είναι αυτο που αρχικά αναφέραμε,μπορούμε δηλαδή να προβούμε χωρίς μεγάλο υπολογιστικό βαθμό δυσκολίας σε προβλέψεις εύκολα και γρήγορα σε μικρό δείγμα παρατηρήσεων.

Εφαρμόσαμε πολλά παραδείγματα προβλέψεων με πραγματικά στοιχεία και αριθμούς και σε όλες τις περιπτώσεις εφαρμόσαμε έλεγχο σφαλμάτων.Αυτο που μπορούμε να πούμε είναι ότι δεν υπάρχει ιδεατό μοντέλο πρόβλεψης.Κάθε μοντέλο πρόβλεψης έχει και κάποια πλεονεκτήματα-μειονεκτήματα έναντι σε ένα άλλο.Γιαυτό τον λόγο εφαρμόσαμε και πολλά παραδείγματα.

Πιο συγκεκριμένα θα μπορούσαμε να πούμε ότι η μέθοδος του Holt-Winters είναι πολύ αξιόπιστη ως προς την αξιοπιστία και την προσαρμοστικότητα,διότι έχει ένα πολυ μεγάλο πλεονέκτημα που αφορά την εποχικότητα.Η εποχικότητα στα δεδομένα των χρονοσειρών είναι ίσως ο πιο σημαντικός παράγοντας και αυτο φάνηκε και απο τα παραδείγματα που αναλύσαμε στα κεφάλαια 2 και 3.Αρκεί να υπενθυμίσουμε τα διαγράμμα των αυτοσυσχετίσεων ACF και μερικών αυτοσυσχετίσεων PACF.

Όσο αφορά τα μοντέλα ARIMA ίσως τελικά να είναι τα πιο αξιόπιστα μοντέλα πρόβλεψης κάτι το οποίο φάνηκε και απο τα δικά μας πειράματα ως προς τα σφάλματα του L-jung Box. Βέβαια όπως είδαμε και στο παράδειγμα μας,η επιλογή κατάλληλου μοντέλου ARIMA είναι μια διαδικασία η οποία είναι ιδιαίτερα χρονοβόρα και δύσκολη,ωστόσο το αποτέλεσμα μας δικαιώνει αν κρίνουμε απο τις τιμές των σφαλμάτων πρόβλεψης.

Βιβλιογραφία

1. Waddell and Sohal, 1994 ,Newbold and Bos ,1994,Klassen and Flores Benito 2001

2.Λέκκας Φραγκίσκος Πανεπιστήμιο Αιγαιόυ

3.Πανεπιστήμιο Αιγαίου 2004,Χρονοσειρές

4.Γ.κοκκολάκης,σημειώσεις,Τομέας μαθηματικών,Αναπληρωτής καθηγητής Σχολή εφαρμοσμένων μαθηματικών

(Γ.παύλος ,Βιβλίο 2)

5.Πολυτεχνείο Κρήτης,Σημειώσεις στοχαστικών Ανελίξεων

6.RobHyndman,forecasting principles and practice

7.RobHyndman,forecasting methods and applications

8.RobHyndman,expotential smoothing

9.Data series library,robjhyndman

10.Bureau of transtortation statistics

11.Δημήτρης Κουγιουμντζής users.auth.timeseries