



Εθνικό Μετσόβιο Πολυτεχνείο  
Σχολή Εφαρμοσμένων Μαθηματικών και Φυσικών Επιστημών  
Τομέας Μαθηματικών

---

**Ανάλυση και Πρόβλεψη Χρονοσειρών:  
Μέθοδοι και Εφαρμογές**

---

Forecasting with Time Series Analysis:  
Methods and Applications

---

Διπλωματική Εργασία  
της  
Γεωργίας Μπίνου

Επιβλέπων: Κολέτσος Ιωάννης  
Επ. Καθηγητής ΕΜΠ

Αθήνα, Ιούλιος 2016





Εθνικό Μετσόβιο Πολυτεχνείο  
Σχολή Εφαρμοσμένων Μαθηματικών και Φυσικών Επιστημών  
Τομέας Μαθηματικών

---

**Ανάλυση και Πρόβλεψη Χρονοσειρών:  
Μέθοδοι και Εφαρμογές**

---

Forecasting with Time Series Analysis:  
Methods and Applications

---

Διπλωματική Εργασία  
της  
Γεωργίας Μπίνου

Εγκρίθηκε από την τριμελή εξεταστική επιτροπή:

.....  
ΚΟΚΚΙΝΗΣ ΒΑΣΙΛΕΙΟΣ  
Επ. ΚΑΘΗΓΗΤΗΣ ΕΜΠ

.....  
ΚΟΛΕΤΣΟΣ ΙΩΑΝΝΗΣ  
Επ. ΚΑΘΗΓΗΤΗΣ ΕΜΠ

.....  
ΤΥΧΟΠΟΥΛΟΣ ΕΥΑΓΓΕΛΟΣ  
Επ. ΚΑΘΗΓΗΤΗΣ ΕΜΠ

Αθήνα, Ιούλιος 2016



“God hides the end of future time in misty night,  
and laughs if mortal worries at what’s right...”

- Horace



## Ευχαριστίες

Η παρούσα διπλωματική αποτελεί το τέλος μιας πορείας ετών που μου έμαθε πολλά και πάνω απ' όλα με έκανε καλύτερο άνθρωπο. Η στήριξη όμως κάποιων ατόμων έπαιξαν καθοριστικό ρόλο σε όλο αυτό. Θα ήθελα να ευχαριστήσω θερμά τον κ. Ιωάννη Κολέτσο, Επ. Καθηγητή ΕΜΠ για την ευκαιρία που μου έδωσε να ασχοληθώ με ένα τόσο ενδιαφέρον θέμα. Η καθοδήγηση και οι συμβουλές του ήταν πολύτιμες καθ' όλη τη διάρκεια της εκπόνησης της διπλωματικής εργασίας. Το μεγαλύτερο όμως ευχαριστώ, το χρωστάω στην οικογένειά μου και τους φίλους μου, που δεν σταμάτησαν ποτέ να πιστεύουν σε μένα και να με στηρίζουν.

Γεωργία Μπίνου

## Περίληψη

Η παρούσα διπλωματική εργασία έχει ως σκοπό τη μελέτη των μεθόδων και εργαλείων που απαιτούνται για τη βέλτιστη επιλογή ενός μοντέλου πρόβλεψης. Τα δεδομένα στα οποία βασίζονται οι προβλέψεις, είναι σε πολλές περιπτώσεις χρονοσειρές. Οι χρονοσειρές βρίσκουν άμεση εφαρμογή σε πολλούς κλάδους ένας εκ των οποίων είναι και η επιχειρησιακή έρευνα.

Στο εισαγωγικό κεφάλαιο γίνεται αναφορά στην επιχειρησιακή έρευνα και στη σημασία που απέκτησε ως κλάδος μέσα από μια ιστορική αναδρομή. Στη συνέχεια γίνεται μια αναφορά στις διάφορες μεθόδους πρόβλεψης και στα βασικά εργαλεία της ανάλυσης δεδομένων. Ακολουθεί η ανάλυση χρονοσειρών, όπου παρουσιάζονται και αναλύονται τρεις βασικές μέθοδοι πρόβλεψων: οι μέθοδοι αποσύνθεσης χρονοσειρών, οι μέθοδοι εξομάλυνσης και η ανάλυση υποδειγμάτων ARIMA. Τέλος γίνεται η εφαρμογή των μεθόδων αυτών σε τρία σύνολα δεδομένων, για την επιλογή του καλύτερου μοντέλου πρόβλεψης



## **Abstract**

This diploma thesis aims to study the methods and tools required for optimal selection of a predictive model. The data underlying the forecasts are in many cases time series. Time series find direct application in many sectors, one of which is operational research.

The introductory chapter, refers to operational research and the importance it gained as an industry through a throwback. Then there is a reference to the different methods of forecasting and basic tools of data analysis. Time series analysis follows, and three basic methods of forecasting are presented and analyzed: the time series decomposition methods, the exponential smoothing methods and the ARIMA model analysis. At the end, applications of these methods regarding three different data sets are presented for the selection of the best predictive model.



# ΠΕΡΙΕΧΟΜΕΝΑ

<b>1</b>	<b>Εισαγωγή</b> .....	<b>1</b>
1.1	Ιστορικά Στοιχεία .....	2
1.2	Μοντέλα και Στάδια Επίλυσης.....	5
<b>2</b>	<b>Πρόβλεψη (Forecasting)</b> .....	<b>7</b>
2.1	Χρονοσειρές .....	8
2.2	Μέθοδοι Πρόβλεψης .....	10
2.2.1	Ποιοτικές.....	11
2.2.2	Ποσοτικές.....	12
2.3	Ορίζοντας πρόβλεψης.....	13
2.4	Βασικά βήματα σε μια διαδικασία πρόβλεψης.....	14
<b>3</b>	<b>Εργαλεία πρόβλεψης</b> .....	<b>15</b>
3.1	Γραφική Αναπαράσταση .....	15
3.1.1	Μοτίβα χρονοσειρών.....	15
3.1.2	Γράφημα εποχικότητας .....	18
3.2	Στατιστικοί Δείκτες .....	19
3.3	Μέτρηση ακρίβειας πρόβλεψης.....	22
3.3.1	Τυπικά στατιστικά μέτρα .....	23
3.3.2	Σύγκριση μεθόδων πρόβλεψης.....	25
3.4	Διαστήματα Πρόβλεψης.....	25
3.5	Μέθοδος ελαχίστων τετραγώνων .....	26
3.6	Τροποποιήσεις και προσαρμογές .....	28
3.6.1	Μαθηματικές τροποποιήσεις.....	28
3.6.2	Ημερολογιακές προσαρμογές.....	29
<b>4</b>	<b>Ανάλυση χρονοσειρών</b> .....	<b>30</b>
4.1	Μέθοδοι Αποσύνθεσης (Time Series Decomposition) .....	30
4.1.1	Κινητοί μέσοι όροι .....	32
4.1.1.1	Απλός κινητός μέσος όρος .....	33
4.1.1.2	Σταθμισμένος κινητός μέσος όρος .....	35
4.1.1.3	Διπλός κινητός μέσος όρος.....	37

4.1.1.4	Κεντρικός κινητός μέσος όρος .....	37
4.1.2	Εξομάλυνση με τοπική γραμμική παλινδρόμηση .....	38
4.1.2.1	Loess .....	40
4.1.3	Κλασική μέθοδος αποσύνθεσης .....	41
4.1.4	Διαμόρφωση προβλέψεων .....	46
4.2	Μέθοδοι Εξομάλυνσης ( Exponential Smoothing Methods).....	47
4.2.1	Μέθοδοι κινητού μέσου όρου .....	49
4.2.2	Μέθοδοι εκθετικής εξομάλυνσης.....	52
4.2.2.1	Απλή εκθετική εξομάλυνση (Simple Exponential Smoothing)....	55
4.2.2.2	Μέθοδος ARSSES .....	59
4.2.2.3	Μέθοδος Holt (Holt's Linear Method).....	63
4.2.2.4	Μέθοδος Holt-Winters (Holt-Winters' trend and seasonality method) 67	
4.3	Ανάλυση Arima .....	75
4.3.1	Συσχέτιση στις χρονοσειρές.....	76
4.3.1.1	Συνάρτηση Αυτοσυσχέτισης (ACF).....	76
4.3.1.2	Portmanteau tests.....	78
4.3.1.3	Συνάρτηση μερικής αυτοσυσχέτισης (PACF).....	80
4.3.1.4	Εντοπισμός εποχικότητας στις χρονοσειρές.....	81
4.3.1.5	Εξέταση στασιμότητας σε δεδομένα χρονοσειράς.....	82
4.3.2	Μοντέλα ARIMA για δεδομένα χρονοσειρών.....	89
4.3.2.1	Αυτοπαλίνδρομα υποδείγματα .....	90
4.3.2.2	Υποδείγματα κινητού μέσου όρου .....	91
4.3.2.3	Υποδείγματα αυτοπαλίνδρομου κινητού μέσου όρου.....	92
4.3.2.4	Υποδείγματα ARIMA.....	92
4.3.2.5	Εποχικά υποδείγματα ARIMA .....	93
4.3.3	Μεθοδολογία Box-Jenkins .....	93
4.3.3.1	Ταυτοποίηση (Identification) .....	93
4.3.3.2	Εκτίμηση (Estimation) .....	94
4.3.3.3	Διαγνωστικός έλεγχος .....	95
4.3.3.4	Πρόβλεψη .....	95
<b>5</b>	<b>Εφαρμογές .....</b>	<b>96</b>
5.1	Δεδομένα τουριστών .....	96
5.1.1	Σύγκριση μεθόδων και επιλογή μοντέλου .....	108

5.2	Δεδομένα κατανάλωσης ηλεκτρικής ενέργειας.....	109
5.2.1	Σύγκριση μεθόδων και επιλογή μοντέλου .....	120
5.3	Δεδομένα πωλήσεων .....	121
5.3.1	Σύγκριση μεθόδων και επιλογή μοντέλου .....	127
5.4	Γενικά συμπεράσματα .....	128



## ΚΑΤΑΛΟΓΟΣ ΠΙΝΑΚΩΝ

Πίνακας 4.1-1: Αποτελέσματα εφαρμογής μεθόδων απλού κινητού μέσου όρου 3,5,7 και 9 περιόδων σε δεδομένα κατανάλωσης ηλεκτρικής ενέργειας στην Ελλάδα τα έτη 2005-2015 .....	34
Πίνακας 4.1-2: Σταθμισμένοι κινητοί μέσοι όροι 3 και 5 για δεδομένα κατανάλωσης ηλεκτρικής ενέργειας στην Ελλάδα τα έτη 2005-2015 .....	36
Πίνακας 4.2-1: Γενικοί τύποι μεθόδων εξομάλυνσης.....	54
Πίνακας 4.2-2: Τύποι για υπολογισμούς προβλέψεων χρησιμοποιώντας την ταξινόμηση κατά Pegel .....	54
Πίνακας 4.2-3: Ενδεικτικά βάρη που αποδίδονται στις παρατηρήσεις με τη χρήση της απλής εκθετικής εξομάλυνσης .....	57
Πίνακας 4.2-4: Αποτελέσματα της Rme την εφαρμογή της απλής εκθετικής εξομάλυνσης σε δεδομένα πωλήσεων.....	58
Πίνακας 4.2-5: Δεδομένα και προβλέψεις πωλήσεων με τη μέθοδο εκθετικής εξομάλυνσης.....	59
Πίνακας 4.2-6 Αποτελέσματα εφαρμογής της μεθόδου ARSES σε δεδομένα πωλήσεων .....	62
Πίνακας 4.2-7: Σφάλματα πρόβλεψης για δοκιμές μοντέλων της μεθόδου του Holt..	66
Πίνακας 4.2-8: Παράμετροι μοντέλου με την εφαρμογή της μεθόδου Holt σε δεδομένα πωλήσεων.....	66
Πίνακας 4.2-9: Εξισώσεις μεθόδου Holt-Wintersγια την περίπτωση πολλαπλασιαστικής εποχικότητας .....	68
Πίνακας 4.2-10: Εξισώσεις μεθόδου Holt-Wintersγια την περίπτωση αθροιστικής εποχικότητας .....	70
Πίνακας 4.2-11: Σφάλματα πρόβλεψης για δοκιμές μοντέλων της μεθόδου Holt-Winters στην περίπτωση αθροιστικής εποχικότητας.....	71
Πίνακας 4.2-12: Παράμετροι μοντέλου με την εφαρμογή της μεθόδου Holt-Winters σε δεδομένα πωλήσεων για την περίπτωση αθροιστικής εποχικότητας .....	72
Πίνακας 4.2-13: Σφάλματα πρόβλεψης για δοκιμές μοντέλων της μεθόδου Holt-Winters στην περίπτωση πολλαπλασιαστικής εποχικότητας.....	73
Πίνακας 4.2-14: Παράμετροι μοντέλου με την εφαρμογή της μεθόδου Holt-Winters σε δεδομένα πωλήσεων για την περίπτωση πολλαπλασιαστικής εποχικότητας .....	74
Πίνακας 4.3-1: Χρονοσειρά με 26 τιμές.....	76
Πίνακας 5.1-1: Εφαρμογή υποδείγματος ARIMA 2,0,01,1,012 στα δεδομένα τουριστών .....	100
Πίνακας 5.1-2: Προβλέψεις για τα δεδομένα τουριστών με το υπόδειγμα ARIMA1,0,02,1,112.....	102
Πίνακας 5.1-3: Μοντέλο περίπτωσης 1 για δεδομένα τουριστών με εφαρμογή της μεθόδου Holt-Winters .....	105

Πίνακας 5.1-4: Μοντέλο περίπτωσης 2 για δεδομένα τουριστών με εφαρμογή της μεθόδου Holt-Winters .....	106
Πίνακας 5.1-5: Στατιστικά μέτρα σφαλμάτων πρόβλεψης για τις 2 περιπτώσεις της μεθόδου Holt-Winters με εφαρμογή σε δεδομένα κατανάλωσης ηλεκτρικής ενέργειας .....	107
Πίνακας 5.2-1: Προβλέψεις για τα δεδομένα ηλεκτρικής κατανάλωσης με το υπόδειγμα ARIMA1,0,10,1,112 .....	113
Πίνακας 5.2-2: Μέθοδος Holt-Winters για δεδομένα κατανάλωσης ηλεκτρικής ενέργειας(περίπτωση 1).....	116
Πίνακας 5.2-3: Μέθοδος Holt-Winters για δεδομένα κατανάλωσης ηλεκτρικής ενέργειας(περίπτωση 2).....	118
Πίνακας 5.2-4: Στατιστικά μέτρα σφαλμάτων πρόβλεψης για τις 2 περιπτώσεις της μεθόδου Holt-Winters με εφαρμογή σε δεδομένα κατανάλωσης ηλεκτρικής ενέργειας .....	119
Πίνακας 5.3-1: Προβλέψεις για τα δεδομένα ηλεκτρικής κατανάλωσης με το υπόδειγμα ARIMA0,1,51,0,212 .....	125



## ΚΑΤΑΛΟΓΟΣ ΣΧΗΜΑΤΩΝ

Εικόνα 2.1-1: Παράδειγμα χρονοσειράς που περιγράφει τιμή μετοχής .....	9
Εικόνα 3.1-1: Γράφημα αριθμού τουριστών (σε χιλιάδες) της Ισπανίας τα έτη 2007-2011 .....	16
Εικόνα 3.1-2: Παραδείγματα τάσης σε μια χρονοσειρά .....	17
Εικόνα 3.1-3: Παραδείγματα εποχικότητας σε μια χρονοσειρά .....	18
Εικόνα 3.1-4: Διάγραμμα εποχικότητας αριθμού τουριστών Ισπανίας τα έτη 2007-2011 .....	19
Εικόνα 3.2-1: Συσχετόγραμμα / Διάγραμμα acf .....	22
Εικόνα 3.5-1: Παράδειγμα εφαρμογής μεθόδου ελαχίστων τετραγώνων .....	28
Εικόνα 4.1-1: Δεδομένα κατανάλωσης ηλεκτρικής ενέργειας στην Ελλάδα τα έτη 2005-2015 .....	32
Εικόνα 4.1-2: Γραφική αναπαράσταση εφαρμογής απλών κινητών μέσων 3,5,7 και 9 περιόδων σε δεδομένα κατανάλωσης ηλεκτρικής ενέργειας στην Ελλάδα τα έτη 2005-2015 .....	35
Εικόνα 4.1-3: Γραφική αναπαράσταση των σταθμισμένων κινητών μέσων 3 και 5 για δεδομένα κατανάλωσης ηλεκτρικής ενέργειας στην Ελλάδα τα έτη 2005-2015 .....	36
Εικόνα 4.1-4: Τοπική παλινδρόμηση Loess σε δεδομένα κατανάλωσης ενέργειας της Ελλάδας το χρονικό διάστημα 2005-2015. ....	41
Εικόνα 4.1-5: Σειρά τάσης κύκλου σε δεδομένα κατανάλωσης ηλεκτρικής ενέργειας .....	44
Εικόνα 4.1-6: Γραφήματα που δείχνουν τη συνιστώσα της εποχικότητας (κάτω) και τα δεδομένα με αφαίρεση της σειράς τάσης κύκλου (πάνω) για τα δεδομένα κατανάλωσης ηλεκτρικής ενέργειας .....	45
Εικόνα 4.1-7: Τυχαίες διακυμάνσεις για τα δεδομένα της κατανάλωσης ηλεκτρικής ενέργειας .....	46
Εικόνα 4.2-1: Σενάριο πρόβλεψης .....	48
Εικόνα 4.2-2: Μηνιαίες πωλήσεις εταιρίας (time plot) .....	51
Εικόνα 4.2-3: Κινητοί μέσοι όροι για τα δεδομένα των πωλήσεων .....	51
Εικόνα 4.2-4: Βέλτιστο αποτέλεσμα με χρήση κινητών μέσων όρων στα δεδομένα πωλήσεων .....	52
Εικόνα 4.2-5: Επέκταση κατηγοριοποίησης κατά Pegel .....	53
Εικόνα 4.2-6 : Εφαρμογή της απλής εκθετικής εξομάλυνσης στο παράδειγμα με τις πωλήσεις για τιμές της σταθεράς εξομάλυνσης $\alpha=0.1$ και $\alpha=0.9$ .....	56
Εικόνα 4.2-7: Προβλέψεις με τη μέθοδο απλής εκθετικής εξομάλυνσης με εφαρμογή στο παράδειγμα των πωλήσεων .....	59
Εικόνα 4.2-8: Γραφική αναπαράσταση των προβλέψεων με εφαρμογή της μεθόδου ARSSES σε δεδομένα πωλήσεων .....	62
Εικόνα 4.2-9: Τιμές της σταθεράς εξομάλυνσης $at$ με εφαρμογή της μεθόδου ARSSES σε δεδομένα πωλήσεων .....	62

Εικόνα 4.2-10: Εφαρμογή μεθόδου Holt σε δεδομένα πωλήσεων για διάφορα initial values.....	65
Εικόνα 4.2-11: Γραφική αναπαράσταση των προβλέψεων με εφαρμογή της μεθόδου Holt σε δεδομένα πωλήσεων για τη βέλτιστη επιλογή της μεθόδου ...	66
Εικόνα 4.2-12: Αποτελέσματα εφαρμογής της μεθόδου Holt για τις συνιστώσες της χρονοσειράς.....	67
Εικόνα 4.2-13: Εφαρμογή μεθόδου Holt-Winters σε δεδομένα πωλήσεων για διάφορα initial values στην περίπτωση αθροιστικής εποχικότητας .....	71
Εικόνα 4.2-14: Γραφική αναπαράσταση των προβλέψεων με εφαρμογή της μεθόδου Holt-Winters σε δεδομένα πωλήσεων για τη βέλτιστη επιλογή της μεθόδου στην περίπτωση πολλαπλασιαστικής εποχικότητας.....	72
Εικόνα 4.2-15: Εφαρμογή μεθόδου Holt-Winters σε δεδομένα πωλήσεων για διάφορα initial values στην περίπτωση πολλαπλασιαστικής εποχικότητας.....	73
Εικόνα 4.2-16: Γραφική αναπαράσταση των προβλέψεων με εφαρμογή της μεθόδου Holt-Winters σε δεδομένα πωλήσεων για τη βέλτιστη επιλογή της μεθόδου στην περίπτωση πολλαπλασιαστικής εποχικότητας.....	74
Εικόνα 4.2-17: Εκτίμηση των συνιστωσών της χρονοσειράς των πωλήσεων για την περίπτωση της πολλαπλασιαστικής εποχικότητας με τη μέθοδο Holt-Winters .....	75
Εικόνα 4.3-1: Συνάρτηση αυτοσυσχέτισης και διάγραμμα τιμών μιας χρονοσειράς..	77
Εικόνα 4.3-2: Συντελεστές μερικής αυτοσυσχέτισης για τη χρονοσειρά του πίνακα 4.3-1.....	81
Εικόνα 4.3-3: Συναρτήσεις αυτοσυσχέτισης και μερικής αυτοσυσχέτισης για δεδομένα πωλήσεων για χρονικές υστερήσεις $k=1,2,\dots,35$ .....	82
Εικόνα 4.3-4: Δεδομένα πωλήσεων μετά την εφαρμογή σειράς διαφορών πρώτης τάξης.....	84
Εικόνα 4.3-5: Συναρτήσεις συσχέτισης και μερικής αυτοσυσχέτισης της πρώτης τάξης σειράς διαφορών των δεδομένων πωλήσεων.....	85
Εικόνα 4.3-6: Δεδομένα πωλήσεων μετά την εφαρμογή εποχιακών διαφορών.....	86
Εικόνα 4.3-7: Συναρτήσεις συσχέτισης και μερικής αυτοσυσχέτισης της πρώτης τάξης σειράς εποχιακών διαφορών των δεδομένων πωλήσεων ( $s=12$ ) .....	87
Εικόνα 4.3-8: Δεδομένα πωλήσεων μετά την εφαρμογή εποχιακών διαφορών και διαφορών πρώτης τάξης.....	88
Εικόνα 4.3-9: Δύο παραδείγματα για αυτοπαλίνδρομα υποδείγματα με διαφορετικές παραμέτρους.....	91
Εικόνα 5.1-1: Δεδομένα τουριστών στην Ελλάδα το χρονικό διάστημα 2001-2016 ..	96
Εικόνα 5.1-2: Διάγραμμα εποχικότητας δεδομένων τουριστών.....	97
Εικόνα 5.1-3: Δεδομένα τουριστών με μετασχηματισμό BoxCox ( $\lambda=0$ ) .....	98
Εικόνα 5.1-4: Συναρτήσεις αυτοσυσχέτισης και μερικής αυτοσυσχέτισης σε μετασχηματισμένα δεδομένα τουριστών .....	98
Εικόνα 5.1-5: Δεδομένα τουριστών μετά από εφαρμογή εποχιακών διαφορών .....	99
Εικόνα 5.1-6: Γράφημα υπολοίπων για δεδομένα τουριστών με εφαρμογή του υποδείγματος ARIMA(1,0,0,2,1,12).....	100

Εικόνα 5.1-7: Προβλέψεις για δεδομένα τουριστών (σε χιλιάδες) με το υπόδειγμα ARIMA1,0,02,1,112.....	101
Εικόνα 5.1-8: Γράφημα των δεδομένων τουριστών και των προβλεπόμενων τιμών του υποδείγματος ARIMA1,0,02,1,112.....	102
Εικόνα 5.1-9: Προβλέψεις δεδομένων τουριστών με τη μέθοδο STL χρησιμοποιώντας τη μέθοδο naïve.....	103
Εικόνα 5.1-10: Αποσύνθεση δεδομένων τουριστών με χρήση μεθόδου STL.....	103
Εικόνα 5.1-11: Προβλέψεις με τη μέθοδο Holt-Winters σε δεδομένα τουριστών (περίπτωση 1).....	105
Εικόνα 5.1-12: Συνάρτηση αυτοσυσχέτισης υπολοίπων για πρόβλεψη περίπτωσης 1 με εφαρμογή της Holt-Winters σε δεδομένα τουριστών.....	105
Εικόνα 5.1-13: Προβλέψεις με τη μέθοδο Holt-Winters σε δεδομένα τουριστών (περίπτωση 2).....	107
Εικόνα 5.1-14: Συνάρτηση αυτοσυσχέτισης υπολοίπων για πρόβλεψη περίπτωσης 2 με εφαρμογή της Holt-Winters σε δεδομένα τουριστών.....	107
Εικόνα 5.2-1: Εποχικό γράφημα δεδομένων κατανάλωσης ηλεκτρικής ενέργειας...	109
Εικόνα 5.2-2: Διαγράμματα ACF και PACF των δεδομένων κατανάλωσης ηλεκτρικής ενέργειας το χρονικό διάστημα Ιαν2005-Οκτ2013 .....	110
Εικόνα 5.2-3: Δεδομένα ηλεκτρικής κατανάλωσης μετά από εφαρμογή εποχικών διαφορών .....	111
Εικόνα 5.2-4: Εφαρμογή υποδείγματος ARIMA 1,0,11,1,112 στα δεδομένα κατανάλωσης ηλεκτρικής ενέργειας .....	112
Εικόνα 5.2-5: Γράφημα υπολοίπων για δεδομένα κατανάλωσης ηλεκτρικής ενέργειας με εφαρμογή του υποδείγματος ARIMA1,0,10,1,112. ....	112
Εικόνα 5.2-6: Προβλέψεις για δεδομένα τουριστών (σε χιλιάδες) με το υπόδειγμα ARIMA1,0,10,1,112.....	113
Εικόνα 5.2-7: Γράφημα των δεδομένων ηλεκτρικής κατανάλωσης και των προβλεπόμενων τιμών του υποδείγματος ARIMA1,0,10,1,112 .....	114
Εικόνα 5.2-8: Προβλέψεις στα δεδομένα τουριστών με τη μέθοδο αποσύνθεσης STL κάνοντας χρήση της μεθόδου naïve. ....	114
Εικόνα 5.2-9: Αποσύνθεση χρονοσειράς τουριστών .....	115
Εικόνα 5.2-10: Προβλέψεις με τη μέθοδο Holt-Winters για δεδομένα κατανάλωσης ηλεκτρικής ενέργειας. ....	117
Εικόνα 5.2-11: Συνάρτηση αυτοσυσχέτισης των υπολοίπων με εφαρμογή της μεθόδου Holt-Winters σε δεδομένα κατανάλωσης ηλεκτρικής ενέργειας για την περίπτωση 1.....	117
Εικόνα 5.2-12: Πρόβλεψη μεθόδου Holt-Winters στα δεδομένα κατανάλωσης ηλεκτρικής ενέργειας(περίπτωση 2) .....	118
Εικόνα 5.2-13: Συνάρτηση αυτοσυσχέτισης των υπολοίπων για την εφαρμογή της μεθόδου Holt-Winters σε δεδομένα κατανάλωσης ηλεκτρικής ενέργειας (περίπτωση 2).....	119
Εικόνα 5.2-14: Προβλέψεις επιλεγμένου μοντέλου με εφαρμογή της μεθόδου Holt-Winters σε δεδομένα κατανάλωσης ηλεκτρικής ενέργειας.....	119

Εικόνα 5.3-1: Εποχικό γράφημα δεδομένων πωλήσεων .....	121
Εικόνα 5.3-2: Διαγράμματα ACF και PACF των δεδομένων πωλήσεων το χρονικό διάστημα Ιαν2002-Φεβ2013.....	122
Εικόνα 5.3-3: Δεδομένα πωλήσεων μετά από εφαρμογή διαφορίσης πρώτης τάξης	123
Εικόνα 5.3-4: Εφαρμογή υποδείγματος ARIMA 2,1,0,0,312 στα δεδομένα κατανάλωσης ηλεκτρικής ενέργειας .....	124
Εικόνα 5.3-5: Γράφημα υπολοίπων για δεδομένα πωλήσεων με εφαρμογή του υποδείγματος ARIMA0,1,51,0,212.....	124
Εικόνα 5.3-6: Προβλέψεις για δεδομένα πωλήσεων με το υπόδειγμα ARIMA0,1,51,0,212.....	125
Εικόνα 5.3-7: Γράφημα των δεδομένων ηλεκτρικής κατανάλωσης και των προβλεπόμενων τιμών του υποδείγματος ARIMA0,1,51,0,212 .....	126
Εικόνα 5.3-8: Προβλέψεις των πωλήσεων με εφαρμογή της μεθόδου αποσύνθεσης STL.....	126
Εικόνα 5.3-9: Συνιστώσες των δεδομένων πωλήσεων με τη μέθοδο αποσύνθεσης STL .....	127

# 1

## Εισαγωγή

Αναφερόμενοι στον όρο Επιχειρησιακή Έρευνα (*Operations Research*) εννοούμε την επιστήμη που ασχολείται με τη βελτιστοποίηση της απόδοσης ενός συστήματος<sup>1</sup>. Πρόκειται για ένα σύνολο από τεχνικές, οι οποίες χρησιμοποιώντας (μαθηματικά) μοντέλα, δημιουργούν μια ποσοτική και ορθολογιστική βάση για τη λήψη αποφάσεων με σκοπό τη βελτιστοποίηση της λειτουργίας του υπό μελέτη συστήματος. Εναλλακτικά, είναι μια επιστημονική μεθοδολογία για την υποστήριξη λήψης διοικητικών αποφάσεων σε περίπλοκα συστήματα που λειτουργούν μέσα σε αβέβαιο περιβάλλον.

Η Επιχειρησιακή Έρευνα γνώρισε ιδιαίτερη ανάπτυξη κυρίως μετά τον Β' παγκόσμιο πόλεμο, και οφείλεται κατά βάση α) στον διαρκή αυξανόμενο ανταγωνισμό μεταξύ των σύγχρονων επιχειρήσεων και β) στην εξέλιξη και διάδοση των Η/Υ μέσω των οποίων γίνονται ταχύτερα οι επιλύσεις περίπλοκων προβλημάτων καθώς και η επεξεργασία μεγάλου όγκου πληροφοριών. Μεταξύ των πεδίων εφαρμογών της περιλαμβάνονται, ενδεικτικά, η βιομηχανία, οι μεταφορές, η χωροθέτηση βιομηχανικών και εμπορικών εγκαταστάσεων, το μαρκετινγκ καθώς και δημόσιοι τομείς όπως αυτοί της υγείας, της εκπαίδευσης και της προστασίας του περιβάλλοντος.

Πολλές φορές οι λήψεις των αποφάσεων λαμβάνονται υπό συνθήκες αβεβαιότητας αφού τα δεδομένα είναι είτε ελλιπή είτε ανακριβή. Ένα εργαλείο που βοηθάει στη μείωση της αβεβαιότητας είναι η πρόβλεψη (*forecasting*). Ως πρόβλεψη, μπορεί να χαρακτηριστεί η προσπάθεια να προβλέψουμε το μέλλον χρησιμοποιώντας ποιοτικές και ποσοτικές μεθόδους. Τέλειες προβλέψεις συνήθως δεν υπάρχουν. Είναι πολλοί οι παράγοντες που δε μπορούν να προβλεφθούν με βεβαιότητα. Αυτό που είναι σημαντικό κατ' επέκταση, είναι να γίνεται μια συνεχής ανασκόπηση των προβλέψεων εφαρμόζοντας την καταλληλότερη μέθοδο με στόχο την ελαχιστοποίηση των σφαλμάτων πρόβλεψης. Ασφαλώς, όσο πιο μεγάλος είναι ο χρονικός ορίζοντας των προβλέψεων, τόσο πιο αναξιόπιστες θα είναι. Συνεπώς, αν θέλουμε να είμαστε όσο πιο ακριβείς γίνεται στις προβλέψεις μας, αυτόματα ανεβαίνει και το κόστος κατασκευής ενός μοντέλου πρόβλεψης.

---

<sup>1</sup> Σύστημα: Σύνολο οντοτήτων (ανθρώπινο δυναμικό, μηχανές, κεφάλαια..κλπ) που αλληλεπιδρούν και συνεργάζονται μεταξύ τους για την επίτευξη κάποιου στόχου (πχ. επιχείρηση, οργανισμός)

Η καλύτερη πρόβλεψη δεν είναι απαραίτητα αυτή που είναι πιο ακριβής ή έχει το χαμηλότερο κόστος. Εξαρτάται και από παράγοντες όπως ο στόχος που έχουμε και η διαθεσιμότητα των δεδομένων στο να καθοριστεί η κατάλληλη ακρίβεια της πρόβλεψης. Απλά παραδείγματα εφαρμογών είναι η αποθήκευση ποσοτήτων που θα πωληθούν στο μέλλον, η αγορά νέου εξοπλισμού για την ικανοποίηση μελλοντικού αγοραστικού ενδιαφέροντος κ.α.

## 1.1 Ιστορικά Στοιχεία

Ο όρος “Επιχειρησιακή Έρευνα” προέρχεται από τη μετάφραση του αγγλικού όρου “*Operational Research*”, ή “*Operations Research*” όπως συνηθίζεται στην Αμερική, με εναλλακτικές ονομασίες, Διοικητική Επιστήμη (*Management Science*), Επιστήμη αποφάσεων (*Decision Science*) και Ανάλυση Συστημάτων (*Systems Analysis*). Η μετάφραση στα ελληνικά του όρου ως «Λειτουργική Έρευνα» είναι ίσως πιο αποδοτική, ωστόσο καθιερώθηκε να αποκαλείται έτσι κυρίως επειδή οι πρώτες εφαρμογές της επιχειρησιακής έρευνας και η ανάπτυξή της αργότερα ξεκίνησε από το στρατό, όπου οι κυριότερες λειτουργίες του ονομάζονται επιχειρήσεις.

Οι ρίζες της επιχειρησιακής έρευνας μπορούν να εντοπισθούν πολλές δεκαετίες πριν με πρωτοπόρο τον *Charles Babbage* (1791- 1871). Η έρευνά του για το κόστος μεταφοράς και το κόστος ταξινόμησης της αλληλογραφίας συντέλεσε στη δημιουργία του γενικού αγγλικού “Ταχυδρομείου της πέννας” το 1840. Το 1917, ο Δανός μηχανικός *A.K. Erlang* (1878-1929) εξέτασε προβλήματα που είχαν σχέση με το χρόνο απασχόλησης των τηλεφωνικών κέντρων. Υπήρξαν βέβαια και άλλες αξιολογές επιστημονικές προσεγγίσεις γύρω από τη διοίκηση των επιχειρήσεων. Ωστόσο η δημιουργία του όρου «Επιχειρησιακή Έρευνα» έγινε με αφορμή τις στρατιωτικές επιχειρήσεις στην Αγγλία στις αρχές του Β’ Παγκοσμίου Πολέμου (1939).

Η ανάγκη της αποτελεσματικότερης διαχείρισης του διαθέσιμου πολεμικού υλικού, αποτέλεσε την αιτία για την εμφάνιση των πρώτων συστηματικών δραστηριοτήτων της επιχειρησιακής έρευνας. Τόσο στην Αγγλία σε πρώτη φάση, όσο και στις ΗΠΑ αργότερα, ομάδες επιστημόνων διαφόρων ειδικοτήτων ( μαθηματικοί, φυσικοί, στατιστικοί, ψυχολόγοι, βιολόγοι, κ.α.), κλήθηκαν να αντιμετωπίσουν από κοινού τα πολύπλοκα προβλήματα στρατηγικής που προέκυπταν, εφαρμόζοντας την επιστημονική μεθοδολογία, βασιζόμενοι σε γνώσεις και εμπειρίες από διάφορους χώρους και επιστήμες. Τρεις από τις πιο χαρακτηριστικές εφαρμογές ήταν:

α) Η μελέτη για την καλύτερη αξιοποίηση της νέας τότε συσκευής ανίχνευσης και εντοπισμού εχθρικών αεροσκαφών (ραντάρ) στη Μεγάλη Βρετανία,

β) Η μελέτη για τον προσδιορισμό τους κατάλληλου βάθους εκρήξεων βομβών βυθού που χρησιμοποιούσε η Αγγλική αεροπορία κατά των Γερμανικών υποβρυχίων,

γ) Η μελέτη μια ομάδας Επιχειρησιακών Ερευνών του Υπουργείου Άμυνας των ΗΠΑ για τον προσδιορισμό του βέλτιστου μεγέθους νηοπομπών (πλήθος φορτηγών πλοίων συνοδευόμενων από πολεμικά) μεταφοράς πολεμικού υλικού και στρατιωτών από τις ΗΠΑ στην Ευρώπη.

Με το τέλος του πολέμου, εμφανίστηκε μια βιομηχανική και εμπορική ανάπτυξη, που είχε σαν αποτέλεσμα τη ραγδαία αύξηση σε μέγεθος και πολυπλοκότητα των επιχειρήσεων και οργανισμών. Αυξήθηκαν οι ανάγκες παραγωγής και παροχής υπηρεσιών, υπήρχε μεγάλος ανταγωνισμός και το περιβάλλον της τεχνολογίας και της αγοράς άλλαζε συνεχώς. Συγκεκριμένα, οι εταιρείες εξαπλώνονταν σε εθνική και πολυεθνική κλίμακα καλύπτοντας όλο και περισσότερο γεωγραφικό χώρο, με αποτέλεσμα να αυξηθεί σημαντικά η κατανομή της εργασίας και ο επιμερισμός των δραστηριοτήτων μεταξύ των ατόμων που διοικούσαν. Η εισαγωγή νέων μορφών τεχνολογίας απαιτούσε εξειδικευμένους τεχνίτες, και αυτό με τη σειρά του εμφάνισε το πρόβλημα του συντονισμού των διαφόρων τμημάτων της επιχείρησης για την επίτευξη κοινού στόχου ως προς το συμφέρον της επιχείρησης.

Τα προβλήματα ήταν πολλά και το ενδιαφέρον της βιομηχανίας για την επιχειρησιακή έρευνα δεν άργησε να έρθει, ύστερα και από τις προφανείς επιτυχείς εφαρμογές της στον στρατιωτικό τομέα. Πολλοί από τους επιστήμονες που μέχρι πρότινος είχαν ασχοληθεί με τα στρατιωτικά προβλήματα, στράφηκαν στον ερευνητικό και ιδιωτικό τομέα. Μάλιστα, στην Αμερική, ιδρύθηκε το *Operations Evaluation Group* (OEG) σε συνεργασία με το MIT (*Massachusetts Institute of Technology*), και υπογράφηκε σύμβαση με την *Douglas Aircraft Company* για το έργο RAND (*Research and Development*), το οποίο αφορούσε την επέκταση της χρήσης των ερευνητών επιχειρησιακής έρευνας για μεγάλο χρονικό διάστημα μετά τη λήξη του πολέμου. Στη Μεγάλη Βρετανία η επιχειρησιακή έρευνα εδραιώθηκε από το 1951 ως επιστημονική μέθοδος λήψης αποφάσεων και το 1957 ιδρύθηκε η Διεθνής Συνομοσπονδία Εταιριών Επιχειρησιακών Ερευνών (*IFORS*). Το 1963 ιδρύθηκε και στην Ελλάδα αντίστοιχη εταιρία, η “Ελληνική Εταιρία Επιχειρησιακών Ερευνών (E.E.E.E)” ή όπως αποκαλείται διεθνώς “*Hellenic Operational Research Society* (HELORS)” και η οποία αποτελεί μέλος της “IFORS” από τις αρχές του έτους 1968.

Πολλοί επιστήμονες προσπάθησαν κατά καιρούς να δώσουν έναν ορισμό που να περιγράφει όσο γίνεται ακριβέστερα, το περιεχόμενο και το πεδίο εφαρμογής της επιχειρησιακής έρευνας. Το 1947, σε ένα επιστημονικό συμπόσιο στο “Dundee” της Σκωτίας, ακούστηκε για πρώτη φορά το όνομα “*Operational Research*” το οποίο προτάθηκε από τον *Watson-Watt*, μαζί με τον *A.P. Rowe*, με τον εξής ορισμό:

«*Η Επιχειρησιακή Έρευνα αποσκοπεί στο να ερευνήσει ποσοτικά εάν ένας οργανισμός παίρνει από τη λειτουργία του εξοπλισμού του, τη βέλτιστη δυνατή συνεισφορά σε σχέση με τον ολικό αντικειμενικό σκοπό του, ποιες αλλαγές σε εξοπλισμό και μεθόδους απαιτούνται για τη βελτίωση των αποτελεσμάτων με το μικρότερο δυνατό κόστος σε προσπάθεια και χρόνο και τέλος σε ποιο βαθμό μεταβολές στους επιμέρους αντικειμενικούς σκοπούς θα συνεισέφεραν στη πιο οικονομική και έγκαιρη εκτέλεση του ολικού στρατηγικού αντικειμενικού σκοπού*».

Ενδεικτικά ακολουθούν μερικοί ακόμα.

Από την Εταιρεία Επιχειρησιακής Έρευνας της Μεγάλης Βρετανίας (*Operational Research Society*) :

«Η Επιχειρησιακή Έρευνα είναι η εφαρμογή της σύγχρονης επιστήμης πάνω σε πολύπλοκα προβλήματα, τα οποία ανακύπτουν στη διεύθυνση και διοίκηση μεγάλων συστημάτων, αποτελούμενων από ανθρώπους, μηχανές, υλικά και κεφάλαια, στη Βιομηχανία, τις Επιχειρήσεις, τις Κυβερνητικές Υπηρεσίες και την Άμυνα.

Η χαρακτηριστική της μεθοδολογία συνίσταται στην ανάπτυξη επιστημονικού μοντέλου του υπό μελέτη συστήματος, που περιλαμβάνει μετρήσεις τυχαίων παραγόντων και με το οποίο προβλέπει και συγκρίνει τα αποτελέσματα εναλλακτικών αποφάσεων, στρατηγικών και ελέγχων.

Ο σκοπός της είναι να βοηθήσει τη διοίκηση να καθορίσει την πολιτική και τις ενέργειές της επιστημονικά (κατά τον καλύτερο δυνατό τρόπο)».

Από τους *Russell Lincoln Ackoff* και *Maurice W. Sasieni*, όπως περιλαμβάνεται στο σύγγραμμά τους με τίτλο “*Fundamentals of Operations Research*”:

«Επιχειρησιακή Έρευνα μπορεί να θεωρηθεί ότι είναι: η εφαρμογή επιστημονικών μεθόδων από μικτές ομάδες σε προβλήματα που αφορούν τον έλεγχο οργανωμένων συστημάτων (αποτελούμενων από ανθρώπους και μηχανές) κατά τρόπο, ώστε να παρέχουν λύσεις που εξυπηρετούν κατά τον καλύτερο δυνατό τρόπο τους σκοπούς του οργανισμού ως συνόλου».

Από την Ε.Ε.Ε.Ε. :

«Επιχειρησιακή Έρευνα είναι η επιστημονική προετοιμασία των αποφάσεων της Διοικήσεως (με την επιστημονική ανάλυση των δεδομένων και τη δημιουργία μαθηματικών προτύπων)».

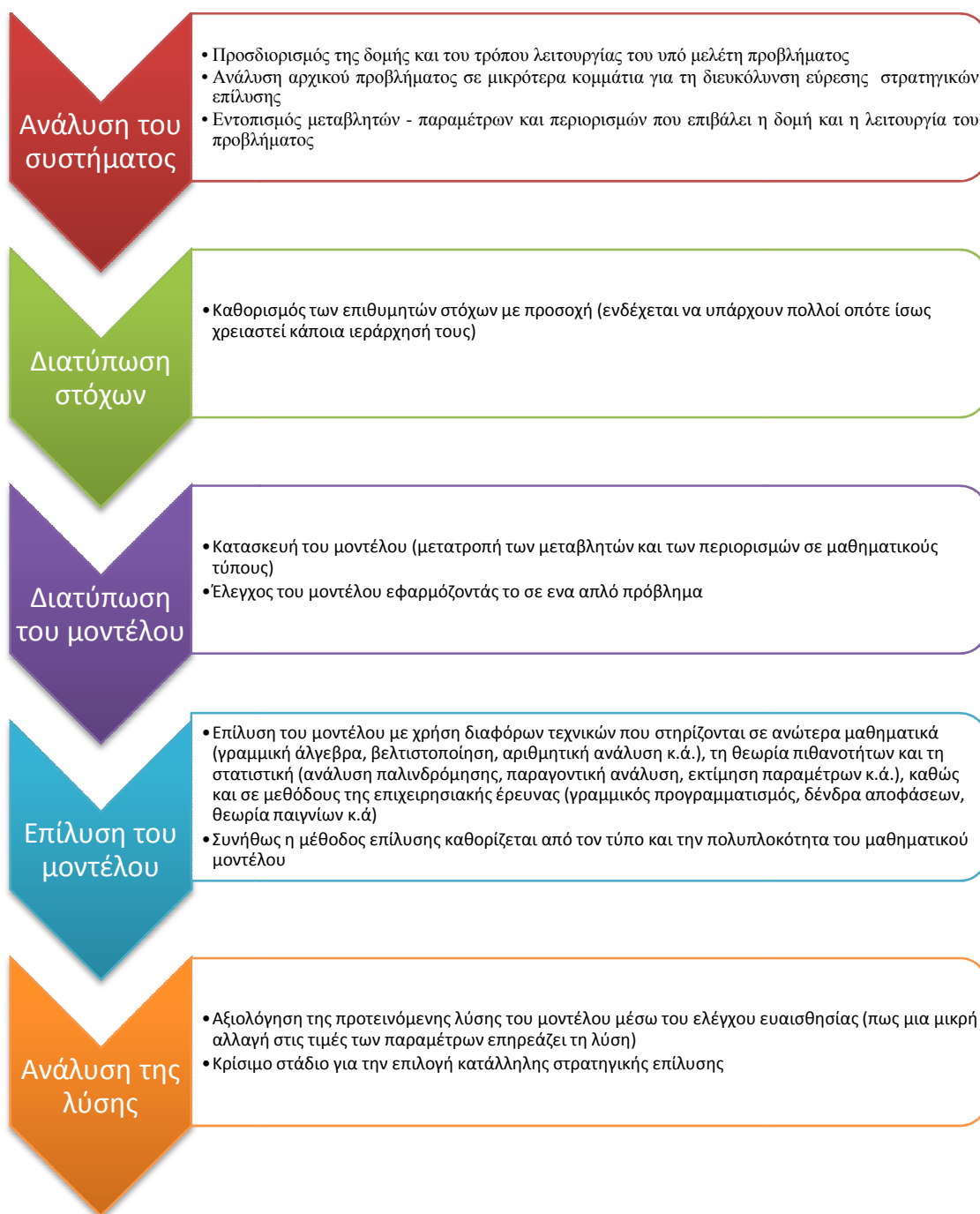
Από τους *Hans G. Daellenbach* και *John A. George* το 1978:

«Επιχειρησιακή έρευνα είναι η συστηματική εφαρμογή ποσοτικών μεθόδων, τεχνικών και εργαλείων στην ανάλυση προβλημάτων που εμπεριέχουν την λειτουργία συστημάτων».



## 1.2 Μοντέλα και Στάδια Επίλυσης

Οι μελέτες της επιχειρησιακής έρευνας αποσκοπούν όπως είδαμε, στην παροχή επιστημονικά θεμελιωμένων προτάσεων για τη λήψη αποφάσεων της διοίκησης μιας επιχείρησης, που έχουν ως στόχο την καλύτερη λειτουργία της. Οι προτάσεις αυτές είναι συνήθως, είτε μια βέλτιστη λύση στο υπό μελέτη πρόβλημα ή ένα σύνολο εναλλακτικών λύσεων κατάλληλα αξιολογημένων. Η επιστημονική μεθοδολογία που χρησιμοποιείται γενικά για την επίλυση ενός προβλήματος αντικατοπτρίζεται στα παρακάτω στάδια:



Τα μοντέλα που κατασκευάζουμε για την επίλυση του εκάστοτε προβλήματος, αποτελούν πρακτικά μια αναπαράσταση του συστήματος που μελετάμε, αποτυπώνοντας τις σχέσεις μεταξύ των παραμέτρων με μαθηματικούς τύπους.

Ωστόσο, οι αποφάσεις που πρέπει να πάρουν τα διοικητικά στελέχη, δεν είναι πάντα εύκολες και λαμβάνονται κάτω από συνθήκες βεβαιότητας, αβεβαιότητας και κινδύνου. Η αβεβαιότητα χαρακτηρίζει την ποιότητα της λύσης. Όταν για παράδειγμα, για τα αποτελέσματα των εναλλακτικών λύσεων δε μπορεί να υπολογιστεί κάποια πιθανότητα σχετικά με την επίτευξή τους, τότε θεωρούνται αβέβια. Έτσι λοιπόν τα μοντέλα μπορούμε να τα χωρίσουμε σε δύο κατηγορίες. Τα **προσδιοριστικά** ή **ντετερμινιστικά (Deterministic)** και τα **πιθανολογικά** ή **στοχαστικά (Stochastic)**. Στην πρώτη κατηγορία έχουμε “σταθερές”<sup>2</sup> τιμές στις παραμέτρους του μοντέλου, ενώ στη δεύτερη οι τιμές των παραμέτρων μεταβάλλονται κατά τυχαίο τρόπο.

Άλλη μία κατηγοριοποίηση που γίνεται είναι ως προς το στόχο που θέλουμε να πετύχουμε κάθε φορά. Για παράδειγμα μπορεί μια επιχείρηση να ψάχνει τρόπους μεγιστοποίησης των κερδών της ή να θέλει να προβλέψει την ποσότητα ενός προϊόντος που θα πουλήσει στο μέλλον. Εδώ έχουμε τρεις κατηγορίες. Τα **μοντέλα βελτιστοποίησης**, τα **περιγραφικά μοντέλα** και τα **μοντέλα πρόβλεψης**. Τα μοντέλα βελτιστοποίησης κοιτούν να μεγιστοποιήσουν ή να ελαχιστοποιήσουν μια μεταβλητή ή την τιμή μιας συνάρτησης που εκφράζει τον επιθυμητό στόχο. Από την άλλη, τα περιγραφικά μοντέλα ενδιαφέρονται για την αναζήτηση και την ποσοτικοποίηση των σχέσεων μεταξύ των μεταβλητών του μοντέλου ώστε να μπορούν να προσδιοριστούν οι αλλαγές που θα προκύψουν σε ένα μέγεθος όταν μεταβληθεί κάποιο από τα μεγέθη που το επηρεάζουν. Το αποτέλεσμα συνήθως απεικονίζεται με γραφική παράσταση των εξεταζόμενων μεγεθών. Τέλος, τα μοντέλα πρόβλεψης είναι κυρίως στατιστικά μοντέλα που αναλύουν ιστορικά στοιχεία, προσδιορίζοντας μια μαθηματική σχέση μεταξύ μιας εξαρτημένης μεταβλητής και ενός ή περισσότερων ανεξάρτητων μεταβλητών που μπορεί να την επηρεάζουν. Με αυτό τον τρόπο μπορούν να προβλέψουν πιθανές τιμές της εξαρτημένης μεταβλητής, με βάση τις μελλοντικές τιμές των ανεξάρτητων μεταβλητών.

---

<sup>2</sup> Ο όρος σταθερές δε σημαίνει ότι οι τιμές των παραμέτρων δεν μεταβάλλονται ποτέ, αλλά ότι δε χαρακτηρίζονται από συνεχείς τυχαίες μεταβολές

# 2

## Πρόβλεψη (Forecasting)

Τι καιρό θα κάνει αύριο; Πόση κίνηση μπορεί να συναντήσω στο δρόμο για τη δουλειά; Είναι απλά καθημερινά ερωτήματα που όλοι κάποια στιγμή έχουμε σκεφτεί. Και συνήθως έχουμε απάντηση. Για τον καιρό συμβουλευόμαστε τις μετεωρολογικές υπηρεσίες ενώ για την κίνηση την εμπειρία μας. Και στις δύο περιπτώσεις όμως έχουμε μια πρόβλεψη. Αντίστοιχες καταστάσεις συναντάμε και στις επιχειρήσεις είτε είναι μικρές, είτε μεγάλες. Για παράδειγμα, ένας επιχειρηματίας που πουλάει ντόνατ χρειάζεται να αποφασίζει καθημερινά την ποσότητα που θα πρέπει να παράξει ώστε να καλύψει τη ζήτηση. Στην περίπτωση που η διαθέσιμη ποσότητα είναι μεγαλύτερη της ζήτησης θα επιβαρυνθεί με το κόστος παραγωγής της απούλητης ποσότητας. Στην αντίθετη περίπτωση θα έχει ζημία λόγω αδυναμίας εξυπηρέτησης των πελατών του, κάτι που μπορεί όμως να έχει και μακροπρόθεσμες συνέπειες.

Μια καλή πρόβλεψη αποτελεί πλεονέκτημα για κάθε επιχείρηση αν θέλει να δει μελλοντικές επιτυχίες. Και όπως γίνεται με όλα τα πράγματα, υπάρχουν καλές και κακές προβλέψεις. Αυτό όμως που τις καθιστά σημαντικές είναι αφενός ότι βοηθούν στην αντιμετώπιση της αβεβαιότητας του μέλλοντος και αφετέρου η βοήθεια που προσφέρουν στη λήψη αποφάσεων γύρω από μια ποικιλία πεδίων στα οποία εφαρμόζονται. Παρακάτω αναφέρονται μερικά πεδία και παραδείγματα εφαρμογής των προβλέψεων:

- **Οικονομία και χρηματοοικονομικά** (πρόβλεψη γενικών δεικτών χρηματιστηρίων και της μεταβλητότητάς τους)
- **Περιβάλλον και κλίμα** (ενεργειακή ζήτηση – πρόβλεψη ζήτησης φορτίου, μετεωρολογία – πρόβλεψη υγρασίας, ρύπανση – πρόβλεψη ατμοσφαιρικών ρύπων)
- **Κοινωνικό περιβάλλον** ( πρόβλεψη δημογραφικών δεδομένων, πρόβλεψη επιδημιολογιών)
- **Τουρισμός** ( εκτίμηση συνολικών αφίξεων τουριστών)
- **Μεταφορές και μετακινήσεις** ( πρόβλεψη του κυκλοφοριακού φόρτου)
- **Ακίνητα και κτηματικές περιουσίες** ( εκτίμηση πραγματικών ή αντικειμενικών αξιών ακινήτων, διαχείριση κινδύνου (*risk management*) σε περιπτώσεις δανειοδοτήσεων)

Υπάρχουν δύο βασικοί λόγοι που είναι αναγκαίο να κάνουμε προβλέψεις. Ο ένας είναι ο σκοπός. Αναφερόμαστε σε οποιαδήποτε ενέργεια που γίνεται στο παρόν και μεριμνεί για κάτι απροσδόκητο που μπορεί να προκύψει στο μέλλον, κάτω από μία

κατάσταση ή ένα σύνολο προϋποθέσεων. Οπότε αυτόματα δημιουργείται ένας στόχος (σκοπός) που βάζουμε ώστε να επωφεληθούμε από αυτές τις μελλοντικές καταστάσεις ή να ελαχιστοποιήσουμε της επίδρασή τους στην περίπτωση που είναι δυσμενείς για εμάς. Ο δεύτερος λόγος είναι ο χρόνος. Για να ετοιμάσουμε ένα πλάνο, να οργανώσουμε τα απαιτούμενα για την υλοποίηση του και να το υλοποιήσουμε χρειαζόμαστε χρόνο. Σε κάποιες περιπτώσεις απαιτείται λίγος χρόνος ενώ σε άλλες μπορεί να πάρει περισσότερο καιρό, ακόμη και χρόνια. Έχοντας όμως μια πρόβλεψη στη διάθεσή μας, τα σχέδιά μας μπορούν να υλοποιηθούν σε πολύ λιγότερο χρόνο.

Ο ρόλος των προβλέψεων, είναι να εκτιμούν ένα μελλοντικό γεγονός συνδυάζοντας και προβάλλοντας συστηματικά και με καθορισμένο τρόπο στοιχεία του παρελθόντος. Μπορεί οι διάφορες τεχνικές πρόβλεψης να διαφέρουν μεταξύ τους ωστόσο υπάρχουν κάποια κοινά χαρακτηριστικά :

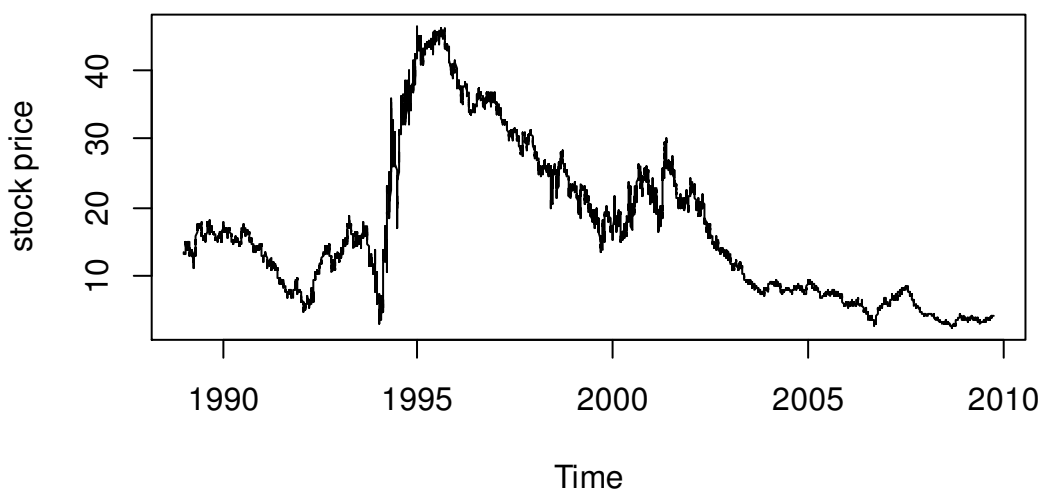
1. Οι μέθοδοι πρόβλεψης υποθέτουν γενικά ότι οι αιτιακές σχέσεις που υπάρχουν στα δεδομένα του παρελθόντος θα συνεχίσουν να εμφανίζονται και μελλοντικά. Δηλαδή οι περισσότερες τεχνικές βασίζονται σε ιστορικά δεδομένα.
2. Οι προβλέψεις είναι σπάνια τέλειες. Για την ακρίβεια είναι συνήθως λανθασμένες. Εμπεριέχουν αβεβαιότητα, οπότε υπάρχει σίγουρα και κάποιο σφάλμα. Συνεπώς στόχος μια διαδικασίας πρόβλεψης είναι η ελαχιστοποίηση του σφάλματος για την καλύτερη δυνατή προσέγγιση της πραγματικής τιμής του υπό μελέτη αντικειμένου.
3. Όσο μεγαλύτερο χρόνο καλύπτει η πρόβλεψη τόσο πιο αναξιόπιστη είναι. Δηλαδή οι βραχυπρόθεσμες προβλέψεις είναι περισσότερο αξιόπιστες από τις μακροπρόθεσμες. Αυτό συμβαίνει γιατί όσο μεγαλώνει ο χρονικός ορίζοντας πρόβλεψης, αυξάνεται η αβεβαιότητα και άρα μεγαλώνει το σφάλμα της πρόβλεψης.
4. Οι προβλέψεις για ομάδες αντικειμένων τείνουν να είναι πιο ακριβείς σε σχέση με τα μεμονωμένα αντικείμενα, αφού τα σφάλματα πρόβλεψης μέσα στην ομάδα τείνουν να αλληλοαναιρούνται.

## 2.1 Χρονοσειρές

Τα ιστορικά δεδομένα στα οποία βασίζονται οι περισσότερες μέθοδοι πρόβλεψης, προέρχονται από μία *χρονοσειρά*. Τέτοια δεδομένα μπορεί να είναι οι μηνιαίες πωλήσεις ενός προϊόντος, οι ημερήσιες τιμές μιας μετοχής, τα ετήσια κέρδη μιας επιχείρησης κ.ά.

Με τον όρο **χρονοσειρά** (ή χρονολογική σειρά), εννοούμε ένα σύνολο διαδοχικών παρατηρήσεων κατά τη διάρκεια ίσων χρονικών περιόδων της τιμής μιας μεταβλητής που μας ενδιαφέρει να μελετήσουμε. Επομένως, αν συμβολίσουμε με  $X_t$  την υπό μελέτη μεταβλητή τη χρονική στιγμή  $i$ , όπου  $i = 1, 2, \dots, t$ , οι χρονικές περιόδους<sup>3</sup> που πήραμε τις παρατηρήσεις της μεταβλητής, τότε το σύνολο  $\{X_1 = x_1, X_2 = x_2, \dots, X_t = x_t\}$  αποτελεί μία χρονοσειρά.

<sup>3</sup> Οι χρονικές περιόδους μπορεί να είναι μέρες, μήνες, χρόνια κλπ



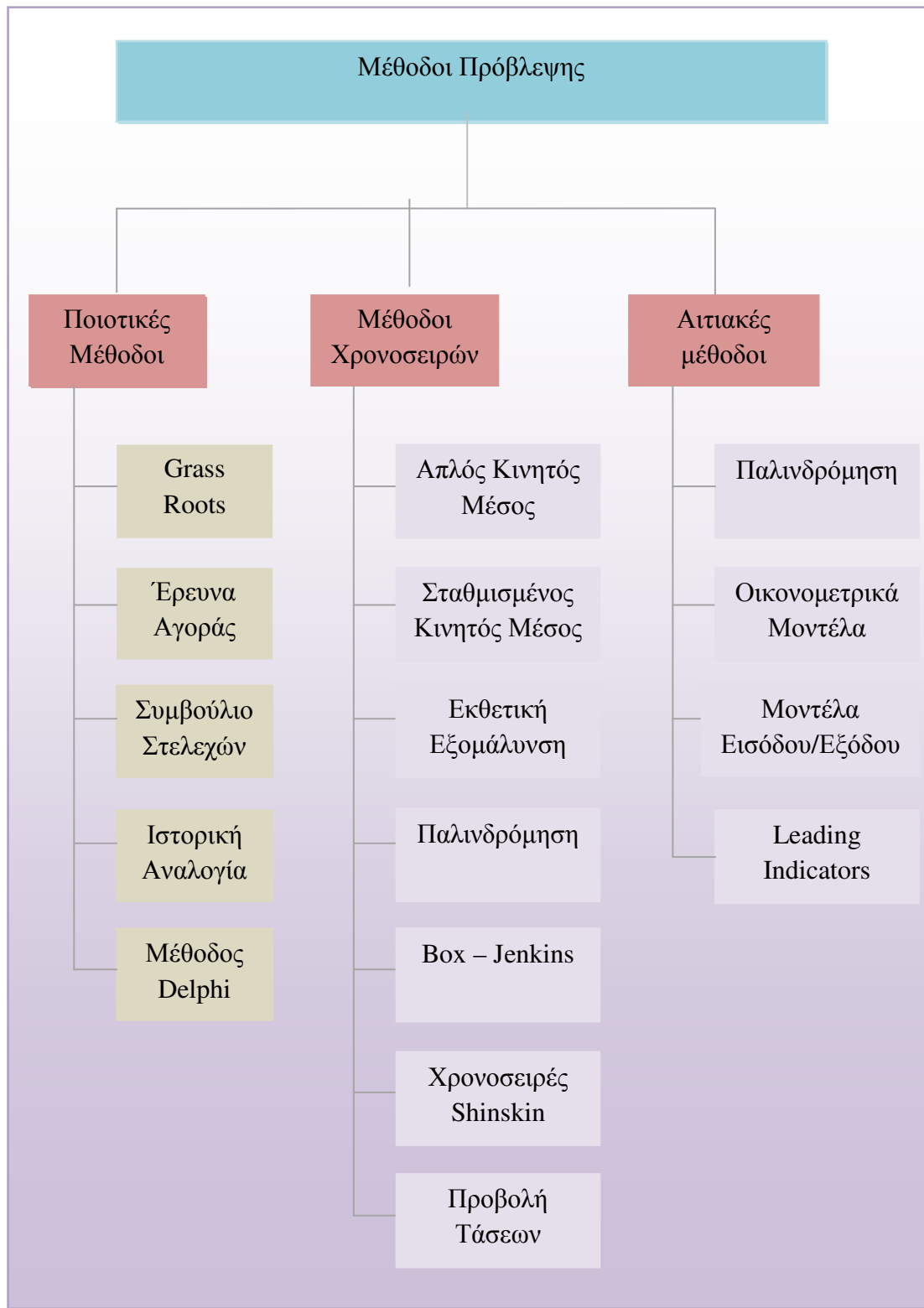
Εικόνα 2.1-1: Παράδειγμα χρονοσειράς που περιγράφει τιμή μετοχής

Ένα χαρακτηριστικό των χρονοσειρών είναι ότι αυτές οι διαδοχικές παρατηρήσεις που μελετάμε δεν είναι ανεξάρτητες μεταξύ τους. Συνεπώς, αν μπορούμε να έχουμε ακριβή εικόνα του μέλλοντος από τα παλαιότερα δεδομένα, δηλαδή ποιοι είναι οι παράμετροι που επηρεάζουν τις τιμές του υπό μελέτη μεγέθους, τότε είναι εύκολο να φτιάξουμε και ένα μαθηματικό μοντέλο που να περιγράφει επ' ακριβώς τη διαχρονική του εξέλιξη. Ένα ντετερμινιστικό μοντέλο για την ακρίβεια. Ωστόσο, στην πραγματικότητα αυτό που συμβαίνει, είναι ότι δε μπορούμε ποτέ να έχουμε μια πλήρη εικόνα του μοντέλου που παράγει τις μελλοντικές τιμές. Και αυτό, γιατί τα μεγέθη που μελετάμε επηρεάζονται από τυχαίους παράγοντες. Οπότε σε αυτή την περίπτωση δουλεύουμε με στοχαστικά μοντέλα.

Συχνά, η επιλογή του κατάλληλου μοντέλου γίνεται παρατηρώντας το μοτίβο της χρονοσειράς. Κάνοντας δηλαδή μια ανάλυση της χρονοσειράς, όπου «σπάμε» τα δεδομένα σε συνιστώσες και τα προβάλλουμε στο μέλλον. Οι 4 πιο συχνές συνιστώσες (*components*) είναι : η **τάση** (*trend*), η **κυκλικότητα** (*cycle*), η **εποχικότητα** (*seasonality*) και οι **μη κανονικές διακυμάνσεις** (*random or irregular variations*).<sup>4</sup>

<sup>4</sup> Θα αναφερθούμε αναλυτικότερα σε αυτά στο κεφάλαιο 3

## 2.2 Μέθοδοι Πρόβλεψης



Στο παραπάνω σχήμα απεικονίζονται διάφορες μέθοδοι πρόβλεψης. Θα δούμε στη συνέχεια πιο αναλυτικά μερικές από αυτές.

Κάθε κατάσταση πρόβλεψης είναι διαφορετική σε σχέση με το χρονικό ορίζοντα που γίνεται η πρόβλεψη, τους παράγοντες που επηρεάζουν το αποτέλεσμα, το μοτίβο που παρουσιάζουν τα δεδομένα και πολλά άλλα. Για αυτό το λόγο έχουν αναπτυχθεί πολλές μέθοδοι πρόβλεψης οι οποίες μπορούν να χωριστούν σε δύο μεγάλες κατηγορίες: τις ποιοτικές και τις ποσοτικές.

### 2.2.1 Ποιοτικές

Οι ποιοτικές μέθοδοι είναι, από τη φύση τους, υποκειμενικές και περιλαμβάνουν τη χρήση αξιολογικής κρίσης, διαίσθησης και συσσωρευμένης γνώσης στην ανάλυση και πρόβλεψη μελλοντικών γεγονότων. Είναι ιδιαίτερα χρήσιμες όταν υπάρχουν λίγα ή καθόλου ιστορικά δεδομένα όπως και σε περιπτώσεις που συμβαίνουν σημαντικά γεγονότα και ενέργειες καθιστώντας τα μέχρι τώρα δεδομένα ακατάλληλα για χρήση στην πρόβλεψη. Η τελική πρόβλεψη μπορεί να είναι αποτέλεσμα της κρίσης και της γνώσης ενός ατόμου ή του συνδυασμού απόψεων μιας ομάδας ατόμων. Παραδείγματα τέτοιων μεθόδων αποτελούν, μια έρευνα καταναλωτών χρησιμοποιώντας ερωτηματολόγιο ή η παρατήρηση των πωλήσεων δύο παρόμοιων προϊόντων όπως τα πλυντήρια και τα στεγνωτήρια. Ακολουθούν κάποιες υπάρχουσες τεχνικές :

1. **Grass Roots** : Η μέθοδος αυτή βασίζεται στην υπόθεση ότι το άτομο που είναι πιο κοντά στο αντικείμενο της πρόβλεψης, ξέρει καλύτερα τις μελλοντικές του ανάγκες. Αυτά είναι τα άτομα που παρέχουν και τα δεδομένα που χρησιμοποιούνται για την πρόβλεψη.
2. **Έρευνα Αγοράς** : Συλλέγονται δεδομένα με διάφορους τρόπους (συνεντεύξεις, ερωτηματολόγια, ...) με στόχο τον έλεγχο συνθηκών της αγοράς.
3. **Συμβούλιο Στελεχών** : Εδώ κυριαρχεί η ιδέα ότι άτομα από διάφορες θέσεις μπορούν να παρέχουν μια πιο αξιόπιστη πρόβλεψη απ' ότι μια μικρή ομάδα ατόμων. Για το λόγο αυτό γίνονται ανοιχτές συνελεύσεις με ελεύθερη ανταλλαγή απόψεων.
4. **Ιστορική αναλογία** : Στη μέθοδο αυτή γίνεται σύνδεση του αντικειμένου πρόβλεψης με ένα παρόμοιο μέγεθος.
5. **Μέθοδος Delphi** : Η μέθοδος αυτή έχει επαναληπτικό χαρακτήρα. Αρχικά μια ομάδα ειδικών από διάφορες θέσεις απαντάει σε ένα ερωτηματολόγιο. Στη συνέχεια αποστέλλεται ένα νέο το οποίο περιλαμβάνει συνοπτικά τις απαντήσεις που δόθηκαν ανώνυμα από όλα τα άτομα στο προηγούμενο ερωτηματολόγιο. Και η διαδικασία επαναλαμβάνεται. Με αυτό τον τρόπο οι συμμετέχοντες μπορούν να δουν και να αξιολογήσουν τις απαντήσεις σχηματίζοντας και διατυπώνοντας την άποψή τους χωρίς να επηρεάζονται από την πίεση των προϊσταμένων. Όταν οι διάφορες απαντήσεις φτάσουν σε ένα ικανοποιητικό εύρος η διαδικασία σταματάει και ο μέσος όρος των απαντήσεων αποτελεί τη ζητούμενη πρόβλεψη.

## 2.2.2 Ποσοτικές

Οι ποσοτικές προβλέψεις μπορούν να εφαρμοσθούν όταν υφίστανται τρεις προϋποθέσεις :

1. Υπάρχουν πληροφορίες για το παρελθόν ώστε να έχουμε επαρκή δεδομένα
2. Οι πληροφορίες αυτές μπορούν να ποσοτικοποιηθούν σε μορφή αριθμητικών δεδομένων
3. Μπορεί να υποτεθεί ότι το μοτίβο των στοιχείων του παρελθόντος θα συνεχίσει να εμφανίζεται μελλοντικά

Η τελευταία προϋπόθεση είναι γνωστή και ως «υπόθεση συνέχειας» (assumption of continuity).

Αυτές οι μέθοδοι είναι ιδιαίτερα χρήσιμες όταν υπάρχουν πολλά ιστορικά δεδομένα για το αντικείμενο που μελετάμε και όταν παρουσιάζονται λίγες ή καθόλου συστηματικές αλλαγές στο υπό μελέτη περιβάλλον. Μπορούμε να τις κατατάξουμε σε δύο βασικές κατηγορίες: την ανάλυση χρονοσειρών και τις αιτιακές μεθόδους.

### Ανάλυση χρονοσειρών

Η ανάλυση χρονοσειρών βασίζεται στην ιδέα ότι οι παρατηρήσεις του παρελθόντος μπορούν να χρησιμοποιηθούν για την πρόβλεψη μελλοντικών παρατηρήσεων και μελετούνται παράγοντες όπως η τάση, η κυκλικότητα και η εποχικότητα. Βασικό τους στόχος είναι η αναγνώριση προτύπων στα δεδομένα του παρελθόντος και η προέκτασή τους στο μέλλον, δεδομένου ότι οι συνθήκες παραμένουν ίδιες. Αν ένα βασικό χαρακτηριστικό αλλάξει επέρχεται αδυναμία πρόβλεψης. Στην κατηγορία αυτή ανήκουν οι τεχνικές: **απλός κινητός μέσος** (*single moving average*), **σταθμισμένος κινητός μέσος** (*weighted moving average*), **εκθετική εξομάλυνση** (*exponential smoothing*), **προβολή τάσεων** (*trend projection*), Box – Jenkins κ.ά.

### Αιτιακές μέθοδοι

Οι αιτιακές μέθοδοι πρόβλεψης βασίζονται στην ιδέα ότι το αντικείμενο πρόβλεψης εξαρτάται με κάποιο τρόπο από έναν ή περισσότερους παράγοντες. Η δυσκολία που παρουσιάζεται είναι να βρούμε τη μαθηματική σχέση που συνδέει το αντικείμενο πρόβλεψης με αυτούς τους παράγοντες. Εδώ ανήκουν τεχνικές όπως η **ανάλυση παλινδρόμησης** (*regression analysis*).

Όποια μέθοδο και αν διαλέξουμε, είναι καλό να χρησιμοποιούμε και την κρίση μας ώστε να αξιοποιήσουμε το αποτέλεσμα με τον καλύτερο δυνατό τρόπο.



## 2.3 Ορίζοντας πρόβλεψης

Ένας από τους παράγοντες που καθορίζουν τη μέθοδο πρόβλεψης που θα χρησιμοποιήσουμε είναι ο ορίζοντας πρόβλεψης. Με τον όρο ορίζοντα πρόβλεψης εννοούμε το χρονικό διάστημα στο μέλλον για το οποίο κάνουμε την πρόβλεψη. Ας το δούμε μέσα από ένα παράδειγμα. Αν βρισκόμαστε στη χρονική περίοδο  $T$  και θέλουμε μια πρόβλεψη για την  $T+4$  τότε ζητάμε μία πρόβλεψη χρονικού ορίζοντα τεσσάρων βημάτων. Το βήμα που θα προχωρήσουμε εξαρτάται από τη συχνότητα των δεδομένων. Δηλαδή αν τα δεδομένα αποτελούνται από μηνιαίες παρατηρήσεις τότε το βήμα που κάνουμε είναι ένας μήνας, αντίστοιχα για ετήσια δεδομένα είναι ένα έτος κλπ.

Ανάλογα με την τιμή του ορίζοντα πρόβλεψης υπάρχουν 3 είδη προβλέψεων:

1. Βραχυπρόθεσμη πρόβλεψη (short-term forecasting)
2. Μεσοπρόθεσμη πρόβλεψη (medium-term forecasting)
3. Μακροπρόθεσμη πρόβλεψη (long-term forecasting)

Οι τιμές του ορίζοντα πρόβλεψης για τη βραχυπρόθεσμη πρόβλεψη είναι σχετικά μικρές και συνήθως δεν ξεπερνούν τις τρεις περιόδους. Η μεσοπρόθεσμη πρόβλεψη αποτελεί τη συνηθέστερη κατηγορία προβλέψεων, όπου ο ορίζοντας πρόβλεψης είναι συνήθως μεγαλύτερος από ένα οικονομικό έτος (12-15 περιόδους για χρονοσειρά με μηνιαίες τιμές). Από την άλλη η μακροπρόθεσμη πρόβλεψη, επειδή αναφέρεται συνήθως σε σχεδιασμούς που δεν έχουν άμεση εφαρμογή, έχει ορίζοντα πρόβλεψης μεγαλύτερο των τριών ετών. Η βιβλιογραφία, μας έχει δείξει, πως οι μέθοδοι εκθετικής εξομάλυνσης δίνουν καλύτερα αποτελέσματα σε βραχυπρόθεσμες και μεσοπρόθεσμες προβλέψεις, ενώ οι μέθοδοι παλινδρόμησης είναι πιο αποδοτικές σε μακροπρόθεσμες προβλέψεις.

Η διαδικασία επιλογής της αποδοτικότερης μεθόδου πρόβλεψης δεν είναι συνήθως κάτι απλό. Σε ότι αφορά τον ορίζοντα πρόβλεψης, από τη στιγμή που θα κάνουμε μια επιλογή μεθόδου πρόβλεψης ακολουθεί η περίοδος ελέγχου όπου επαληθεύουμε τα πραγματικά δεδομένα με αυτά της μεθόδου πρόβλεψης και βλέπουμε αν υπάρχουν αποκλίσεις. Μόλις τελειώσει ο χρονικός ορίζοντας της πρόβλεψης και υπολογίσουμε τα σφάλματα πρόβλεψης μπορούμε να τα χρησιμοποιήσουμε μαζί με τα νέα δεδομένα για να κάνουμε νέες προβλέψεις. Ωστόσο μπορεί να παρατηρήσουμε ότι τα πραγματικά δεδομένα αποκλίνουν συστηματικά σε μεγάλο βαθμό από τα προβλεπόμενα, πολύ πριν περάσει ο ορίζοντας πρόβλεψης, και τότε η διαδικασία ελέγχου σταματάει.

Από τα παραπάνω είναι εύκολο να συμπεράνουμε ότι ο ορίζοντας πρόβλεψης είναι σημαντικός για τουλάχιστον 2 λόγους. Ο ένας είναι ότι οι προβλέψεις αλλάζουν ανάλογα με τον ορίζοντα πρόβλεψης. Ο δεύτερος είναι ότι τα μοντέλα πρόβλεψης αλλάζουν. Δηλαδή αν έχω ένα μοντέλο που περιγράφει πολύ καλά τις προβλέψεις βραχυπρόθεσμα αυτό δε σημαίνει ότι το ίδιο θα κάνει και μακροπρόθεσμα. Ενδεχομένως να χρειαστούν άλλες παράμετροι.

## 2.4 Βασικά βήματα σε μια διαδικασία πρόβλεψης

Υπό την προϋπόθεση ότι έχουμε στα χέρια μας ποσοτικά δεδομένα, η διαδικασία πρόβλεψης που μπορούμε να ακολουθήσουμε, κατά τους Μακρυδάκη, Wheelright και Hyndman (1998), αποτελείται από πέντε βήματα.

- **Βήμα 1: Καθορισμός του προβλήματος**  
Σε αυτό το κομμάτι αντιμετωπίζεται συχνά η μεγαλύτερη δυσκολία στη διαδικασία της πρόβλεψης. Χρειάζεται πολύ καλή κατανόηση σε ότι αφορά το τι θέλουμε να προβλέψουμε, πως θέλουμε να γίνει αυτό και από ποιους θα χρησιμοποιηθούν οι προβλέψεις.
- **Βήμα 2: Συλλογή των δεδομένων**  
Είναι πολύ σημαντικό να συλλεχθούν δεδομένα των αντικειμένων που μας ενδιαφέρουν. Αυτά μπορούν να είναι είτε στατιστικά (συνήθως αριθμητικά δεδομένα), είτε συσσωρευμένη γνώση και εμπειρία των εργαζομένων.
- **Βήμα 3: Προκαταρκτική (διερευνητική) ανάλυση**  
Σε αυτό το βήμα προσπαθούμε να πάρουμε μια αίσθηση του τι μας λένε τα δεδομένα. Αναλύοντάς τα χρησιμοποιώντας γραφικά μέσα και στατιστική μεθοδολογία, μπορούμε να αναγνωρίσουμε πιθανά λανθάνοντα πρότυπα, αν υπάρχει σημαντική τάση ή εποχικότητα καθώς και ακραίες τιμές που να μπορούν να εξηγηθούν από ειδικευμένα άτομα.
- **Βήμα 4: Επιλογή μεθόδων πρόβλεψης**  
Φτάνουμε πια στην επιλογή του καταλληλότερου μοντέλου πρόβλεψης και των παραμέτρων του, χρησιμοποιώντας τα ιστορικά δεδομένα που έχουμε στη διάθεσή μας και βασιζόμενοι στις ιδιότητες των διαφόρων μεθόδων πρόβλεψης.
- **Βήμα 5: Χρήση και αξιολόγηση των μοντέλων πρόβλεψης**  
Έχοντας επιλέξει το μοντέλο που θα χρησιμοποιήσουμε, παράγουμε τις προβλέψεις. Όταν ξεκινήσει η περίοδος πρόβλεψης και τα πρώτα δεδομένα γίνουν διαθέσιμα, τότε μπορεί να ξεκινήσει και η αξιολόγηση του μοντέλου καθώς και η μέτρηση της ακρίβειας των προβλέψεων.

# 3

## Εργαλεία πρόβλεψης

Σε αυτό το κεφάλαιο θα αναφερθούμε στη μεθοδολογία που ακολουθείται για την ανάλυση των δεδομένων, όσον αφορά τις ποσοτικές μεθόδους, όπως αυτή προκύπτει από την εμπειρία αλλά και τη στατιστική. Συγκεκριμένα θα ασχοληθούμε με: α) τις γραφικές μεθόδους που υπάρχουν ώστε να δούμε οπτικά τα δεδομένα και τι συμπεράσματα μπορούμε να βγάλουμε, β) βασικούς στατιστικούς δείκτες που μας δίνουν συνοπτικά κάποια χαρακτηριστικά του δείγματος, γ) μέτρα που μας βοηθούν να αξιολογήσουμε την ακρίβεια των προβλέψεών μας και δ) το πως διορθώνουμε τα αρχικά δεδομένα όταν αυτό κρίνεται απαραίτητο.

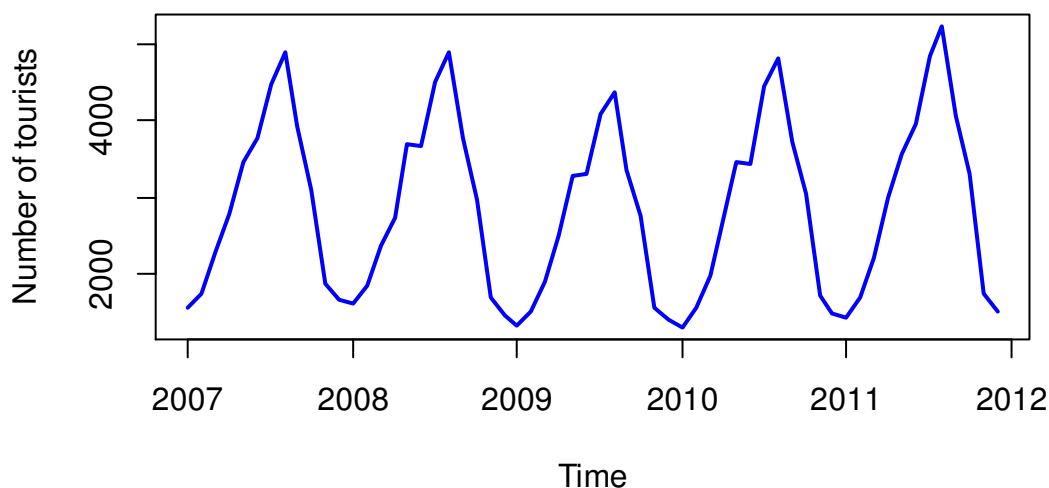
### 3.1 Γραφική Αναπαράσταση

Οι γραφικές αναπαραστάσεις των δεδομένων, αποτελούν μια πολύ σημαντική διαδικασία που πρέπει να ακολουθήσουμε για να αποκτήσουμε μια πρώτη εικόνα των δεδομένων που έχουμε να αναλύσουμε. Μας βοηθούν να εντοπίσουμε, αν υπάρχουν, μοτίβα ή ακραίες παρατηρήσεις, καθώς και να εξηγήσουμε όπου είναι δυνατόν, τις διακυμάνσεις που παρατηρούνται στα δεδομένα. Ο τύπος των δεδομένων καθορίζει και ποιο είναι το καταλληλότερο γράφημα. Στις προβλέψεις μας απασχολούν κυρίως δύο τύποι δεδομένων: τα διαστρωματικά στοιχεία (*cross-sectional data*) που αποτελούνται από παρατηρήσεις διαφόρων μεταβλητών αναφερόμενες όλες στην ίδια χρονική περίοδο και τις χρονολογικές σειρές ή χρονοσειρές δεδομένων (*time series data*) που αποτελούνται από μια αλληλουχία διαχρονικών παρατηρήσεων μίας μεταβλητής.

#### 3.1.1 Μοτίβα χρονοσειρών

Όπως ήδη έχουμε αναφέρει, θα συναντήσουμε αρκετά συχνά δεδομένα σε μορφή χρονοσειράς. Και ακριβώς επειδή η χρονοσειρά αποτελεί μια ακολουθία κατανεμημένων παρατηρήσεων ως προς το χρόνο, το πρώτο πράγμα που συνίσταται να κάνουμε είναι μια γραφική αναπαράσταση των δεδομένων ως προς το χρόνο (*time plot*). Το γράφημα αυτό θα μας βοηθήσει να εντοπίσουμε σε πρώτη φάση τυχόν τάσεις στα δεδομένα ή εποχικότητα και άλλα συστηματικά χαρακτηριστικά, ώστε να μπορέσουμε αργότερα να τα ενσωματώσουμε στο στατιστικό μοντέλο που προορίζεται για την πρόβλεψη.

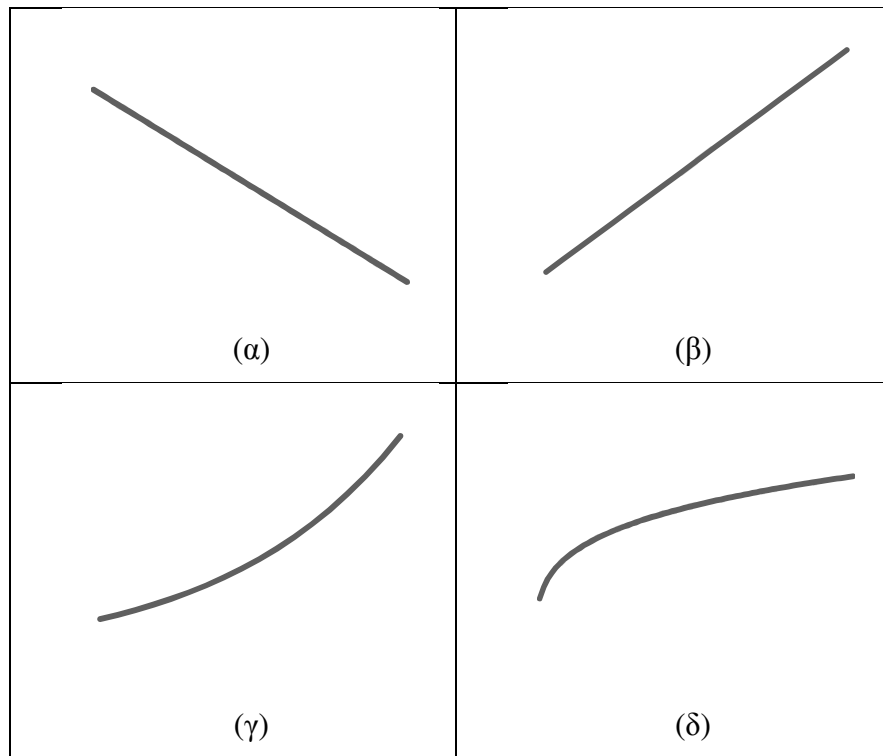
Το παρακάτω γράφημα απεικονίζει ένα time plot των δεδομένων για τον τουρισμό της Ισπανίας τα έτη 2007 έως 2011. Με μια πρώτη ματιά είναι ευδιάκριτα το εύρος των δεδομένων καθώς και τα σημεία στα οποία παρουσιάζονται μέγιστες τιμές (*peaks*). Όπως είναι αναμενόμενο, τις καλοκαιρινές περιόδους επισκέπτονται τη χώρα πολλοί τουρίστες με τους περισσότερους να καταφθάνουν τον Αύγουστο.



Εικόνα 3.1-1: Γράφημα αριθμού τουριστών (σε χιλιάδες) της Ισπανίας τα έτη 2007-2011

Σε επόμενη φάση πρέπει να αναγνωρίσουμε τα μοτίβα που κυριαρχούν στα δεδομένα μας, ώστε να μπορέσουμε αργότερα να κάνουμε και την επιλογή του καταλληλότερου μοντέλου πρόβλεψης. Μπορεί να υπάρχει ένα μοτίβο, αλλά μπορεί και κάποιος συνδυασμός τους. Τέσσερις κατηγορίες μοτίβων μπορούμε να διακρίνουμε: 1)τάση, 2)κυκλικότητα, 3)εποχικότητα, και 4) μη κανονικές διακυμάνσεις.

Τάση: Σχετίζεται με τις μακροπρόθεσμες μεταβολές, με τάση προς τα πάνω ή προς τα κάτω, που παρατηρούνται στα δεδομένα. Ενδέχεται, σε κάποιες περιπτώσεις, να χρειαστούν δεδομένα μεγάλου χρονικού διαστήματος προκειμένου να καταφέρουμε να τις εντοπίσουμε και να τις περιγράψουμε. Ωστόσο αυτές οι μακροπρόθεσμες αυξήσεις ή μειώσεις δε χρειάζεται να είναι γραμμικές. Μπορεί να παρατηρηθούν αυξητικές τάσεις που ακολουθούνται από πτωτικές τάσεις και το αντίστροφο. Τέτοιες μεταβολές συνήθως εμφανίζονται από την εξέλιξη της τεχνολογίας, τις μεταβολές της παραγωγικότητας, τον πληθωρισμό και τις αλλαγές στον πληθυσμό.

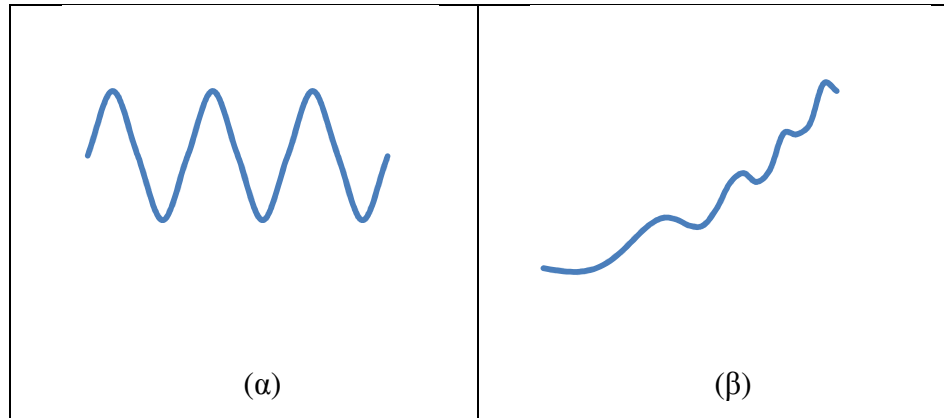


Εικόνα 3.1-2: Παραδείγματα τάσης σε μια χρονοσειρά

**Κυκλικότητα:** Εδώ παρατηρούνται επαναλαμβανόμενες διακυμάνσεις κυματοειδής μορφής στα δεδομένα, με συχνότητα μεγαλύτερη από ένα έτος, που δεν έχουν σταθερή περίοδο εμφάνισης. Δηλαδή βλέπουμε περιόδους διαστολής που ακολουθούνται από περιόδους συστολής. Αυτό αποδίδεται κυρίως στο φαινόμενο των επιχειρηματικών κύκλων (business cycles)<sup>5</sup>, η περίοδος των οποίων μπορεί να κυμαίνεται από ένα έτος έως και τριάντα.

**Εποχικότητα:** Εντοπίζεται όταν έχουμε περιοδικές ή επαναλαμβανόμενες διακυμάνσεις των σημείων των δεδομένων πάνω ή κάτω από την τιμή τάσης, με συχνότητα συνήθως έως ένα έτος. Ένα σημαντικό χαρακτηριστικό της εποχικότητας είναι ότι η περίοδος που εμφανίζεται είναι πάντα σταθερή και γνωστή. Τέτοια μοτίβα προκύπτουν, όταν η χρονοσειρά επηρεάζεται από εποχιακούς παράγοντες όπως οι μήνες, το τρίμηνο του έτους, μια συγκεκριμένη μέρα της εβδομάδας κ.ά. Στην κατηγορία αυτή εμπίπτουν οι καταναλωτικές συμπεριφορές συγκεκριμένων προϊόντων όπως τα αναψυκτικά, τα ψυγεία και τα μάλλινα ρούχα, η κατανάλωση ηλεκτρικού ρεύματος των νοικοκυριών κλπ.

<sup>5</sup> Αναφερόμενοι σε business cycles εννοούμε φαινόμενα όπως η ύφεση, ο πληθωρισμός, η ανεργία και η ευημερία.



Εικόνα 3.1-3: Παραδείγματα εποχικότητας σε μια χρονοσειρά

Μη κανονικές διακυμάνσεις: Αναφέρονται κυρίως στις μικρές ασταθείς διακυμάνσεις που παρατηρούνται αφού έχουν απομονωθεί οι υπόλοιπες συνιστώσες (τάση, κυκλικότητα και εποχικότητα). Δεν ακολουθούν κάποιο συγκεκριμένο μοτίβο και προέρχονται συνήθως από απρόβλεπτα ή μη επαναλαμβανόμενα γεγονότα. Τέτοια γεγονότα μπορεί να είναι οι πλημμύρες, οι απεργίες, οι εκλογές, οι κλιματικές αλλαγές, ακόμα και αλλαγές στη νομοθεσία. Είναι φυσικό αυτή η τυχαιότητα που έχουν να τα καθιστά δύσκολα στην πρόβλεψη.

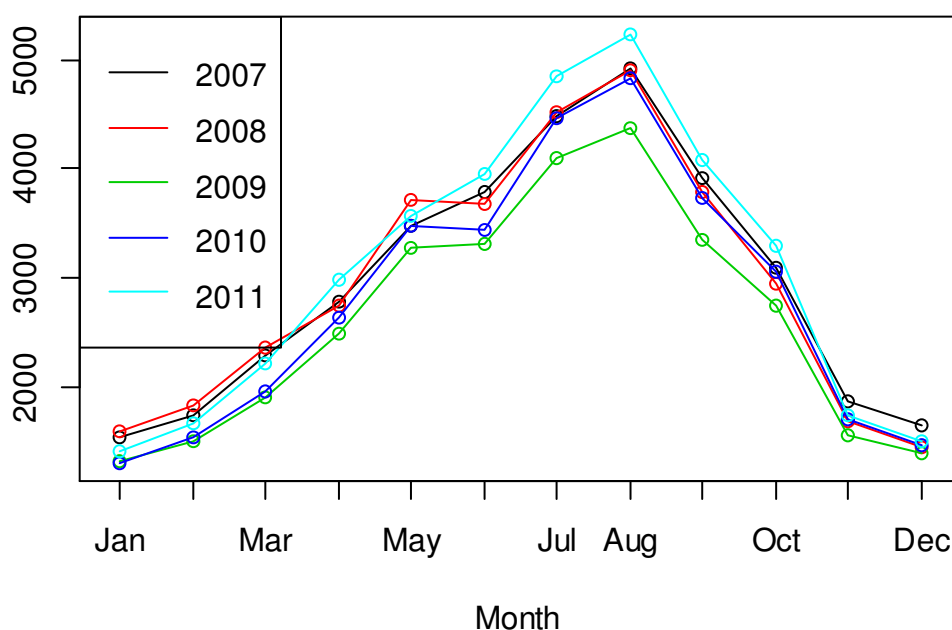
Υπάρχει και μία άλλη κατηγορία χρονοσειρών. Σε αυτές απουσιάζουν τα παραπάνω μοτίβα. Βασικό χαρακτηριστικό τους είναι ότι οι τιμές των δεδομένων κυμαίνονται γύρω από μια μέση τιμή (*level*) και δεν παρατηρείται κάποια συστηματική αλλαγή στο μέσο όρο ή τη διασπορά των παρατηρήσεων (Σχήμα (α)). Σε αυτή την κατηγορία ανήκουν και τα προϊόντα των οποίων οι πωλήσεις δε μεταβάλλονται με το χρόνο και έχουν κατά συνέπεια σταθερή ζήτηση (Σχήμα (β)). Τέτοιες χρονοσειρές καλούνται στάσιμες (*stationary*).

Πολύ πιθανό είναι να συναντήσουμε και ασυνέχειες (*outliers*) οι οποίες οφείλονται κυρίως σε ασυνήθιστες τιμές ή αλλαγή του επιπέδου των τιμών (*level shift*).

### 3.1.2 Γράφημα εποχικότητας

Στην περίπτωση που στη χρονοσειρά παρατηρήσουμε μοτίβο εποχικότητας είναι χρήσιμο να κάνουμε ένα γράφημα όπου θα λαμβάνουμε υπόψη την εποχικότητα (*Seasonal plot*). Στην εικόνα 3.1-4 παρουσιάζεται ένα τέτοιο γράφημα για τον τουρισμό της Ισπανίας που είδαμε προηγουμένως.

### Seasonal plot: ts



Εικόνα 3.1-4: Διάγραμμα εποχικότητας αριθμού τουριστών Ισπανίας τα έτη 2007-2011

Στο παραπάνω γράφημα μπορούμε να παρατηρήσουμε ότι τα δεδομένα που έχουμε συλλέξει είναι σχεδιασμένα ανά περιόδους. Εδώ ως περίοδο έχουμε επιλέξει το ένα έτος. Μας βοηθάει να δούμε πιο καθαρά την εποχικότητα που υποβόσκει στα δεδομένα καθώς και αν υπήρξε κάποια απόκλιση από το μοτίβο σε κάποιο έτος. Για παράδειγμα φαίνεται ότι το έτος 2011 και το μήνα Αύγουστο η επισκεψιμότητα των τουριστών ξεπέρασε κάθε προηγούμενο, σε αντίθεση με το 2009 που παρατηρείται η χαμηλότερη τιμή.

### 3.2 Στατιστικοί Δείκτες

Κάτι που μας βοηθάει πολύ στο να αποκτήσουμε μια δεύτερη εικόνα των δεδομένων είναι κάποιοι βασικοί στατιστικοί δείκτες. Πρόκειται για μια συνοπτική παρουσίαση των δεδομένων με τη βοήθεια αριθμητικών μεθόδων. Όλα όσα θα αναφέρουμε αφορούν τη μελέτη μιας μεταβλητής ή μιας χρονοσειράς  $X$  μεγέθους  $n$  παρατηρήσεων ( $X_1, X_2, \dots, X_n$ ). Να σημειώσουμε ότι μπορούμε να υπολογίσουμε με μεγάλη ευκολία τα παρακάτω μεγέθη χρησιμοποιώντας διάφορα στατιστικά πακέτα όπως είναι η R, το Minitab, το SPSS κ.ά.

#### ΜΕΣΗ ΤΙΜΗ (AVERAGE)

Είναι ένας γραμμικός μέσος όρος των παρατηρήσεων που μας δείχνει το επίπεδο γύρω από το οποίο κινούνται οι τιμές. Συμβολίζεται ως  $\bar{X}$  και υπολογίζεται από τη σχέση:

$$\begin{aligned}\bar{X} &= (X_1 + X_2 + \dots + X_n)/n \\ &= \frac{1}{n} * \sum_{i=1}^n X_i\end{aligned}$$

#### ΔΙΑΜΕΣΟΣ (*MEDIAN*)

Αποτελεί τη μεσαία παρατήρηση ενός δείγματος. Για να την υπολογίσουμε διατάσσουμε τα δεδομένα σε αύξουσα σειρά και παίρνουμε τη μεσαία παρατήρηση. Βέβαια στην περίπτωση που το μέγεθος που δείγματος είναι άρτιος αριθμός τότε η διάμεσος είναι το ημιάθροισμα των 2 μεσαίων παρατηρήσεων.

Τόσο η μέση τιμή όσο και η διάμεσος μας δείχνουν τη θέση του κέντρου των παρατηρήσεων. Ένα μειονέκτημα όμως της μέσης τιμής είναι ότι επηρεάζεται από τις ακραίες παρατηρήσεις, κάτι που δεν ισχύει για τη διάμεσο.

#### ΜΕΓΙΣΤΗ ΚΑΙ ΕΛΑΧΙΣΤΗ ΤΙΜΗ (*MAXIMUM / MINIMUM*)

Δείχνουν ποιες είναι οι ακραίες παρατηρήσεις και βοηθούν να αναγνωρίσουμε το εύρος των δεδομένων. Συνήθως υπολογίζονται εκχωρώντας σε κάποιο στατιστικό πακέτο τις εντολές max και min αντίστοιχα.

Πολύ χρήσιμο είναι να δούμε πόσο αποκλίνουν οι παρατηρήσεις από τη μέση τιμή ( $X_i - \bar{X}$ ) ώστε να καταλάβουμε αν οι παρατηρήσεις βρίσκονται σχετικά κοντά ή είναι διασκορπισμένες σε μεγάλο εύρος. Το άθροισμα αυτών των διαφορών θα είναι πάντα μηδέν, οπότε για να μπορέσουμε να αξιοποιήσουμε την πληροφορία που μας δίνουν ή παίρνουμε την απόλυτη τιμή τους ή τις υψώνουμε στο τετράγωνο. Δύο τέτοια μέτρα είναι η διασπορά και η τυπική απόκλιση.

#### ΔΙΑΣΠΟΡΑ (*VARIANCE*)

Χρησιμεύει στο να εκφράσουμε το πόσο μακριά βρίσκονται οι παρατηρήσεις από τη μέση τιμή. Συμβολίζεται με  $S^2$  και υπολογίζεται από τη σχέση:

$$S^2 = \frac{1}{n-1} * \sum_{i=1}^n (X_i - \bar{X})^2$$

#### ΤΥΠΙΚΗ ΑΠΟΚΛΙΣΗ (*STANDARD DEVIATION*)

Είναι η ρίζα της διασποράς. Ο λόγος που παίρνουμε τη ρίζα είναι για να εκφράσουμε το αποτέλεσμα σε μονάδες ίδιες με τις παρατηρήσεις μας. Συμβολίζεται με  $S$  και δίνεται από τον τύπο:

$$S = \sqrt{S^2}$$

Όταν οι παρατηρήσεις είναι διεσπαρμένες τότε θα βρίσκονται μακριά από τη μέση τιμή. Συνεπώς οι τιμές της διασποράς και της τυπικής απόκλισης θα είναι μεγάλες.

Στην περίπτωση των χρονοσειρών μπορούμε να εξετάσουμε πόσο σχετίζονται οι παρατηρήσεις μιας περιόδου με άλλες διαφορετικής χρονικής περιόδου. Δηλαδή αν πάρω την παρατήρηση στο χρόνο  $t$ ,  $Y_t$ , και την παρατήρηση στον χρόνο  $t-1$ ,  $Y_{t-1}$ , μπορώ να δω πως σχετίζονται μεταξύ τους. Το ίδιο γίνεται και για παρατηρήσεις που διαφέρουν 2 χρονικές περιόδους, 3 χρονικές περιόδους κλπ.



#### ΣΥΝΤΕΛΕΣΤΗΣ ΑΥΤΟΔΙΑΣΠΟΡΑΣ (*AUTOCOVARIANCE COEFFICIENT*)

Περιγράφει τη γραμμική συσχέτιση μεταξύ των τιμών μιας χρονοσειράς με χρονική υστέρηση  $k$  περιόδους. Συμβολίζεται με  $c_k$  και δίνεται από τη σχέση:

$$c_k = \frac{1}{n} * \sum_{t=k+1}^n (Y_t - \bar{Y}) * (Y_{t-k} - \bar{Y})$$

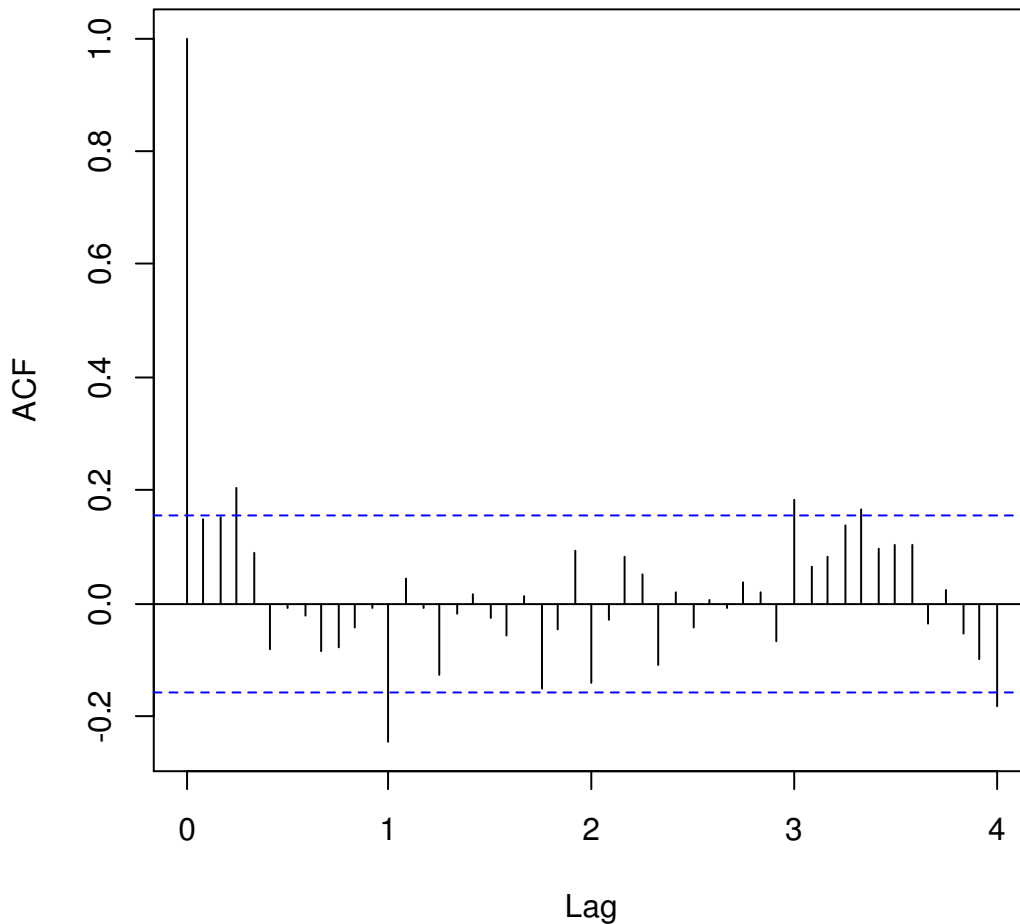
#### ΣΥΝΤΕΛΕΣΤΗΣ ΑΥΤΟΣΥΣΧΕΤΙΣΗΣ (*AUTOCORRELATION COEFFICIENT*)

Όπως ο συντελεστής αυτοδιασποράς, έτσι και ο συντελεστής αυτοσυσχέτισης περιγράφει τη γραμμική συσχέτιση μεταξύ των παρατηρήσεων με χρονική υστέρηση  $k$  περιόδους και παίρνει τιμές στο διάστημα  $[-1,1]$ . Αν η τιμή του είναι κοντά στο μηδέν τότε δεν υφίσταται συσχέτιση μεταξύ των παρατηρήσεων, ενώ αν είναι κοντά στο 1 ή το  $-1$  τότε υπάρχει ισχυρή συσχέτιση (θετική ή αρνητική αντίστοιχα). Επίσης, είναι πολύ χρήσιμος στον εντοπισμό εποχιακής συμπεριφοράς στα δεδομένα. Μια διαφορά με τον συντελεστή αυτοδιασποράς είναι ότι δεν έχει μονάδα μέτρησης. Για δεδομένα λοιπόν, με χρονική υστέρηση  $k$ , ο συντελεστής αυτοσυσχέτισης συμβολίζεται ως  $r_k$  και προσδιορίζεται από τη σχέση:

$$r_k = \frac{\sum_{t=k+1}^n (Y_t - \bar{Y}) * (Y_{t-k} - \bar{Y})}{\sum_{t=1}^n (Y_t - \bar{Y})^2} = \frac{c_k}{c_0}$$

Μπορεί να υπολογιστεί και με τη χρήση υπολογιστικού φύλλου μέσω της συνάρτησης *acf*.

## ACF diagram



Εικόνα 3.2-1: Συσχετόγραμμα / Διάγραμμα acf

Αν συγκεντρώσουμε τις αυτοσυσχετίσεις για κάθε περίοδο υστέρησης (1,2,3, ...), τότε παίρνουμε τη *συνάρτηση αυτοσυσχέτισης* (ACF). Έτσι αντί να επεξεργαζόμαστε τις τιμές, μπορούμε να έχουμε μια πιο γρήγορη εικόνα κάνοντας ένα γράφημα των συντελεστών αυτοσυσχέτισης συναρτήσει της χρονικής υστέρησης  $k$  (συσχετόγραμμα / correlogram). Μέσα από ένα τέτοιο γράφημα μπορούμε να ελέγξουμε την ύπαρξη εποχικότητας, κυκλικότητας και άλλων μοτίβων χρονοσειρών. Ένα παράδειγμα φαίνεται στην Εικόνα 3.2-1. Θα δούμε αναλυτικά την εφαρμογή της στην ενότητα 4.3.

### 3.3 Μέτρηση ακρίβειας πρόβλεψης

Στόχος μας πάντα είναι να βρούμε το καταλληλότερο μοντέλο πρόβλεψης για τα δεδομένα μας. Ένας πολύ σημαντικός παράγοντας είναι η ακρίβεια των προβλέψεων. Πόσο κοντά δηλαδή βρίσκονται οι προβλέψεις που μας δίνει το μοντέλο με τις

πραγματικές τιμές. Σε αυτή την ενότητα θα αναφερθούμε στις διάφορους δείκτες που μετράνε την ακρίβεια των προβλέψεων.

### 3.3.1 Τυπικά στατιστικά μέτρα

Για την επιλογή της κατάλληλης μεθόδου πρόβλεψης χρησιμοποιούνται κάποια μέτρα αξιολόγησης, τα οποία βασίζονται στις τιμές των αποκλίσεων των προβλεπόμενων τιμών από τις αντίστοιχες τιμές μιας χρονοσειράς. Τα λεγόμενα σφάλματα της πρόβλεψης. Αν  $F_t$  είναι η τιμή της πρόβλεψης και  $Y_t$  η τιμή της χρονοσειράς για μία περίοδο, τότε το σφάλμα της πρόβλεψης (*forecast error*) συμβολίζεται με  $e_t$  και δίνεται από τον τύπο:

$$e_t = Y_t - F_t$$

Κάθε  $F_t$  έχει υπολογιστεί με εφαρμογή μοντέλου πρόβλεψης ενός βήματος (*one-step forecast*) στις παρατηρήσεις  $Y_1, Y_2, Y_3, \dots, Y_{t-1}$ .

Γενικά αν  $Y_1, Y_2, \dots, Y_n$  είναι τα δεδομένα που εφαρμόζουμε το μοντέλο πρόβλεψης τότε οι προβλεπόμενες τιμές που αναμένουμε είναι  $F_1, F_2, \dots, F_{n+h}$ , όπου  $h$  ο ορίζοντας πρόβλεψης. Οι πρώτες παρατηρήσεις αποτελούν την περίοδο προσαρμογής του μοντέλου στα δεδομένα, ενώ οι υπόλοιπες τις μελλοντικές προβλέψεις. Είναι προφανές ότι στην περίπτωση των σφαλμάτων λαμβάνονται υπόψη μόνο οι τιμές της πρόβλεψης για τις οποίες υπάρχουν ήδη δεδομένα προς επαλήθευση.

Αν λοιπόν έχουμε παρατηρήσεις και προβλέψεις για  $n$  περιόδους τότε έχουμε και πόρους για τα σφάλματα και μπορούμε να ορίσουμε τα παρακάτω στατιστικά μέτρα:

#### **ΜΕΣΟ ΣΦΑΛΜΑ (MEAN ERROR)**

Αποτελεί ένα μέτρο συστηματικότητας του σφάλματος και στην ουσία βρίσκει το μέσο όρο των σφαλμάτων. Όταν δεν υπάρχουν συστηματικά σφάλματα η τιμή του είναι κοντά στο μηδέν, αφού τα σφάλματα θα αναιρούν το ένα το άλλο καθώς άλλα θα είναι θετικά και άλλα αρνητικά. Η βασική του ιδιότητα είναι να δείχνει αν υπάρχει bias στα δεδομένα και δε δίνει ιδιαίτερες πληροφορίες για το μέγεθος των σφαλμάτων. Θετικές τιμές του μέσου σφάλματος δηλώνουν ότι οι προβλέψεις ήταν μεγαλύτερες των πραγματικών τιμών, ενώ στην περίπτωση αρνητικών τιμών μικρότερες. Υπολογίζεται από τον τύπο:

$$ME = \frac{1}{n} * \sum_{t=1}^n e_t$$

#### **ΜΕΣΟ ΑΠΟΛΥΤΟ ΣΦΑΛΜΑ (MEAN ABSOLUTE ERROR)**

Αυτό το μέτρο διορθώνει κατά κάποιο τρόπο το πρόβλημα με τα πρόσημα των σφαλμάτων παίρνοντας για κάθε σφάλμα την απόλυτη τιμή του και δείχνει τη μέση αστοχία της πρόβλεψης. Όσο μεγαλύτερη είναι η τιμή του τόσο χειρότερη είναι και η ακρίβεια της πρόβλεψης. Δίνεται από τον τύπο:

$$MAE = \frac{1}{n} * \sum_{t=1}^n |e_t|$$

### **ΜΕΣΟ ΤΕΤΡΑΓΩΝΙΚΟ ΣΦΑΛΜΑ (*MEAN SQUARED ERROR*)**

Αποτελεί και αυτό ένα μέτρο ακρίβειας της πρόβλεψης με τη διαφορά από τα προηγούμενα ότι υπολογίζει το μέσο όρο των τετραγώνων των σφαλμάτων. Κατά συνέπεια η μονάδα μέτρησής του είναι εκφρασμένη στη μονάδα μέτρησης των παρατηρήσεων υψωμένη στο τετράγωνο. Βασικό του μειονέκτημα είναι ότι επηρεάζεται από τα μεγάλα σφάλματα που προκύπτουν από τις ακραίες παρατηρήσεις. Δίνεται από τον τύπο:

$$MSE = \frac{1}{n} * \sum_{t=1}^n e_t^2$$

### **ΡΙΖΑ ΜΕΣΟΥ ΤΕΤΡΑΓΩΝΙΚΟΥ ΣΦΑΛΜΑΤΟΣ (*ROOT MEAN SQUARED ERROR*)**

Αυτό το μέτρο χρησιμοποιείται για να διορθωθεί το πρόβλημα με τη μονάδα μέτρησης του MSE. Για να εκφραστεί στις ίδιες μονάδες μέτρησης με τις τιμές της χρονοσειράς, παίρνουμε τη θετική τιμή της τετραγωνικής ρίζας του . Έτσι προκύπτει το RMSE που έχει τις ίδιες ιδιότητες με το MSE και δίνεται από τον τύπο:

$$RMSE = \sqrt{MSE} = \sqrt{\frac{1}{n} * \sum_{t=1}^n e_t^2}$$

Καθένα από τα παραπάνω 4 μέτρα εκφράζει την ακρίβεια της πρόβλεψης με βάση την κλίμακα των δεδομένων χωρίς να βοηθούν ιδιαίτερα σε περιπτώσεις σύγκρισης δύο ή περισσότερων χρονοσειρών καθώς και σε συγκρίσεις ανάμεσα σε διαφορετικά χρονικά διαστήματα. Για παράδειγμα άλλο βάρος έχει ένα σφάλμα όταν αναφέρεται σε μηνιαίες προβλέψεις και άλλο το ίδιο σφάλμα σε ετήσιες. Αυτό βέβαια αντιμετωπίζεται αν δουλέψουμε με ποσοστιαία σφάλματα. Όσο μικρότερο το ποσοστό, τόσο καλύτερη η ακρίβεια της πρόβλεψης.

Ένα ποσοστιαίο σφάλμα ορίζεται από τον τύπο:

$$PE_t = \left( \frac{Y_t - F_t}{Y_t} \right) * 100$$

Δουλεύοντας με αυτά τα σφάλματα διακρίνουμε τα παρακάτω 2 μέτρα:

### **ΜΕΣΟ ΠΟΣΟΣΤΙΑΙΟ ΣΦΑΛΜΑ (*MEAN PERCENTAGE ERROR*)**

Όπως και το ME, χρησιμοποιείται όταν μας ενδιαφέρει να προσδιορίσουμε αν η μέθοδος πρόβλεψης είναι μεροληπτική, δηλαδή αν έχουμε προβλέψεις που είναι συστηματικά μεγαλύτερες ή μικρότερες από τις αντίστοιχες πραγματικές. Αν η απόλυτη τιμή του είναι μεγάλη τότε σίγουρα υπάρχει μεροληψία οπότε όσο πιο κοντά στο μηδέν είναι τόσο πιο αμερόληπτη θα είναι η μέθοδος. Δίνεται από τον τύπο:

$$MPE = \frac{1}{n} * \sum_{t=1}^n PE_t$$

**ΜΕΣΟ ΑΠΟΛΥΤΟ ΠΟΣΟΣΤΙΑΙΟ ΣΦΑΛΜΑ (MEAN ABSOLUTE PERCENT ERROR)**

$$MAPE = \frac{1}{n} * \sum_{t=1}^n |PE_t|$$

Αυτό το μέτρο δε μπορεί να χρησιμοποιηθεί όταν έχουμε χρονοσειρές διακοπτόμενης ζήτησης (δηλαδή πολλές μηδενικές τιμές) γιατί καταλήγει σε απροσδιοριστία.

### 3.3.2 Σύγκριση μεθόδων πρόβλεψης

Για να καταλάβουμε πόσο καλά δουλεύει μια μέθοδος δεν αρκεί να δούμε τα παραπάνω στατιστικά μέτρα που αναφέραμε. Χρειάζεται να έχουμε και ένα μέτρο σύγκρισης ως προς τα οφέλη που κερδίζουμε με το να εφαρμόσουμε τη συγκεκριμένη μέθοδο. Εξάλλου δεν είναι λίγες οι φορές που στις χρονοσειρές εφαρμόζονται περισσότερες από μια μέθοδοι πρόβλεψης.

Μια μέθοδος αναφοράς για αυτό το σκοπό είναι η απλή μέθοδος Naive( ή NF1). Η μέθοδος αυτή, χρησιμοποιεί μόνο την τελευταία διαθέσιμη παρατήρηση της χρονοσειράς για να κάνει την πρόβλεψη. Δεν την ενδιαφέρουν οι παλαιότερες παρατηρήσεις. Δηλαδή  $F_t = Y_{t-1}$ . Οπότε συγκρίνοντας για παράδειγμα τα MAPE που προκύπτουν από τη μέθοδο αυτή και οποιαδήποτε άλλη εφαρμόσουμε μπορούμε να έχουμε ένα καλύτερο μέτρο σύγκρισης ως προς την αποδοτικότητα πρόβλεψης της μεθόδου.

Στην περίπτωση που έχουμε δεδομένα με εποχικότητα μπορούμε να χρησιμοποιήσουμε μία άλλη εξίσου απλή μέθοδο Naive την NF2. Αυτή εφαρμόζεται σε δεδομένα που έχουν αποεποχικοποιηθεί. Οπότε σε πρώτο βήμα αφαιρούμε την εποχικότητα από τη χρονοσειρά και σε δεύτερο βήμα εφαρμόζουμε τη μέθοδο όπως και την NF1. Χρησιμοποιεί και αυτή την τελευταία παρατήρηση για να κάνει την πρόβλεψη και όπως και οι παρατηρήσεις που παίρνει σαν όρισμα, έτσι και οι προβλέψεις δεν περιέχουν εποχικότητα. Ένα πλεονέκτημα που έχει, είναι στο να αποφασίσουμε αν η βελτίωση που ακολουθεί πέρα από μια απλή προσαρμογή της εποχικότητας, αξίζει το χρόνο που θα αφιερώσουμε.

## 3.4 Διαστήματα Πρόβλεψης

Όταν κάνουμε προβλέψεις συνήθίζεται να μην παρέχουμε μόνο την τιμή της πρόβλεψης αλλά και ένα διάστημα εμπιστοσύνης της πρόβλεψης. Η πιθανότητα η τιμή να συμπεριλαμβάνεται στο διάστημα εμπιστοσύνης καθορίζεται από το επίπεδο εμπιστοσύνης (*confidence level*). Αν για παράδειγμα θέλω να υπολογίσω ένα 95% διάστημα εμπιστοσύνης, αυτό σημαίνει ότι αν πάρω για την παράμετρο που θέλω να εκτιμήσω πολλά διαφορετικά τυχαία δείγματα και κατασκευάσω τα διαστήματα

εμπιστοσύνης για το καθένα από αυτά, τότε το ποσοστό αυτών που θα περιέχει την πραγματική τιμή της παραμέτρου θα είναι περίπου ίσο με 95%. Αυτό το ποσοστό αποτελεί και το επίπεδο εμπιστοσύνης. Με αυτό τον τρόπο παρέχεται η εικόνα της «χειρότερης» ή «καλύτερης» εκτίμησης αφήνοντας περιθώρια λάθους. Άλλωστε οι προβλέψεις όπως έχουμε ήδη αναφέρει δεν είναι τέλειες και τα διαστήματα πρόβλεψης εκφράζουν ακριβώς αυτό.

Πριν ξεκινήσουμε την κατασκευή των διαστημάτων πρόβλεψης υποθέτουμε ότι τα σφάλματα πρόβλεψης ακολουθούν την κανονική κατανομή με μέση τιμή μηδέν. Υπό αυτή την υπόθεση μια προσέγγιση των διαστημάτων πρόβλεψης είναι η ακόλουθη:

$$F_{n+1} \pm z * \sqrt{MSE}$$

Όπου  $z$  η τιμή της τυποποιημένης κανονικής κατανομής η οποία καθορίζει το μήκος και την πιθανότητα του διαστήματος πρόβλεψης. Παρακάτω ακολουθεί πίνακας με διάφορες τιμές της ζανάλογα με το επίπεδο εμπιστοσύνης.

Z	Confidence Level
2.576	99%
1.96	95%
1.645	90%
1.282	80%
1.15	75%
1	68%
0.674	50%

Τα παραπάνω διαστήματα πρόβλεψης αναφέρονται σε προβλέψεις του επόμενου βήματος (one-step forecasting). Αν έχουμε προβλέψεις πολλαπλών βημάτων (multi-step forecasting) χρειάζεται τροποποίηση του MSE. Μια προσέγγιση για h-step είναι η εξής:

$$MSE_h = \frac{1}{n-h} \sum_{t=h+1}^n (e_t^{(h)})^2$$

Όπου το  $e_t^{(h)}$  αποτελεί το σφάλμα της πρόβλεψης μιας τιμής για h βήματα μελλοντικά τη χρονική στιγμή  $t$ . Αν πάλι θεωρήσουμε ότι τα σφάλματα αυτά ακολουθούν την κανονική κατανομή με μέση τιμή μηδέν τότε προκύπτει το παρακάτω διάστημα εμπιστοσύνης:

$$F_{n+h} \pm z * \sqrt{MSE_h}$$

### 3.5 Μέθοδος ελαχίστων τετραγώνων

Τα δεδομένα που έχουμε κάθε φορά στη διάθεσή μας αποτελούνται από 2 βασικά κομμάτια. Το ένα αφορά τη λειτουργική σχέση που διέπει το σύστημα που μελετάμε (δηλαδή αν έχουμε να κάνουμε με δεδομένα χρονοσειράς ή άλλα που περιλαμβάνουν

την αλληλεπίδραση πολλών μεταβλητών) ή αλλιώς το υπόδειγμα (*pattern*) που ακολουθούν. Το άλλο αφορά την τυχαιότητα που περιέχουν ή αλλιώς τα σφάλματα. Για να γίνουν ορθότερα οι προβλέψεις, καλούμαστε να βρούμε έναν τρόπο να ξεχωρίσουμε το υπόδειγμα από τα σφάλματα (*error*). Σε αυτό μας βοηθάει η μέθοδος των ελαχίστων τετραγώνων. Με τη μέθοδο αυτή προσπαθούμε να προσαρμόσουμε στα δεδομένα μια συνάρτηση που ελαχιστοποιεί το κομμάτι των σφαλμάτων. Η απομόνωση των σφαλμάτων γίνεται μέσω της παρακάτω εξίσωσης:

$$error = data - pattern$$

ή εναλλακτικά

$$e_i = Y_i - \hat{Y}_i$$

Όπου  $e_i$  είναι η κατακόρυφη απόκλιση της παρατήρησης  $i$  από την εκτιμημένη συνάρτηση (καλείται και *υπόλοιπο* / residual),  $Y_i$  η τιμή της παρατήρησης  $i$  και  $\hat{Y}_i$  το εκτιμημένο υπόδειγμα για την αντίστοιχη παρατήρηση.

Στόχος της μεθόδου ελαχίστων τετραγώνων είναι η ελαχιστοποίηση της παράστασης:

$$SSE = \sum_{i=1}^n (Y_i - \hat{Y}_i)^2$$

Γνωστή και ως άθροισμα τετραγώνων των υπολοίπων.

Στις περιπτώσεις των χρονοσειρών τα δεδομένα αποτελούνται από μια εξαρτημένη μεταβλητή (για παράδειγμα τις μηνιαίες πωλήσεις ενός προϊόντος, την ημερήσια τιμή μιας μετοχής κλπ) και την ανεξάρτητη μεταβλητή που είναι ο χρόνος. Θα υποθέσουμε ότι υπάρχει γραμμική συσχέτιση μεταξύ τους της μορφής:

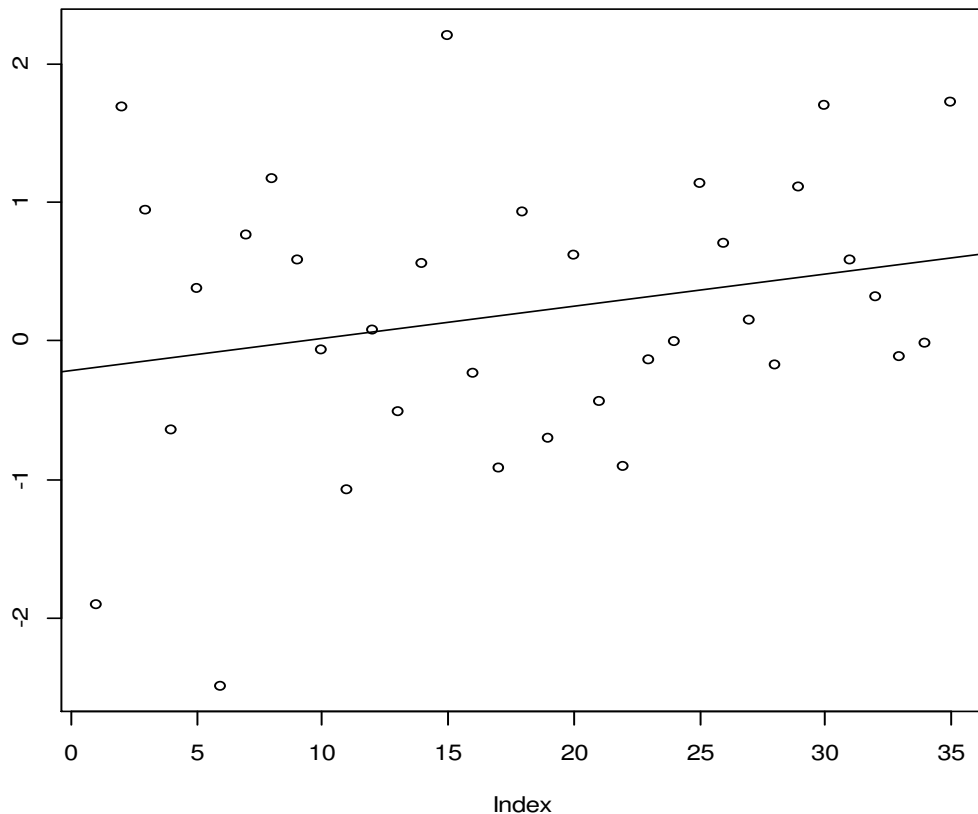
$$Y = \alpha + \beta x + \varepsilon$$

όπου  $\alpha$  η τιμή όταν τέμνει τον άξονα των  $y$ ,  $\beta$  η κλίση της και  $\varepsilon$  το τυχαίο σφάλμα. Με τη μέθοδο ελαχίστων τετραγώνων θέλουμε να υπολογίσουμε τις τιμές των  $\alpha$  και  $\beta$  ώστε να έχουμε την καλύτερη προσαρμογή της ευθείας  $\hat{Y} = \hat{\alpha} + \hat{\beta}x$  στα δεδομένα. Εφαρμόζοντας μερικές παραγώγους στον παραπάνω τύπο ως προς τα  $\alpha$  και  $\beta$  και θέτοντας τις παραγώγους ίσες με μηδέν, προκύπτουν οι εκτιμημένες τιμές των  $\alpha$  και  $\beta$ . Δηλαδή:

$$\hat{\beta} = \frac{\sum_{i=1}^n (x_i - \bar{X})(Y_i - \bar{Y})}{\sum_{i=1}^n (x_i - \bar{X})^2} \quad \text{και} \quad \hat{\alpha} = \bar{Y} - \hat{\beta}\bar{X}$$

Όπου

$$\bar{X} = \frac{\sum_{i=1}^n x_i}{n} \quad \text{και} \quad \bar{Y} = \frac{\sum_{i=1}^n Y_i}{n}$$



Εικόνα 3.5-1: Παράδειγμα εφαρμογής μεθόδου ελαχίστων τετραγώνων

## 3.6 Τροποποιήσεις και προσαρμογές

Η προσαρμογή των ιστορικών δεδομένων αποτελεί μια λύση που οδηγεί μερικές φορές σε πιο απλά και κατανοητά μοντέλα πρόβλεψης. Θα αναφερθούμε σε μαθηματικές εφαρμογές και ημερολογιακές προσαρμογές.

### 3.6.1 Μαθηματικές τροποποιήσεις

Είναι πολύ πιθανό να δουλεύουμε με δεδομένα που παρουσιάζουν εξαιρετικά μεγάλο εύρος τιμών. Να ξεκινούν από πολύ χαμηλές τιμές και όσο περνάει ο χρόνος να αυξάνονται με αποτέλεσμα την αύξηση της διασποράς των τιμών. Σε τέτοιες περιπτώσεις μας είναι πολύ χρήσιμη μια μαθηματική τροποποίηση.

Δύο πολύ χρήσιμες είναι να εφαρμόσουμε στα δεδομένα τη συνάρτηση της ρίζας ή τη λογαριθμική συνάρτηση. Στην πρώτη περίπτωση παίρνουμε τις ρίζες των τιμών των ιστορικών δεδομένων και στη δεύτερη το λογάριθμό τους. Έτσι πετυχαίνουμε μείωση της διασποράς και ειδικά με τη λογαριθμική συνάρτηση έχουμε καλύτερη ερμηνεία των δεδομένων. Όποια και αν είναι όμως τελικά η τροποποίηση, επειδή οι προβλέψεις και τα διαστήματα πρόβλεψης θέλουμε να είναι πάνω στα αρχικά μας δεδομένα χρειάζεται να αντιστρέψουμε τις τιμές στην αρχική κλίμακα.



Πολλές μέθοδοι πρόβλεψης δίνουν μεγάλο βάρος στις πιο πρόσφατες παρατηρήσεις, οπότε οι τροποποιήσεις δεν αλλάζουν και πολύ το αποτέλεσμα της ακρίβειας των προβλέψεων. Βέβαια δεν ισχύει το ίδιο σε δεδομένα που η διασπορά αλλάζει εξαιρετικά γρήγορα.

### 3.6.2 Ημερολογιακές προσαρμογές

Όταν έχουμε να κάνουμε με δεδομένα που οι παρατηρήσεις σχετίζονται με πλήθος ημερών (εβδομαδιαίες, μηνιαίες κ.ά. ), ένα μέρος της διακύμανσης των τιμών της χρονοσειράς οφείλεται στις διακυμάνσεις του πλήθους των ημερών κάθε μήνα, των αργιών και των ημερών συναλλαγών (trading days).

Σε ότι έχει να κάνει με το πλήθος των ημερών κάθε μήνα, η διαφορά από μήνα σε μήνα που είναι περίπου  $\frac{31-28}{30}=10\%$  η μέγιστη, εμφανίζεται στα δεδομένα ως εποχιακό μοτίβο και κάνει δύσκολη την ερμηνεία τους. Η προσαρμογή μπορεί να γίνει πολύ εύκολα με τη βοήθεια του παρακάτω τύπου:

$$W_t = Y_t * \frac{\text{no. of days in an average month}}{\text{no. of days in month } t} = \frac{365.25/12}{\text{no. of days in month } t}$$

Η προσαρμογή αντίστοιχα για τις ημέρες συναλλαγών είναι παρόμοια με την παραπάνω διαδικασία. Αυτή η τροποποίηση είναι σημαντική καθώς οι μέρες συναλλαγών διαφέρουν από μήνα σε μήνα και ασκούν ιδιαίτερη επιρροή σε κλάδους όπως η βιομηχανία και οι τράπεζες. Όπως δείξαμε και παραπάνω έχουμε τον αντίστοιχο τύπο:

$$W_t = Y_t * \frac{\text{no. of trading days in an average month}}{\text{no. of trading days in month } t}$$

Εδώ το  $Y_t$  είναι ήδη προσαρμοσμένο στο πλήθος των ημερών ανά μήνα.

# 4

## Ανάλυση χρονοσειρών

Η διερεύνηση της συμπεριφοράς μιας χρονοσειράς καθώς και η πρόβλεψη των μελλοντικών τιμών της αποτελούν τους βασικούς στόχους της ανάλυσης χρονοσειρών. Οι μέθοδοι που χρησιμοποιούνται για τις προβλέψεις μπορούν να χωριστούν σε 3 μεγάλες κατηγορίες:

- Μέθοδοι Αποσύνθεσης
- Μέθοδοι Εξομάλυνσης
- Ανάλυση ARIMA

Κοινό χαρακτηριστικό όλων είναι ότι οι προβλέψεις βασίζονται στις παλαιότερες παρατηρήσεις και μόνο, χωρίς να γίνεται προσπάθεια εύρεσης παραγόντων που μπορεί να επηρεάζουν το αποτέλεσμα. Έτσι έχουμε μια σχέση της μορφής:

$$Y_{t+1} = f(Y_t, Y_{t-1}, Y_{t-2}, \dots, error)$$

Όπου το  $t$  συμβολίζει την χρονική περίοδο της παρατήρησης στο παρόν, το  $t-1$  μια χρονική περίοδο πριν, το  $t+1$  μια χρονική περίοδο μετά κλπ. Το σφάλμα υπάρχει γιατί δε θα μπορούμε πάντα να υπολογίζουμε με ακρίβεια την πρόβλεψη.

Το γιατί μπορεί να θέλουμε να εστιάσουμε σε κάτι τέτοιο γίνεται για 2 βασικούς λόγους. Ο πρώτος, είναι γιατί ακόμα και αν βρούμε τους παράγοντες που υποθέτουμε ότι επηρεάζουν το αποτέλεσμα, μπορεί να έχουμε μεγάλη δυσκολία στο να βρούμε ποσοτικά την επίδραση αυτή. Ο δεύτερος, είναι γιατί κατά βάση μπορεί να μας ενδιαφέρει να προβλέψουμε τι αναμένεται να συμβεί χωρίς να μας απασχολεί το γιατί συμβαίνει αυτό.

### 4.1 Μέθοδοι Αποσύνθεσης (Time Series Decomposition)

Οι μέθοδοι αποσύνθεσης βρίσκονται ανάμεσα στις παλαιότερες μεθόδους ανάλυσης χρονοσειρών. Εφαρμόζοντας απλές μαθηματικές σχέσεις, προσπαθούν να απομονώσουν τις τέσσερις βασικές συνιστώσες των χρονοσειρών (τάση, κυκλικότητα, εποχικότητα, μη κανονικές διακυμάνσεις) με όση περισσότερη ακρίβεια γίνεται. Αυτό συμβάλει στον προσδιορισμό του βαθμού που επηρεάζει η καθεμία συνιστώσα τις παρατηρήσεις της χρονοσειράς. Βασίζονται στην ιδέα ότι τα δεδομένα έχουν την εξής μορφή:

$$\text{δεδομένα} =$$

$$f(\text{τάση, κυκλικότητα, εποχικότητα, μη κανονικές διακυμάνσεις}) + \text{σφάλμα}$$

Η εναλλακτικά:

$$Y_t = f(T_t, S_t, C_t, R_t, error)$$

όπου:

$Y_t$  = παρατήρηση τη χρονική περίοδο  $t$

$T_t$  = συνιστώσα της τάσης τη χρονική περίοδο  $t$

$S_t$  = συνιστώσα της εποχικότητας τη χρονική περίοδο  $t$

$C_t$  = συνιστώσα της κυκλικότητας τη χρονική περίοδο  $t$

$R_t$  = συνιστώσα τυχαιότητας τη χρονική περίοδο  $t$

Βλέπουμε ότι εκτός από τις τέσσερις βασικές συνιστώσες υπάρχει και το στοιχείο του σφάλματος. Το σφάλμα αυτό είναι η διαφορά ανάμεσα στο συνδυασμό των συνιστωσών και των παρατηρήσεων της χρονοσειράς. Αναμενόμενο, αφού δε μπορούμε ποτέ να προβλέψουμε με ακρίβεια τις μελλοντικές τιμές.

Η ακριβής μορφή της συνάρτησης, που φανερώνει τον τρόπο με τον οποίο οι παρατηρήσεις της χρονοσειράς προσδιορίζονται από τις συνιστώσες της, εξαρτάται από τη μέθοδο αποσύνθεσης που εφαρμόζεται. Ωστόσο μια συνήθης προσέγγιση που γίνεται είναι να υποθέσουμε ότι έχει προσθετική (additive) ή πολλαπλασιαστική (multiplicative) μορφή.

Το προσθετικό μοντέλο έχει τη μορφή:

$$Y_t = T_t + S_t + C_t + R_t$$

δηλαδή κάθε παρατήρηση προκύπτει από το άθροισμα των συνιστωσών.

Εναλλακτικά, το πολλαπλασιαστικό μοντέλο έχει τη μορφή:

$$Y_t = T_t \times S_t \times C_t \times R_t$$

δηλαδή κάθε παρατήρηση προκύπτει από το γινόμενο των συνιστωσών.

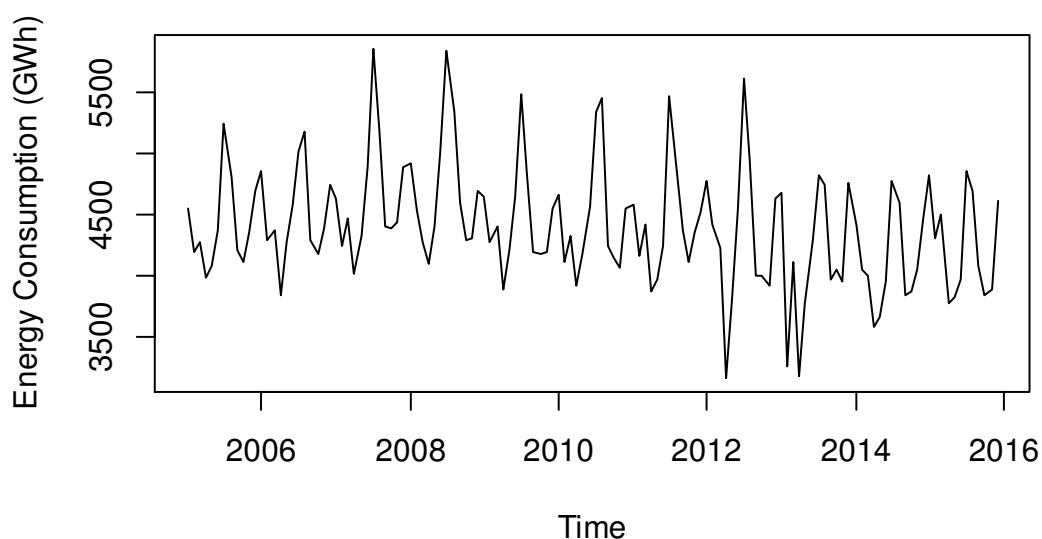
Από τα δύο παραπάνω μοντέλα, αυτό που χρησιμοποιείται συχνότερα είναι το πολλαπλασιαστικό μοντέλο. Είναι καταλληλότερο σε περιπτώσεις που οι διακυμάνσεις εξαρτώνται από το επίπεδο των τιμών της χρονοσειράς και κατά συνέπεια οι συνιστώσες αλληλεπιδρούν. Για παράδειγμα, η αύξηση της συνιστώσας της τάσης να προκαλεί και ανάλογη αύξηση της συνιστώσας της εποχικότητας. Από την άλλη, στην περίπτωση του προσθετικού μοντέλου, οι συνιστώσες είναι ανεξάρτητες μεταξύ τους. Συνεπώς μια αύξηση της συνιστώσας της τάσης δε θα επηρεάσει και αυτή της εποχικότητας. Βέβαια δεν προτιμάται τόσο γιατί συνήθως είναι δύσκολο να αναλυθεί και ισχύει σπάνια σε περιπτώσεις πρόβλεψης οικονομικών και επιχειρησιακών δεδομένων. Παρόλα αυτά εφαρμόζεται καλά σε περιπτώσεις που οι διακυμάνσεις είναι σταθερές.

Οι μέθοδοι αποσύνθεσης προσπαθούν συνήθως να αναγνωρίσουν δύο συνιστώσες που χαρακτηρίζουν τις οικονομικές και επιχειρησιακές σειρές. Η μία είναι η συνιστώσα τάσης-κύκλου (*Trend-Cycle*) και η άλλη η εποχικότητα (*Seasonality*). Η συνιστώσα της εποχικότητας σχετίζεται όπως έχουμε δει με τις περιοδικές

διακυμάνσεις σταθερού μήκους που παρατηρούνται στις τιμές της χρονοσειράς. Η συνιστώσα τάσης-κύκλου αναφέρεται στις μακροχρόνιες αλλαγές του επιπέδου των τιμών μιας χρονοσειράς. Μπορεί να χωριστεί και σε 2 ξεχωριστές, τη συνιστώσα τάσης και τη συνιστώσα κυκλικότητας. Ωστόσο επειδή και ο διαχωρισμός αυτός είναι κάπως τεχνητός, οι περισσότερες μέθοδοι αποσύνθεσης αναγνωρίζουν την τάση και την κυκλικότητα ως μία συνιστώσα.

Οι προβλέψεις εφαρμόζονται συνήθως στην καμπύλη τάσης-κύκλου (TC). Επομένως, για να υπολογίσουμε την καμπύλη τάσης-κύκλου, πρέπει σε πρώτο βήμα να απομονώσουμε την εποχικότητα και την τυχαιότητα από τα δεδομένα. Κάτι τέτοιο απαιτεί την εξομάλυνση ή αποσύνθεση των δεδομένων. Σε επόμενο βήμα και εφόσον έχουν παραχθεί οι προβλέψεις, αυτές εποχικοποιούνται, ώστε να προκύψουν οι τελικές στατιστικές προβλέψεις.

Σε όλη την ενότητα των μεθόδων αποσύνθεσης, χρησιμοποιήθηκε για εφαρμογή των μεθόδων που αναφέρονται, μια χρονοσειρά που περιγράφει τη μηνιαία κατανάλωση ηλεκτρικής ενέργειας στην Ελλάδα το χρονικό διάστημα Ιαν2005-Δεκ2015<sup>6</sup>. Μια εικόνα των δεδομένων φαίνεται στο παρακάτω γράφημα:



Εικόνα 4.1-1: Δεδομένα κατανάλωσης ηλεκτρικής ενέργειας στην Ελλάδα τα έτη 2005-2015

### 4.1.1 Κινητοί μέσοι όροι

Οι κινητοί μέσοι όροι αποτελούν απλές μεθόδους εξομάλυνσης των δεδομένων, με στόχο να απομονωθούν τα πρότυπα από τις τυχαίες διακυμάνσεις και να υπολογιστεί η καμπύλη τάσης-κύκλου. Θα αναφερθούμε σε τέσσερις μεθόδους που ανήκουν σε

<sup>6</sup>Πηγή: [https://www.quandl.com/data/ENTSOE/GR\\_CONS-Electricity-Consumption-Greece](https://www.quandl.com/data/ENTSOE/GR_CONS-Electricity-Consumption-Greece)

αυτή την κατηγορία: τον απλό κινητό μέσο όρο, τον σταθμισμένο κινητό μέσο όρο, τον κεντρικό κινητό μέσο όρο και το διπλό κινητό μέσο όρο.

#### 4.1.1.1 Απλός κινητός μέσος όρος

Η ιδέα πίσω από τη μέθοδο των κινητών μέσων όρων είναι ότι οι τιμές που βρίσκονται χρονικά κοντά είναι πιθανό να έχουν και τιμές που δε διαφέρουν σημαντικά. Οπότε με το να πάρουμε το μέσο όρο των τιμών που βρίσκονται κοντά σε μια παρατήρηση θα έχουμε μια καλή εκτίμηση της σειράς τάσης-κύκλου για την παρατήρηση αυτή, αφού ο μέσος όρος διώχνει σε ένα βαθμό την τυχαιότητα των δεδομένων.

Στην περίπτωση του απλού κινητού μέσου όρου (*simple moving average*) υπολογίζουμε το μέσο όρο  $k$  τιμών της χρονοσειράς γύρω από την παρατήρηση που θέλουμε να γίνει ο υπολογισμός της τάσης-κύκλου ( $k$  MA<sup>7</sup>). Για λόγους συμμετρίας ο αριθμός  $k$  πρέπει να είναι περιττός. Σε κάθε επόμενο υπολογισμό του μέσου όρου αφαιρούμε την παλαιότερη χρονικά παρατήρηση και συμπεριλαμβάνουμε την επόμενη νεότερη χρονικά. Αυτή η διαδικασία συνεχίζεται σε όλες τις παρατηρήσεις που είναι δυνατό να εφαρμοστεί. Το αποτέλεσμα της παραπάνω διαδικασίας είναι η σειρά τάσης-κύκλου. Αν λοιπόν θέλουμε να υπολογίσουμε έναν απλό κινητό μέσο όρο  $k$  περιόδων τότε η μαθηματική σχέση που τον περιγράφει είναι η ακόλουθη:

$$TC_t = \frac{1}{k} * \sum_{i=-m}^m Y_{t+i}$$

Όπου μείναι το πλήθος των παρατηρήσεων που βρίσκονται αριστερά και δεξιά αντίστοιχα από την κεντρική παρατήρηση που γίνεται ο υπολογισμός τάσης-κύκλου και εκφράζεται μαθηματικά από τη σχέση  $m = (k - 1)/2$ . Αντίστοιχα  $TC_t$  η σειρά τάσης-κύκλου τη χρονική στιγμή  $t$  και  $Y_t$  η παρατήρηση της χρονοσειράς τη χρονική στιγμή  $t$ . Οπότε για έναν μέσο όρο 5 περιόδων το  $m$  έχει τιμή 2, για ένα μέσο όρο 3 περιόδων έχει τιμή 1 κλπ.

Αξίζει να σημειώσουμε ότι ο αριθμός των στοιχείων που θα πάρουμε για τον υπολογισμό του μέσου όρου παίζει πολύ σημαντικό ρόλο καθώς επηρεάζει την εξομάλυνση. Μεγάλος αριθμός στοιχείων σημαίνει ότι αυξάνεται η πιθανότητα να εξαλειφθεί η τυχαιότητα, άρα υπάρχει μεγάλη εξομάλυνση. Ωστόσο αν πάρουμε πολύ μεγάλο αριθμό στοιχείων χάνεται αρκετή πληροφορία αφού υπάρχουν κενά στην αρχή και στο τέλος της σειράς τάσης-κύκλου. Επίσης όταν παίρνουμε ένα μέσο όρο  $k$  περιόδων χρειαζόμαστε  $m$  γειτονικές παρατηρήσεις της τρέχουσας παρατήρησης. Οπότε είναι αδύνατος ο υπολογισμός της τάσης-κύκλου στην αρχή και στο τέλος της χρονοσειράς. Οι  $m$  όροι που χάνονται στην αρχή δεν έχουν τόσο μεγάλη σημασία. Δεν ισχύει το ίδιο όμως και για τους  $m$  όρους στο τέλος αφού αποτελούν το αρχικό σημείο της πρόβλεψης της κυκλικότητας. Αυτό το πρόβλημα αντιμετωπίζεται αν πάρουμε ένα μικρότερου μήκους μέσο όρο. Μια προσέγγιση είναι να πάρουμε για τις

---

<sup>7</sup>k MA= k-point moving average

τιμές που έχουν απομείνει μέσο όρο με μικρότερο αριθμό κ. Στις περιπτώσεις που υπάρχει έντονη εποχικότητα στη χρονοσειρά συνίσταται η εφαρμογή ενός μέσου όρου ίσου ή μεγαλύτερου μήκους της εποχιακής περιοδικότητας. Για παράδειγμα αν έχουμε μηνιαίες προβλέψεις τότε  $k=12$ .

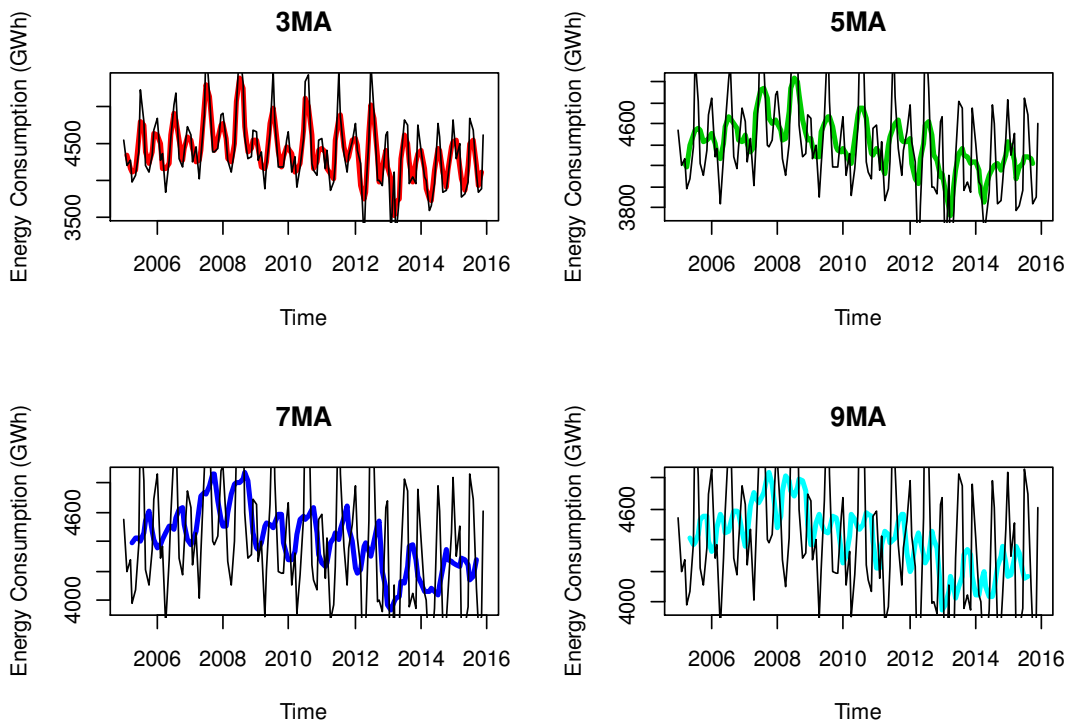
Η εφαρμογή της μεθόδου σε δεδομένα κατανάλωσης ηλεκτρικής ενέργειας αποτυπώνεται στον παρακάτω πίνακα<sup>8</sup>.

TIME	3MA	5MA	7MA	9MA
Jan-05	-	-	-	-
Feb-05	4338	-	-	-
Mar-05	4152	4215	-	-
Apr-05	4111	4181.8	4385.714	-
May-05	4146.667	4391.6	4424	4414.111
Jun-05	4566.333	4499.8	4426.429	4365.667
Jul-05	4812.667	4545.2	4403.143	4382.667
Aug-05	4757.667	4552.2	4455	4429.111
Sep-05	4378.667	4546.4	4543	4525.333
Oct-05	4224.333	4435.2	4610.714	4548.667
Nov-05	4383	4443.2	4473.286	4548
Dec-05	4631	4457.2	4410.429	4391
Jan-06	4609	4510	4356	4331.889
Feb-06	4503.333	4406.6	4380.571	4371.778
Mar-06	4163.333	4324.8	4412.429	4473.111

Πίνακας 4.1-1: Αποτελέσματα εφαρμογής μεθόδων απλού κινητού μέσου όρου 3,5,7 και 9 περιόδων σε δεδομένα κατανάλωσης ηλεκτρικής ενέργειας στην Ελλάδα τα έτη 2005-2015

Επειδή το δείγμα είναι αρκετά μεγάλο αναφέρονται ενδεικτικά οι πρώτες 15 παρατηρήσεις. Μια γραφική αναπαράσταση μπορούμε να δούμε στην εικόνα 4.1-2. Παρατηρείται ότι όσο μεγαλώνουν οι όροι που συμπεριλαμβάνονται στο μέσο όρο τόσο πιο πολύ αυξάνεται η εξομάλυνση.

<sup>8</sup> Εντολές της R για τον υπολογισμό περιέχονται στο αρχείο elec\_ma.r του παραρτήματος



Εικόνα 4.1-2: Γραφική αναπαράσταση εφαρμογής απλών κινητών μέσων 3,5,7 και 9 περιόδων σε δεδομένα κατανάλωσης ηλεκτρικής ενέργειας στην Ελλάδα τα έτη 2005-2015

#### 4.1.1.2 Σταθμισμένος κινητός μέσος όρος

Ο σταθμισμένος κινητός μέσος όρος (*weighted moving average*) μοιάζει αρκετά με τον απλό κινητό μέσο όρο με τη διαφορά ότι οι παρατηρήσεις που συμμετέχουν στο υπολογισμό της σειράς τάσης-κύκλου έχουν άνισα βάρη. Για να λειτουργεί σωστά ο σταθμισμένος κινητός μέσος όρος πρέπει τα βάρη αυτά να έχουν άθροισμα τη μονάδα και να είναι συμμετρικά ( $a_i = a_{-i}$ ). Μεγαλύτερο βάρος δίνεται στις παρατηρήσεις που γειτνιάζουν με την τρέχουσα παρατήρηση και μικρότερο στις πιο μακρινές παρατηρήσεις. Αν υποθέσουμε ότι συμμετέχουν  $k$  παρατηρήσεις, τότε ο σταθμισμένος κινητός μέσος όρος εκφράζεται μαθηματικά από τη σχέση:

$$TC_t = \sum_{i=-m}^m a_i * Y_{t+i}$$

Όπου  $a_i$  είναι τα βάρη για κάθε γειτονική παρατήρηση και  $m = (k - 1)/2$ . Ένα μεγάλο πλεονέκτημα της μεθόδου είναι πως καταλήγει σε μια πιο εξομαλυμένη σειρά τάσης-κύκλου σε σχέση με τον απλό κινητό μέσο όρο. Φυσικά και εδώ οι παρατηρήσεις που συμμετέχουν στο μέσο όρο είναι περιττός αριθμός.

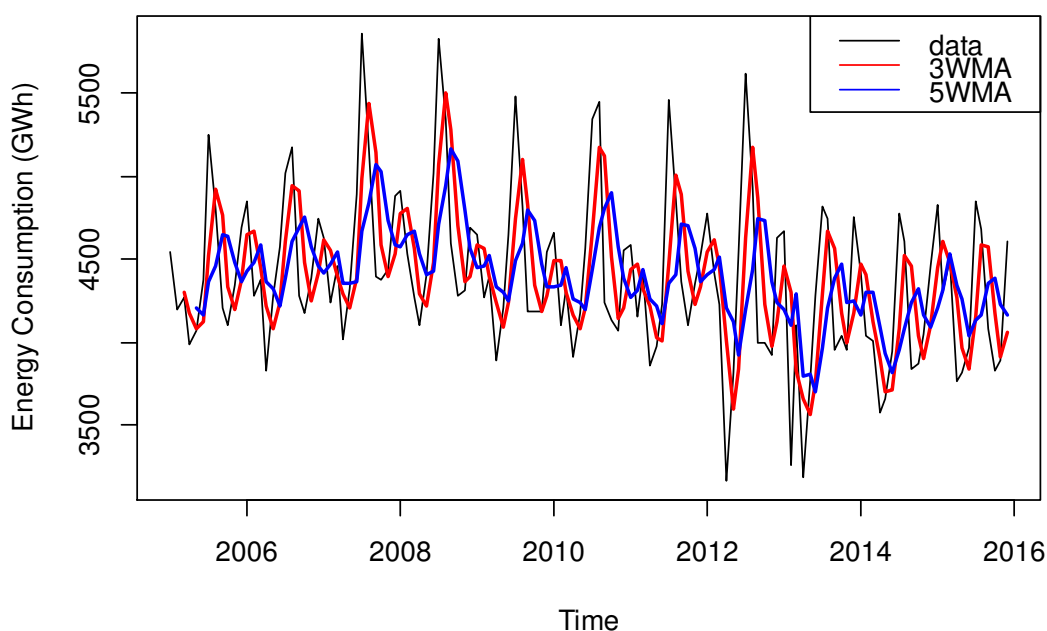
Υπάρχουν κάποια σύνολα από βάρη που χρησιμοποιούνται ευρέως και έχουν πάρει το όνομα των ατόμων που τα δημιούργησαν. Για παράδειγμα ο Spencer(1904) πρότεινε ένα 5x4x4 κινητό μέσο όρο που ακολουθείται από έναν σταθμισμένο κινητό μέσο όρο μήκους 5 με βάρη  $\alpha_0 = 1$ ,  $\alpha_1 = \alpha_{-1} = 3/4$ ,  $\alpha_2 = \alpha_{-2} = -3/4$ . Το

παραπάνω είναι ισοδύναμο από έναν σταθμισμένο κινητό μέσο όρο μήκους 15 με βάρη  $-.0009, -.019, -.016, .009, .066, .144, .209, .231, .209, .144, .066, .099, -.016, -.019, -.009$ .

Στον παρακάτω πίνακα<sup>9</sup> φαίνεται η εφαρμογή ενός σταθμισμένου κινητού μέσου όρου 5 με βάρη  $(1/5, 2/5, 3/5, 2/5, 1/5)$  και ενός σταθμισμένου κινητού μέσου όρου 3 με βάρη  $(1/3, 2/3, 1/3)$ . Η εφαρμογή έγινε στα δεδομένα ηλεκτρικής κατανάλωσης.

TIME	WMA3	WMA5	TIME	WMA3	WMA5
Jan-05	-	-	Aug-05	4957.333	4460.444
Feb-05	4313	-	Sep-05	4413.667	4645.667
Mar-05	4247	-	Oct-05	4144	4638
Apr-05	4082	-	Nov-05	4269.667	4480.333
May-05	4045	4210	Dec-05	4576.667	4366.444
Jun-05	4277.333	4170.444	Jan-06	4798.667	4430.889
Jul-05	4957	4371	Feb-06	4473.667	4477.778

Πίνακας 4.1-2: Σταθμισμένοι κινητοί μέσοι όροι 3 και 5 για δεδομένα κατανάλωσης ηλεκτρικής ενέργειας στην Ελλάδα τα έτη 2005-2015



Εικόνα 4.1-3: Γραφική αναπαράσταση των σταθμισμένων κινητών μέσων 3 και 5 για δεδομένα κατανάλωσης ηλεκτρικής ενέργειας στην Ελλάδα τα έτη 2005-2015

<sup>9</sup> Εντολές της R για τον υπολογισμό περιέχονται στο αρχείο elec\_wma.r του παραρτήματος



### 4.1.1.3 Διπλός κινητός μέσος όρος

Ο διπλός κινητός μέσος όρος (*double moving average*) αποτελεί ένα παράδειγμα του πως ένας κινητός μέσος όρος μπορεί να εξομαλυνθεί από έναν άλλο κινητό μέσο όρο. Οποιοσδήποτε συνδυασμός απλών κινητών μέσων όρων μπορεί να εφαρμοστεί (με ίσα ή άνισα μήκη  $n \times m$ ). Για παράδειγμα αν θέλουμε να υπολογίσουμε έναν διπλό κινητό μέσο όρο μήκους  $3 \times 3$ , πρέπει πρώτα να υπολογίσουμε τον απλό κινητό μέσο όρο μήκους 3 και πάνω σε αυτόν να εφαρμόσουμε έναν απλό κινητό μέσο όρο μήκους 3. Θα δούμε ότι αυτός είναι ισοδύναμος με έναν σταθμισμένο κινητό μέσο όρο μήκους 5.

Ένας απλός κινητός μέσος όρος μήκους 3 δίνεται από τις παρακάτω σχέσεις<sup>10</sup>:

$$TC_{t-1} = \frac{Y_{t-2} + Y_{t-1} + Y_t}{3}$$

$$TC_t = \frac{Y_{t-1} + Y_t + Y_{t+1}}{3}$$

$$TC_{t+1} = \frac{Y_t + Y_{t+1} + Y_{t+2}}{3}$$

Τότε εφαρμόζοντας έναν απλό κινητό μέσο όρο μήκους 3 σε αυτές τις τιμές που υπολογίσαμε έχουμε:

$$\begin{aligned} TC''_t &= \frac{TC_{t-1} + TC_t + TC_{t+1}}{3} \\ &= \left( \frac{Y_{t-2} + Y_{t-1} + Y_t}{3} + \frac{Y_{t-1} + Y_t + Y_{t+1}}{3} + \frac{Y_t + Y_{t+1} + Y_{t+2}}{3} \right) / 3 \\ &= (Y_{t-2} + 2Y_{t-1} + 3Y_t + 2Y_{t+1} + Y_{t+2}) / 9 \end{aligned}$$

Πρόεκυψε λοιπόν ένας σταθμισμένος κινητός μέσος όρος μήκους 5 με βάρη .1111, .2222, .3333, .2222, .1111 για τους όρους  $Y_{t-2}, Y_{t-1}, Y_t, Y_{t+1}, Y_{t+2}$  αντίστοιχα.

### 4.1.1.4 Κεντρικός κινητός μέσος όρος

Ας υποθέσουμε τώρα ότι θέλουμε να πάρουμε έναν κινητό μέσο όρο με άρτιο αριθμό παρατηρήσεων. Σε αυτό μας βοηθάει η μέθοδος κεντρικού κινητού μέσου όρου (*centered moving average*) που αποτελείται από έναν συνδυασμό απλού και διπλού κινητού μέσου όρου.

Έστω ότι έχουμε μια χρονοσειρά με μηνιαίες παρατηρήσεις και θέλουμε να εφαρμόσουμε έναν απλό κινητό μέσο όρο μήκους 4. Ο Ιανουάριος αντιστοιχεί στην περίοδο 1, ο Φεβρουάριος στην περίοδο 2 κλπ. Αντίστοιχα αν περάσει ένα έτος ο Ιανουάριος του επόμενου έτους αντιστοιχεί στην περίοδο 13, ο Φεβρουάριος στην περίοδο 14 κλπ. Αν αναζητούμε τον απλό κινητό μέσο όρο μήκους 4 για την περίοδο 3 τότε έχουμε 2 επιλογές:

<sup>10</sup> Ενδεικτικά υπολογίσαμε 3 τιμές της σειράς τάσης-κύκλου ώστε να εφαρμόσουμε έναν απλό κινητό μέσο όρο μήκους 3

$$TC_{2.5} = \frac{Y_1 + Y_2 + Y_3 + Y_4}{4}$$

ή

$$TC_{3.5} = \frac{Y_2 + Y_3 + Y_4 + Y_5}{4}$$

Βλέπουμε ότι δε μπορούμε να αντιστοιχίσουμε τις υπολογισμένες τιμές σε κάποια συγκεκριμένη παρατήρηση αλλά στο ενδιάμεσο 2 παρατηρήσεων. Η πρώτη έχει κέντρο το 2.5 (δηλαδή μισή περίοδο πριν) και η δεύτερη το 3.5 (δηλαδή μισή περίοδο μετά). Το πρόβλημα αυτό αντιμετωπίζεται με τη χρήση ενός απλού κινητού μέσου όρου 2 περιόδων πάνω στις ήδη υπολογισμένες τιμές του απλού κινητού μέσου όρου 4 περιόδων. Αυτός ο κεντρικός κινητός μέσος όρος συμβολίζεται ως 2x4. Το αποτέλεσμα εφαρμογής του είναι το ακόλουθο:

$$\begin{aligned} TC''_3 &= \frac{TC_{2.5} + TC_{3.5}}{2} \\ &= \frac{1}{2} \left( \frac{Y_1 + Y_2 + Y_3 + Y_4}{4} + \frac{Y_2 + Y_3 + Y_4 + Y_5}{4} \right) \\ &= (Y_1 + 2Y_2 + 2Y_3 + 2Y_4 + Y_5)/8 \end{aligned}$$

Οπότε οι 2 ακριανοί όροι έχουν βάρη  $1/8=0.125$  και οι υπόλοιποι το διπλάσιο βάρος δηλαδή  $1/4=0.25$ . Προκύπτει λοιπόν ότι ένας κεντρικός κινητός μέσος όρος 2x4 είναι ισοδύναμος με έναν σταθμισμένο κινητό μέσο όρο μήκους 5. Γενικά παρατηρείται ότι ένας 2xk κεντρικός κινητός μέσος όρος είναι ισοδύναμος με έναν σταθμισμένο κινητό μέσο όρο μήκους k+1 με βάρη  $1/k$  για όλες τις παρατηρήσεις εκτός από τις 2 ακριανές που έχουν βάρη  $1/2k$ .

Ο κεντρικός κινητός μέσος όρος είναι πολύ χρήσιμος σε χρονοσειρές που παρουσιάζουν εποχιακή συμπεριφορά με άρτιο αριθμό περιόδων ανά έτος (μηνιαίες, τριμηνιαίες, εξαμηνιαίες παρατηρήσεις). Ειδικότερα οι κεντρικοί κινητοί μέσοι όροι 2x4 και 2x12 χρησιμοποιούνται συχνά για την εκτίμηση της σειράς τάσης-κύκλου με εξομάλυνση της εποχικότητας σε μηνιαίες και τριμηνιαίες παρατηρήσεις.

#### 4.1.2 Εξομάλυνση με τοπική γραμμική παλινδρόμηση

Η ιδέα πίσω από αυτή τη μέθοδο εξομάλυνσης, είναι η προσαρμογή μιας ευέλικτης καμπύλης τάσης-κύκλου στα δεδομένα. Στο κεφάλαιο 3 είδαμε τη μέθοδο ελαχίστων τετραγώνων η οποία κάνει προσαρμογή μιας ευθείας γραμμής στα δεδομένα. Η ευθεία αυτή θα έχει τη μορφή:

$$TC_t = \alpha + \beta t$$

Όπου τα  $\alpha$  και  $\beta$  αποτελούν τις παραμέτρους του μοντέλου και οι τιμές τους υπολογίζονται από την ελαχιστοποίηση του αθροίσματος των τετραγώνων των υπολοίπων ,δηλαδή από την παράσταση:

$$\sum_{t=1}^n (Y_t - \alpha - \beta t)^2$$

Η υπόθεση της γραμμικής τάσης στα δεδομένα εξυπηρετεί ένα μεγάλο κομμάτι χρονοσειρών, ωστόσο υπάρχουν πολλές χρονοσειρές όπου μια καμπύλη είναι καταλληλότερη. Με την τοπική γραμμική παλινδρόμηση πετυχαίνουμε να προσαρμόσουμε μια σειρά από ευθείες γραμμές σε υποσύνολα των δεδομένων αντί για μια ευθεία γραμμή σε όλα τα δεδομένα που κάνει η γραμμική παλινδρόμηση.

Η εκτιμημένη σειρά τάσης-κύκλου τη χρονική περίοδο τείναι  $TC_t = \alpha + \beta t$  όπου τα  $\alpha$  και  $\beta$  έχουν επιλεγεί ώστε να ελαχιστοποιούν την παράσταση:

$$\sum_{i=-m}^m a_i (Y_{t+i} - \alpha - b(t+i))^2$$

Σε αυτή την περίπτωση οι τιμές των  $\alpha$  και  $\beta$  είναι διαφορετικές για κάθε χρονική περίοδο  $t$ . Οπότε στην ουσία έχουμε την προσαρμογή διαφορετικής ευθείας γραμμής σε κάθε παρατήρηση.

Τα βήματα υπολογισμού που ακολουθούνται για την εκτίμηση μιας παρατήρησης είναι τα εξής:

#### ΒΗΜΑ 1

Διαλέγουμε τον αριθμό των σημείων που θέλουμε να συμπεριλάβουμε στην σταθμισμένη παλινδρόμηση, δηλαδή αυτά που βρίσκονται κοντά στην παρατήρηση που εστιάζουμε, αριστερά και δεξιά από αυτήν.

Το νούμερο αυτό παίζει πολύ σημαντικό ρόλο στο αποτέλεσμα της προσαρμογής. Γενικά όσο πιο πολλές παρατηρήσεις συμπεριλαμβάνουμε τόσο πιο ομαλή θα είναι και η προσαρμοσμένη ευθεία. Βέβαια ένα πολύ μεγάλο νούμερο μπορεί να αλλοιώσει το αποτέλεσμα.

#### ΒΗΜΑ 2

Σε κάθε παρατήρηση αντιστοιχίζεται ένα βάρος σύμφωνα με κάποια συνάρτηση βάρους. Εδώ μπορούμε να χρησιμοποιήσουμε ενδεικτικά την τεταρτοβάθμια συνάρτηση:

$$Q(i, m) = \begin{cases} (1 - (i/m)^2)^2 & \text{για } -m < i < m \\ 0 & \text{αλλιώς} \end{cases}$$

Τότε κάθε βάρος υπολογίζεται από τη σχέση:

$$a_i = \frac{Q(i, m)}{\sum_{i=-m}^m Q(i, m)}$$

Με αυτό τον τρόπο φροντίζουμε το άθροισμα όλων το βαρών να ισούται με τη μονάδα. Η κεντρική τιμή έχει το μεγαλύτερο βάρος και όσο απομακρυνόμαστε από αυτή μειώνονται και τα βάρη.

#### ΒΗΜΑ 3

Γίνεται η προσαρμογή της ευθείας στα δεδομένα με χρήση των κατάλληλων μεταβλητών  $\alpha$  και  $\beta$  που ελαχιστοποιούν τη σχέση που είδαμε παραπάνω.

Αυτή η διαδικασία ακολουθείται για κάθε παρατήρηση και προκύπτει η καμπύλη τάσης-κύκλου. Στις ακριανές παρατηρήσεις βέβαια χρησιμοποιούνται λιγότερες τιμές για τον υπολογισμό της προσαρμοσμένης τιμής της καμπύλης τάσης-κύκλου. Ένα μεγάλο πλεονέκτημα της μεθόδου αυτής είναι ότι αποφεύγεται η μεροληψία προς το τέλος της χρονοσειράς ειδικά αν εκεί παρατηρείται έντονο το στοιχείο της κυκλικότητας.

#### 4.1.2.1 Loess

Μία πολύ γνωστή μέθοδος τοπικής παλινδρόμησης είναι η Loess. Το μοντέλο γραμμικής παλινδρόμησης που θεωρείται είναι το:

$$y_i = f(x_i) + \varepsilon_i$$

Όπου  $f(\cdot)$  η άγνωστη παράμετρος. Η συνάρτηση παλινδρόμησης εκτιμάται σε ένα σημείο  $x_0$ , με εφαρμογή πολυωνυμικής παλινδρόμησης  $p$ -τάξης της γπάνω στη χτων τοπικά σταθμισμένων ελαχίστων τετραγώνων:

$$y_i = a + \beta_1(x_i - x_0) + \beta_2(x_i - x_0)^2 + \dots + \beta_p(x_i - x_0)^p + \varepsilon_i$$

Με τη συνάρτηση αυτή οι παρατηρήσεις σταθμίζονται σε σχέση με το πόσο κοντά βρίσκονται στο σημείο  $x_0$ . Η εκτίμηση γίνεται φυσικά για καθεμία από τις παρατηρήσεις  $x_i$ , πέραν της  $x_0$ . Μια ευρέως διαδεδομένη συνάρτηση στάθμισης που χρησιμοποιεί αυτή η μέθοδος είναι η τρικυβική συνάρτηση:

$$W(z) = \begin{cases} (1 - |z|^3)^3 & \text{για } |z| < 1 \\ 0 & \text{για } |z| \geq 1 \end{cases}$$

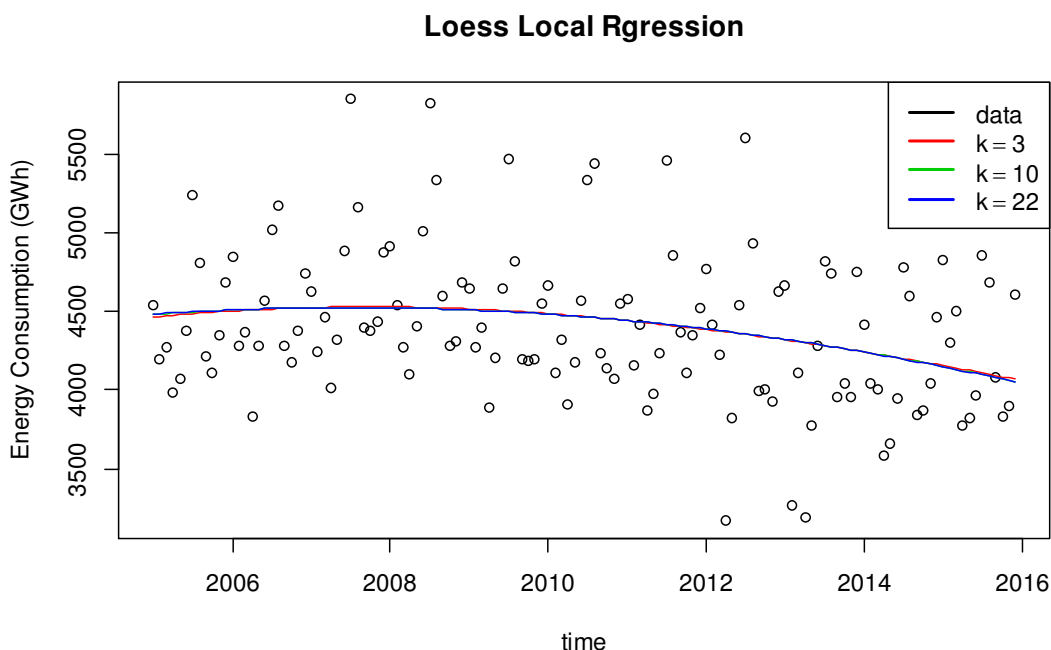
Το βάρος για ένα συγκεκριμένο σημείο γίνεται υπολογίζοντας την τιμή της συνάρτησης στάθμισης πάνω στην απόσταση του σημείου από το σημείο εκτίμησης, αφού πρώτα έχουν κλιμακοποιηθεί οι αποστάσεις χρησιμοποιώντας τη μεγαλύτερη απόσταση, ώστε η μέγιστη απόλυτη τιμή της μεγαλύτερης απόστασης να είναι μονάδα.

Βασικό πλεονέκτημα που παρουσιάζει η μέθοδος αυτή είναι ότι δεν προϋποθέτει τη γνώση των παραμέτρων του πολυωνύμου με βάση το οποίο γίνεται η παλινδρόμηση. Το μόνο που χρειάζεται, είναι να δώσουμε εμείς την τιμή της παραμέτρου εξομάλυνσης  $k$  και την τάξη του πολυωνύμου. Βέβαια απαιτεί μεγάλο αριθμό δεδομένων για τον υπολογισμό όσο το δυνατόν καλύτερης εκτιμημένης καμπύλης και παρουσιάζει ευαισθησία στις ακραίες παρατηρήσεις όπως και κάθε άλλη μέθοδος ελαχίστων τετραγώνων.

Υπάρχει όμως μια παραλλαγή της που παρουσιάζει πλεονέκτημα σε ότι αφορά τις ακραίες τιμές παρέχοντας ένα μοντέλο πιο ισχυρό. Για κάθε εκτίμηση υπολογίζει τη συνιστώσα τυχαιότητας μέσω των διαφορών:  $\hat{E}_t = Y_t - \widehat{TC}_t$  όπου  $Y_t$  η παρατήρηση

τη χρονική στιγμή και  $\widehat{TC}_t$  η εκτιμημένη τιμή της καμπύλης τάσης-κύκλου. Μετά εφαρμόζεται πάλι η τοπική παλινδρόμηση όπου τα βάρη αυτή τη φορά είναι τέτοια ώστε οι παρατηρήσεις με μεγάλο σφάλμα να έχουν μικρά βάρη και αντίστροφα. Υπολογίζεται πάλι για κάθε παρατήρηση η συνιστώσα της τυχαιότητας και η διαδικασία συνεχίζεται ώσπου να σταθεροποιηθεί η εκτιμημένη καμπύλη και να μη χρειάζεται περαιτέρω επανάληψη.

Εφαρμογή της τοπικής παλινδρόμησης Loess στα δεδομένα κατανάλωσης ενέργειας<sup>11</sup>:



Εικόνα 4.1-4: Τοπική παλινδρόμηση Loess σε δεδομένα κατανάλωσης ενέργειας της Ελλάδας το χρονικό διάστημα 2005-2015.

### 4.1.3 Κλασική μέθοδος αποσύνθεσης

Η κλασική μέθοδος αποσύνθεσης (*classical decomposition*) είναι μια απλή μαθηματικά διαδικασία διάσπασης μιας χρονοσειράς στις 4 συνιστώσες της. Θα δούμε αναλυτικά τη μέθοδο με την υπόθεση του πολλαπλασιαστικού μοντέλου η οποία συχνά καλείται μέθοδος του λόγου της πραγματικής τιμής προς τον κινητό μέσο ή στην αγγλική ορολογία “*ratio-to-moving average*” και στο τέλος θα περιγράψουμε τη μέθοδο με τη χρήση του προσθετικού μοντέλου εφαρμόζοντάς τη στα δεδομένα της ηλεκτρικής κατανάλωσης.

#### ΒΗΜΑ 1

Εφαρμόζουμε έναν κινητό μέσο όρο που να στηρίζεται στην εποχικότητα των δεδομένων (για παράδειγμα αν έχουμε μηνιαίες προβλέψεις συνίσταται ένας κινητός

<sup>11</sup> Ο κώδικας της R περιέχεται στο αρχείο elec\_Loess.r του παραρτήματος

μέσος όρος μήκους 12). Με αυτό τον τρόπο εξαλείφεται η εποχικότητα και μειώνεται σημαντικά η τυχαιότητα των δεδομένων. Έτσι προκύπτει μια καλή εκτίμηση της συμπεριφοράς της χρονοσειράς σε ότι αφορά την τάση και την κυκλικότητα ( $TC_t = T_t \times C_t$ ). Για δεδομένα με εποχιακή συμπεριφορά και άρτιο αριθμό περιόδων ανά έτος χρησιμοποιείται ο κεντρικός κινητός μέσος όρος.

#### ΒΗΜΑ 2

Διαιρούμε τις παρατηρήσεις της χρονοσειράς,  $Y_t$ , με τις αντίστοιχες τιμές του κινητού μέσου όρου που υπολογίσαμε στο προηγούμενο βήμα και προκύπτει ο λόγος εποχικότητας:

$$\frac{Y_t}{TC_t} = \frac{S_t \times T_t \times C_t \times R_t}{TC_t} = S_t \times R_t$$

Αυτός ο λόγος απομονώνει τη συνιστώσα της τυχαιότητας και της εποχικότητας.

#### ΒΗΜΑ 3

Επόμενο βήμα είναι η απαλοιφή της τυχαιότητας από τους λόγους εποχικότητας του προηγούμενου βήματος. Για να το πετύχουμε αυτό βρίσκουμε τη μέση τιμή των λόγων εποχικότητας που αναφέρονται στις αντίστοιχες χρονικές περιόδους. Για παράδειγμα, στην περίπτωση μηνιαίων δεδομένων, υπολογίζουμε το μέσο όρο όλων των λόγων εποχικότητας που αντιστοιχούν στο μήνα Ιανουάριο, αντίστοιχα στο μήνα Φεβρουάριο κλπ. Οι μέσοι όροι αυτοί αποτελούν τους δείκτες εποχικότητας της αρχικής χρονοσειράς. Το άθροισμα των δεικτών εποχικότητας πρέπει να είναι ίσο με το άθροισμα των εποχικών περιόδων ανά έτος. Για να διορθωθεί αυτό στην περίπτωση που δεν ισχύει, θα πρέπει να γίνει κανονικοποίηση ώστε οι τιμές για την ίδια χρονική περίοδο να είναι ίσες για όλα τα έτη που μελετάμε τη χρονοσειρά (για μηνιαίες παρατηρήσεις θα πρέπει το άθροισμα ανά 12μηνο να κάνει 12). Οπότε κάθε δείκτης εποχικότητας πολλαπλασιάζεται με το πηλίκο :

$$Multiplier = \text{αριθμός εποχιακών περιόδων} / \text{άθροισμα δεικτών εποχικότητας}$$

Εναλλακτικά χρησιμοποιούμε διαμέσους. Αυτό γίνεται σε περιπτώσεις μεγάλης τυχαιότητας και ασυνήθιστων τιμών. Έτσι για κάθε χρονική περίοδο και για όλες τις διαθέσιμες τιμές των δεικτών εποχικότητας όλων των ετών βρίσκουμε τη διάμεσο. Μετά προσθέτουμε τις διαμέσους των δεικτών εποχικότητας και βρίσκουμε τον ακόλουθο πολλαπλασιαστή:

$$Multiplier = \text{αριθμός εποχιακών περιόδων}(k) / \sum_{i=1}^k Median.$$

Τέλος πολλαπλασιάζουμε κάθε δείκτη εποχικότητας με τον πολλαπλασιαστή και παίρνουμε τις κανονικοποιημένες τιμές των δεικτών.

#### ΒΗΜΑ 4

Διαιρούμε κάθε τιμή της χρονοσειράς,  $Y_t$ , με τον αντίστοιχο δείκτη εποχικότητας για να απαλείψουμε την εποχικότητα ως εξής :

$$SAY_t = \frac{Y_t}{S_t} = \frac{T_t \times C_t \times S_t \times R_t}{S_t} = T_t \times C_t \times R_t$$

Όπου  $SAY_t$  συμβολίζουμε τις τιμές της χρονοσειράς απαλλαγμένες από την εποχικότητα (*seasonally adjusted data*) για την περίοδο  $t$ . Οι νέες αυτές τιμές περιέχουν την τάση, την κυκλικότητα και την τυχαιότητα.

#### ΒΗΜΑ 5

Αφού απαλείψαμε την εποχικότητα τώρα θα γίνει και απαλοιφή της τυχαιότητας. Αυτό επιτυγχάνεται με την εφαρμογή στα αποεποχικοποιημένα πλέον δεδομένα ενός κινητού μέσου όρου μήκους 3 ή 6 παρατηρήσεων. Το αποτέλεσμα είναι μια σειρά τάσης-κύκλου. Για καλύτερη εξομάλυνση συνιστάται η χρήση διπλού κινητού μέσου όρου  $3 \times 3$ . Ο υπολογισμός της συνιστώσας της τυχαιότητας γίνεται με απλή διαίρεση των αποεποχικοποιημένων δεδομένων  $SAY_t$  με τη σειρά τάσης-κύκλου που προέκυψε :

$$\frac{SAY_t}{\text{κινητό μέσο όρο } 3 \times 3} = \frac{T_t \times C_t \times R_t}{T_t \times C_t} = R_t$$

#### ΒΗΜΑ 6

Σειρά έχει ο υπολογισμός της συνιστώσας της τάσης από τη σειρά τάσης-κύκλου που βρήκαμε στο προηγούμενο βήμα. Αν βέβαια κάτι τέτοιο δεν είναι επιθυμητό μπορούμε και να παραβλέψουμε το βήμα αυτό. Για να γίνει σωστά ο διαχωρισμός πρέπει να επιλεγεί το καταλληλότερο μοντέλο που περιγράφει τη χρονοσειρά. Αυτό μπορεί να είναι γραμμικό, παραβολικό, εκθετικό κλπ. Εμείς θα εστιάσουμε στο γραμμικό. Υποθέτουμε δηλαδή ότι η χρονοσειρά εκφράζεται από ένα γραμμικό μοντέλο. Τότε ο υπολογισμός της συνιστώσας της τάσης γίνεται με τη μέθοδο ελαχίστων τετραγώνων<sup>12</sup>. Η εκτιμηθείσα γραμμική εξίσωση ορίζεται ως:

$$\hat{T}_t = a + bt$$

Όπου:

$$\alpha = \overline{TC} - b \cdot \bar{t} \text{ και } \beta = \frac{\frac{\sum_{i=1}^n t_i \cdot TC_i}{n} - \bar{t} \cdot \overline{TC}}{\frac{\sum_{i=1}^n t_i^2}{n} - \bar{t}^2}$$

$$\overline{TC} = \frac{\sum_{i=1}^n TC_i}{n} \quad \bar{t} = \frac{\sum_{i=1}^n t_i}{n}$$

Το  $\hat{T}_t$  δηλώνει την εκτιμημένη τιμή της σειράς τάσης τη χρονική στιγμή  $t$ . Το  $a$  είναι η τιμή της τάσης όταν  $t=0$ . Το  $\beta$  είναι η κλίση της ευθείας και εκφράζει τη μεταβολή της τιμής της τάσης όταν το  $t$  μεταβάλλεται κατά 1 μονάδα. Αρνητικές τιμές του  $\beta$  δηλώνουν πτωτική τάση και θετικές τιμές του  $\beta$  ανοδική τάση. Το  $TC$  δηλώνει τις τιμές της σειράς τάσης κύκλου και αποτελεί την εξαρτημένη μεταβλητή της γραμμικής παλινδρόμησης, ενώ το  $t$  δηλώνει την χρονική περίοδο και αποτελεί την ανεξάρτητη μεταβλητή.

Η συνιστώσα του κύκλου εφόσον έχει υπολογιστεί η συνιστώσα της τάσης προκύπτει από τον λόγο:

<sup>12</sup> Κάνουμε δηλαδή εφαρμογή της απλής γραμμικής παλινδρόμησης

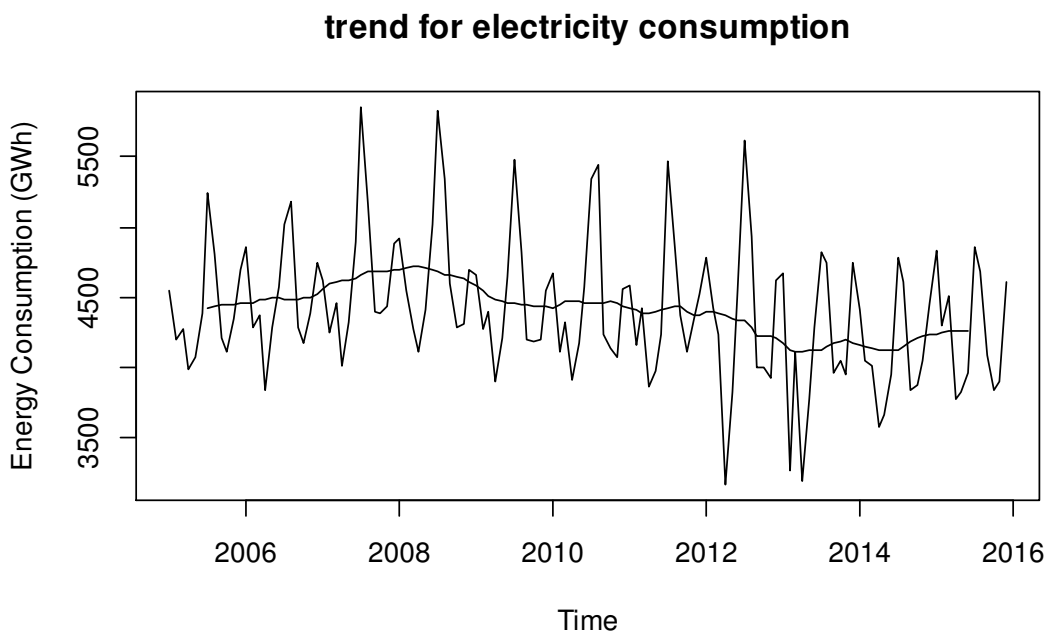
$$C_t = \frac{T_t \times C_t}{T_t}$$

### ΜΕΘΟΔΟΣ ΚΛΑΣΣΙΚΗΣ ΑΠΟΣΥΝΘΕΣΗΣ ΜΕ ΧΡΗΣΗ ΤΟΥ ΠΡΟΣΘΕΤΙΚΟΥ ΜΟΝΤΕΛΟΥ<sup>13</sup>

Υποθέτουμε σε αυτό το σημείο ότι η χρονοσειρά έχει προσθετική μορφή ως προς τις συνιστώσες της. Θα εφαρμόσουμε τη μέθοδο αποσύνθεσης στα δεδομένα τιμών κατανάλωσης ηλεκτρικής ενέργειας. Ο κώδικας της μεθόδου στην R περιέχεται στο παράρτημα στο αρχείο elec\_classical.r.

#### ΒΗΜΑ 1

Για να υπολογίσουμε τη σειρά τάσης κύκλου, εφαρμόζουμε έναν κεντρικό κινητό μέσο όρο μήκους 12. Αυτό το κάνουμε για να απαλείψουμε την εποχικότητα από τα δεδομένα επειδή είναι εποχικά.



Εικόνα 4.1-5: Σειρά τάσης κύκλου σε δεδομένα κατανάλωσης ηλεκτρικής ενέργειας

#### ΒΗΜΑ 2

Αφαιρούμε από τις παρατηρήσεις τη σειρά τάσης κύκλου που προέκυψε από το προηγούμενο βήμα ώστε να μείνουν η εποχικότητα και οι τυχαίες διακυμάνσεις:

$$Y_t - T_t = S_t + E_t$$

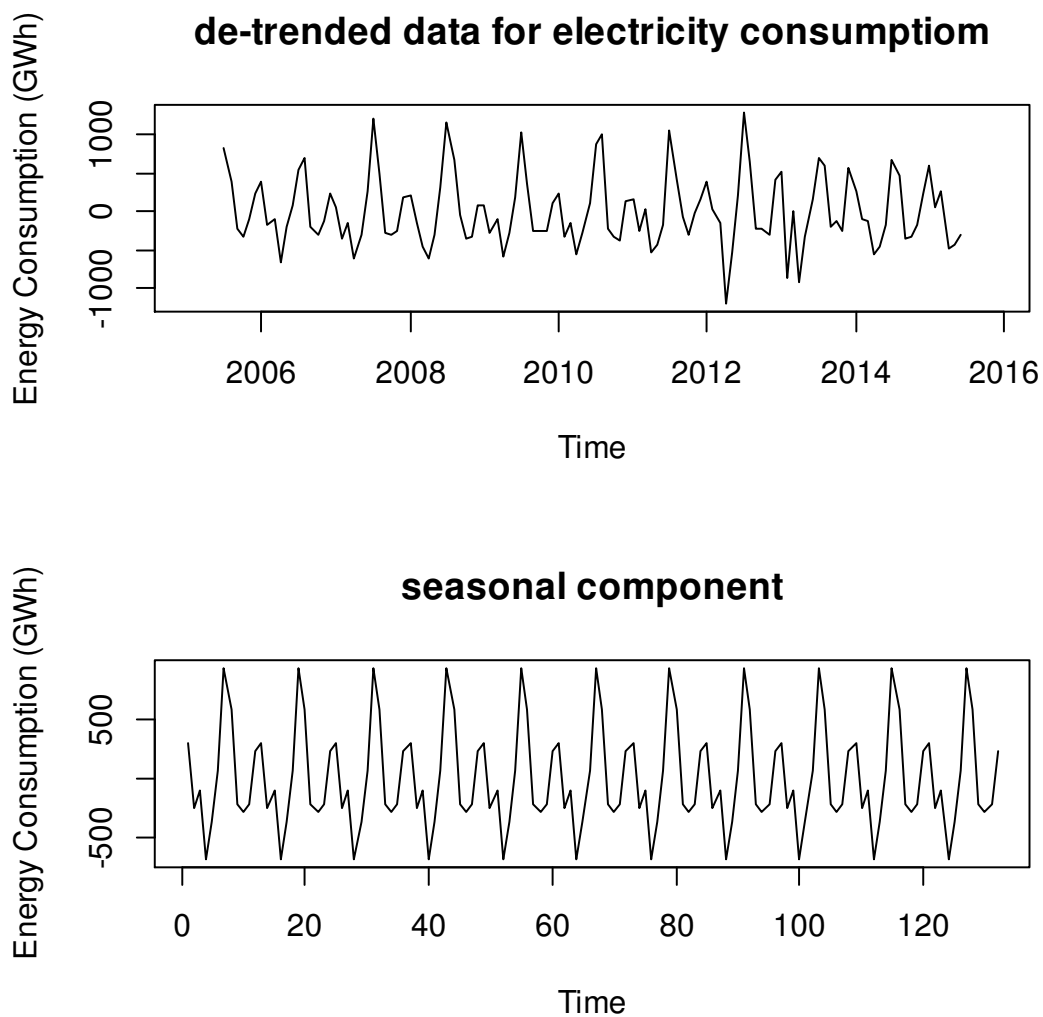
#### ΒΗΜΑ 3

Υπολογίζουμε τη συνιστώσα της εποχικότητας. Η μέθοδος της κλασικής αποσύνθεσης θεωρεί ότι η εποχικότητα έχει σταθερή επίδραση κάθε χρόνο. Έτσι η συνιστώσα της εποχικότητας αποτελείται από 12 τιμές γνωστές ως δείκτες εποχικότητας και αντιστοιχούν ένας για κάθε μήνα του έτους. Οπότε ας πάρουμε σαν παράδειγμα το δείκτη εποχικότητας του Ιανουαρίου (για τον υπολογισμό θα

<sup>13</sup> Η επεξεργασία των δεδομένων με τη χρήση της Réγιναν με τη βοήθεια της πηγής: <https://anomaly.io/seasonal-trend-decomposition-in-r/>



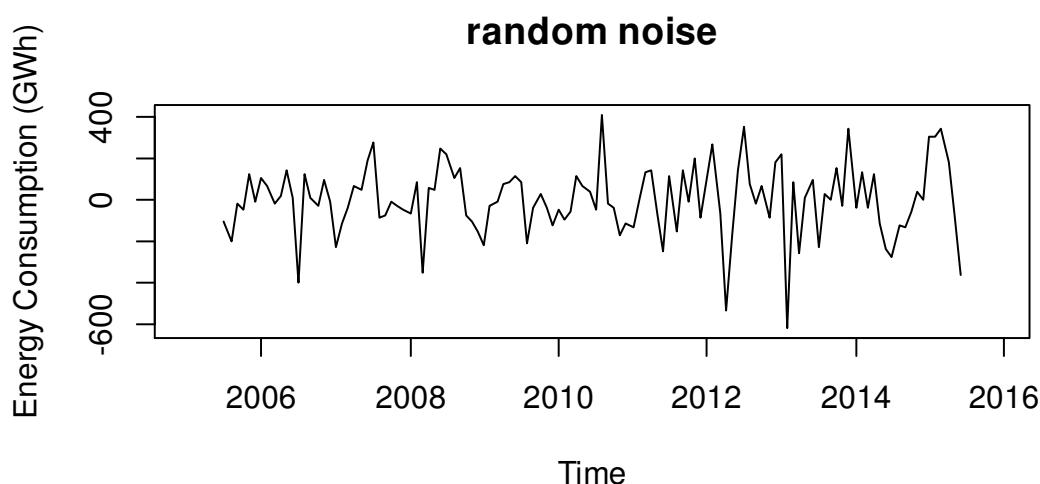
δουλέψουμε με τις τιμές που υπολογίσαμε στο προηγούμενο βήμα και περιέχουν μόνο την εποχικότητα και τις τυχαίες διακυμάνσεις). Ο δείκτης του Ιανουαρίου υπολογίζεται ως η μέση τιμή όλων των τιμών που αντιστοιχούν στο μήνα Ιανουάριο. Αντίστοιχη διαδικασία γίνεται για τους υπόλοιπους μήνες. Στην εικόνα 4.1-6 μπορούμε να δούμε τη συνιστώσα της εποχικότητας καθώς και τα δεδομένα που εξεργαστήκαμε για να την υπολογίσουμε.



Εικόνα 4.1-6: Γραφήματα που δείχνουν τη συνιστώσα της εποχικότητας (κάτω) και τα δεδομένα με αφαίρεση της σειράς τάσης κύκλου (πάνω) για τα δεδομένα κατανάλωσης ηλεκτρικής ενέργειας

#### ΒΗΜΑ 4

Τέλος, οι τυχαίες διακυμάνσεις υπολογίζονται αν αφαιρέσουμε από τις παρατηρήσεις της χρονοσειράς όλες τις συνιστώσες που υπολογίσαμε παραπάνω.



Εικόνα 4.1-7: Τυχαίες διακυμάνσεις για τα δεδομένα της κατανάλωσης ηλεκτρικής ενέργειας

#### 4.1.4 Διαμόρφωση προβλέψεων

Η αποσύνθεση αποτελεί μία πολύ καλή μέθοδο στο να αναλύσουμε μια χρονοσειρά και να κατανοήσουμε τη δομή της, ωστόσο δε βοηθάει ιδιαίτερα στο κομμάτι της πρόβλεψης. Είναι ένα εργαλείο που μας βοηθάει να δούμε τις συνιστώσες μιας χρονοσειράς και να καταλάβουμε καλύτερα τη δομή της και να αποφασίσουμε μια κατάλληλη μέθοδο για την πρόβλεψη. Η συνιστώσα της τάσης-κύκλου έχει μεγάλη δυσκολία να προβλεφθεί. Και αυτό οφείλεται κυρίως στο γεγονός ότι πολλές φορές παρόλο που υποθέτουμε ότι το μοντέλο που την περιγράφει είναι γραμμικό ή οποιοδήποτε άλλο παραμετρικό, αυτό δεν επαρκεί για να περιγράψει με ακρίβεια τη συμπεριφορά της. Η συνιστώσα της εποχικότητας από την άλλη, υπολογίζεται ευκολότερα. Φυσικά οι τυχαίες διακυμάνσεις είναι αδύνατον να προβλεφθούν. Παρόλα αυτά μπορούμε να κάνουμε προβλέψεις με τη μέθοδο αποσύνθεσης με τον εξής τρόπο:

Στην περίπτωση ενός πολλαπλασιαστικού μοντέλου, οι προβλεπόμενες τιμές της χρονοσειράς σε  $m$  περιόδους μελλοντικά περιγράφονται από την παρακάτω σχέση :

$$F_{t+m} = T_{t+m} \times S_{t+m} \times C_{t+m}$$

Η τιμή της τάσης  $T_{t+m}$  για τη μελλοντική περίοδο  $m$  προκύπτει από τη σχέση:

$$T_{t+m} = \alpha + \beta(t + m)$$

Για τιμή του δείκτη εποχικότητας  $S_{t+m}$  χρησιμοποιούμε την τιμή του προσαρμοσμένου (κανονικοποιημένου) δείκτη εποχικότητας που αντιστοιχεί στη περίοδο του έτους που αναφέρεται η ημελλοντική περίοδος. Η τιμή της κυκλικότητας  $C_{t+m}$  από την άλλη, στις περιπτώσεις που οι κυκλικές διακυμάνσεις είναι μικρές (κοντά στη μονάδα) δεν περιλαμβάνεται στην πρόβλεψη. Σε κάθε άλλη περίπτωση δίνεται από τη σχέση  $\frac{TC_{t+m}}{T_{t+m}}$ .

συμπεριλαμβάνονται στην πρόβλεψη (η τιμή της συνιστώσας είναι 1).

Στην περίπτωση ενός πολλαπλασιαστικού μοντέλου, οι προβλεπόμενες τιμές της χρονοσειράς σε  $m$  περιόδους μελλοντικά περιγράφονται από την παρακάτω σχέση:

$$F_{t+m} = T_{t+m} + S_{t+m}$$

Οι ποσότητες  $T_{t+m}$  προκύπτουν με εφαρμογή κάποιας μεθόδου πρόβλεψης που δεν λαμβάνει υπόψη την εποχικότητα. Τέτοια είναι η μέθοδος απλής εκθετικής εξομάλυνσης που θα δούμε παρακάτω. Η τιμή του  $S_{t+m}$  από την άλλη, είναι ο δείκτης εποχικότητας που αντιστοιχεί στην περίοδο  $m$

Στο κεφάλαιο 5 θα παρουσιάσουμε τις προβλέψεις διαφόρων δεδομένων που θα μελετήσουμε. Η μέθοδος που θα χρησιμοποιήσουμε όμως είναι η μέθοδος αποσύνθεσης STL (Seasonal Trend Decomposition using Loess)<sup>14</sup>. Είναι ένας αλγόριθμος ειδικά σχεδιασμένος ώστε να διασπά τη χρονοσειρά στις 3 συνιστώσες της τάσης, της εποχικότητας και των τυχαίων διακυμάνσεων. Ο λόγος που επιλέχθηκε έναντι της μεθόδου κλασικής αποσύνθεσης είναι ότι είναι πιο αποτελεσματική ειδικά στο κομμάτι των προβλέψεων. Η μέθοδος κλασικής αποσύνθεσης υστερεί σε 2 σημεία:

- Υποθέτει ότι η συνιστώσα της εποχικότητας παραμένει αμετάβλητη με την πάροδο του χρόνου
- Η εκτίμηση της τάσης χάνει παρατηρήσεις στην αρχή και στο τέλος. Αυτό συμβαίνει επειδή για την εκτίμηση χρησιμοποιούμε κινητούς μέσους όρους.

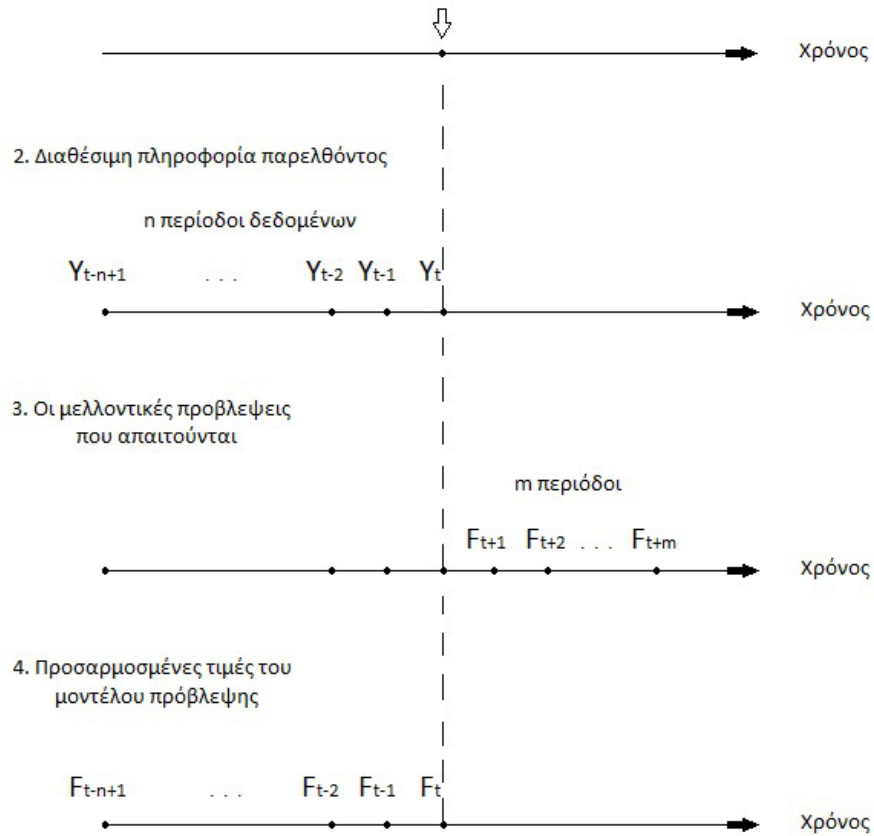
## 4.2 Μέθοδοι Εξομάλυνσης ( Exponential Smoothing Methods)

Πριν αναφερθούμε αναλυτικά στις μεθόδους εξομάλυνσης να δούμε πως είναι ένα σενάριο πρόβλεψης και πως γίνεται η αξιολόγηση μιας μεθόδου. Όπως φαίνεται και στην εικόνα 4.2-1, ξεκινάμε πάντα από ένα συγκεκριμένο σημείο στο χρόνο που είναι το σημείο αναφοράς μας. Μόλις επιλέξουμε το μοντέλο που θα εφαρμόσουμε, γίνεται η προσαρμογή του μοντέλου στα δεδομένα με την κατάλληλη επιλογή παραμέτρων και προκύπτουν οι τιμές προσαρμογής. Μετά εκτιμούμε τα σφάλματα προσαρμογής και εξετάζουμε την καταλληλότητα του μοντέλου χρησιμοποιώντας τα μέτρα που έχουμε περιγράψει. Τέλος, όταν ξεκινήσουμε να λαμβάνουμε μελλοντικές τιμές τις χρονοσειράς μπορούμε να εξετάσουμε και τα σφάλματα πρόβλεψης.

---

<sup>14</sup> The Seasonal Trend The methodology was presented by Robert Cleveland, William Cleveland, Jean McRae and Irma Terpenning in the Journal of Official Statistics in 1990. The STL is available within R via the `stl` function.

1. Σημείο Αναφοράς



5. Σφάλματα προσαρμογής

$$(Y_{t-n+1} - F_{t-n+1}), \dots, (Y_{t-1} - F_{t-1}), (Y_t - F_t)$$

6. Σφάλματα πρόβλεψης

(όταν γίνουν διαθέσιμες οι τιμές  $Y_{t+1}, Y_{t+2}$ , κλπ)

$$(Y_{t+1} - F_{t+1}), (Y_{t+2} - F_{t+2}), \dots$$

Εικόνα 4.2-1: Σενάριο πρόβλεψης

Όποια μέθοδος πρόβλεψης και αν εφαρμοστεί το πρώτο βήμα που πρέπει να γίνει είναι η διάσπαση της χρονοσειράς σε 2 τμήματα. Το ένα θα αποτελεί το αρχικό κομμάτι της μεθόδου και το άλλο το κομμάτι που θα γίνεται η δοκιμή της μεθόδου. Μόλις διαλέξουμε τη μέθοδο εξομάλυνσης, χρησιμοποιούνται τα δεδομένα από το αρχικό κομμάτι, ώστε να ξεκινήσει η μέθοδος και να υπολογιστούν όπου χρειάζεται οι συνιστώσες της τάσης, της εποχικότητας και οποιεσδήποτε παράμετροι. Μετά

παράγονται οι προβλέψεις και ελέγχονται με το κομμάτι δοκιμής που αναφέραμε για να δούμε την απόδοση της μεθόδου. Στο στάδιο αυτό γίνονται και όποιες αλλαγές παραμέτρων χρειάζονται. Στο τελικό στάδιο γίνεται η αξιολόγηση της μεθόδου ως προς την καταλληλότητά της για τα διάφορα είδη προτύπων των δεδομένων. Για τα πρότυπα αυτά θα αναφερθούμε αναλυτικά στη συνέχεια.

Οι μέθοδοι εξομάλυνσης αποτελούν τεχνικές εύκολα εφαρμόσιμες με την ικανότητα να παρέχουν σχετικά γρήγορα προβλέψεις και έχουν εφαρμογή αποκλειστικά σε δεδομένα χρονοσειρών. Είναι πολύ χρήσιμες σε περιπτώσεις που τα δεδομένα έχουν ελάχιστη πληροφορία (για παράδειγμα μόνο 4 τιμές) και σε περιπτώσεις που υπάρχει τεράστια πληροφορία που παράγεται σε μεγάλη συχνότητα όπου υπάρχει η ανάγκη μιας σχετικά αυτοματοποιημένης προσέγγισης των προβλέψεων. Οπότε μπορούμε να δούμε ότι το μέγεθος της πληροφορίας που διαθέτουμε δεν αποτελεί ανασταλτικό παράγοντα για την εφαρμογή τους. Με τη χρήση των μεθόδων εξομάλυνσης δεν αναζητούμε την καλύτερη προσαρμογή ενός μοντέλου στα δεδομένα αλλά στην ουσία υπάρχει ένα προδιαγεγραμμένο μοντέλο που «αναγκάζουμε» τη μέθοδο να χρησιμοποιήσει. Αυτό που πετυχαίνουμε, είναι την απομόνωση σε ένα βαθμό των πραγματικών προτύπων από τις τυχαίες διακυμάνσεις και άρα καλύτερες προβλέψεις.

Στην ενότητα αυτή θα αναφερθούμε σε δύο κατηγορίες μεθόδων εξομάλυνσης. Η πρώτη αφορά μεθόδους που χρησιμοποιούν μέσους όρους σταθερού αριθμού δεδομένων με ίσα βάρη γνωστές και ως «μέθοδοι κινητού μέσου όρου». Η δεύτερη κατηγορία αφορά μεθόδους που αποδίδουν άνισα βάρη στα δεδομένα, βάρη τα οποία τείνουν να μειώνονται με εκθετικό τρόπο από την πιο κοντινή στην πιο μακρινή παρατήρηση. Για το λόγο αυτό καλούνται μέθοδοι εκθετικής εξομάλυνσης.

## 4.2.1 Μέθοδοι κινητού μέσου όρου

Οι μέθοδοι που θα αναλύσουμε σε αυτή την ενότητα κάνουν χρήση του μέσου όρου των πιο πρόσφατων παρατηρήσεων για την πρόβλεψη της επόμενης μελλοντικής τιμής της χρονοσειράς<sup>15</sup> και είναι ιδιαίτερα αποτελεσματικές σε περιπτώσεις που τα δεδομένα παρουσιάζουν τυχαιότητα (δηλαδή όταν απουσιάζουν οι συνιστώσες της τάσης, εποχικότητας και κυκλικότητας). Θεωρείται ότι οι πιο πρόσφατες παρατηρήσεις είναι αντιπροσωπευτικότερες για την πρόβλεψη σε σχέση με τις παλαιότερες τιμές. Έτσι κάθε φορά που μια νέα παρατήρηση είναι διαθέσιμη, προκύπτει ένας νέος μέσος όρος αφαιρώντας την παλαιότερη παρατήρηση και προσθέτοντας τη νεότερη. Ο αριθμός, φυσικά, των παρατηρήσεων που χρησιμοποιούνται στον υπολογισμό κάθε μέσου όρου, παραμένει σταθερός. Οπότε πρέπει να έχουμε τουλάχιστον τον απαιτούμενο αριθμό παρατηρήσεων για να εφαρμοστεί η μέθοδος.

<sup>15</sup> Αυτοί οι μέθοδοι διαφέρουν από αυτούς που είδαμε στην προηγούμενη ενότητα. Οι κινητοί μέσοι όροι που χρησιμοποιούνται στις μεθόδους αποσύνθεσης κάνουν εκτίμηση της καμπύλης τάσης-κύκλου παίρνοντας το μέσο όρο γειτονικών παρατηρήσεων.

Ένας κινητός μέσος όρος τάξης  $k$ , είναι η μέση τιμή  $k$  διαδοχικών τιμών μιας χρονοσειράς και δίνεται από τη σχέση:

$$F_{t+1} = \frac{1}{k} \sum_{i=t-k+1}^t Y_i \quad (4.0)$$

Όπου:

$F_{t+1}$  = η πρόβλεψη της επόμενης περιόδου

$Y_i$  = η τιμή της παρατήρησης της χρονοσειράς

$k$  = ο αριθμός παραγόντων του κινητού μέσου όρου

Για τη βέλτιστη επιλογή του αριθμού  $k$ , χρησιμοποιούνται συνήθως κάποιο από τα στατιστικά μέτρα που έχουμε αναφέρει, όπως το μέσο τετραγωνικό σφάλμα (MSE). Επιλέγοντας διάφορες τιμές του  $k$  και εφαρμόζοντας τη μέθοδο κινητού μέσου όρου, καταλήγουμε στην τιμή που ελαχιστοποιεί το MSE. Στην ακραία περίπτωση που  $k=1$  θα πάρουμε ως πρόβλεψη την τιμή της προηγούμενης περιόδου δηλαδή:  $F_{t+1} = Y_t$ .

Πιο αναλυτικά, εφαρμόζοντας την παραπάνω μαθηματική σχέση στις  $n-k$  τελευταίες παρατηρήσεις της χρονοσειράς, όπου  $n$  το πλήθος των παρατηρήσεων, προκύπτουν οι εξομαλυμένες τιμές που μας δίνει η μέθοδος. Οι τιμές αυτές αποτελούν τις προβλέψεις της αντίστοιχης περιόδου. Οπότε συγκρίνοντας τις παρατηρήσεις με τις προβλέψεις έχουμε και τα αντίστοιχα σφάλματα πρόβλεψης.

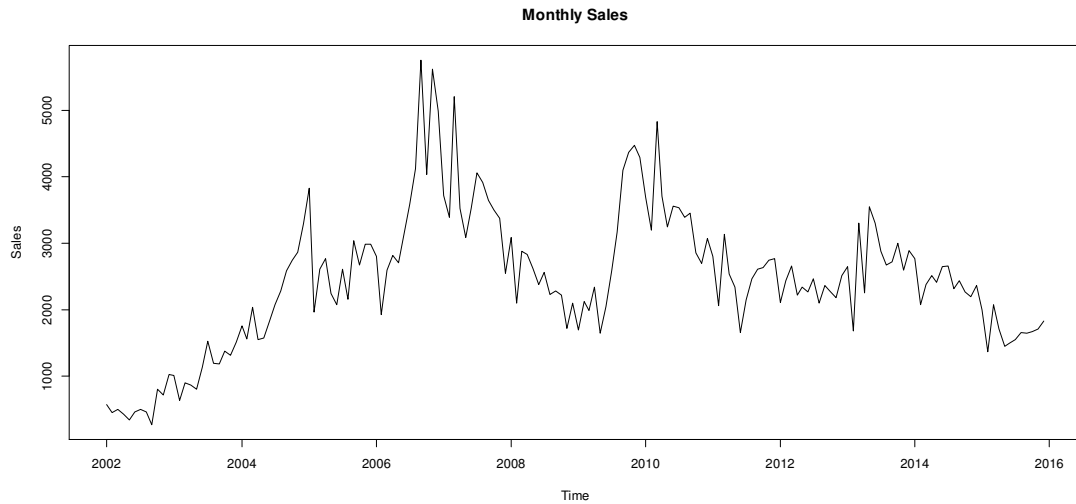
Η εξίσωση (4.0) μπορεί να γραφεί εναλλακτικά:

$$F_{t+1} = F_t + \frac{1}{k}(Y_t - Y_{t-k})$$

Εδώ φαίνεται καλύτερα πως προσαρμόζεται κάθε νέα πρόβλεψη με βάση την προηγούμενη. Βλέπουμε λοιπόν ότι για μεγάλες τιμές του  $k$ , η προσαρμογή για κάθε νέα χρονική περίοδο είναι μικρή άρα το αποτέλεσμα της εξομάλυνσης μεγάλο.

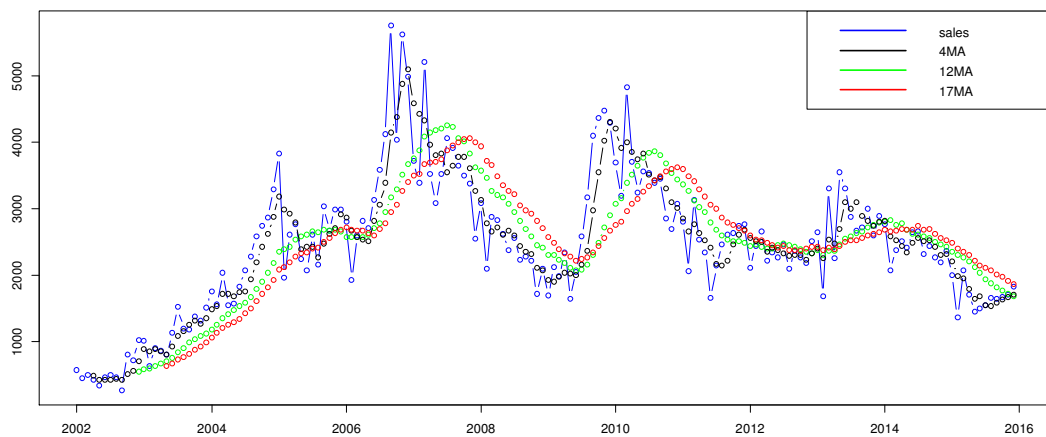
Μπορούμε να παρατηρήσουμε 2 βασικά μειονεκτήματα. Το πρώτο είναι ότι χρειάζεται να αποθηκεύουν πολλή πληροφορία αφού κάθε φορά χρειάζονται  $k$  παρατηρήσεις για τον υπολογισμό του μέσου όρου. Το δεύτερο είναι ότι δε μπορούν να χειριστούν καλά δεδομένα με εποχικότητα και τάση. Οι απότομες αλλαγές στα δεδομένα δεν είναι κάτι που ακολουθούν με σχετική ευκολία με αποτέλεσμα να παρουσιάζεται χρονική υστέρηση στις προβλέψεις. Οπότε για δεδομένα που υπάρχει κάποιο συγκεκριμένο πρότυπο συμπεριφοράς συνίσταται μικρή τιμή του  $k$ , ώστε να επιτρέπει στον κινητό μέσο να ακολουθήσει το πρότυπο των παρατηρήσεων. Από την άλλη, για δεδομένα που κυριαρχεί η τυχαιότητα η χρήση μεγάλου αριθμού περιόδων  $k$  οδηγεί στην ομαλοποίηση των διακυμάνσεων που σχετίζονται με την τυχαιότητα των παρατηρήσεων και άρα σε μεγαλύτερη εξομάλυνση.

Ας δούμε ένα παράδειγμα που αφορά τις μηνιαίες πωλήσεις μιας εταιρείας το διάστημα 2002-2015<sup>16</sup>. Στην εικόνα 4.2-2 φαίνεται η γραφική παράσταση των τιμών συναρτήσεως του χρόνου. Τα δεδομένα φαίνεται να παρουσιάζουν τάση και έντονη κυκλικότητα με μέσο μήκος τα 3 έτη.



Εικόνα 4.2-2: Μηνιαίες πωλήσεις εταιρείας (time plot)

Θα εφαρμόσουμε κινητούς μέσους όρους διαφόρων τάξεων για να δούμε την εξομάλυνση που πραγματοποιείται στα δεδομένα (Εικόνα 4.2-3). Ενδεικτικά έχουμε υπολογίσει για  $k=4,12,17$ .

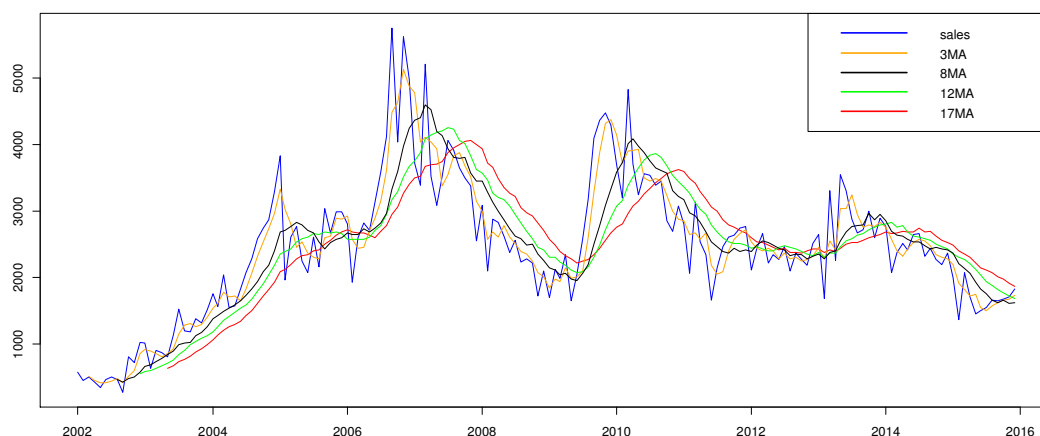


Εικόνα 4.2-3: Κινητοί μέσοι όροι για τα δεδομένα των πωλήσεων

Αυτό που παρατηρείται είναι ότι για μικρές τιμές του  $k$  ο κινητός μέσος όρος ακολουθεί το μοτίβο των δεδομένων καλύτερα από ότι για μεγάλες τιμές. Γενικά ισχύει ότι όσο μεγαλώνει το  $k$  τόσο μεγαλύτερη είναι και η εξομάλυνση και συνήθως οι μεγάλες σχετικά τιμές του  $k$  συνίσταται για δεδομένα που παρουσιάζουν έντονο το στοιχείο της τυχαιότητας. Από την άλλη για δεδομένα με έντονη εποχικότητα

<sup>16</sup> Πηγή: <http://www.adata.com/en/about/content-sales-1/>

συνίσταται η τιμή του  $k$  να συμπίπτει με τον αριθμό των περιόδων σε ένα έτος. Δηλαδή  $k=12$  για μηνιαίες παρατηρήσεις,  $k=3$  για τετραμηνιαίες παρατηρήσεις κλπ. Αυτό βέβαια εξομαλύνει τη σειρά ως προς την εποχικότητα και δυσκολεύει την πρόβλεψη. Παρόλα αυτά φαίνεται στο παράδειγμα αυτό, ότι αυτή η μέθοδος δείχνει να παρουσιάζει βέλτιστη απόδοση για  $k=3$ . Η αναπαράσταση δίνεται στο παρακάτω σχήμα. Εντοπίζει το μοτίβο αλλά πάντα μένει πίσω μία με δύο περιόδους κάτι που είναι αναμενόμενο. Και φαίνεται ότι στα σημεία που παρουσιάζονται απότομες αλλαγές ακολουθεί καλύτερα το μοτίβο σε σύγκριση με το να παίρναμε έναν κινητό μέσο όρο μεγαλύτερης τάξης.



Εικόνα 4.2-4: Βέλτιστο αποτέλεσμα με χρήση κινητών μέσων όρων στα δεδομένα πωλήσεων

## 4.2.2 Μέθοδοι εκθετικής εξομάλυνσης

Η περίπτωση των κινητών μέσων που μελετήσαμε έχει το «κακό» ότι αποδίδει ίσα βάρη σε όλες τις παρατηρήσεις. Και συνήθως οι πιο πρόσφατες παρατηρήσεις είναι αυτές που μας δίνουν μία επίγνωση της κατάστασης και αποτελούν μια καλή βάση για μελλοντικές προβλέψεις. Το γεγονός αυτό εκμεταλλεύονται οι μέθοδοι εκθετικής εξομάλυνσης, οι οποίες παράγουν προβλέψεις με τη χρήση κάποιου σταθμικού μέσου όρου, δίνοντας μεγαλύτερη βαρύτητα στις πιο πρόσφατες παρατηρήσεις και μικρότερη στις παλαιότερες. Είναι μέθοδοι που χρησιμοποιούνται ιδιαίτερα σε περιπτώσεις που θέλουμε να κάνουμε βραχυπρόθεσμες προβλέψεις, δηλαδή ο ορίζοντας πρόβλεψης που έχουμε είναι μικρός

Οι μέθοδοι εκθετικής εξομάλυνσης είναι από τις πιο διαδομένες και εφαρμόσιμες μεθόδους πρόβλεψης. Οι λόγοι που χρησιμοποιούνται τόσο είναι μεταξύ άλλων:

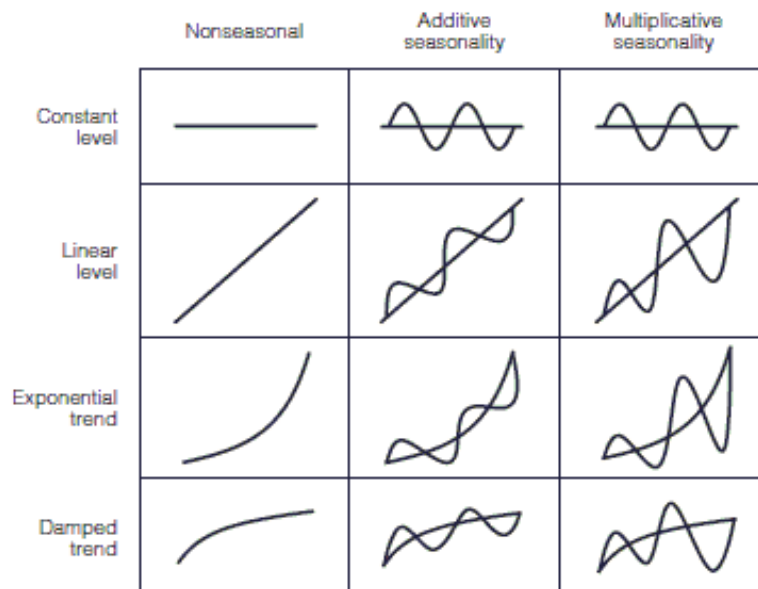
- ✓ Παρουσιάζουν μεγάλη αξιοπιστία καθώς είναι εξαιρετικά ακριβείς στις προβλέψεις
- ✓ Διαθέτουν ευκολία διατύπωσης και κατανόησης του εκθετικού μοντέλου
- ✓ Δεν έχουν απαίτηση αποθήκευσης μεγάλου όγκου πληροφορίας
- ✓ Χρειάζονται λίγους υπολογισμούς



- ✓ Η ακρίβεια των προβλέψεων είναι πολύ εύκολο να υπολογιστεί

Είναι πολύ σημαντικό όταν δουλεύουμε με μεθόδους εξομάλυνσης να μπορούμε να αναγνωρίσουμε τα πρότυπα των ιστορικών δεδομένων. Μπορεί να γίνει μια κατηγοριοποίηση των μεθόδων η οποία προτάθηκε από τον Pegel (1969) (Pegel's classification scheme). Σύμφωνα με αυτή, υπάρχουν 2 κριτήρια ταξινόμησης των μεθόδων. Το ένα αναφέρεται στην ύπαρξη ή όχι συστηματικής συνιστώσας τάσης ή/και εποχικότητας. Το άλλο αφορά τον τρόπο με τον οποίο κάθε συνιστώσα επιδρά στη μεταβολή των υπό εξέταση δεδομένων, δηλαδή αν είναι αθροιστική (additive) ή πολλαπλασιαστική (multiplicative). Κάθε περίπτωση καθώς και οι εξισώσεις που περιγράφουν τις μεθόδους φαίνονται αναλυτικά στον πίνακα 4.2-2. Στην Εικόνα 4.2-5 παρουσιάζεται η επεκταμένη μορφή της κατηγοριοποίησης του Pegel από τον Everette S. Gardner (1985). Φαίνεται να υπάρχουν τέσσερα μοντέλα τάσης που συνδυάζονται με τρία εποχιακά μοντέλα.

Το μοντέλο σταθερού επιπέδου (constant level), υποθέτει ότι απουσιάζει η τάση από τα δεδομένα. Δηλαδή οι χρονοσειρές ακολουθούν ένα σταθερό επίπεδο τιμών. Τα μοντέλα αυτά χρησιμοποιούνται για προβλέψεις ενός βήματος ή όταν κυριαρχεί τυχαιότητα και μεγάλος θόρυβος στη χρονοσειρά. Τέτοιο μοντέλο χρησιμοποιεί η μέθοδος απλής εκθετικής εξομάλυνσης. Όταν συνυπάρχουν και στοιχεία τάσης και εποχικότητας, απαιτούνται μοντέλα που να μπορούν να λάβουν υπόψη τις επιδράσεις αυτών των στοιχείων. Τέτοια μοντέλα χρησιμοποιούν οι μέθοδοι Holt και Holt-Winters που θα αναφερθούμε στη συνέχεια.



Εικόνα 4.2-5: Επέκταση κατηγοριοποίησης κατά Pegel

Οι παρακάτω τύποι αναπαριστούν μαθηματικά και με συνοπτικό τρόπο όλες τις περιπτώσεις που αναφέρονται στην Εικόνα 4.2-5 :

$$L_t = \alpha P_t + (1 - \alpha)Q_t \quad (4.1)$$

$$b_t = \beta R_t + (1 - \beta)b_{t-1} \quad (4.2)$$

$$S_t = \gamma T_t + (1 - \gamma)S_{t-s} \quad (4.3)$$

Πίνακας 4.2-1: Γενικοί τύποι μεθόδων εξομάλυνσης

Τα  $P, Q, R, T$  είναι διαφορετικά ανάλογα με την περίπτωση της μεθόδου που αναφέρονται. Η εξίσωση 4.1 περιγράφει την εκτίμηση της εκάστοτε μεθόδου για το επίπεδο της χρονοσειράς. Αποτελεί μια διαδικασία εξομάλυνσης των στοιχείων P και Q με μια σταθερά εξομάλυνσης  $\alpha$ . Κατά τον ίδιο τρόπο, οι εξισώσεις 4.2 και 4.3 περιγράφουν την εκτίμηση της τάσης και της εποχικότητας της χρονοσειράς αντίστοιχα με συντελεστές εξομάλυνσης  $\beta$  και  $\gamma$ .

Η κατηγοριοποίηση κατά Pegel παρουσιάζεται αναλυτικά στον παρακάτω πίνακα. Η ερμηνεία του θα γίνει κατανοητή μέσα από την εξέταση κάποιων μεθόδων που θα δούμε στη συνέχεια.

Τάση	Συνιστώσα εποχικότητας		
	1 (καμία)	2 (προσθετική)	3 (πολλαπλασιαστική)
A (καμία)	$P_t = Y_t$ $Q_t = L_{t-1}$ $F_{t+m} = L_t$	$P_t = Y_t - S_{t-s}$ $Q_t = L_{t-1}$ $T_t = Y_t - L_t$ $F_{t+m} = L_t + S_{t+m-s}$	$P_t = Y_t/S_{t-s}$ $Q_t = L_{t-1}$ $T_t = Y_t/L_t$ $F_{t+m} = L_t S_{t+m-s}$
B (προσθετική)	$P_t = Y_t$ $Q_t = L_{t-1} + b_{t-1}$ $R_t = L_t - L_{t-1}$ $F_{t+m} = L_t + mb_t$	$P_t = Y_t - S_{t-s}$ $Q_t = L_{t-1} + b_{t-1}$ $R_t = L_t - L_{t-1}$ $T_t = Y_t - L_t$ $F_{t+m} = L_t + mb_t + S_{t+m-s}$	$P_t = Y_t/S_{t-s}$ $Q_t = L_{t-1} + b_{t-1}$ $R_t = L_t - L_{t-1}$ $T_t = Y_t/L_t$ $F_{t+m} = (L_t + mb_t) S_{t+m-s}$
Γ (πολλαπλασιαστική)	$P_t = Y_t$ $Q_t = L_{t-1}b_{t-1}$ $R_t = L_t/L_{t-1}$ $F_{t+m} = L_t b_t^m$	$P_t = Y_t - S_{t-s}$ $Q_t = L_{t-1}b_{t-1}$ $R_t = L_t/L_{t-1}$ $T_t = Y_t - L_t$ $F_{t+m} = L_t b_t^m + S_{t+m-s}$	$P_t = Y_t/S_{t-s}$ $Q_t = L_{t-1}b_{t-1}$ $R_t = L_t/L_{t-1}$ $T_t = Y_t/L_t$ $F_{t+m} = L_t b_t^m S_{t+m-s}$

Πίνακας 4.2-2: Τύποι για υπολογισμούς προβλέψεων χρησιμοποιώντας την ταξινόμηση κατά Pegel

### 4.2.2.1 Απλή εκθετική εξομάλυνση (Simple Exponential Smoothing<sup>17</sup>)

Έστω ότι θέλουμε να προβλέψουμε την επόμενη μελλοντική τιμή της χρονοσειράς ( $F_{t+1}$ ). Με τη μέθοδο της απλής εκθετικής εξομάλυνσης χρειάζονται τρεις πληροφορίες για τον υπολογισμό της: η πιο πρόσφατη πρόβλεψη ( $F_t$ ), η πραγματική τιμή της χρονοσειράς για την πρόβλεψη αυτή ( $Y_t$ ) και μια σταθερά εξομάλυνσης  $\alpha$ . Η διαδικασία που ακολουθεί είναι η εξής :

$$\text{Νέα πρόβλεψη} = \text{πρόσφατη πρόβλεψη} + \alpha \cdot (\text{σφάλμα της πρόσφατης πρόβλεψης})$$

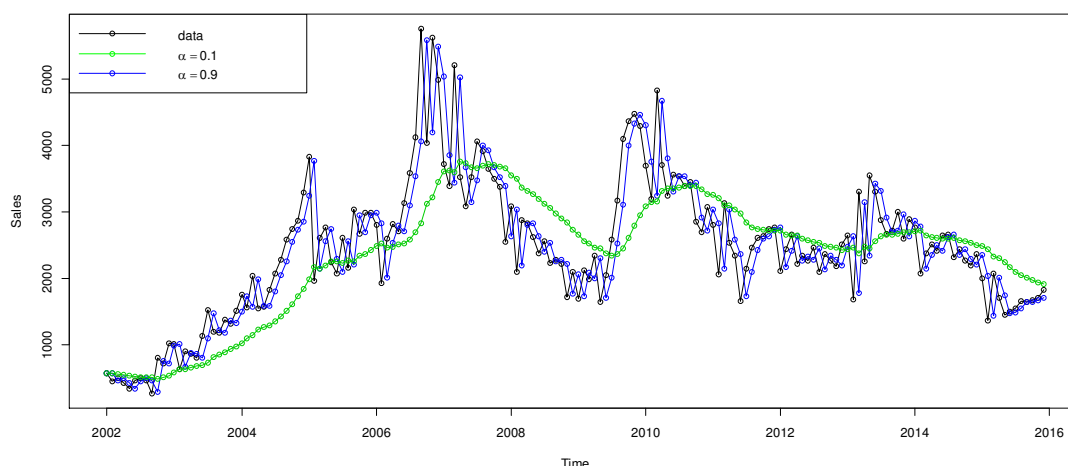
Δηλαδή χρησιμοποιεί την πρόβλεψη της προηγούμενης περιόδου και την προσαρμόζει χρησιμοποιώντας το σφάλμα της πρόβλεψης για την περίοδο αυτή. Η μαθηματική σχέση για το μοντέλο της απλής εκθετικής εξομάλυνσης είναι:

$$\begin{aligned} F_{t+1} &= F_t + \alpha(Y_t - F_t) \\ &= \alpha Y_t + (1 - \alpha)F_t \end{aligned} \quad (4.4)$$

Επειδή υπάρχει μόνο μία σταθερά, το μοντέλο αυτό λέγεται και μοντέλο εκθετικής εξομάλυνσης μιας παραμέτρου. Η σταθερά εξομάλυνσης  $\alpha$  παίρνει τιμές στο διάστημα  $[0,1]$ . Όταν η τιμή της πλησιάζει το 1 τότε η νέα πρόβλεψη θα περιέχει μεγάλο ποσοστό του σφάλματος πρόβλεψης για την προηγούμενη παρατήρηση, άρα θα παρουσιάζει μεγαλύτερη ευαισθησία και μεγαλύτερη ανταπόκριση στις μεταβολές. Ενώ στην περίπτωση που πλησιάζει το 0, θα περιέχει μικρό ποσοστό του σφάλματος οπότε η πρόβλεψη θα είναι πιο σταθερή. Αυτή η τεχνική που εφαρμόζει, κάνει την απλή εκθετική εξομάλυνση να δίνει προβλέψεις που ακολουθούν τα πρότυπα στις παρατηρήσεις της χρονοσειράς με τον περιορισμό ότι στα σημεία καμψής των παρατηρήσεων, οι προβλέψεις που δίνει παρουσιάζουν καθυστέρηση (είναι πιο έντονο για μικρές τιμές της  $\alpha$ ). Μία εφαρμογή των παραπάνω φαίνεται στην Εικόνα 4.2-6.

Αξίζει να παρατηρήσουμε ότι η μέθοδος αυτή μας δίνει αυτό-διορθούμενες προβλέψεις αφού διαθέτει μηχανισμό διόρθωσης σφαλμάτων. Πιο αναλυτικά, αν το σφάλμα είναι αρνητικό (δηλαδή η προηγούμενη πρόβλεψη ήταν μεγαλύτερη της πραγματικής τιμής), τότε η πρόβλεψη για την επόμενη χρονική περίοδο μειώνεται. Ενώ στην περίπτωση που το σφάλμα είναι θετικό (δηλαδή η προηγούμενη πρόβλεψη ήταν μικρότερη της πραγματικής τιμής), τότε η πρόβλεψη για την επόμενη χρονική περίοδο αυξάνεται.

<sup>17</sup> Σε κάποια βιβλία αναφέρεται και ως Single Exponential Smoothing



Εικόνα 4.2-6 : Εφαρμογή της απλής εκθετικής εξομάλυνσης στο παράδειγμα με τις πωλήσεις για τιμές της σταθεράς εξομάλυνσης  $\alpha=0.1$  και  $\alpha=0.9$

Η μέθοδος απλής εκθετικής εξομάλυνσης χρησιμοποιεί έναν σταθμισμένο κινητό μέσο όρο αποδίδοντας στις παρατηρήσεις βάρη τα οποία φθίνουν εκθετικά από την πιο πρόσφατη στην παλαιότερη παρατήρηση. Πράγματι, αν αναπτύξουμε τη σχέση 4.4 για περισσότερες περιόδους μπορούμε να δούμε ότι:

$$F_{t+1} = \alpha Y_t + \alpha(1 - \alpha)Y_{t-1} + \alpha(1 - \alpha)^2 Y_{t-2} + \dots + \alpha(1 - \alpha)^{t-1} Y_1 + (1 - \alpha)^t F_1$$

Υπάρχουν δύο σημεία που θέλουν την προσοχή μας:

- Το ένα έχει να κάνει με τι θα επιλέξουμε ως αρχικό επίπεδο για την έναρξη της μεθόδου, δηλαδή ποια θα είναι η τιμή του  $F_1$ . Εδώ έχουμε διάφορες επιλογές. Αρχικά μπορούμε να πάρουμε ως τιμή, την τιμή της πρώτης παρατήρησης ( $Y_1$ ). Μια άλλη επιλογή είναι ο μέσος όρος όλων των παρατηρήσεων. Μπορούμε βέβαια να περιοριστούμε και σε μέσο όρο  $n$  παρατηρήσεων. Όποια και αν είναι η τιμή του αρχικού επιπέδου, θα επηρεάσει όλες τις μετέπειτα προβλέψεις. Αυτό έχει σαν αποτέλεσμα, στην περίπτωση που η τιμή δεν είναι αντιπροσωπευτική να υπάρχει πρόβλημα. Βέβαια το βάρος που παίρνει είναι συνήθως μικρό ( $(1 - \alpha)^t$ ). Παρόλα αυτά, όσο μικρότερο είναι το  $\alpha$  τόσο πιο καθοριστικό ρόλο παίζει στην πρόβλεψη η τιμή της αρχικής πρόβλεψης.
- Το δεύτερο σημείο που πρέπει να προσέξουμε είναι η τιμή που θα δώσουμε στο  $\alpha$ . Μεγάλη τιμή της σταθεράς εξομάλυνσης δίνει μικρή εξομάλυνση στις προβλέψεις και αντίστροφα. Για  $\alpha=1$  η απλή εκθετική εξομάλυνση ταυτίζεται με τη μέθοδο NF1, δηλαδή χρησιμοποιεί για πρόβλεψη την τελευταία παρατήρηση. Επίσης όταν υπάρχει τυχαιότητα στα δεδομένα συνίσταται μικρή τιμή του  $\alpha$ , ενώ αν παρουσιάζεται τάση στα δεδομένα συνίσταται μεγάλη τιμή (το μοντέλο παρακολουθεί καλύτερα τις μεταβολές).

Η βέλτιστη τιμή που θα πάρει η σταθερά  $\alpha$  λοιπόν είναι καθοριστικής σημασίας. Καθορίζει το μέγεθος της εξομάλυνσης και την ταχύτητα που αντιδρά η μέθοδος στις διαφορές των παρατηρήσεων με τις προβλέψεις προηγούμενων περιόδων. Οι τιμές της  $\alpha$  που δουλεύουν καλά είναι συνήθως μεταξύ 0.05 και 0.30. Τιμές της  $\alpha$  που είναι

μεγαλύτερες από 0.30 συνήθως δείχνουν ότι κάποιο εναλλακτικό μοντέλο πρόβλεψης θα είναι περισσότερο κατάλληλο.

Ο υπολογισμός θα μπορούσε να γίνει με την επιλογή εκείνης της τιμής για την σταθερά εξομάλυνσης που ελαχιστοποιεί κάποιο από τα MSE ή MAPE, μια επιλογή που γίνεται μέσα από μια διαδικασία δοκιμής και λάθους (trial and error). Εφαρμόζοντας το μέτρο εκτίμησης σφάλματος σε ένα κομμάτι των δεδομένων για διάφορες τιμές του  $\alpha$ , επιλέγεται τελικά εκείνη η τιμή του  $\alpha$  που δίνει την ελάχιστη τιμή για τα σφάλματα. Μια σχετικά απλή προσέγγιση για τον υπολογισμό της βέλτιστης τιμής της σταθεράς εξομάλυνσης είναι να πάρουμε τυχαία 2 τιμές για το  $\alpha$  και να δούμε ποια δίνει μικρότερα σφάλματα. Μετά να παίξουμε με τιμές  $\pm 0,17$  γύρω από την τιμή που βρήκαμε ότι μας δίνει μικρότερο σφάλμα. Μετά σε αυτή που θα βρούμε να μειώσουμε ακόμα παραπάνω στα αντίστοιχα διαστήματα ( $\pm 0,08, \pm 0,04, \pm 0,02, \pm 0,015, \text{κλπ}$ ) μέχρι να φτάσουμε σε κάτι μικρότερο από 1%. Όλο αυτό με την προϋπόθεση ότι το MSE ή οποιοδήποτε άλλο στατιστικό μέτρο ελέγχουμε, παρουσιάζει ελάχιστο. Τα αποτελέσματα για καθένα από τα στατιστικά μέτρα που αναφέραμε ενδέχεται να είναι διαφορετικά.

Ας υποθέσουμε ότι δίνουμε τις ακόλουθες τιμές στο  $\alpha$  : 0.2, 0.4, 0.6, 0.8. Τότε τα βάρη που αναθέτονται στις παλαιότερες παρατηρήσεις αποτυπώνονται στον παρακάτω πίνακα:

Παρατηρήσεις	$\alpha=0.2$	$\alpha=0.4$	$\alpha=0.6$	$\alpha=0.8$
$Y_t$	0.2	0.4	0.6	0.8
$Y_{t-1}$	0.16	0.24	0.24	0.16
$Y_{t-2}$	0.128	0.144	0.096	0.032
$Y_{t-3}$	0.1024	0.0864	0.0384	0.0064
$Y_{t-4}$	$(0.2)(0.8)^4$	$(0.4)(0.6)^4$	$(0.6)(0.4)^4$	$(0.8)(0.2)^4$

Πίνακας 4.2-3: Ενδεικτικά βάρη που αποδίδονται στις παρατηρήσεις με τη χρήση της απλής εκθετικής εξομάλυνσης

Το άθροισμα όλων των συντελεστών βαρύτητας ισούται με τη μονάδα.

Αν θέλουμε να υπολογίσουμε περισσότερες από μία προβλέψεις, ( $h>1$ ) τότε όλες οι επόμενες θα πάρουν την τιμή της τελευταίας υπολογισμένης πρόβλεψης ( $F_{t+h} = F_{t+1}$ ). Αυτό συμβαίνει γιατί το μοντέλο θεωρεί την ύπαρξη ενός σταθερού επιπέδου (απουσία μοτίβων).

Ας πάρουμε πάλι το παράδειγμα με τις πωλήσεις. Η χρονοσειρά που μελετάμε αφορά της μηνιαίες παρατηρήσεις μιας εταιρίας τα έτη 2002 έως 2015. Ενδεικτικά θα δείξουμε τις πρώτες 17 τιμές που έχουμε στη διάθεσή μας. Θα εφαρμόσουμε τη μέθοδο απλής εκθετικής εξομάλυνσης χρησιμοποιώντας το στατιστικό πακέτο R. Οι υπολογισμοί και οι εντολές που χρησιμοποιήθηκαν αναφέρονται στο παράρτημα κάτω από το αρχείο saleses.r. Τα αποτελέσματα φαίνονται στους παρακάτω πίνακες.

---

### Forecast method: Simple exponential smoothing

---

---

**Model Information:****Simple exponential smoothing****Smoothing parameters:****alpha = 0.5558****Initial states:****I = 521.0755**

---

Πίνακας 4.2-4: Αποτελέσματα της Rμε την εφαρμογή της απλής εκθετικής εξομάλυνσης σε δεδομένα πωλήσεων

Ο Πίνακας 4.2-4 δείχνει τις βέλτιστες τιμές της σταθεράς εξομάλυνσης και του αρχικού επιπέδου. Βλέπουμε ότι η βέλτιστη τιμή της σταθεράς είναι 0.56 περίπου, τιμή που ξεπερνά το 0.30. Αυτό είναι ένδειξη ότι ίσως πρέπει να αναζητήσουμε ένα καλύτερο μοντέλο πρόβλεψης. Η εκτίμηση των σφαλμάτων με τη χρήση του MSE δίνει την τιμή 252014.3 ως την ελάχιστη δυνατή.

Περίοδος	Πωλήσεις	Απλή εκθετική εξομάλυνση		
		$\alpha=0.5558$	$\alpha=0.1$	$\alpha=0.9$
Jan 2002	574	-	-	-
Feb 2002	449	550.49308	574	574
Mar 2002	499	494.0791	561.5	461.5
Apr 2002	424	496.81434	555.25	495.25
May 2002	338	456.34117	542.125	431.125
Jun 2002	467	390.56234	521.7125	347.3125
Jul 2002	500	433.0495	516.2413	455.0313
Aug 2002	467	470.26331	514.6171	495.5031
Sep 2002	274	468.44943	509.8554	469.8503
Oct 2002	803	360.36654	486.2699	293.585
Nov 2002	724	606.40019	517.9429	752.0585
Dec 2002	1022	671.76694	538.5486	726.8059
Jan 2003	1011	866.4407	586.8937	992.4806
Feb 2003	641	946.79263	629.3044	1009.148
Mar 2003	905	776.82067	630.4739	677.8148
Apr 2003	865	848.06795	657.9265	882.2815
May 2003	805	857.47947	678.6339	866.7281

	h=1	Forecast		
Jan 2016	-	1766.134	1910.472	1817.142
MSE		252014.3	492328.4	288137.3
MAPE		15.13217	22.299	16.26489

Πίνακας4.2-5: Δεδομένα και προβλέψεις πωλήσεων με τη μέθοδο εκθετικής εξομάλυνσης

Ο Πίνακας4.2-5 περιέχει τα δεδομένα και τα αποτελέσματα που προκύπτουν μετά την εφαρμογή της μεθόδου απλής εκθετικής εξομάλυνσης με τιμές σταθεράς  $\alpha=0.1$ ,  $\alpha=0.9$  και της βέλτιστης τιμής όπως προέκυψε να είναι  $\alpha=0.5558$ . Οι τιμές 0.1 και 0.9 είναι δύο ακραίες περιπτώσεις που επιλέχθηκαν για να δούμε πως φαίνεται η εξομάλυνση. Στην πρώτη περίπτωση που η τιμή της  $\alpha$  είναι μεγάλη αναμένουμε η καμπύλη πρόβλεψης να ακολουθεί αυτή των δεδομένων ενώ στην άλλη περίπτωση που η τιμή της  $\alpha$  είναι σχετικά μικρή αναμένουμε μια πιο σταθερή καμπύλη (βλέπε Εικόνα 4.2-6). Οι μελλοντικές προβλέψεις φαίνονται γραφικά στην Εικόνα 4.2-7. Διαλέξαμε ένα τμήμα των δεδομένων, δηλαδή για την περίοδο 2012-2015 ώστε να είναι πιο ευδιάκριτες και οι τιμές της μελλοντικής πρόβλεψης για τις διάφορες τιμές της  $\alpha$  σε σχέση με τα δεδομένα.



Εικόνα 4.2-7: Προβλέψεις με τη μέθοδο απλής εκθετικής εξομάλυνσης με εφαρμογή στο παράδειγμα των πωλήσεων

### 4.2.2.2 Μέθοδος ARSSES

Η μέθοδος ARSSES ή αλλιώς μέθοδος της απλής εκθετικής εξομάλυνσης με αναπροσαρμόζόμενο ρυθμό ανταπόκρισης αποτελεί μία παραλλαγή της μεθόδου απλής εκθετικής εξομάλυνσης. Η διαφορά έγκειται στη σταθερά εξομάλυνσης  $\alpha$ . Η ARSSES έχει το πλεονέκτημα ότι μπορεί να μεταβάλλει την τιμή της  $\alpha$  καθώς αλλάζει

το μοτίβο των δεδομένων. Το χαρακτηριστικό αυτό την καθιστά ιδιαίτερα ελκυστική κυρίως σε περιπτώσεις όπου οι τιμές που χρειάζονται να προβλεφθούν είναι πολλές και όταν απουσιάζουν από τα δεδομένα μοτίβα τάσης και εποχικότητας..

Η αλγεβρική διατύπωση της μεθόδου είναι:

$$F_{t+1} = a_t Y_t + (1 - a_t) F_t \quad (4.6)$$

Όπου:

$$a_{t+1} = \left| \frac{A_t}{M_t} \right| \quad (4.7)$$

$$\text{(σφάλμα εξομάλυνσης)} \quad A_t = \beta E_t + (1 - \beta) A_{t-1} \quad (4.8)$$

$$\text{(απόλυτο σφάλμα εξομάλυνσης)} \quad M_t = \beta |E_t| + (1 - \beta) A_{t-1} \quad (4.9)$$

$$\text{(σφάλμα)} \quad E_t = Y_t - F_t \quad (4.10)$$

Η παράμετρος  $\beta$  παίρνει τιμές στο διάστημα (0,1). Οι συνήθεις τιμές που παίρνει είναι στην περιοχή 0.1 ή 0.2 γιατί μειώνουν την επίδραση προηγούμενων σφαλμάτων και η ανταπόκριση της μεθόδου στα δεδομένα γίνεται σταδιακά.

Η εξίσωση (4.6) μας δείχνει ότι η πρόβλεψη της επόμενης περιόδου είναι, όπως και στην περίπτωση της απλής εκθετικής εξομάλυνσης, ένας γραμμικός συνδυασμός της πραγματικής τιμής της προηγούμενης περιόδου και της αντίστοιχης πρόβλεψης. Βλέπουμε από τη σχέση (4.7) ότι η σταθερά εξομάλυνσης  $a$  είναι ο λόγος 2 ποσοτήτων κατ' απόλυτη τιμή. Των  $A_t$  και  $M_t$ , τα οποία και θα αναλύσουμε αμέσως. Η ποσότητα που βρίσκεται στον αριθμητή ( $A_t$ ) είναι το σφάλμα εξομάλυνσης. Ο τρόπος υπολογισμού του φαίνεται στη σχέση (4.8). Είναι ένας σταθμισμένος μέσος όρος του σφάλματος της πρόβλεψης την περίοδο που εξετάζουμε ( $E_t$ ) και του σφάλματος εξομάλυνσης της προηγούμενης περιόδου ( $A_{t-1}$ ). Αντίστοιχα, η ποσότητα που βρίσκεται στον παρανομαστή ( $M_t$ ), είναι η εξομαλυμένη εκτίμηση για το απόλυτο σφάλμα της πρόβλεψης (απόλυτο σφάλμα εξομάλυνσης). Από τη σχέση (4.9) μπορούμε να δούμε ότι και αυτό είναι ένας σταθμισμένος μέσος όρος αλλά του απολύτου σφάλματος της πρόβλεψης την περίοδο που εξετάζουμε ( $E_t$ ) και του σφάλματος εξομάλυνσης της προηγούμενης περιόδου ( $M_{t-1}$ ). Στην ουσία και οι 2 ποσότητες αποτελούν εκτιμήσεις απλής εκθετικής εξομάλυνσης.

Αν λοιπόν θέλαμε να υπολογίσουμε την πρόβλεψη για την περίοδο  $t+2$ , θα χρησιμοποιούσαμε ως τιμή της  $a$  το λόγο  $\left| \frac{A_t}{M_t} \right|$ . Ο λόγος που προτιμούμε να έχουμε στην εξίσωση 4.8 την τιμή της  $a$  για την περίοδο  $t+1$  και όχι για την  $t$ , είναι γιατί η μέθοδος ARSSES ανταποκρίνεται αμέσως στις αλλαγές στα δεδομένα, παραπάνω από όσο θα ήταν επιθυμητό, και με αυτό τον τρόπο εισάγουμε μια καθυστέρηση που επιτρέπει στο σύστημα να σταθεροποιηθεί. Αν τα σφάλματα  $E_i = Y_i - F_i$  είναι συνεχώς μεγάλα κατ' απόλυτη τιμή, τότε η  $a_t$  θα τείνει να πάρει την τιμή 1, κάνοντας έτσι τη μέθοδο να ανταποκρίνεται πολύ γρήγορα στις αλλαγές. Αντιθέτως, στην περίπτωση που τα σφάλματα πρόβλεψης είναι συνεχώς μικρά, τότε οι τιμές της  $a_t$  θα τείνουν κοντά στο 0 και άρα η μέθοδος θα εξομαλύνει τις τυχαίες διακυμάνσεις.



Επειδή οι τιμές της  $a_t$  αλλάζουν πολύ εύκολα με κάθε προσαρμογή στα δεδομένα, διαφορετικές τιμές του αρχικού επιπέδου της μεθόδου είναι δυνατόν να δημιουργήσουν και διαφορετική σειρά τιμών της  $a_t$ . Για να το αντιμετωπίσουμε αυτό, θα μπορούσε να μπει κάποιος περιορισμός ως προς το εύρος των τιμών που μπορεί να πάρει.

Η διαδικασία ορισμού του αρχικού επιπέδου της μεθόδου είναι λίγο πιο ιδιαίτερη από αυτή της απλής εκθετικής εξομάλυνσης. Θα το δούμε μέσα από ένα παράδειγμα. Ας εφαρμόσουμε τη μέθοδο στα δεδομένα των πωλήσεων που ήδη έχουμε αναφερθεί.

Για να ορίσουμε το αρχικό επίπεδο της μεθόδου υποθέτουμε τα εξής:

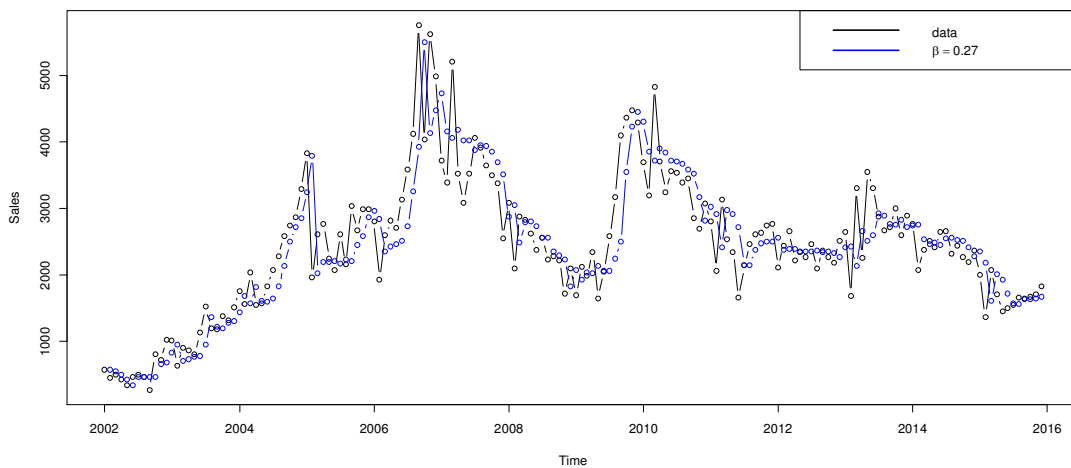
$$A_1 = M_1 = 0, \quad F_2 = Y_1, \quad a_2 = a_3 = 0.2, \quad \beta = 0.27$$

Οι προβλέψεις με τη μέθοδο ARSSES φαίνονται στον πίνακα 4.2-6. Ενδεικτικά θα δείξουμε τις πρώτες 17 τιμές από το σύνολο των δεδομένων που έχουμε στη διάθεσή μας. Θα εφαρμόσουμε τη μέθοδο ARSSES χρησιμοποιώντας το στατιστικό πακέτο R. Ο κώδικας που ακολουθήθηκε για την κατασκευή της μεθόδου στην R βρίσκεται στο παράρτημα κάτω από το αρχείο salesARSSES.r

Period	Observed values (Sales)	Forecast	Error	Smoothed Error	Absolute smoothed error	
t	$Y_t$	$F_t$	$E_t$	$A_t$	$M_t$	$a_t$
Jan 2002	574	-	-	-	-	-
Feb 2002	449	574	-125	-25	25	0.2
Mar 2002	499	549	-50	-30	30	0.2
Apr 2002	424	499	-75	-39	39	1
May 2002	338	424	-86	-48.4	48.4	1
Jun 2002	467	338	129	-12.92	64.52	1
Jul 2002	500	467	33	-3.736	58.216	1
Aug 2002	467	473.6082	-6.60818	-4.3104367	47.89444	0.200248
Sep 2002	274	473.1841	-199.184	-43.28517	78.15237	0.0641748
Oct 2002	803	455.2578	347.7422	34.920305	132.0703	0.0899987
Nov 2002	724	647.857	76.14305	43.164853	120.8849	0.5538561
Dec 2002	1022	667.9897	354.0103	105.33394	167.51	0.2644069
Jan 2003	1011	794.3976	216.6024	127.58764	177.3285	0.357074
Feb 2003	641	930.6019	-289.602	44.149721	199.7832	0.628822
Mar 2003	905	722.2336	182.7664	71.873048	196.3798	0.719499
Apr 2003	865	762.6229	102.3771	77.973867	177.5793	0.2209882
May 2003	805	800.0919	4.908129	63.360719	143.045	0.36599

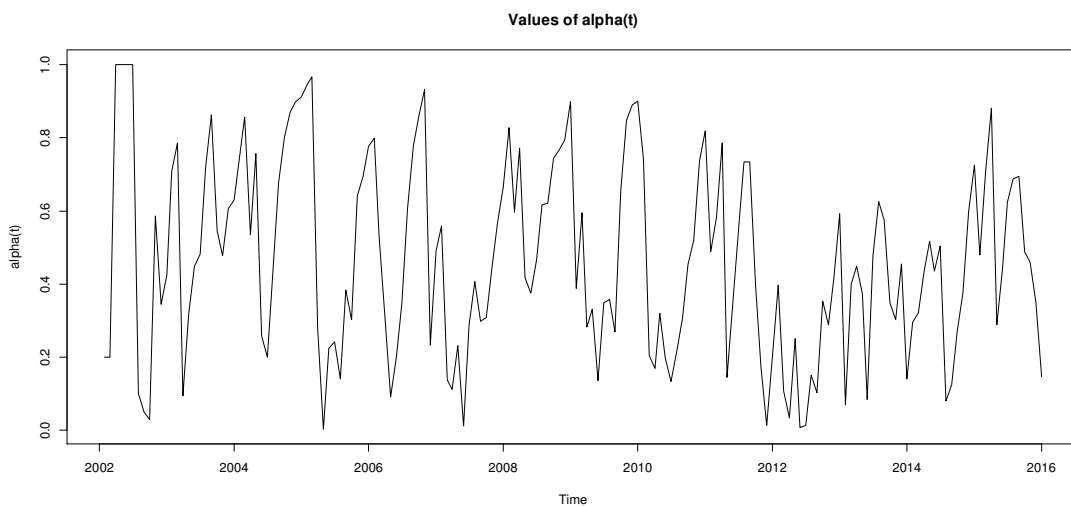
<b>h=1</b>	Forecast
<b>Jan 2016</b>	1696.513
<b>MSE</b>	266127.1
<b>MAPE</b>	15.7957

Πίνακας 4.2-6 Αποτελέσματα εφαρμογής της μεθόδου ARSSES σε δεδομένα πωλήσεων



Εικόνα 4.2-8: Γραφική αναπαράσταση των προβλέψεων με εφαρμογή της μεθόδου ARSSES σε δεδομένα πωλήσεων

Στο παρακάτω σχήμα μπορούμε να δούμε και τις τιμές της  $a_t$  κατά τον υπολογισμό των προβλέψεων με τη μέθοδο ARSSES.



Εικόνα 4.2-9: Τιμές της σταθεράς εξομάλυνσης  $a_t$  με εφαρμογή της μεθόδου ARSSES σε δεδομένα πωλήσεων

Παρόλο που η μέθοδος ARSSES υπερέχει σε σύγκριση με την SES στο κομμάτι της αυτοματοποιημένης διαδικασίας παραγωγής των σταθερών εξομάλυνσης, υπάρχουν ενδείξεις πειραματικές που δείχνουν ότι η μέθοδος αυτή δεν έχει καλύτερη απόδοση από την SES μακροπρόθεσμα. Αυτό οφείλεται στη μακροπρόθεσμη αστάθεια που παρουσιάζει μερικές φορές και μάλιστα μια κακή χρήση της μεθόδου, μπορεί να φέρει το αντίθετο από το αναμενόμενο αποτέλεσμα, με μεγάλες τιμές στα σφάλματα πρόβλεψης.

Οι μέθοδοι SES και ARSSES αποτελούν ειδικές περιπτώσεις σταθμισμένων μέσων όρων με βάρη που έχουν θετικές τιμές. Αυτό τους προσδίδει και το χαρακτηριστικό ότι τείνουν να παράγουν προβλέψεις που υπερεκτιμούν τη χρονοσειρά όταν εμφανίζεται στα δεδομένα αρνητική τάση και την υποτιμούν όταν εμφανίζεται θετική τάση.

### 4.2.2.3 Μέθοδος Holt (Holt's Linear Method)

Μια επέκταση της απλής εκθετικής εξομάλυνσης αποτελεί το μοντέλο γραμμικής τάσης που πρότεινε ο Holt (1957) και το οποίο μπορεί να διαχειριστεί τη συνιστώσα της τάσης που παρατηρείται συχνά στα δεδομένα. Το μοντέλο περιέχει 2 παραμέτρους  $\alpha$  και  $\beta$ , που παίρνουν τιμές στο διάστημα  $[0,1]$  και περιγράφεται από τις παρακάτω εξισώσεις:

$$L_t = \alpha Y_t + (1 - \alpha)(L_{t-1} + b_{t-1}) \quad (4.11)$$

$$b_t = \beta(L_t - L_{t-1}) + (1 - \beta)b_{t-1} \quad (4.12)$$

$$F_{t+m} = L_t + b_t m \quad (4.13)$$

Όπου:

$L_t$  = εκτίμηση του επιπέδου της χρονοσειράς τη χρονική στιγμή  $t$

$b_t$  = εκτίμηση της τάσης (κλισης) της χρονοσειράς τη χρονική στιγμή  $t$

$F_{t+m}$  = πρόβλεψη της τιμής της χρονοσειράς  $m$  περιόδους μελλοντικά

$\alpha$  = συντελεστής εξομάλυνσης για το επίπεδο

$\beta$  = συντελεστής εξομάλυνσης για την τάση

Η εξίσωση 4.11 προσαρμόζει το  $L_t$  για την τάση της προηγούμενης περιόδου,  $b_{t-1}$ , προσθέτοντάς το στην εξομαλυνθείσα τιμή για την προηγούμενη περίοδο,  $L_{t-1}$ . Αυτό βοηθάει στη μείωση της καθυστέρησης της εξομάλυνσης και προσαρμόζει άμεσα το  $L_t$  στα νέα δεδομένα. Η τάση από την άλλη, ανανεώνεται όπως φαίνεται στη σχέση 4.12 και εκφράζεται ως διαφορά ανάμεσα στις δύο τελευταίες εξομαλυνθείσες τιμές. Αυτό μπορούμε να το καταλάβουμε καλύτερα αν σκεφτούμε ότι στη περίπτωση που υπάρχει τάση στα δεδομένα, κάθε νέα παρατήρηση θα είναι

είτε μεγαλύτερη είτε μικρότερη από τις προηγούμενες. Οπότε αυτό δημιουργεί μια χωριστή χρονοσειρά που αποτελείται από τιμές του βήματος της τάσης, οι οποίες κυμαίνονται γύρω από μία μέση τιμή. Και επειδή θα υπάρχει κάποια τυχαιότητα, για να τη μειώσουμε εξομαλύνουμε την τιμή της τάσης της τελευταίας περιόδου ( $L_t - L_{t-1}$ ) κατά  $\beta$  και την προσθέτουμε στην προηγούμενη εκτίμηση της τάσης,  $b_{t-1}$ , πολλαπλασιάζοντας με  $(1-\beta)$ . Έχοντας εκτιμήσει την τάση και το επίπεδο, τότε οι προβλέψεις παράγονται όπως φαίνεται στη σχέση 4.13. Η τελευταία εκτιμημένη τάση,  $b_t$ , πολλαπλασιάζεται με τον αριθμό των μελλοντικών περιόδων,  $m$ , που θέλουμε να προβλέψουμε και προσθέτουμε σε αυτή, την τιμή της τελευταίας εκτίμησης του επιπέδου,  $L_t$ .

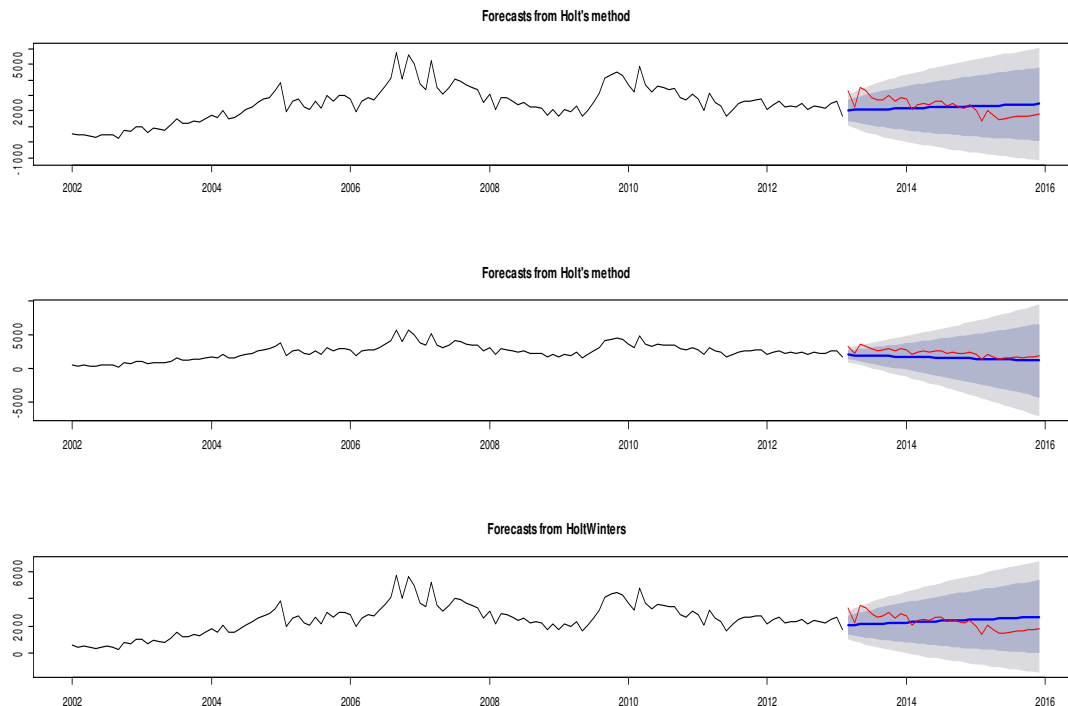
Ας δούμε την εφαρμογή της μεθόδου του Holt στα δεδομένα των πωλήσεων. Έχουμε ένα δείγμα 168 παρατηρήσεων των μηνιαίων πωλήσεων μιας εταιρίας το διάστημα 2002-2015. Θα αναζητήσουμε τις προβλεπόμενες τιμές για 5 περιόδους μπροστά. Δηλαδή, Ιανουάριος 2016, Φεβρουάριος, 2016, Μάρτιος 2016, Απρίλιος 2016 και Μάιος 2016. Η επεξεργασία της μεθόδου έγινε με το στατιστικό πρόγραμμα R. Οι εντολές που χρησιμοποιήθηκαν βρίσκονται στο παράρτημα κάτω από το αρχείο salesHolt.r.

Για να ξεκινήσει η μέθοδος του Holt χρειάζεται 2 εκτιμήσεις. Η μία αφορά την εξομαλυμένη τιμή του επιπέδου,  $L_1$ , και η άλλη την εξομαλυμένη τιμή της τάσης,  $b_1$ . Μία εναλλακτική είναι να υποθέτουμε ότι:

$$L_1 = Y_1 \text{ και } b_1 = Y_2 - Y_1 \text{ ή } b_1 = \frac{Y_4 - Y_1}{3}$$

Ως  $L_1$  ορίζεται η αρχική παρατήρηση και ως  $b_1$  έχουμε 2 περιπτώσεις. Η πρώτη είναι να πάρουμε τη διαφορά των 2 πρώτων παρατηρήσεων. Η δεύτερη είναι να πάρουμε τη διαφορά της  $v$ -στης παρατήρησης και της αρχικής, διαιρεμένες με  $v-1$ . Η άλλη εναλλακτική είναι να εφαρμόσουμε ένα μοντέλο απλής γραμμικής παλινδρόμησης στις πρώτες παρατηρήσεις της χρονοσειράς και να πάρουμε για  $b_1$  τη σταθερά της κλίσης και  $L_1$  το intercept. Είναι πολύ σημαντικό οι τιμές που θα πάρουμε να είναι προσεκτικά επιλεγμένες γιατί θα επηρεάσουν το μοντέλο της πρόβλεψης και κατά συνέπεια και τα σφάλματα πρόβλεψης. Η τελική επιλογή προκύπτει από τα ποιοτικά χαρακτηριστικά της εκάστοτε χρονοσειράς.

Θα παρουσιάσουμε τις 2 εναλλακτικές που αναφέραμε παραπάνω καθώς και τη βέλτιστη που υπολογίζει η μέθοδος.



Εικόνα 4.2-10: Εφαρμογή μεθόδου Holtσε δεδομένα πωλήσεων για διάφορα initial values

Στην Εικόνα 4.2-10 βλέπουμε τα αποτελέσματα εφαρμογής της μεθόδου Holt για 3 διαφορετικές επιλογές αρχικοποίησης της μεθόδου. Έχουμε χωρίσει τα δεδομένα σε 2 τμήματα. Στο αρχικό, που αποτελείται από 134 παρατηρήσεις, γίνεται η εφαρμογή του μοντέλου (το λεγόμενο fitting) και καθορίζονται οι τιμές των παραμέτρων. Στο δεύτερο τμήμα, που αποτελεί το test set (34 παρατηρήσεις), ελέγχεται η απόδοση του κάθε μοντέλου σύμφωνα με τα σφάλματα πρόβλεψης. Τα στοιχεία με την μπλε γραμμή είναι οι προβλέψεις για κάθε μέθοδο και τα στοιχεία με την κόκκινη γραμμή είναι οι παρατηρήσεις από το test set που ξεχωρίσαμε. Η μωβ περιοχή δείχνει τα 80% και 95% διαστήματα εμπιστοσύνης αντίστοιχα. Η πρώτη γραφική παράσταση αφορά το μοντέλο που προκύπτει με τη βέλτιστη επιλογή της μεθόδου R. Η δεύτερη γραφική παράσταση αφορά το μοντέλο που προκύπτει έχοντας επιλέξει ως παραμέτρους τις τιμές των  $\alpha$  και  $\beta$  που ελαχιστοποιούν το MSE και οι αρχικές τιμές του επιπέδου και της τάσης έχουν οριστεί ως  $L_1 = Y_1$  και  $b_1 = Y_2 - Y_1$ . Η τρίτη γραφική παράσταση έχει ως αρχικές τιμές επιπέδου και τάσης το intercept και slope που προκύπτει με εφαρμογή της μεθόδου ελαχίστων τετραγώνων στα δεδομένα των 2 πρώτων ετών της χρονοσειράς.

Τα αποτελέσματα για τα σφάλματα πρόβλεψης φαίνονται στον παρακάτω πίνακα:

	ME	RMSE	MAE	MAPE	MASE	ACF1	Theil's U
Model_1	51.79	667.65	-5.27	26.12	0.58	0.76	1.76
Model_2	-64.92	726.77	-11.29	29.39	0.63	0.78	2.05
Model_3	703.45	796.85	28.37	28.67	0.73	0.39	1.78

Πίνακας 4.2-7: Σφάλματα πρόβλεψης για δοκιμές μοντέλων της μεθόδου του Holt

Η σύγκριση γίνεται για το στατιστικό μέτρο RMSE. Βλέπουμε ότι την καλύτερη επίδοση έχει το μοντέλο που μας παρείχε η R μέσα από τη μέθοδο βελτιστοποίησης που κάνει.

Η βέλτιστη επιλογή όπως προκύπτει και από τα σφάλματα είναι:

---

**Smoothing parameters:**

**alpha = 0.5807**

**beta = 1e-04**

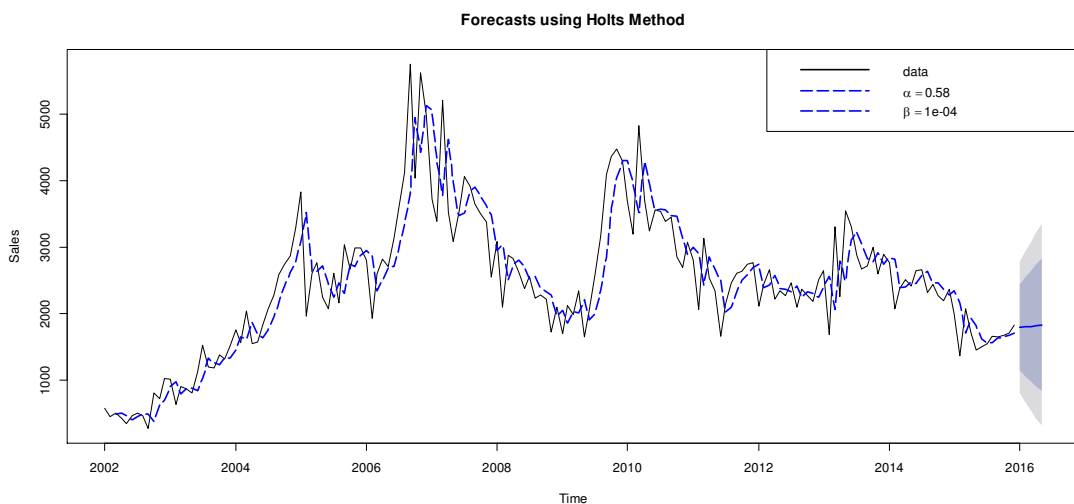
**Initial states:**

**l = 477.234**

**b = 11.5558**

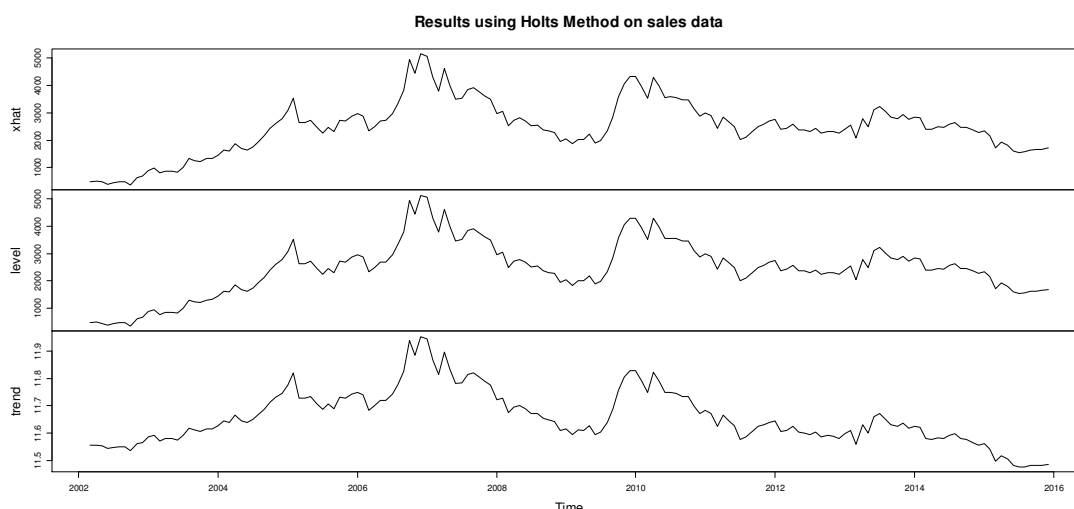
---

Πίνακας 4.2-8: Παράμετροι μοντέλου με την εφαρμογή της μεθόδου Holt σε δεδομένα πωλήσεων



Εικόνα 4.2-11: Γραφική αναπαράσταση των προβλέψεων με εφαρμογή της μεθόδου Holt σε δεδομένα πωλήσεων για τη βέλτιστη επιλογή της μεθόδου

Ο Πίνακας 4.2-8: Παράμετροι μοντέλου με την εφαρμογή της μεθόδου Holt σε δεδομένα πωλήσεων απεικονίζει τις τιμές των παραμέτρων και τις τιμές αρχικών συνθηκών του μοντέλου Holt που δίνουν το βέλτιστο αποτέλεσμα. Η τελική επιλογή προέρχεται από τη χρήση αλγορίθμου βελτιστοποίησης της R. Παρακάτω φαίνεται η εκτίμηση κάθε στοιχείου της χρονοσειράς και η πρόβλεψη που κάνει το μοντέλο που διαλέξαμε ως το καλύτερο δυνατό.



Εικόνα 4.2-12: Αποτελέσματα εφαρμογής της μεθόδου Holt για τις συνιστώσες της χρονοσειράς.

#### 4.2.2.4 Μέθοδος Holt-Winters (Holt-Winters' trend and seasonality method)

Μέχρι τώρα αναπτύξαμε μεθόδους που μπορούν να αντιμετωπίσουν πολλά είδη δεδομένων. Για δεδομένα με την απουσία τάσης και εποχικότητας, έχουν καλή εφαρμογή οι μέθοδοι κινητών μέσων και απλής εκθετικής εξομάλυνσης. Αν τα δεδομένα εμφανίζουν γραμμική τάση, τότε η μέθοδος Holt είναι η πιο κατάλληλη. Αν όμως τα δεδομένα παρουσιάζουν έντονη εποχικότητα, τότε όλες οι μέθοδοι που έχουμε αναπτύξει δε θα έχουν καλή απόδοση γιατί δεν λαμβάνουν υπόψη το στοιχείο της εποχικότητας. Ο Winters (1960) ανέπτυξε μια μέθοδο που υπολογίζει και αυτή τη συνιστώσα και αποτελεί μια επέκταση της μεθόδου του Holt. Γι' αυτό και ονομάζεται μέθοδος Holt-Winters. Είναι μια μέθοδος εκθετικής εξομάλυνσης με προσαρμογή στη γραμμική τάση και την εποχικότητα. Η μέθοδος αυτή, μπορεί να περιγραφεί από 4 βασικές εξισώσεις. Τρεις εξισώσεις εξομάλυνσης, που αφορούν καθεμία χωριστά τα στοιχεία του επιπέδου, της τάσης και της εποχικότητας, και μία εξίσωση πρόβλεψης. Ανάλογα με τον τρόπο που αλληλεπιδρά η εποχικότητα με τα άλλα στοιχεία (προσθετικά ή πολλαπλασιαστικά), η μέθοδος Holt-Winters έχει 2 διαφορετικές μορφές εξισώσεων. Στην περίπτωση που υπάρχει πολλαπλασιαστική επίδραση, το μοντέλο περιγράφεται από τις εξισώσεις που φαίνονται στον παρακάτω πίνακα.

Επίπεδο:	$L_t = a \frac{Y_t}{S_{t-s}} + (1 - a)(L_{t-1} + b_{t-1})$	(4.14)
Τάση:	$b_t = \beta(L_t - L_{t-1}) + (1 - \beta)b_{t-1}$	(4.15)
Εποχικότητα:	$S_t = \gamma \frac{Y_t}{L_t} + (1 - \gamma)S_{t-s}$	(4.16)

Πρόβλεψη:

$$F_{t+m} = (L_t + b_t m) S_{t-s+m}$$

(4.17)

Πίνακας 4.2-9: Εξισώσεις μεθόδου Holt-Winters για την περίπτωση πολλαπλασιαστικής εποχικότητας

Όπου:

$L_t$  = εκτίμηση του επιπέδου της χρονοσειράς την περίοδο  $t$

$b_t$  = εκτίμηση της τάσης (κλισης) της χρονοσειράς την περίοδο  $t$

$S_t$  = εκτίμηση της εποχικότητας της χρονοσειράς την περίοδο  $t$

$F_{t+m}$  = πρόβλεψη της τιμής της χρονοσειράς  $m$  περιόδους μελλοντικά

$a$  = συντελεστής εξομάλυνσης για το επίπεδο

$\beta$  = συντελεστής εξομάλυνσης για την τάση

$\gamma$  = συντελεστής εξομάλυνσης για την εποχικότητα

$s$  = εύρος της εποχικότητας (π.χ. ο αριθμός μηνών ή ο αριθμός τετραμήνων σε ένα έτος)

Αν συγκρίνουμε το μοντέλο αυτό με το μοντέλο του Holt θα παρατηρήσουμε ότι η εξίσωση για την εκτίμηση της τάσης παραμένει ίδια. Σε ότι αφορά τις υπόλοιπες εξισώσεις θα δούμε αναλυτικά τι γίνεται.

Η εξίσωση 4.16 δίνει τις εκτιμηθείσες τιμές της εποχιακής συνιστώσας μέσω μιας διαδικασίας εξομάλυνσης με χρήση της σταθεράς  $\gamma$ . Η εκτίμηση της πρόσφατης εποχιακής συνιστώσας  $Y_t/L_t$  πολλαπλασιάζεται με τη σταθερά  $\gamma$  και η παλαιότερη

εκτίμηση πολλαπλασιάζεται με  $(1-\gamma)$ . Αυτό γίνεται για να μειώσουμε όσο γίνεται την τυχαιότητα από τις παρατηρήσεις. Μπορούμε να παρατηρήσουμε ότι η εποχικότητα εκφράζεται πιο πολύ ως δείκτης παρά ως απόλυτο μέγεθος, από το πηλίκο της τιμής της χρονοσειράς,  $Y_t$ , προς την εξομαλυσμένη τιμή της χρονοσειράς για το επίπεδο,  $L_t$ . Αυτό επιτρέπει την εύρεση των νέων εποχικών εκτιμήσεων με βάση τον εποχικό δείκτη της προηγούμενης περιόδου. Έτσι λοιπόν, αν  $Y_t > L_t$  τότε ο λόγος τους θα έχει τιμή μεγαλύτερη της μονάδας, ενώ αν  $Y_t < L_t$  τότε ο λόγος τους θα έχει τιμή μικρότερη της μονάδας. Να σημειώσουμε σε αυτό το σημείο ότι οι εξομαλυσμένες τιμές του επιπέδου  $L_t$  είναι απαλλαγμένες από την εποχικότητα. Δηλαδή στην ουσία έχουμε αποεποχικοποιημένες τιμές.

Όταν μελετάμε δεδομένα που περιέχουν εποχικότητα και αναζητούμε βραχυχρόνιες προβλέψεις, είναι απαραίτητο να γίνεται απαλοιφή της συνιστώσας της εποχικότητας. Αν παρατηρήσουμε την εξίσωση 4.14 θα δούμε ότι στο πρώτο κομμάτι της εξίσωσης, κάθε παρατήρηση  $Y_t$  διαιρείται με τον εποχιακό δείκτη  $S_{t-s}$ , δηλαδή τον δείκτη του προηγούμενου έτους για την αντίστοιχη περίοδο. Αυτό γίνεται για να αποεποχικοποιήσουμε τα δεδομένα (να θυμηθούμε ότι η εποχικότητα στην περίπτωση που μελετάμε έχει πολλαπλασιαστικό χαρακτήρα). Ο λόγος που χρησιμοποιούμε τον δείκτη  $S_{t-s}$  και όχι τον  $S_t$ , είναι γιατί όταν κάνουμε τους



υπολογισμούς ο  $S_t$  δεν είναι ακόμα διαθέσιμος αφού χρειάζεται την τιμή της  $L_t$  για να υπολογιστεί.

Η πρόβλεψη της περιόδου  $t+m$  προκύπτει όπως φαίνεται και στη σχέση 4.17 από την πρόβλεψη που δίνει η μέθοδος Holt πολλαπλασιασμένη με τον εποχιακό δείκτη  $S_{t-s+m}$ .

Αν κάθε έτος στα δεδομένα που μελετάμε αποτελείται από  $s$  περιόδους τότε θα υπάρχουν και  $s$  εποχιακοί δείκτες, ένας για κάθε περίοδο. Οπότε αν  $S_t$  είναι ο εποχιακός δείκτης για την περίοδο  $t$  τότε ο  $S_{t-s}$  είναι ο εποχιακός δείκτης της αντίστοιχης περιόδου ένα χρόνο πριν. Επίσης κάθε συντελεστής εξομάλυνσης λαμβάνει τιμές στο διάστημα  $[0,1]$ .

Πολύ σημαντική διαδικασία και εδώ αποτελούν η εύρεση της βέλτιστης τιμής των αρχικών συνθηκών και των συντελεστών εξομάλυνσης.

#### ΑΡΧΙΚΟΠΟΙΗΣΗ

Για να ξεκινήσει η μέθοδος χρειαζόμαστε αρχικές τιμές για το επίπεδο  $L_t$ , την τάση  $b_t$  και τους εποχιακούς δείκτες  $S_t$ . Για την εκτίμηση των εποχιακών δεικτών, χρειαζόμαστε τουλάχιστον  $s$  παρατηρήσεις (το νούμερο αντιστοιχεί στις παρατηρήσεις ενός έτους). Οπότε η αρχικοποίηση ξεκινάει από τη χρονική περίοδο  $s$ .

Η αρχική τιμή της τάσης θα είναι ο μέσος όρος των πρώτων παρατηρήσεων. Δηλαδή:

$$L_s = \frac{1}{s} (Y_1 + Y_2 + \dots + Y_s) \quad (4.18)$$

Επειδή τα δεδομένα είναι εποχιακά και η τιμή του μέσου όρου όσο το εύρος της εποχικότητας, ο μέσος εξαλείφει την εποχικότητα από τα δεδομένα. Για την αρχική τιμή της τάσης, βολεύει να δουλέψουμε με τιμές των δεδομένων 2 ετών μέσα από την παρακάτω διαδικασία:

$$b_t = \frac{1}{s} \left[ \frac{Y_{s+1} - Y_1}{s} + \frac{Y_{s+2} - Y_2}{s} + \dots + \frac{Y_{s+s} - Y_s}{s} \right] \quad (4.19)$$

Κάθε όρος του αθροίσματος αποτελεί την εκτίμηση της τάσης για ένα έτος, οπότε ο μέσος όρος των όρων αυτών αποτελεί την αρχική εκτίμηση της τάσης. Οι αρχικές τιμές των εποχιακών δεικτών προκύπτουν από το λόγο των παρατηρήσεων του πρώτου έτους προς τη μέση τιμή των παρατηρήσεων του πρώτου έτους ( $L_s$ ), δηλαδή:

$$S_1 = \frac{Y_1}{L_s}, \quad S_2 = \frac{Y_2}{L_s}, \dots, \quad S_s = \frac{Y_s}{L_s} \quad (4.20)$$

Υπάρχουν και άλλες εναλλακτικές αρχικοποίησης. Μια περίπτωση είναι να πάρουμε τους συντελεστές επιπέδου και τάσης από την ευθεία που προκύπτει αν εφαρμόσουμε τη μέθοδο ελαχίστων τετραγώνων, για τις αρχικές τιμές του επιπέδου και της τάσης του μοντέλου μας. Όσο για τις αρχικές τιμές των εποχιακών δεικτών,

μπορούμε να χρησιμοποιήσουμε τους εποχιακούς δείκτες που προκύπτουν με εφαρμογή της κλασικής μεθόδου αποσύνθεσης της χρονοσειράς στις συνιστώσες της.

Η επιλογή των συντελεστών εξομάλυνσης  $\alpha$ ,  $\beta$ ,  $\gamma$  μπορεί να γίνει με στόχο την ελαχιστοποίηση κάποιου μέτρου εκτίμησης σφαλμάτων όπως το MSE ή το MAPE.

Στην περίπτωση που η εποχικότητα είναι προσθετική στο μοντέλο, κάτι που βέβαια δεν είναι και τόσο συχνό, τότε οι εξισώσεις που περιγράφουν τη μέθοδο Holt-Winters είναι οι παρακάτω:

Επίπεδο:	$L_t = a(Y_t - S_{t-s}) + (1 - a)(L_{t-1} + b_{t-1})$	(4.21)
Τάση:	$b_t = \beta(L_t - L_{t-1}) + (1 - \beta)b_{t-1}$	(4.22)
Εποχικότητα:	$S_t = \gamma(Y_t - L_t) + (1 - \gamma)S_{t-s}$	(4.23)
Πρόβλεψη:	$F_{t+m} = L_t + b_t m + S_{t-s+m}$	(4.24)

Πίνακας 4.2-10: Εξισώσεις μεθόδου Holt-Winters για την περίπτωση αθροιστικής εποχικότητας

Η διαφορά με το μοντέλο της πολλαπλασιαστικής εποχικότητας παρατηρείται στον τρόπο που χρησιμοποιούνται και εκτιμούνται οι δείκτες εξομάλυνσης στις διάφορες εξισώσεις. Αντί να παίρνουμε λόγους και γινόμενα, χρησιμοποιούμε προσθέσεις και αφαιρέσεις. Σε ότι έχει να κάνει με τις αρχικές τιμές, η αρχικοποίηση για το επίπεδο και την τάση είναι ίδιες. Οι αρχικοί δείκτες εξομάλυνσης όμως, υπολογίζονται από τις ακόλουθες σχέσεις:

$$S_1 = Y_1 - L_s, S_2 = Y_2 - L_s, \dots, S_s = Y_s - L_s \quad (4.25)$$

Θα εφαρμόσουμε τη μέθοδο Holt-Winters στα δεδομένα των πωλήσεων. Έχουμε ένα δείγμα 168 παρατηρήσεων των μηνιαίων πωλήσεων μιας εταιρίας το διάστημα 2002-2015. Θα αναζητήσουμε τις προβλεπόμενες τιμές για 5 περιόδους μπροστά. Δηλαδή, Ιανουάριος 2016, Φεβρουάριος, 2016, Μάρτιος 2016, Απρίλιος 2016 και Μάιος 2016. Η επεξεργασία της μεθόδου έγινε με το στατιστικό πρόγραμμα R. Οι εντολές που χρησιμοποιήθηκαν βρίσκονται στο παράρτημα κάτω από το αρχείο salesHoltWinters.r.

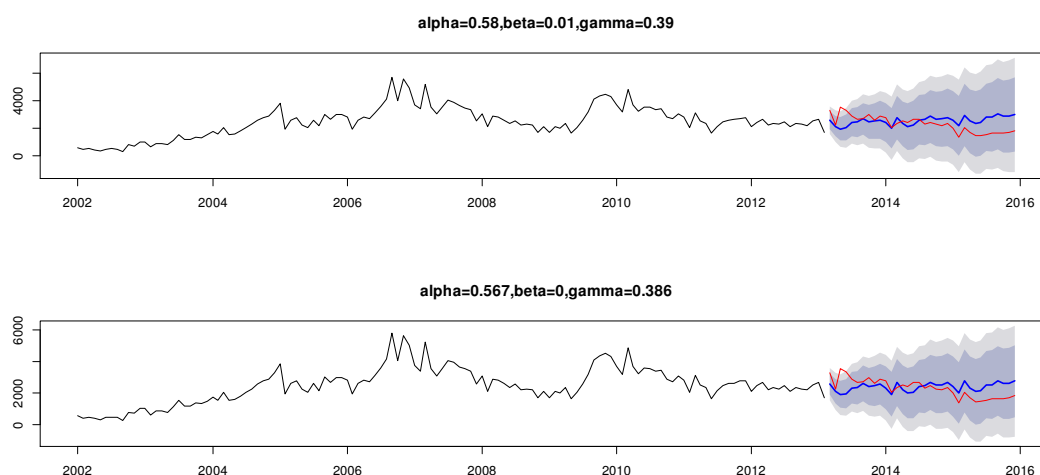
Τα συγκεκριμένα δεδομένα δε φαίνεται να είναι και τόσο κατάλληλα για αυτή τη μέθοδο γιατί δεν είναι έντονο το στοιχείο της εποχικότητας. Παρόλα αυτά, θα εξετάσουμε και τα 2 είδη της μεθόδου Holt-winters (αθροιστικής εποχικότητας και πολλαπλασιαστικής εποχικότητας) για να δούμε ποιο εφαρμόζει καλύτερα. Για κάθε είδος, θα εξετάσουμε 2 μοντέλα. Στο πρώτο μοντέλο, η αρχικοποίηση γίνεται με τον εναλλακτικό τρόπο που έχουμε αναφέρει. Δηλαδή η εκτίμηση της κλίσης και του επιπέδου γίνεται με χρήση των ελαχίστων τετραγώνων και με χρήση της μεθόδου αποσύνθεσης η αρχική εκτίμηση για τους δείκτες εποχικότητας.<sup>18</sup> Στο δεύτερο μοντέλο, η αρχικοποίηση γίνεται με τη διαδικασία που περιγράψαμε παραπάνω

<sup>18</sup> Είναι η αυτόματη διαδικασία εκτίμησης παραμέτρων στην R

κάνοντας χρήση των σχέσεων (4.18)-(4.20) για την περίπτωση της αθροιστικής εποχικότητας και των αντίστοιχων εξισώσεων για την περίπτωση της πολλαπλασιαστικής εποχικότητας.

### Περίπτωση 1: Αθροιστική Εποχικότητα

Στην παρακάτω εικόνα βλέπουμε γραφικά την απόδοση των 2 μοντέλων.



Εικόνα 4.2-13:Εφαρμογή μεθόδου Holt-Winters σε δεδομένα πωλήσεων για διάφορα initial values στην περίπτωση αθροιστικής εποχικότητας

Έχουμε χωρίσει τα δεδομένα σε 2 τμήματα. Στο αρχικό, που αποτελείται από 134 παρατηρήσεις, γίνεται η εφαρμογή του μοντέλου και καθορίζονται οι τιμές των παραμέτρων. Στο δεύτερο τμήμα, που αποτελεί το test set (34 παρατηρήσεις), ελέγχεται η απόδοση του κάθε μοντέλου σύμφωνα με τα σφάλματα πρόβλεψης. Τα στοιχεία με την μπλε γραμμή είναι οι προβλέψεις για κάθε μέθοδο και τα στοιχεία με την κόκκινη γραμμή είναι οι παρατηρήσεις από το test set που ξεχωρίσαμε. Η μωβ περιοχή δείχνει τα 80% και 95% διαστήματα εμπιστοσύνης αντίστοιχα. Η πρώτη γραφική παράσταση αφορά το μοντέλο που χρησιμοποιήθηκαν η μέθοδος ελαχίστων τετραγώνων και η μέθοδος αποσύνθεσης για τον καθορισμό των αρχικών τιμών. Η δεύτερη γραφική παράσταση αφορά το μοντέλο με αρχικές τιμές όπως προκύπτουν με χρήση των σχέσεων 4.18, 4.19, 4.25.

Τα αποτελέσματα των σφαλμάτων πρόβλεψης φαίνονται στον παρακάτω πίνακα:

	ME	RMSE	MAE	MPE	MAPE	ACF1	Theil's U
Model_1	-241.39	765.10	627.91	-18.64	31.33	0.81	2.35
Model_2	-80.22	665.29	557.91	-10.45	26.56	0.78	1.91

Πίνακας 4.2-11: Σφάλματα πρόβλεψης για δοκιμές μοντέλων της μεθόδου Holt-Winters στην περίπτωση αθροιστικής εποχικότητας

Έχοντας ως μέτρο σύγκρισης το RMSE, βλέπουμε ότι καλύτερα αποτελέσματα δίνει το δεύτερο μοντέλο. Οπότε η βέλτιστη επιλογή είναι:

---

## Holt-Winters exponential smoothing with trend and additive seasonal component.

### Call:

HoltWinters(x = y\_i, l.start = l[12], b.start = b[12], s.start = S[1:12])

### Smoothing parameters:

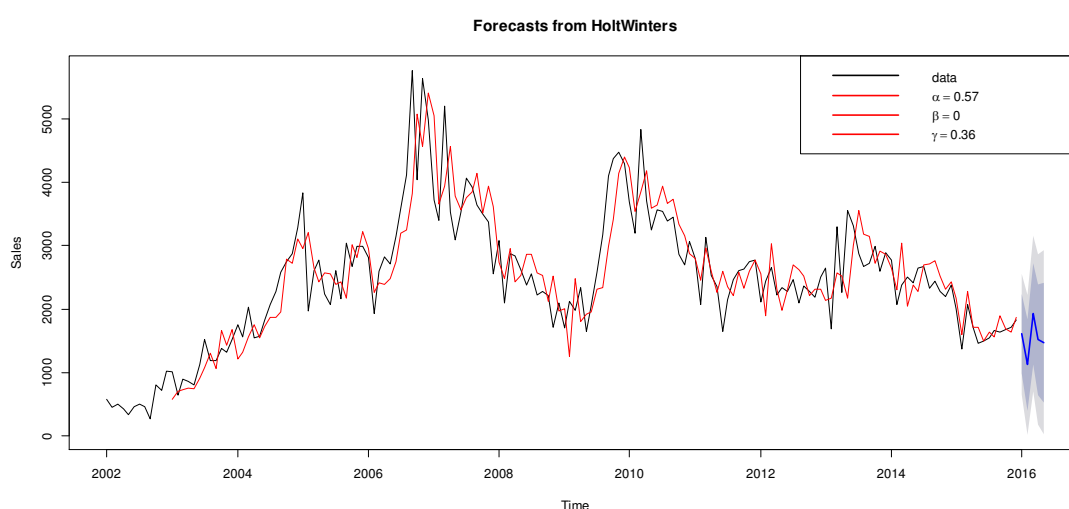
alpha: 0.5665786

beta : 0

gamma: 0.3859994

---

Πίνακας 4.2-12: Παράμετροι μοντέλου με την εφαρμογή της μεθόδου Holt-Winters σε δεδομένα πωλήσεων για την περίπτωση αθροιστικής εποχικότητας

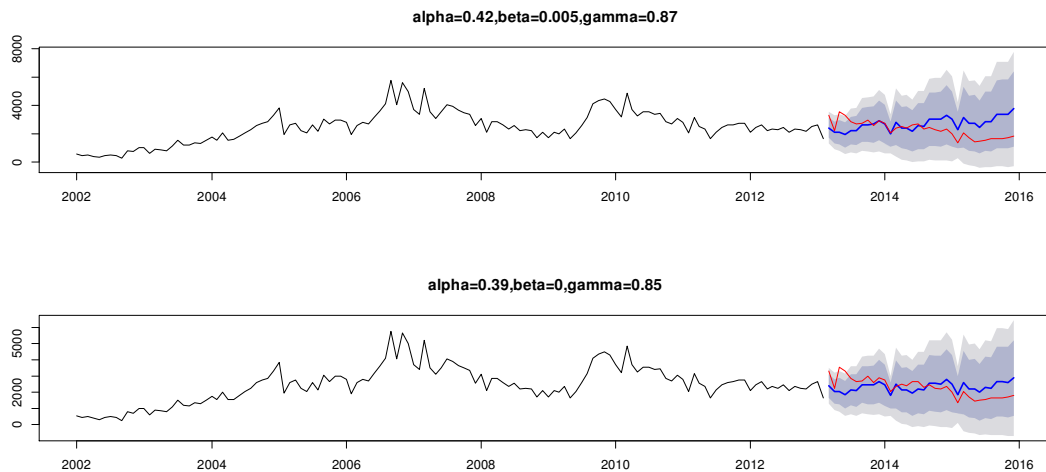


Εικόνα 4.2-14: Γραφική αναπαράσταση των προβλέψεων με εφαρμογή της μεθόδου Holt-Winters σε δεδομένα πωλήσεων για τη βέλτιστη επιλογή της μεθόδου στην περίπτωση πολλαπλασιαστικής εποχικότητας

Στην Εικόνα 4.2-14: Γραφική αναπαράσταση των προβλέψεων με εφαρμογή της μεθόδου Holt-Winters σε δεδομένα πωλήσεων για τη βέλτιστη επιλογή της μεθόδου στην περίπτωση πολλαπλασιαστικής εποχικότητας αποτυπώνεται η πρόβλεψη 5 περιόδων της βέλτιστης μεθόδου για την περίπτωση της αθροιστικής εποχικότητας μαζί με τα 80% και 95% διαστήματα εμπιστοσύνης. Με την κόκκινη γραμμή βλέπουμε την εφαρμογή του μοντέλου πρόβλεψης που επιλέξαμε.

### Περίπτωση 2: Πολλαπλασιαστική εποχικότητα

Ακολουθούμε την ίδια διαδικασία με την πρώτη περίπτωση. Στην παρακάτω εικόνα βλέπουμε γραφικά την απόδοση των 2 μοντέλων.



Εικόνα 4.2-15: Εφαρμογή μεθόδου Holt-Winters σε δεδομένα πωλήσεων για διάφορα initial values στην περίπτωση πολλαπλασιαστικής εποχικότητας

Και σε αυτή την περίπτωση, για να είναι δυνατή και η σύγκριση, χωρίσαμε τα δεδομένα σε 2 τμήματα. Στο αρχικό, που αποτελείται από 134 παρατηρήσεις, γίνεται η εφαρμογή του μοντέλου και καθορίζονται οι τιμές των παραμέτρων. Στο δεύτερο τμήμα, που αποτελεί το test set (34 παρατηρήσεις), ελέγχεται η απόδοση του κάθε μοντέλου σύμφωνα με τα σφάλματα πρόβλεψης. Τα στοιχεία με την μπλε γραμμή είναι οι προβλέψεις για κάθε μέθοδο και τα στοιχεία με την κόκκινη γραμμή είναι οι παρατηρήσεις από το test set που ξεχωρίσαμε. Η μωβ περιοχή δείχνει τα 80% και 95% διαστήματα εμπιστοσύνης αντίστοιχα. Η πρώτη γραφική παράσταση αφορά το μοντέλο που χρησιμοποιήθηκαν η μέθοδος ελαχίστων τετραγώνων και η μέθοδος αποσύνθεσης για τον καθορισμό των αρχικών τιμών. Η δεύτερη γραφική παράσταση αφορά το μοντέλο με αρχικές τιμές όπως προκύπτουν με χρήση των σχέσεων 4.18, 4.19, 4.20.

Τα αποτελέσματα των σφαλμάτων πρόβλεψης φαίνονται στον παρακάτω πίνακα:

	ME	RMSE	MAE	MPE	MAPE	ACF1	Theil's U
Model_1	-402.83	956.59	765.71	-27.16	38.94	0.83	2.96
Model_2	-25.99	658.48	558.04	-7.83	26.06	0.79	1.79

Πίνακας 4.2-13: Σφάλματα πρόβλεψης για δοκιμές μοντέλων της μεθόδου Holt-Winters στην περίπτωση πολλαπλασιαστικής εποχικότητας

Εδώ όποιο μέτρο σύγκρισης για τα σφάλματα και να πάρουμε θα διαπιστώσουμε ότι το δεύτερο μοντέλο είναι καλύτερο από το πρώτο. Οπότε η βέλτιστη επιλογή είναι:

**Holt-Winters exponential smoothing with trend and multiplicative seasonal component.**

**Call:**

**HoltWinters(x = y\_i, l.start = l[12], b.start = b[12], s.start = S[1:12])**

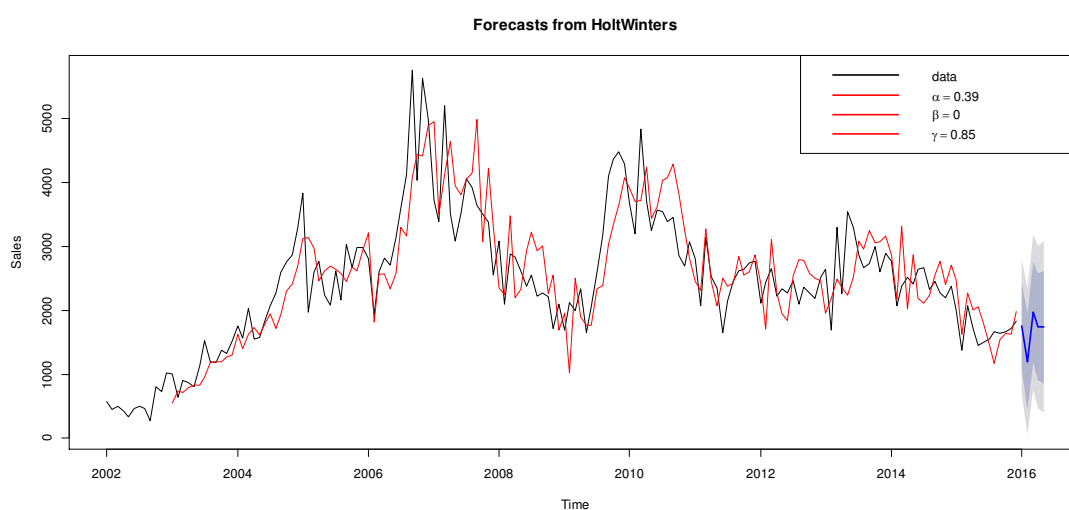
**Smoothing parameters:**

**alpha: 0.3914628**

**beta : 0**

**gamma: 0.8528836**

Πίνακας 4.2-14: Παράμετροι μοντέλου με την εφαρμογή της μεθόδου Holt-Winters σε δεδομένα πωλήσεων για την περίπτωση πολλαπλασιαστικής εποχικότητας



Εικόνα 4.2-16: Γραφική αναπαράσταση των προβλέψεων με εφαρμογή της μεθόδου Holt-Winters σε δεδομένα πωλήσεων για τη βέλτιστη επιλογή της μεθόδου στην περίπτωση πολλαπλασιαστικής εποχικότητας

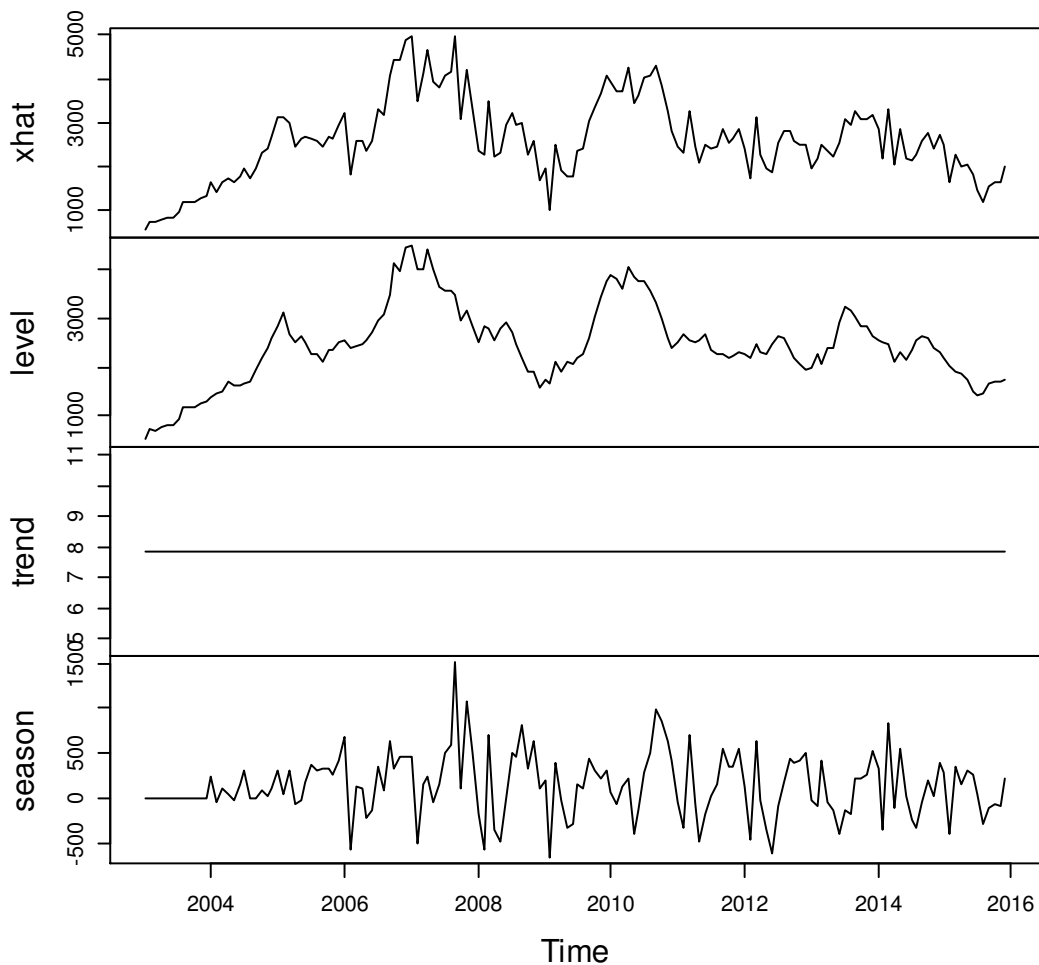
Στην Εικόνα 4.2-16 αποτυπώνεται η πρόβλεψη 5 περιόδων της βέλτιστης μεθόδου για την περίπτωση της αθροιστικής εποχικότητας (μπλε γραμμή) μαζί με τα 80% και 95% διαστήματα εμπιστοσύνης. Με την κόκκινη γραμμή βλέπουμε την εφαρμογή του μοντέλου πρόβλεψης που επιλέξαμε.

Ας δούμε τώρα τον πίνακα των σφαλμάτων για κάθε ένα από τα βέλτιστα μοντέλα που επιλέξαμε:

	ME	RMSE	MAE	MPE	MAPE	ACF1	Theil's U
Additive	-80.22	665.29	557.91	-10.45	26.56	0.78	1.91
Multiplicative	-25.99	658.48	558.04	-7.83	26.06	0.79	1.79

Σε γενικές γραμμές το πολλαπλασιαστικό μοντέλο φαίνεται να έχει καλύτερη απόδοση ως προς τις προβλέψεις. Στην παρακάτω εικόνα φαίνονται και οι εκτιμήσεις που κάνει για κάθε συνιστώσα.

## Estimation of time series components



Εικόνα 4.2-17: Εκτίμηση των συνιστωσών της χρονοσειράς των πωλήσεων για την περίπτωση της πολλαπλασιαστικής εποχικότητας με τη μέθοδο Holt-Winters

### 4.3 Ανάλυση Arima

Η συντομογραφία ARIMA χρησιμοποιείται για να περιγράψει τα ολοκληρωμένα αυτοπαλίνδρομα υποδείγματα κινητού μέσου όρου (Autoregressive/ Integrated/ Moving Average - ARIMA). Το AR αναφέρεται στην αυτοπαλινδρόμηση (autoregression), το I αναφέρεται στο διαφορισμό των δεδομένων (integrated), και το MA αναφέρεται στον κινητό μέσο (moving average). Είναι γραμμικά στατιστικά μοντέλα που χρησιμοποιούνται για την περιγραφή της διαχρονικής εξέλιξης ενός φυσικού μεγέθους. Το όνομα των μοντέλων αυτών είναι άρρηκτα συνδεδεμένο με τα άτομα που τα έκαναν διάσημα, George Box και Gwilym Jenkins. Η ανάπτυξη και η κατασκευή υποδειγμάτων ARIMA είναι γνωστή ως μεθοδολογία Box-Jenkins. Στην ενότητα αυτή θα δούμε πως μπορούμε να τα χρησιμοποιήσουμε για τη δημιουργία προβλέψεων.

### 4.3.1 Συσχέτιση στις χρονοσειρές

Πριν προχωρήσουμε στην άμεση περιγραφή των μοντέλων ARIMA χρειάζεται να δούμε κάποιες τεχνικές που μας βοηθούν να αναλύσουμε τα δεδομένα και να αποκτήσουμε μια εικόνα για το πιο μπορεί να είναι ένα κατάλληλο μοντέλο πρόβλεψης.

#### 4.3.1.1 Συνάρτηση Αυτοσυσχέτισης (ACF)

Όπως είδαμε και στη ενότητα 3.2 ένας πολύ χρήσιμος στατιστικός δείκτης είναι ο συντελεστής αυτοσυσχέτισης. Για μια μεταβλητή  $Y$ , η σχέση που περιγράφει το συντελεστή για τιμές της μεταβλητής με χρονική στέρηση  $k$  είναι:

$$r_k = \frac{\sum_{t=k+1}^n (Y_t - \bar{Y}) * (Y_{t-k} - \bar{Y})}{\sum_{t=1}^n (Y_t - \bar{Y})^2}$$

Στην περίπτωση των χρονοσειρών, περιγράφει τη γραμμική συσχέτιση μεταξύ των παρατηρήσεων,  $Y_i$ , της χρονοσειράς με χρονική υστέρηση 1, 2 ή περισσότερων περιόδων. Ο συντελεστής  $r_1$  υποδεικνύει το πως σχετίζονται μεταξύ τους οι διαδοχικές παρατηρήσεις  $Y_i$  μιας χρονοσειράς. Αντίστοιχα, ο συντελεστής  $r_2$  υποδεικνύει το πώς σχετίζονται μεταξύ τους οι παρατηρήσεις με 2 περιόδους διαφορά. Αν συγκεντρώσουμε αυτές τις τιμές για  $k=1,2,\dots$  και κάνουμε τη γραφική τους παράσταση τότε σχηματίζεται μια συνάρτηση που καλείται συνάρτηση αυτοσυσχέτισης (ACF).

Ας δούμε ένα παράδειγμα. Έστω ότι έχουμε μια χρονοσειρά με τιμές για 24 περιόδους όπως φαίνεται στον Πίνακα 4.3-1: Χρονοσειρά με 26 τιμές<sup>19</sup>. Αν οι τιμές δε σχετίζονται μεταξύ τους, περιμένουμε κάθε συντελεστής αυτοσυσχέτισης να είναι κοντά στο 0.

Περίοδος	Τιμή	Περίοδος	Τιμή	Περίοδος	Τιμή
1	78	9	25	17	33
2	78	10	35	18	63
3	79	11	12	19	22
4	15	12	80	20	48
5	24	13	14	21	80
6	96	14	54	22	84
7	55	15	58	23	17
8	4	16	35	24	53

Πίνακας 4.3-1: Χρονοσειρά με 26 τιμές

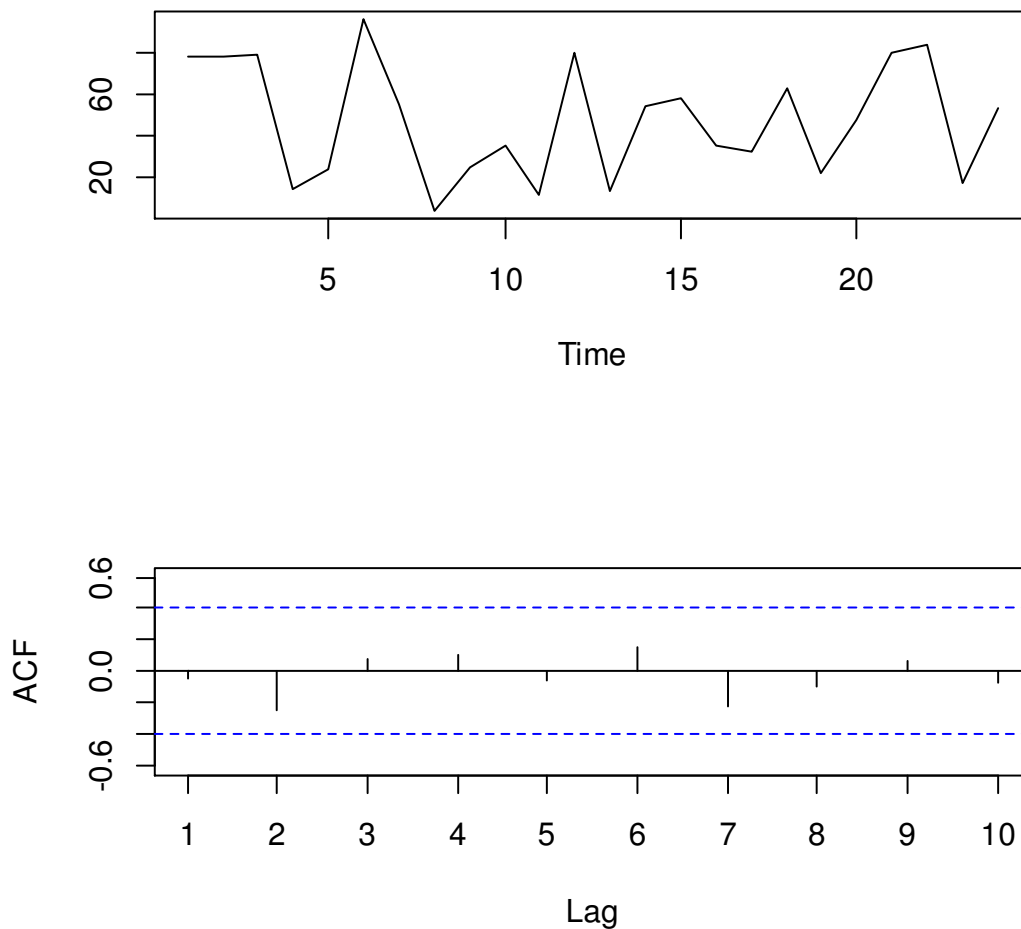
<sup>19</sup> Η χρονοσειρά κατασκευάστηκε επιλέγοντας τυχαία 24 τιμές στο διάστημα (0:100)



Οι τιμές του συντελεστή αυτοσυσχέτισης για χρονική υστέρηση  $k=1,2,\dots,10$  περιόδων είναι:

$r_1 = -0.04642$	$r_2 = -0.25336$	$r_3 = 0.073728$	$r_4 = 0.098995$	$r_5 = -0.06723$
$r_6 = 0.152819$	$r_7 = -0.23133$	$r_8 = -0.09636$	$r_9 = 0.067522$	$r_{10} = -0.07571$

Στην εικόνα 4.3-1 μπορούμε να δούμε το γράφημα της χρονοσειράς καθώς και το γράφημα της συνάρτησης αυτοσυσχέτισης για  $k=1,2,\dots,10$ .



Εικόνα 4.3-1: Συνάρτηση αυτοσυσχέτισης και διάγραμμα τιμών μιας χρονοσειράς

Για να καταλάβουμε καλύτερα το θεωρητικό κομμάτι γύρω από τους συντελεστές αυτοσυσχέτισης θα εξετάσουμε την ειδική περίπτωση χρονοσειράς που περιέχει μόνο τυχαίες τιμές και περιγράφεται από το μοντέλο λευκού θορύβου (*white noise*). Το μοντέλο αυτό προσδιορίζει κάθε παρατήρηση  $Y_t$  με 2 τιμές: μία σταθερά  $c$  και ένα τυχαίο σφάλμα  $e_t$  (τα σφάλματα δεν παρουσιάζουν συσχέτιση από περίοδο σε περίοδο). Η σχέση που το περιγράφει είναι :

$$Y_t = c + e_t$$

Το μοντέλο λευκού θορύβου αποτελεί πρότυπο για τα σφάλματα πρόβλεψης κάθε μοντέλου που έχει καλή εφαρμογή. Για να δούμε αν το μοντέλο που έχει επιλεγεί ως το καλύτερο είναι όντως καλό, αρκεί να κοιτάξουμε τη συνάρτηση αυτοσυσχέτισης για τα σφάλματα πρόβλεψης και να παρατηρήσουμε αν υπάρχει κάποιο μοτίβο. Αν τα σφάλματα είναι τυχαία, δηλαδή ακολουθούν το μοντέλο του λευκού θορύβου, τότε οι τιμές των συντελεστών αυτοσυσχέτισης θα είναι όλες κοντά στο μηδέν και αυτό αποτελεί ένδειξη ενός καλού μοντέλου. Αν όμως διακρίνουμε για κάποιο/ους συντελεστή/ες σχετικά μεγάλη τιμή τότε ενδέχεται να υπάρχει κάποιο μοτίβο στα δεδομένα και σίγουρα το μοντέλο χρειάζεται βελτίωση.

Θεωρητικά όλοι οι συντελεστές αυτοσυσχέτισης για τυχαίες τιμές θα πρέπει να είναι μηδέν. Γενικά όμως, παρατηρείται ότι οι τιμές των συντελεστών δεν είναι ποτέ μηδέν. Έχει δειχθεί από τους Anderson (1942), Barlett (1946), Quenouille (1949) και άλλους, ότι οι συντελεστές αυτοσυσχέτισης για τιμές που ακολουθούν το μοντέλο του λευκού θορύβου, ακολουθούν κατανομή που μπορεί να προσεγγιστεί από μια κανονική κατανομή με μέση τιμή μηδέν και διασπορά  $1/\sqrt{n}$ , όπου  $n$  ο αριθμός των παρατηρήσεων. Επομένως, για μια τυχαία χρονοσειρά, το 95% των συντελεστών αυτοσυσχέτισης θα πρέπει να βρίσκονται σε ένα εύρος γύρω από τη μέση τιμή  $\pm 1.96/\sqrt{n}$ . Αν δε συμβαίνει αυτό τότε η χρονοσειρά δεν είναι λευκός θόρυβος. Γι' αυτό το λόγο στα γραφήματα της ACF υπάρχουν 2 γραμμές πάνω και κάτω από τη μέση τιμή. Είναι οι τιμές  $\pm 1.96/\sqrt{n}$  και είναι γνωστές ως *critical values*.

Ας δούμε το παράδειγμα με την χρονοσειρά στον πίνακα 4.3-1. Το μέγεθος των παρατηρήσεων είναι  $n=24$ , οπότε τα *critical values* θα έχουν τιμή 0.4. Αυτό φαίνεται και στην εικόνα 4.3-1. Παρατηρούμε ότι όλοι οι συντελεστές είναι μέσα στα όρια που ορίζουν τα *critical values* επιβεβαιώνοντας ότι πρόκειται για δεδομένα που ακολουθούν το πρότυπο του λευκού θορύβου. Βλέπουμε ότι η συνάρτηση αυτοσυσχέτισης αποτελεί ένα από τα βασικά εργαλεία της ανάλυσης χρονοσειρών και παίζει σημαντικό ρόλο στην αξιολόγηση και επιλογή των μοντέλων πρόβλεψης.

### 4.3.1.2 Portmanteau tests

Ένας εναλλακτικός τρόπος ελέγχου της τυχειότητας είναι να ελέγξουμε από κοινού όλους τους συντελεστές συσχέτισης. Δηλαδή να φτιάξουμε ένα σύνολο που περιέχει πολλές τιμές  $r_k$  και να εξετάσουμε αν το σύνολο αυτό διαφέρει στατιστικά σημαντικά από ένα μηδενικό σύνολο. Οπότε αντί να κοιτάμε την τυχειότητα ξεχωριστά, στις διάφορες χρονικές υστερήσεις  $k$ , εξετάζουμε αν σαν σύνολο έχει τυχειότητα. Τα κριτήρια που βασίζονται σε αυτή τη λογική καλούνται *portmanteau tests*. Τέτοια κριτήρια είναι των Box-Pierce (1970) και Ljung-Box (1978).

Ο έλεγχος γίνεται κάτω από την αρχική υπόθεση:

$$H_0: r_1 = r_2 = \dots = r_h = 0$$

Έναντι της εναλλακτικής:

$$H_1: \text{όχι όλα τα } r_i = 0$$

### BOX-PIERCE

Η στατιστική των Box-Pierce (*Box-Pierce Q statistic*) ορίζεται ως εξής:

$$Q = n \sum_{k=1}^h r_k^2 \quad (4.26)$$

Όπου  $h$  ο μέγιστος αριθμός χρονικής υστέρησης που λαμβάνουμε υπόψη και  $n$  ο αριθμός των παρατηρήσεων. Συνήθως επιλέγεται  $h \approx 20$ .

Σύμφωνα με τη σχέση 4.26, αν κάθε  $r_k$  είναι μικρό με τιμή κοντά στο 0, τότε η τιμή του  $Q$  θα είναι πολύ μικρή. Στην περίπτωση όμως που μερικά  $r_k$  είναι σχετικά μεγάλα (είτε θετικά, είτε αρνητικά) τότε και η τιμή του  $Q$  θα είναι μεγάλη. Το κριτήριο αυτό, σχεδιάστηκε το 1970 για τον έλεγχο των υπολοίπων στα μοντέλα πρόβλεψης. Αν τα υπόλοιπα είναι λευκός θόρυβος, τότε το στατιστικό  $Q$  ακολουθεί την κατανομή  $\chi^2$  με  $(h-m)$  βαθμούς ελευθερίας, όπου  $m$  οι αριθμοί των παραμέτρων του μοντέλου πρόβλεψης.

Στο παράδειγμα του πίνακα 4.3-1 η τιμή της στατιστικής των Box-Pierce είναι:

$$Q = 24 \sum_{k=1}^{10} r_k^2 = 4.381$$

Για να δούμε πόσο αξιόπιστο είναι κοιτάμε την τιμή  $p$ -value. Η τιμή αυτή αντιπροσωπεύει έναν δείκτη της αξιοπιστίας ενός αποτελέσματος ή αλλιώς την πιθανότητα λάθους της αρχικής υπόθεσης. Συνήθως η αρχική υπόθεση απορρίπτεται όταν  $p$ -value < 0.05. Στο παράδειγμά μας, προέκυψε  $p$ -value = 0.9285 πολύ μεγαλύτερο της τιμής 0.05. Αυτό πρακτικά σημαίνει ότι η αρχική μας υπόθεση ( $H_0$ ) έχει πιθανότητα 93% να μην ισχύει. Άρα έχουμε ισχυρές ενδείξεις ότι τα δεδομένα είναι λευκός θόρυβος.

### LJUNGBOX

Ένα εναλλακτικό κριτήριο είναι αυτό των Ljung-Box (1978) (*Ljung-Box Q\* statistic*). Η στατιστική των Ljung-Box αν και ακολουθεί την ίδια κατανομή  $\chi^2$ , με αυτή της  $Q$  δίνει καλύτερα αποτελέσματα όταν εφαρμόζεται σε μικρά δείγματα και ορίζεται ως εξής:

$$Q^* = n(n+2) \sum_{k=1}^h (n-k)^{-1} r_k^2 \quad (4.27)$$

Στο παράδειγμα του πίνακα 4.3-1 το αντίστοιχο στατιστικό  $Q^*$  είναι:

$$Q^* = 24 * (26) \sum_{k=1}^{10} \frac{1}{24-k} r_k^2 = 6.0759$$

Η τιμή  $p$ -value προκύπτει να είναι 0.81, πολύ μεγαλύτερη από 0.05. Άρα και μέσω αυτού του κριτηρίου πιστοποιείται η ένδειξη ότι οι παρατηρήσεις της χρονοσειράς είναι λευκός θόρυβος.

Φυσικά τα κριτήρια είναι για να μας δίνουν τις απαραίτητες ενδείξεις και δεν αποτελούν σε καμία περίπτωση αποκλειστικό κριτήριο για την επιλογή ή όχι ενός μοντέλου πρόβλεψης.

### 4.3.1.3 Συνάρτηση μερικής αυτοσυσχέτισης (PACF)

Οι συντελεστές μερικής αυτοσυσχέτισης (*partial autocorrelation coefficients*) χρησιμοποιούνται για να μετρήσουν το βαθμό συσχέτισης μεταξύ των  $Y_t$  και  $Y_{t-k}$  όταν οι επιδράσεις από όλες τις υπόλοιπες χρονικές υστερήσεις (1,2,...,k-1) έχουν αφαιρεθεί.

Θα δούμε πιο αναλυτικά πως γίνεται αυτό. Ας υποθέσουμε ότι υπάρχει μια σημαντική συσχέτιση μεταξύ των  $Y_t$  και  $Y_{t-1}$ . Τότε θα υπάρχει επίσης ισχυρή συσχέτιση μεταξύ των  $Y_{t-1}$  και  $Y_{t-2}$ , αφού και αυτές οι παρατηρήσεις έχουν την ίδια χρονική υστέρηση. Κατά συνέπεια θα υπάρχει συσχέτιση και μεταξύ των  $Y_t$  και  $Y_{t-2}$  γιατί και οι δύο σχετίζονται με την  $Y_{t-1}$ . Οπότε για να μετρήσουμε τη σχέση που έχουν οι  $Y_t$  και  $Y_{t-2}$  πρέπει να αφαιρέσουμε την επίδραση της  $Y_{t-1}$ .

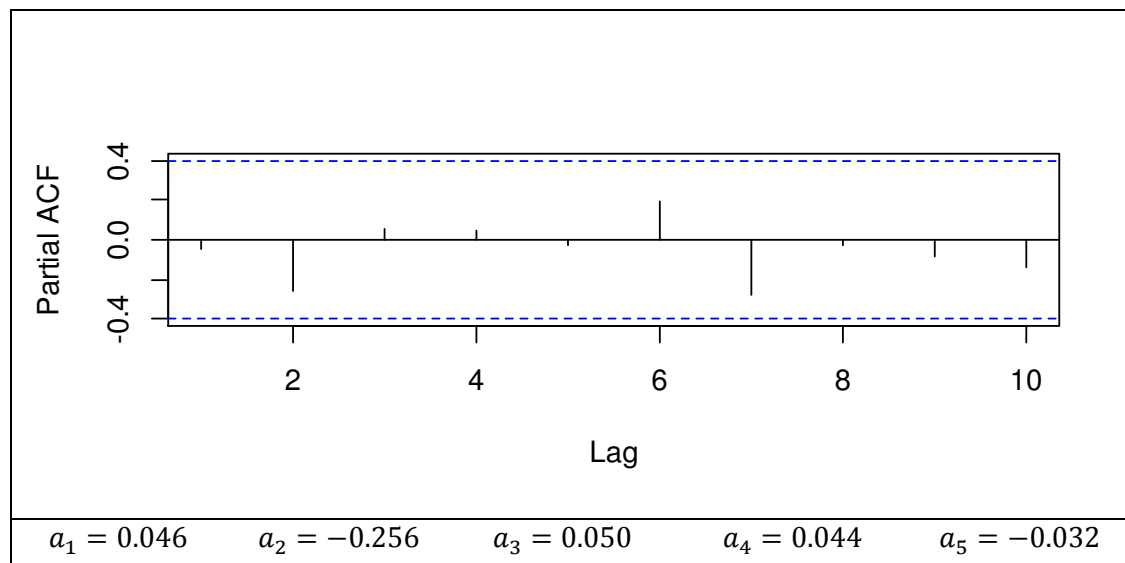
Ο συντελεστής μερικής αυτοσυσχέτισης τάξης k συμβολίζεται με  $a_k$  και μπορεί να υπολογισθεί εφαρμόζοντας τη μέθοδο πολλαπλής γραμμικής παλινδρόμησης έχοντας ως ανεξάρτητη μεταβλητή την  $Y_t$  και εξαρτημένες τις  $Y_{t-1}, \dots, Y_{t-k}$ :

$$Y_t = b_0 + b_1 Y_{t-1} + b_2 Y_{t-2} + \dots + b_k Y_{t-k}$$

Ο συντελεστής μερικής αυτοσυσχέτισης  $a_k$  είναι ίδιος με τον εκτιμημένο συντελεστή  $b_k$  και ο πρώτος συντελεστής μερικής αυτοσυσχέτισης,  $a_1$ , είναι πάντα ίσος με τον πρώτο συντελεστή αυτοσυσχέτισης  $r_1$ .

Όπως και με την ACF, οι μερικές αυτοσυσχετίσεις πρέπει να είναι κοντά στο μηδέν αν τα δεδομένα ακολουθούν το μοτίβο του λευκού θορύβου. Και οι κρίσιμες τιμές (critical values) είναι και αυτές  $\pm 1.96/\sqrt{n}$ .

Ας δούμε μια εφαρμογή στο παράδειγμα της χρονοσειράς του πίνακα 4.3-1.



$$a_6 = 0.193 \quad a_7 = -0.279 \quad a_8 = -0.024 \quad a_9 = -0.082 \quad a_{10} = -0.136$$

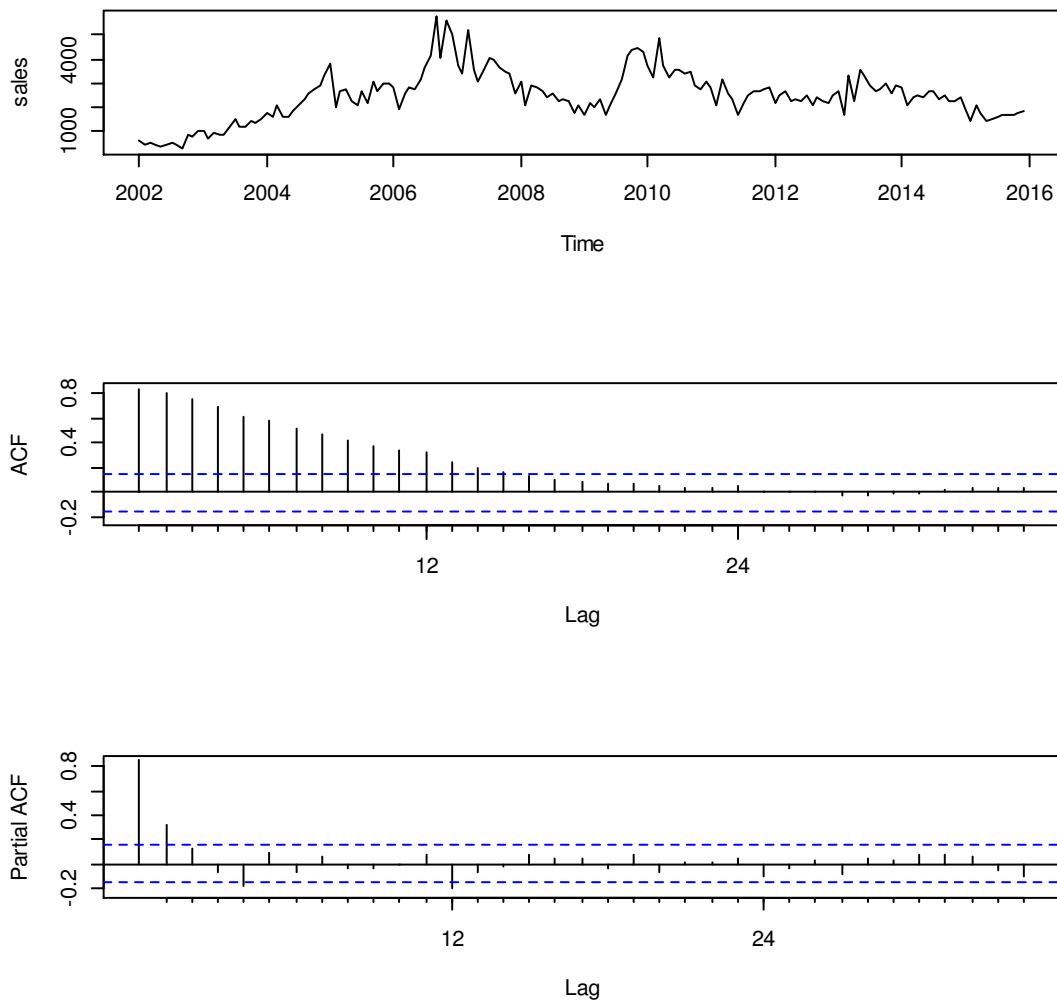
Εικόνα 4.3-2: Συντελεστές μερικής αυτοσυσχέτισης για τη χρονοσειρά του πίνακα 4.3-1

#### 4.3.1.4 Εντοπισμός εποχικότητας στις χρονοσειρές

Η εποχικότητα είναι ένα μοτίβο που επαναλαμβάνεται με σταθερή περίοδο. Αν λοιπόν έχουμε δεδομένα που παρουσιάζουν εποχικότητα τότε το διάγραμμα ACF θα έχει μια μεγάλη θετική τιμή σε κάθε περίοδο επανεμφάνισης. Για παράδειγμα αν το φαινόμενο γίνεται κάθε 12 μήνες τότε θα πρέπει να παρουσιάζονται υψηλές τιμές στους δείκτες αυτοσυσχέτισης με χρονική υστέρηση 12 μηνών.

Αν τα δεδομένα παρουσιάζουν και άλλο μοτίβο τότε είναι πιο δύσκολο να αναγνωριστεί η εποχικότητα μέσω της συνάρτησης αυτοσυσχέτισης. Γι' αυτό πρέπει πρώτα τα δεδομένα να μετασχηματιστούν σε σταθερή σειρά και στη συνέχεια να εξεταστεί η παρουσία εποχικότητας.

Ας δούμε τα δεδομένα των πωλήσεων που αναλύσαμε στην προηγούμενη ενότητα.



Εικόνα 4.3-3: Συναρτήσεις αυτοσυσχέτισης και μερικής αυτοσυσχέτισης για δεδομένα πωλήσεων για χρονικές υστερήσεις  $k=1,2,\dots,35$

Από τις συναρτήσεις αυτοσυσχέτισης και μερικής αυτοσυσχέτισης μπορούμε να σίγουρα να καταλάβουμε ότι η χρονοσειρά δεν αποτελείται από τυχαίες παρατηρήσεις. Δε μπορούμε όμως να βγάλουμε κάποιο συμπέρασμα για την εποχικότητα. Εκ πρώτης όψεως φαίνεται να μην υπάρχει κάτι έντονο που να μας υποδεικνύει την εποχικότητα. Αυτό όμως θα φανεί καλύτερα όταν μετασχηματίσουμε τα δεδομένα.

#### 4.3.1.5 Εξέταση στασιμότητας σε δεδομένα χρονοσειράς

Η στασιμότητα (*stationarity*) σε μια χρονοσειρά σημαίνει ότι δεν υπάρχουν στοιχεία ανάπτυξης ή απόκλισης. Δηλαδή τα δεδομένα πρέπει να κυμαίνονται σταθερά γύρω από το μέσο (ή το επίπεδο) της χρονοσειράς και η διασπορά των διακυμάνσεων να είναι σχετικά σταθερή με την πάροδο του χρόνου. Αν η χρονοσειρά είναι μη σταθερή (*non-stationary*), τότε θα ακολουθεί κάποιο πρότυπο. Στην

περίπτωση που υπάρχει μεταβολή στο επίπεδο τότε λέμε ότι παρουσιάζει τάση στο μέσο όρο και στην περίπτωση που η μεταβλητότητα αλλάζει διαχρονικά τότε λέμε ότι παρουσιάζει τάση στη διακύμανση.

Αυτό μπορεί να γίνει εύκολα αντιληπτό με μια γραφική παράσταση των δεδομένων (time plot). Επίσης η διαφορά στη μέση τιμή (δηλαδή η εμφάνιση τάσης) μπορεί να φανεί και από τη συνάρτηση αυτοσυσχέτισης. Στην περίπτωση μη σταθερής χρονοσειράς, οι συντελεστές αυτοσυσχέτισης των πρώτων χρονικών υστερήσεων είναι στατιστικά διάφοροι του μηδενός και σταδιακά πέφτουν στο μηδέν ή δείχνουν ένα πλασματικό πρότυπο καθώς αυξάνεται ο αριθμός των χρονικών περιόδων.

Στην εικόνα 4.3-3 αν κοιτάξουμε την πρώτη γραφική παράσταση είναι εύκολα αντιληπτό ότι η χρονοσειρά δεν είναι στάσιμη. Ακόμα αν δούμε και τη συνάρτηση αυτοσυσχέτισης παρατηρείται το πρότυπο που περιγράψαμε παραπάνω. Ξεκινάει με στατιστικά σημαντικούς συντελεστές αυτοσυσχέτισης που φθίνουν. Ο συντελεστής αυτοσυσχέτισης για χρονική υστέρηση μιας περιόδου,  $r_1$ , είναι αρκετά μεγάλος και θετικός. Ο συντελεστής αυτοσυσχέτισης για χρονική υστέρηση 2 περιόδων,  $r_2$ , είναι επίσης μεγάλος αλλά πιο μικρός από τον  $r_1$ , επειδή χρησιμοποιείται ένας όρος λιγότερος στον υπολογισμό του συντελεστή αυτοσυσχέτισης με αποτέλεσμα να υποεκτιμείται το  $r_2$  και το στατιστικό σφάλμα να αυξάνει. Η συνάρτηση μερικής αυτοσυσχέτισης φαίνεται και αυτή να ακολουθεί την τυπική μορφή για μη στάσιμες χρονοσειρές ξεκινώντας με μια πολύ μεγάλη τιμή κοντά στο 1 για τη χρονική υστέρηση μιας περιόδου.

Οι περισσότερες χρονοσειρές δεν έχουν χαρακτήρα στάσιμων διαδικασιών. Μπορούν όμως να μετατραπούν σε στάσιμες με τις διαδικασίες που αναφέρουμε παρακάτω. Ο λόγος που χρειάζεται να γίνει η μετατροπή είναι για να αποφύγουμε το πρόβλημα της φαινομενικής ή νόθου παλινδρόμησης.

#### ΕΞΑΛΕΙΨΗ ΤΗΣ ΤΑΣΗΣ ΑΠΟ ΤΑ ΔΕΔΟΜΕΝΑ

Οι πλασματικές συσχετίσεις που εμφανίζονται, είναι αποτέλεσμα της ύπαρξης τάσης στα δεδομένα. Γι' αυτό η τάση πρέπει να εξαλειφθεί, ώστε να μπορέσουμε να προχωρήσουμε σε περαιτέρω ανάλυση της χρονοσειράς. Ένας τρόπος να γίνει αυτό είναι με τη χρήση της μεθόδου διαφορίσης (differencing).

Κάθε σειρά διαφορών ορίζεται ως η διαφορά μεταξύ 2 παρατηρήσεων. Η σειρά διαφορών πρώτης τάξης είναι της μορφής:

$$Y'_t = Y_t - Y_{t-1}$$

Οι σειρές διαφορών πρώτης τάξης έχουν n-1 τιμές, όπου ποι παρατηρήσεις της χρονοσειράς. Αν η τάση στα αρχικά δεδομένα είναι γραμμική τότε η νέα σειρά  $Y'_t$  θα είναι στάσιμη. Στην περίπτωση που δεν είναι στάσιμη, τότε εφαρμόζουμε πάλι τη μέθοδο διαφορών και προκύπτει η σειρά διαφορών δεύτερης τάξης:

$$Y''_t = Y'_t - Y'_{t-1} = (Y_t - Y_{t-1}) - (Y_{t-1} - Y_{t-2}) = Y_t - 2Y_{t-1} + Y_{t-2}$$

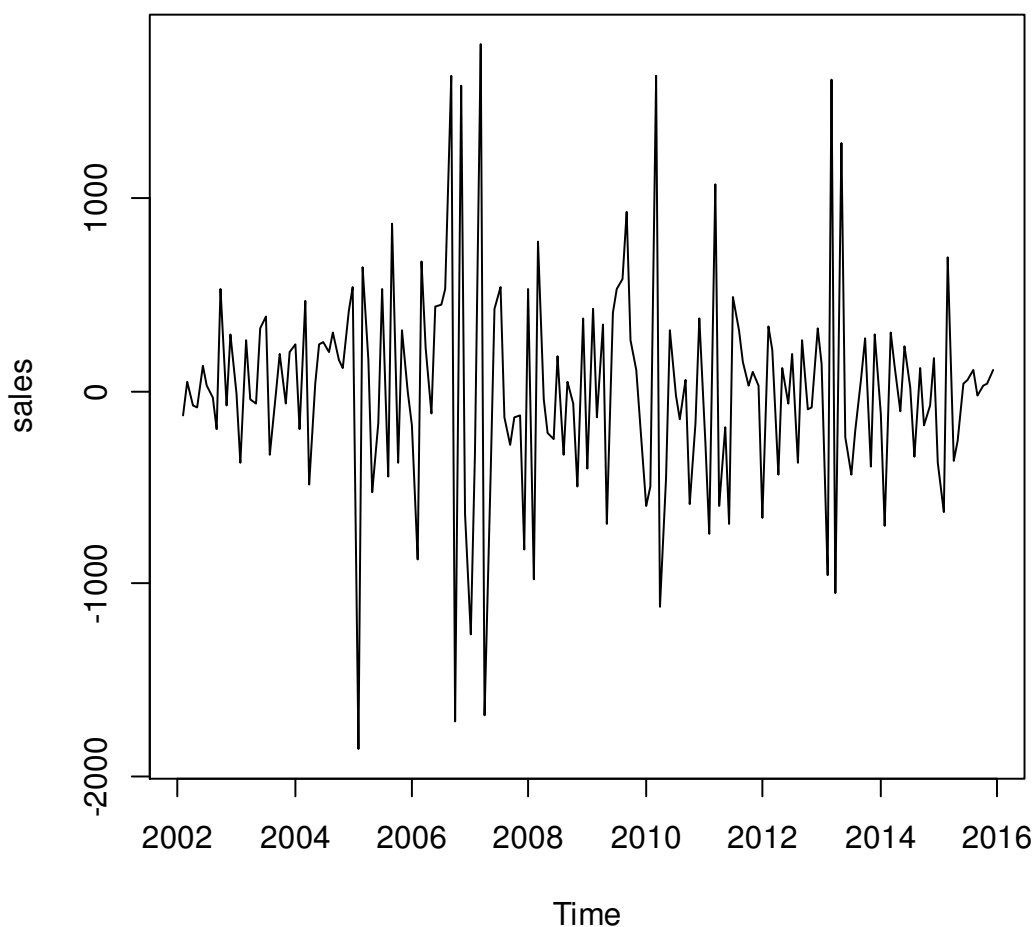
Αυτή η σειρά θα έχει n-2 τιμές.

Ένα πολύ γνωστό μοντέλο μη σταθερής χρονοσειράς είναι αυτό του τυχαίου περιπάτου (*random walk*):

$$Y_t = Y_{t-1} + e_t$$

Χρησιμοποιείται συχνά στην ανάλυση οικονομικών και τιμών μετοχής σειρών και ο λόγος είναι γιατί παρουσιάζει μεγάλες περιόδους τάσης (αυξητικής ή πτωτικής) που μπορούν να αλλάξουν κατεύθυνση απρόβλεπτα.

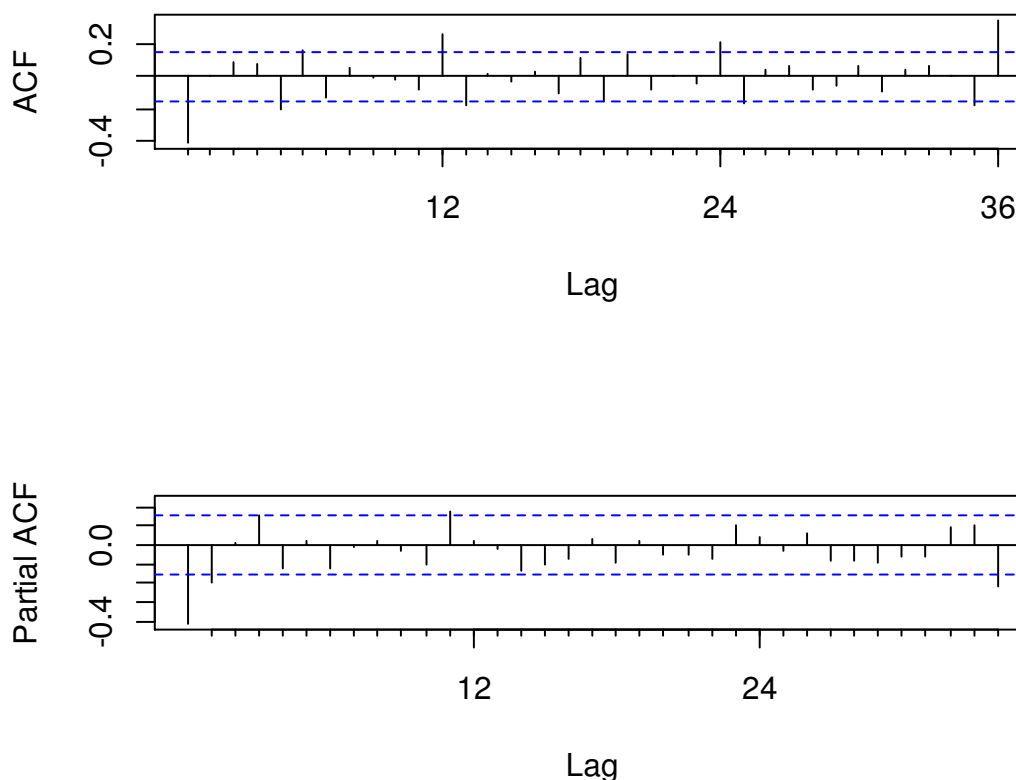
Ας δούμε πάλι το παράδειγμα με τα δεδομένα των πωλήσεων. Εφαρμόζοντας τις διαφορές πρώτης τάξης προκύπτει η εικόνα:



Εικόνα 4.3-4: Δεδομένα πωλήσεων μετά την εφαρμογή σειράς διαφορών πρώτης τάξης

Σε επόμενο βήμα θα αναλύσουμε τις συναρτήσεις αυτοσυσχέτισης και μερικής αυτοσυσχέτισης.





Εικόνα 4.3-5: Συναρτήσεις συσχέτισης και μερικής αυτοσυσχέτισης της πρώτης τάξης σειράς διαφορών των δεδομένων πωλήσεων

Από τα γραφήματα στις εικόνες 4.3-4 και 4.3-5 φαίνεται ότι η σειρά είναι στάσιμη. Ωστόσο παρατηρείται ένα μοτίβο εποχικότητας με περίοδο ένα έτος.

#### ΕΞΑΛΕΙΨΗ ΤΗΣ ΕΠΟΧΙΚΟΤΗΤΑΣ

Για εποχιακά δεδομένα που δεν είναι στάσιμα, απαιτείται η εξάλειψη του εποχικού προτύπου για την περαιτέρω ανάλυση της χρονοσειράς. Αυτό επιτυγχάνεται με την εφαρμογή εποχιακών διαφορών. Μια εποχιακή διαφορά είναι η διαφορά ανάμεσα σε μια παρατήρηση και την αντίστοιχη παρατήρηση του προηγούμενου έτους. Αν έχουμε εποχιακά δεδομένα μηνιαίων παρατηρήσεων τότε:

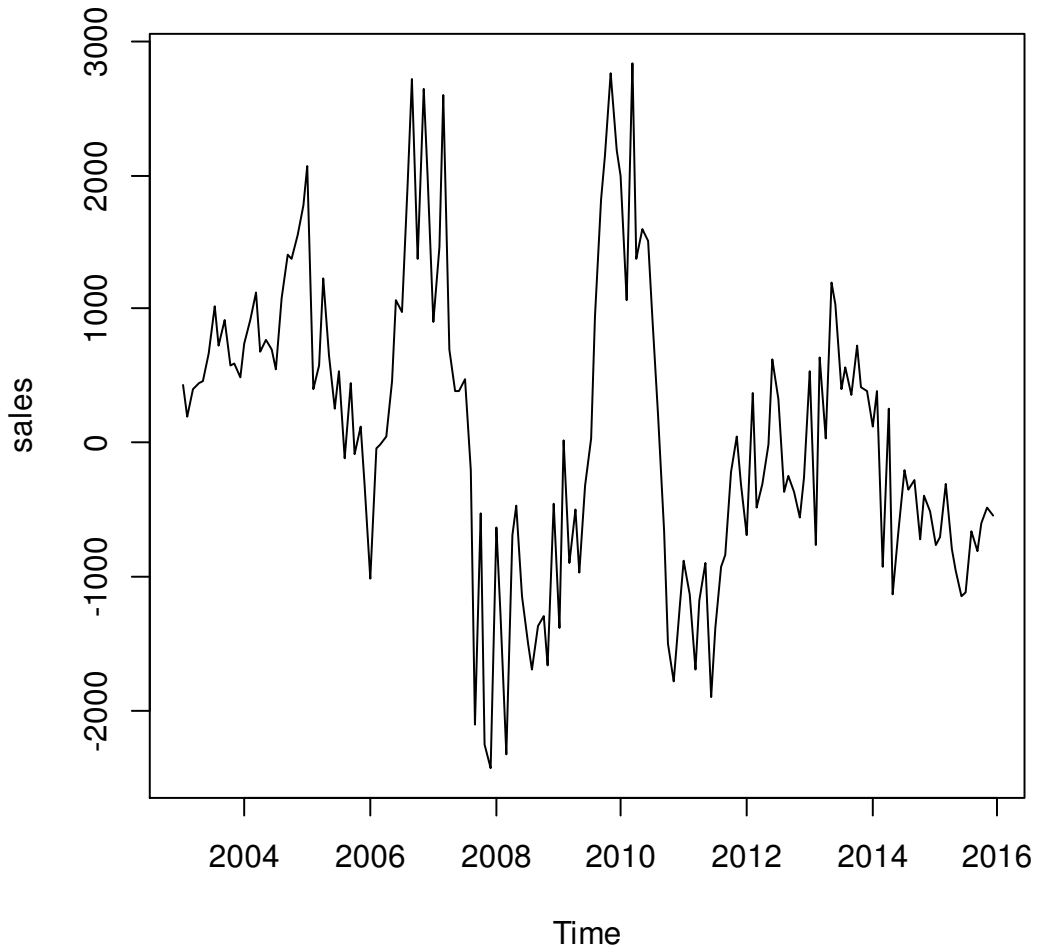
$$Y'_t = Y_t - Y_{t-12}$$

Γενικά οι εποχιακές διαφορές πρώτης τάξης ορίζονται ως:

$$Y'_t = Y_t - Y_{t-s}$$

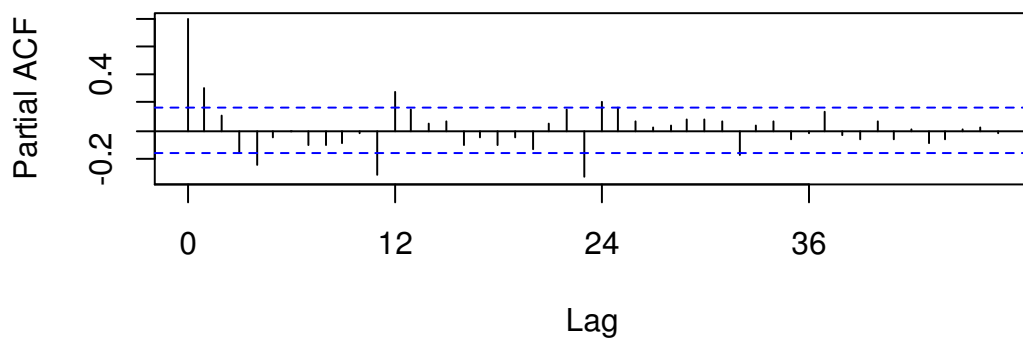
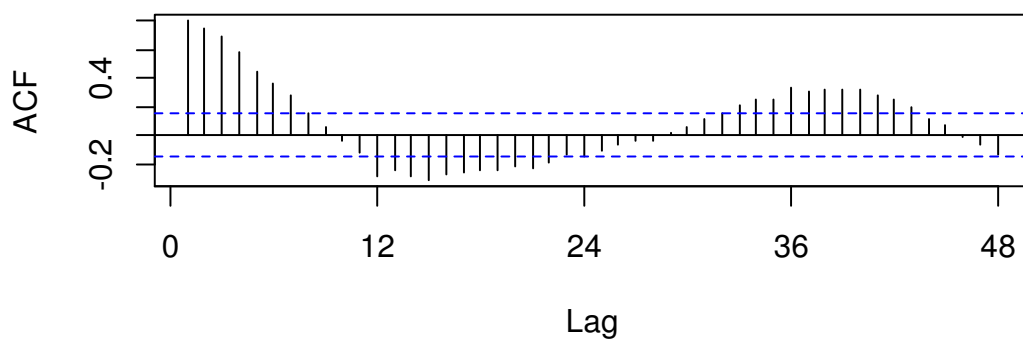
Όπου  $s$  είναι ο αριθμός των περιόδων σε ένα έτος. Για τριμηνιαία δεδομένα,  $s=4$ , για τετραμηνιαία  $s=3$  κλπ. Για εποχιακά δεδομένα συνηθίζεται να γίνεται πρώτα η εφαρμογή των εποχιακών διαφορών γιατί μπορεί η σειρά να παρουσιάσει στασιμότητα και να μη χρειαστεί και η εφαρμογή των διαφορών πρώτης τάξης.

Ας πάρουμε πάλι το παράδειγμα των πωλήσεων εφαρμόζοντας πρώτα τις εποχικές διαφορές. Τα αποτελέσματα φαίνονται στο παρακάτω γράφημα:



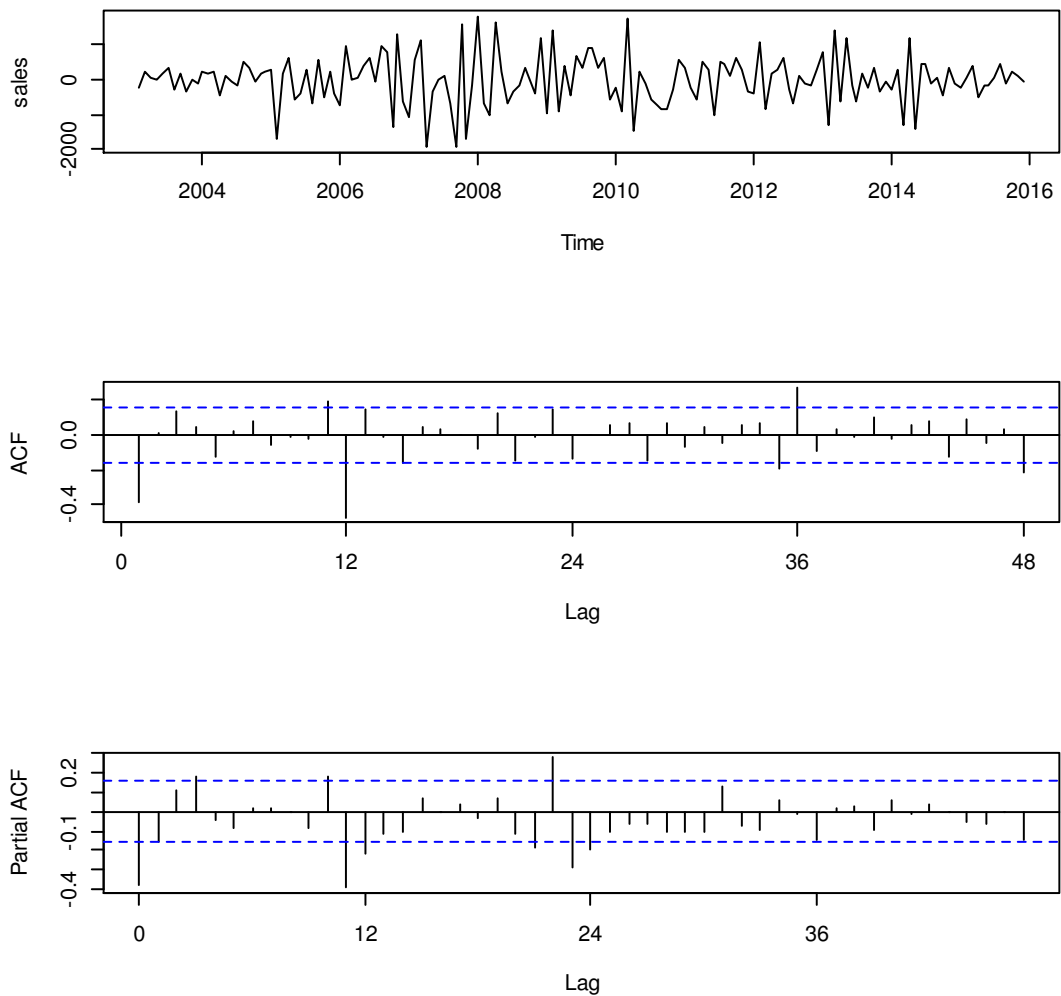
Εικόνα 4.3-6: Δεδομένα πωλήσεων μετά την εφαρμογή εποχιακών διαφορών

Και τα αντίστοιχα γραφήματα των συναρτήσεων αυτοσυσχέτισης και μερικής αυτοσυσχέτισης είναι:



Εικόνα 4.3-7: Συναρτήσεις συσχέτισης και μερικής αυτοσυσχέτισης της πρώτης τάξης σειράς εποχιακών διαφορών των δεδομένων πωλήσεων ( $s=12$ )

Η σειρά δεν είναι ακόμα στάσιμη. Παρόλα αυτά φαίνεται η εποχικότητα.



Εικόνα 4.3-8: Δεδομένα πωλήσεων μετά την εφαρμογή εποχιακών διαφορών και διαφορών πρώτης τάξης

### ΤΕΛΕΣΤΗΣ ΟΛΙΣΘΗΣΗΣ Β

Η έκφραση και η κατανόηση των μοντέλων ARIMA μπορεί να γίνει πιο εύκολη με τη χρήση ενός τελεστή ολίσθησης B που ορίζεται ως εξής:

$$BY_t = Y_{t-1}$$

Η δράση του B πάνω στο  $Y_t$ , έχει ως αποτέλεσμα τη μετατόπιση των δεδομένων μια περίοδο πίσω. Η δράση του B στο  $Y_t$  δεύτερη φορά μετατοπίζει τα δεδομένα 2 περιόδους πίσω:

$$B(BY_t) = B^2Y_t = Y_{t-2}$$

Αν η δράση του B στο  $Y_t$  γίνει m φορές τότε ο γενικός τύπος είναι:

$$B^m Y_t = Y_{t-m}$$

Για μηνιαία δεδομένα αν θέλουμε να δώσουμε προσοχή στο ίδιο μήνα αλλά το προηγούμενο έτος η περιγραφή γίνεται:  $B^{12}Y_t = Y_{t-12}$

Με τον τελεστή ολίσθησης  $B$  γίνεται εύκολη η περιγραφή των μεθόδων διαφορών. Μια διαφορά πρώτης τάξης μπορεί να γραφεί:

$$Y'_t = Y_t - Y_{t-1} = Y_t - BY_t = (1 - B)Y_t$$

Για δεύτερης τάξης διαφορές:

$$\begin{aligned} Y''_t &= Y'_t - Y'_{t-1} = (Y_t - Y_{t-1}) - (Y_{t-1} - Y_{t-2}) = (1 - 2B + B^2)Y_t \\ &= (1 - B)^2 Y_t \end{aligned}$$

Γενικότερα οι διαφορές  $d$ -τάξης μπορούν να περιγραφούν ως  $(1 - B)^d Y_t$ .

Η εποχιακή διαφορά πρώτης τάξης δεδομένων με μηνιαία εποχικότητα μπορεί να περιγραφεί ως:

$$Y'_t = Y_t - Y_{t-12} = Y_t - B^{12}Y_t = (1 - B^{12})Y_t$$

Η εποχιακή διαφορά πρώτης τάξης για  $s$  περιόδους που ακολουθείται από διαφορά πρώτης τάξης γράφεται:  $(1 - B)(1 - B^s)Y_t$

### 4.3.2 Μοντέλα ARIMA για δεδομένα χρονοσειρών

Πριν προχωρήσουμε στα μοντέλα ARIMA θα μας χρειαστεί μια μικρή αναφορά στα μοντέλα παλινδρόμησης.

Οι μέθοδοι παλινδρόμησης εξετάζουν τη συσχέτιση που υπάρχει μεταξύ μιας εξαρτημένης μεταβλητής  $Y$  και πολλών ανεξάρτητων μεταβλητών  $X_i$ . Αυτές οι ανεξάρτητες μεταβλητές είναι παράγοντες που υποθέτουμε ότι μπορεί να επηρεάζουν τη διακύμανση στην ανεξάρτητη μεταβλητή  $Y$ . Ένα μοντέλο παλινδρόμησης είναι μια γραμμική συνάρτηση που έχει την ακόλουθη μορφή:

$$Y = b_0 + b_1 X_1 + \dots + b_p X_p + e \quad (4.28)$$

Ας υποθέσουμε τώρα ότι για κάθε  $X_i$  ισχύει:  $X_1 = Y_{t-1}$ ,  $X_2 = Y_{t-2}$ , ...,  $X_p = Y_{t-p}$ , η εξίσωση 4.28 γίνεται:

$$Y_t = b_0 + b_1 Y_{t-1} + \dots + b_p Y_{t-p} + e_t \quad (4.29)$$

Η εξίσωση 4.29 που προέκυψε συνεχίζει να είναι μια εξίσωση μοντέλου παλινδρόμησης με τη διαφορά ότι οι μεταβλητές στη δεξιά μεριά (επεξηγηματικές μεταβλητές) είναι παλαιότερες παρατηρήσεις της εξαρτημένης μεταβλητής  $Y_t$ . Η παλινδρόμηση αυτής της μορφής καλείται αυτοπαλινδρόμηση (*AR - autoregression*).

Η ειδική μεταχείριση τέτοιων μοντέλων οφείλεται στο ότι είναι δύσκολο να προσδιοριστεί ο ακριβής αριθμός των παλαιών τιμών της μεταβλητής  $Y_t$  που θα συμπεριληφθούν στο μοντέλο και στο ότι παραβιάζεται πολύ συχνά η υπόθεση ανεξαρτησίας των υπολοίπων (σφάλματα πρόβλεψης).

Υπάρχουν υποδείγματα (μοντέλα) χρονοσειρών που έχουν για επεξηγηματικές μεταβλητές τα παλαιότερα σφάλματα:

$$Y_t = b_0 + b_1 e_{t-1} + \dots + b_q e_{t-q} + e_t \quad (4.30)$$

Τέτοια υποδείγματα καλούνται κινητοί μέσοι όροι (MA). Εδώ δεν πρέπει να συγχέουμε την έννοια κινητός μέσος όρος με την έννοια που είχε η φράση στις προηγούμενες ενότητες. Εκεί είχαμε το μέσο όρο των παρατηρήσεων

Όταν τα υποδείγματα αυτοπαλινδρόμησης (AR) συνδέονται με τα υποδείγματα κινητού μέσου όρου (MA), τότε προκύπτουν τα υποδείγματα αυτοπαλινδρόμησης κινητού μέσου όρου (ARMA) τα οποία μπορούν να χρησιμοποιηθούν μόνο όταν τα δεδομένα είναι στάσιμα. Η επέκταση αυτών των υποδειγμάτων που επιτρέπει και τις διαφορίσεις είναι τα ολοκληρωμένα αυτοπαλίνδρομα υποδείγματα κινητού μέσου όρου (ARIMA) (Box and Jenkins (1970)).

Η ποικιλία των υποδειγμάτων ARIMA προκύπτει από το γενικό μη-εποχικό υπόδειγμα ARIMA(p,d,q):

AR:  $p =$  τάξη του αυτοπαλίνδρομου υποδείγματος

I:  $d =$  τάξης της διαφορίσης για την επιτευξη σταθερότητας

MA:  $q =$  τάξη του υποδείγματος κινητού μέσου όρου

Το μοντέλο του λευκού θορύβου μπορεί οριστεί ως υπόδειγμα ARIMA(0,0,0) γιατί δεν υπάρχει τίποτα από τα p, d, q. Είναι μια τυχαία χρονοσειρά. Από την άλλη το μοντέλο του τυχαίου περιπάτου ορίζεται ως ARIMA(0,1,0).

### 4.3.2.1 Αυτοπαλίνδρομα υποδείγματα

Η γενική μορφή ενός αυτοπαλίνδρομου υποδείγματος p-τάξης, AR(p) ή ARIMA(p,0,0), ορίζεται ως:

$$Y_t = c + \varphi_1 Y_{t-1} + \varphi_2 Y_{t-2} + \dots + \varphi_p Y_{t-p} + e_t \quad (4.31)$$

όπου c σταθερός όρος,  $\varphi_i$  οι παράμετροι και  $e_t$  ο όρος του σφάλματος τη χρονική στιγμή t. Οι παράμετροι του υποδείγματος υπόκεινται στους παρακάτω περιορισμούς:

για p=1 :  $-1 < \varphi_1 < 1$

για p=2 :  $-1 < \varphi_2 < 1$  ,  $\varphi_1 + \varphi_2 < 1$  ,  $\varphi_2 - \varphi_1 < 1$

Για μεγαλύτερες τιμές του p γίνονται πιο περίπλοκοι.

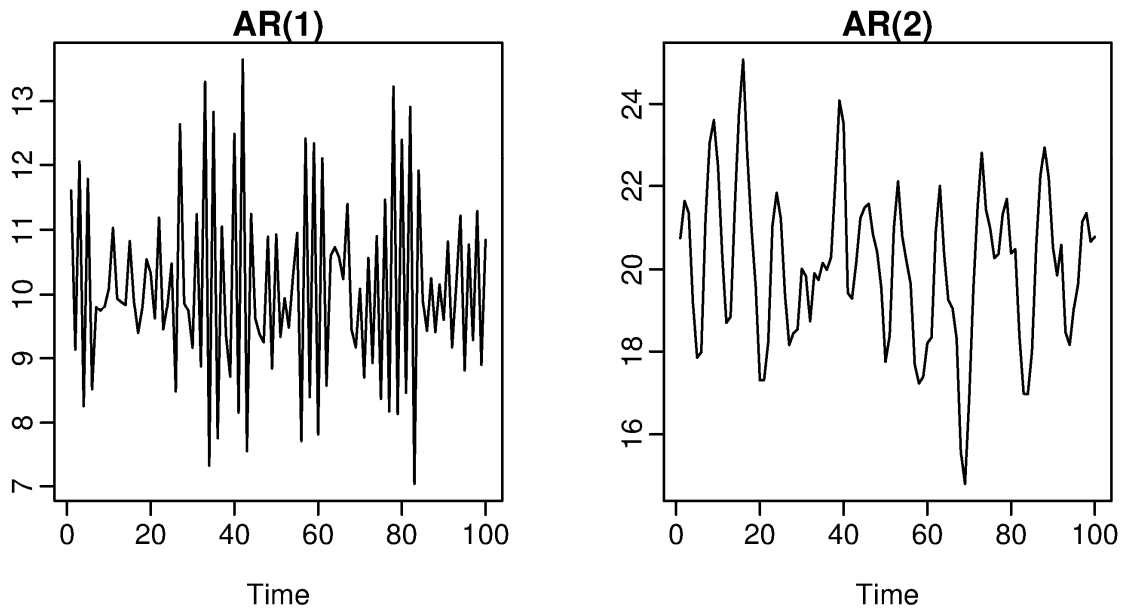
Αν η τάξη του αυτοπαλίνδρομου υποδείγματος είναι p=1 τότε συμβολίζεται ως AR(1) ή ARIMA(1,0,0) και περιγράφεται από την εξίσωση:

$$Y_t = c + \varphi_1 Y_{t-1} + e_t$$

Όταν  $\varphi_1=0$  είναι ισοδύναμο με το μοντέλο του λευκού θορύβου ενώ όταν  $\varphi_1=1$  και  $c=0$  είναι ισοδύναμο με το μοντέλο του τυχαίου περιπάτου. Όταν  $\varphi_1 < 0$ , η  $Y_t$  δείχνει να ταλαντώνεται μεταξύ θετικών και αρνητικών τιμών. Με τη χρήση του τελεστή ολίσθησης B, η εξίσωση (4.31) εκφράζεται ως:

$$Y_t - \varphi_1 Y_{t-1} - \varphi_2 Y_{t-2} - \dots - \varphi_p Y_{t-p} = c + e_t$$

$$(1 - \varphi_1 B - \varphi_2 B^2 - \dots - \varphi_p B^p) Y_t = c + e_t$$



Εικόνα 4.3-9: Δύο παραδείγματα για αυτοπαλίνδρομα υποδείγματα με διαφορετικές παραμέτρους.

Στην παραπάνω εικόνα έχουμε αριστερά: AR(1) με  $Y_t = 18 - 0.8Y_{t-1} + e_t$  και δεξιά: AR(2) με  $Y_t = 8 + 1.3Y_{t-1} - Y_{t-2} + e_t$ . Και στις 2 περιπτώσεις τα σφάλματα  $e_t$  ακολουθούν κανονική κατανομή με μέση τιμή 0 και διασπορά 1.

Μπορούμε να χρησιμοποιήσουμε τα διαγράμματα ACF και PACF για να διακρίνουμε το είδος υποδείματος ARIMA. Για τα αυτοπαλίνδρομα υποδείγματα AR(p) ισχύει ότι:

- i. Οι συντελεστές αυτοσυσχέτισης φθίνουν εκθετικά στο μηδέν
- ii. Υπάρχουν ακριβώς p στατιστικά σημαντικοί συντελεστές μερικής αυτοσυσχέτισης.

Αν  $p \geq 2$ , το ACF μπορεί να εμφανίσει οι συντελεστές εξομάλυνσης να φθίνουν εκθετικά ή να φθίνουν με ημιτονοειδή τρόπο

### 4.3.2.2 Υποδείγματα κινητού μέσου όρου

Η γενική μορφή ενός υποδείματος μέσου όρου q-τάξης, MA(q) ή ARIMA(0,0,q), ορίζεται ως:

$$Y_t = c + e_t - \theta_1 e_{t-1} - \dots - \theta_q e_{t-q} \quad (4.32)$$

όπου c σταθερός όρος,  $\theta_i$  οι παράμετροι και  $e_t$  ο όρος του σφάλματος τη χρονική στιγμή t. Οι παράμετροι του υποδείματος υπόκεινται στους παρακάτω περιορισμούς:

για q=1 :  $-1 < \theta_1 < 1$

για q=2 :  $-1 < \theta_2 < 1$  ,  $\theta_1 + \theta_2 < 1$  ,  $\theta_2 - \theta_1 < 1$

Για μεγαλύτερες τιμές του q γίνονται πιο περίπλοκοι.

Με τη χρήση του τελεστή ολίσθησης B, η εξίσωση (4.32) εκφράζεται ως:

$$Y_t = c + (1 - \theta_1 B - \theta_2 B^2 - \dots - \theta_q B^q) e_t$$

Αν η τάξη του υποδείματος κινητού μέσου όρου είναι  $q=1$  τότε συμβολίζεται ως MA(1) ή ARIMA(0,0,1) και περιγράφεται από την εξίσωση:

$$Y_t = e_t - \theta_1 e_{t-1}$$

Αντίστοιχα με τα αυτοπαλίνδρομα υποδείγματα AR(p), για να χρησιμοποιήσουμε ένα υπόδειγμα κινητού μέσου όρου MA(q), χρειάζεται να εμφανίζονται τα εξής στα διαγράμματα ACF και PACF:

- i. Οι συντελεστές μερικής αυτοσυσχέτισης φθίνουν εκθετικά στο μηδέν
- ii. Υπάρχουν ακριβώς q στατιστικά σημαντικοί συντελεστές αυτοσυσχέτισης.

Αν  $q \geq 2$ , το PACF μπορεί να παρουσιάσει οι συντελεστές μερικής εξομάλυνσης να φθίνουν εκθετικά ή να φθίνουν με ημιτονοειδή τρόπο.

### 4.3.2.3 Υποδείγματα αυτοπαλίνδρομου κινητού μέσου όρου

Όπως αναφέρθηκε και πριν τα βασικά στοιχεία των υποδειγμάτων AR και MA μπορούν να συνδυαστούν προκύπτοντας πολλά είδη μοντέλων για την πρόβλεψη στάσιμων χρονοσειρών. Για παράδειγμα μια εξίσωση που συνδυάζει τα MA(1) και AR(1) ονομάζεται υπόδειγμα ARMA(1,1) ή ARIMA(1,0,1) και είναι η ακόλουθη:

$$Y_t = c + \varphi_1 Y_{t-1} + e_t - \theta_1 e_{t-1}$$

Στη γενική του μορφή έχουμε το αυτοπαλίνδρομο υπόδειγμα κινητού μέσου ARMA(p,q):

$$Y_t = c + \varphi_1 Y_{t-1} + \varphi_2 Y_{t-2} + \dots + \varphi_p Y_{t-p} + e_t - \theta_1 e_{t-1} - \theta_2 e_{t-2} - \dots - \theta_q e_{t-q}$$

Η με τη χρήση του τελεστή ολίσθησης B:

$$(1 - \varphi_1 B + \varphi_2 B^2 - \dots - \varphi_p B^p) Y_t = c + (1 - \theta_1 B - \theta_2 B^2 - \dots - \theta_q B^q) e_t$$

Ισχύουν και εδώ οι περιορισμοί για τις παραμέτρους ακριβώς όπως στα MA και AR.

Τα μεικτά υποδείγματα ARMA έχουν θεωρητικές ACF και με τα 2 χαρακτηριστικά AR και MA, ενώ οι PACF δεν διακόπτονται μετά από κάποιες χρονικές υστερήσεις παρά φθίνουν προς το μηδέν

### 4.3.2.4 Υποδείγματα ARIMA

Αν ένα υπόδειγμα ARMA δεν παρουσιάζει στασιμότητα, τότε παράγεται το γενικό υπόδειγμα ARIMA(p,d,q). Τα ολοκληρωμένα αυτοπαλίνδρομα υποδείγματα κινητού μέσου όρου ARIMA(p,d,q), χρησιμοποιούνται για να περιγράψουν χρονοσειρές που δεν είναι στάσιμες αλλά μπορούν να γίνουν με εφαρμογή της μεθόδου διαφορίσης.



Η εξίσωση για ένα υπόδειγμα ARIMA(1,1,1) είναι η ακόλουθη:

$$(1 - \phi_1 B)(1 - B)Y_t = c + (1 - \theta_1 B)e_t$$

Επειδή υπάρχουν πάρα πολλά υποδείγματα κάτω από αυτή τη γενική κατηγορία δεν μπορεί να υπάρξει ένας γενικός κανόνας για τα μοτίβα των ACF και PACF. Συνήθως οι τιμές που παίρνουν τα  $p, d, q$  είναι 0, 1 ή 2.

### 4.3.2.5 Εποχικά υποδείγματα ARIMA

Μέχρι στιγμής έχουμε μελετήσει μη εποχικά δεδομένα και μη εποχικά υποδείγματα ARIMA. Σε περιπτώσεις που τα δεδομένα είναι εποχικά τότε υπάρχουν και τα αντίστοιχα εποχικά υποδείγματα ARIMA που προέρχονται από τα υποδείγματα ARIMA που έχουμε δει με προσθήκη σε αυτά κάποιων επιπλέον όρων που αφορούν την εποχικότητα. Συμβολίζονται ως ARIMA  $(p,d,q)(P,D,Q)_s$ , όπου  $s$  είναι ο αριθμός περιόδων σε ένα έτος. Η πρώτη παρένθεση αναφέρεται στο μη εποχικό κομμάτι της χρονοσειράς και η δεύτερη στο εποχικό.

Ένα υπόδειγμα ARIMA(1,1,1)(1,1,1)<sub>4</sub> είναι:

$$(1 - \phi_1 B) (1 - \Phi_1 B^4) (1 - B) (1 - B^4) y_t = (1 + \theta_1 B) (1 + \Theta_1 B^4) e_t.$$

$\uparrow$   
 (Non-seasonal  
AR(1))

$\uparrow$   
 (Non-seasonal  
difference)

$\uparrow$   
 (Non-seasonal  
MA(1))

$\uparrow$   
 (Seasonal  
AR(1))

$\uparrow$   
 (Seasonal  
difference)

$\uparrow$   
 (Seasonal  
MA(1))

Το κομμάτι της εποχικότητας θα φανεί στις χρονικές υστερήσεις που αντιστοιχούν στην εποχικότητα των δεδομένων μέσα από τα διαγράμματα των ACF και PACF.

### 4.3.3 Μεθοδολογία Box-Jenkins

Η μεθοδολογία Box-Jenkins για την ανάλυση χρονολογικών σειρών αναφέρεται στη μέθοδο εύρεσης ενός στατιστικού υποδείγματος ARIMA που να περιγράφει ικανοποιητικά τη στοχαστική διαδικασία από την οποία προήλθαν τα δεδομένα. Περιλαμβάνει τέσσερα στάδια: Την ταυτοποίηση (identification), την εκτίμηση (estimation), το διαγνωστικό έλεγχο, (diagnostic checking) και την πρόβλεψη (forecasting).

#### 4.3.3.1 Ταυτοποίηση (Identification)

Λόγω της μεγάλης ποικιλίας υποδειγμάτων ARIMA, είναι πολύ δύσκολο να καταλήξουμε στο καταλληλότερο υπόδειγμα για τα δεδομένα που εξετάζουμε. Υπάρχουν όμως κάποια βήματα που διευκολύνουν την επιλογή.

Παρατηρούμε τα δεδομένα εξετάζοντας αν υπάρχουν ασυνήθιστες παρατηρήσεις. Επίσης εφαρμόζουμε όποια τροποποίηση χρειάζεται (αν χρειάζεται) για να σταθεροποιήσουμε τη διασπορά των παρατηρήσεων (πολύ χρήσιμος είναι ο μετασχηματισμός Box-Cox).

#### ΒΗΜΑ 2 – ΕΞΕΤΑΣΗ ΣΤΑΣΙΜΟΤΗΤΑΣ

Μέσα από το διάγραμμα των δεδομένων και τις γραφικές παραστάσεις των ACF και PACF, αναζητούμε ενδείξεις στασιμότητας της χρονοσειράς.

#### ΒΗΜΑ 3 – ΜΕΤΑΤΡΟΠΗ ΜΗ ΣΤΑΣΙΜΩΝ ΔΕΔΟΜΕΝΩΝ ΣΕ ΣΤΑΣΙΜΑ

Αν τα δεδομένα δεν είναι στάσιμα τα μετατρέπουμε με χρήση της διαφορίσης. Για μη εποχικά δεδομένα εφαρμόζουμε τις σειρές διαφορών πρώτης τάξης. Για εποχικά δεδομένα εφαρμόζουμε τις εποχικές σειρές διαφορών. Αν δεν είναι στάσιμα εφαρμόζουμε πάλι τις διαφορές πρώτης τάξης. Συνήθως δυο εφαρμογές είναι αρκετές για να μετατρέψουν τα δεδομένα σε στάσιμα.

#### ΒΗΜΑ 4 – ΕΞΕΤΑΣΗ ΑΥΤΟΣΥΣΧΕΤΙΣΗΣ

Αφού έχει επιτευχθεί η στασιμότητα, ελέγχουμε αν υπάρχουν άλλα μοτίβα στα δεδομένα με τους συντελεστές αυτοσυσχέτισης. Υπάρχουν 3 πιθανά ενδεχόμενα:

- Ένδειξη εποχικότητας. Δηλαδή οι αυτοσυσχετίσεις και ίσως οι μερικές αυτοσυσχετίσεις παρουσιάζουν στατιστικά σημαντικούς συντελεστές.
- Εμφάνιση υποδειγμάτων AR ή MA. Δηλαδή το μοτίβο των συναρτήσεων ACF ή PACF υποδεικνύει κάποιο από τα 2 υποδείγματα. Αν δεν υπάρχουν στατιστικά σημαντικοί συντελεστές συσχέτισης μετά από  $q$  χρονικές υστερήσεις, τότε είναι πιθανό να είναι κατάλληλο ένα υπόδειγμα MA( $q$ ). Αν δεν υπάρχουν στατιστικά σημαντικοί συντελεστές μερικής συσχέτισης μετά από  $p$  χρονικές υστερήσεις, τότε είναι πιθανό να είναι κατάλληλο ένα υπόδειγμα AR( $p$ ).
- Αν δεν υπάρχει ισχυρή ένδειξη κάποιων AR ή MA τότε πιθανόν να είναι απαραίτητος κάποιος συνδυασμός τους.

Από την παραπάνω διαδικασία παράγεται ένα δοκιμαστικό υπόδειγμα ARIMA.

### 4.3.3.2 Εκτίμηση (Estimation)

Αφού έχει προσδιοριστεί το κατάλληλο υπόδειγμα και η τάξη του, σειρά έχει η εκτίμηση των παραμέτρων  $\varphi_1, \varphi_2, \dots, \varphi_p$  της αυτοπαλίνδρομης διαδικασίας και  $\theta_1, \theta_2, \dots, \theta_q$  της διαδικασίας κινητού μέσου. Αν η σειρά είναι μόνο αυτοπαλίνδρομη τότε για τον υπολογισμό τους μπορεί να χρησιμοποιηθεί η μέθοδος ελαχίστων τετραγώνων. Αν όμως περιέχει και όρους κινητού μέσου χρειάζεται άλλη προσέγγιση μέσω μιας μη γραμμικής μεθόδου όπως η ελαχιστοποίηση του MSE ή η μέθοδος μέγιστης πιθανοφάνειας. Υπάρχουν και πολλά υπολογιστικά προγράμματα που βοηθούν μέσα από μια αυτοματοποιημένη διαδικασία που έχουν εντοπίζοντας τις βέλτιστες τιμές των παραμέτρων του υποδείγματος.

Σε αυτό το στάδιο είναι καλό να εξετάσουμε αν το υπόδειγμα που έχει εκτιμηθεί επιφέρει κάποια βελτίωση. Μπορεί κάποια παράμετρος να μην είναι στατιστικά σημαντική ή να είναι προτιμότερο ένα μεικτό υπόδειγμα ARIMA ή ακόμα και να

έχουμε να επιλέξουμε ανάμεσα σε 2 μοντέλα το καλύτερο δυνατό. Τόσο η μέθοδος μέγιστης πιθανοφάνειας όσο και η ελαχιστοποίηση του MSE δεν είναι πάντα αποτελεσματικά κριτήρια επιλογής. Πολλές φορές το MSE μπορεί να γίνει μεγάλο και η μέγιστη πιθανοφάνεια μικρή με την πρόσθεση περισσότερων όρων στο μοντέλο. Για να αντιμετωπισθεί αυτό συγκρίνεται το υπόδειγμα με άλλο μεγαλύτερης τάξης και χρησιμοποιείται το κριτήριο του Akaike (AIC) (έλεγχος τάξης του μοντέλου). Με το κριτήριο αυτό βάζουμε ένα κάτω φράγμα στη μέγιστη πιθανοφάνεια. ώστε αν προσθέτοντας έναν όρο η μέγιστη πιθανοφάνεια δε βελτιωθεί πάνω από αυτό το φράγμα, τότε δεν αξίζει να συμπεριληφθεί αυτός ο όρος στο υπόδειγμα.

Έστω  $m = p + q + P + Q$  ο αριθμός των όρων προς εκτίμηση στο υπόδειγμα. Η επιλογή των  $p, q, P, Q$  μπορεί να γίνει ελαχιστοποιώντας την παρακάτω ποσότητα (Akaike's Information Criterion):

$$AIC = -2\log L + 2m$$

όπου  $L$  η συνάρτηση πιθανοφάνειας.

#### 4.3.3.3 Διαγνωστικός έλεγχος

Στο στάδιο αυτό γίνεται ο έλεγχος της καλής προσαρμογής του υποδείματος. Δηλαδή ελέγχεται πόσο καλά ταιριάζει το εκτιμώμενο υπόδειγμα με τα δεδομένα και αν υπάρχει ενδεχομένως κάποιο άλλο που να ταιριάζει καλύτερα.

Για τον έλεγχο της καταλληλότητας θα μελετήσουμε τα υπόλοιπα (*residuals*). Τα υπόλοιπα κάθε καλού εκτιμώμενου μοντέλου ακολουθούν το μοντέλο του λευκού θορύβου. Οπότε μελετώντας τους συντελεστές αυτοσυσχέτισης και μερικής αυτοσυσχέτισης των υπολοίπων, πρέπει όλοι να βρίσκονται ανάμεσα στα *critical values*.

Ένας άλλος τρόπος για να εξετάσουμε την καλή προσαρμογή του εκτιμώμενου υποδείματος είναι χρησιμοποιώντας τη στατιστική  $Q$  των Box-Pierce.

Αν κάποια από τις παραπάνω μεθόδους δε δείξει το επιθυμητό αποτέλεσμα της καλής εφαρμογής, τότε μια ένδειξη για το ποιο θα μπορούσε να είναι κάποιο άλλο υπόδειγμα ARIMA είναι να δούμε της συναρτήσεις ACF και PACF. Αν για παράδειγμα οι τιμές που παρουσιάζουν στατιστικά σημαντικούς συντελεστές είναι στις χρονικές υστερήσεις της εποχικότητας των δεδομένων τότε καλό θα ήταν να προστεθεί η συνιστώσα της εποχικότητας στο υπόδειγμα ARIMA.

#### 4.3.3.4 Πρόβλεψη

Αφού επιλέξουμε το υπόδειγμα που περιγράφει καλύτερα τη χρονοσειρά, προχωράμε στη δημιουργία προβλέψεων.

# 5

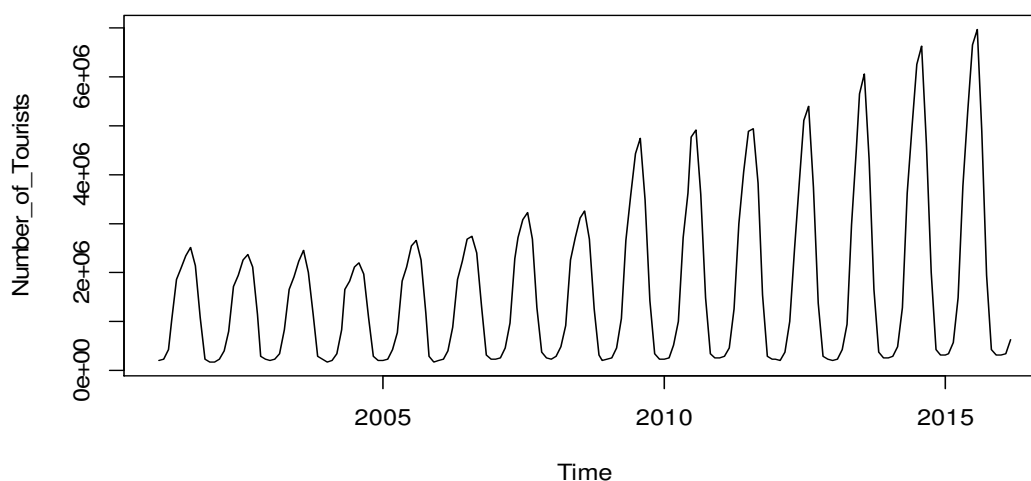
## Εφαρμογές

Στο κεφάλαιο αυτό θα ασχοληθούμε με την εφαρμογή των μεθόδων πρόβλεψης που αναφέραμε σε διάφορα δεδομένα χρονοσειρών. Συγκεκριμένα θα αναλύσουμε 3 σύνολα δεδομένων και θα προσπαθήσουμε να βρούμε την πιο αποδοτική μέθοδο πρόβλεψης. Η σύγκριση θα γίνει ανάμεσα στις μεθόδους πρόβλεψης ARIMA, Holt-Winters, STL Decomposition και Naïve.

Ο τρόπος επεξεργασίας των δεδομένων έγινε χρησιμοποιώντας το στατιστικό πακέτο R. Για την αποτελεσματική σύγκριση των μεθόδων κάθε σύνολο δεδομένων χωρίστηκε σε δύο τμήματα σε ποσοστά 80-20. 80% για το κομμάτι που θα γίνεται η εφαρμογή της μεθόδου και η εκτίμηση του μοντέλου και 20% για το κομμάτι απόδοσης της μεθόδου και παραγωγής σφαλμάτων πρόβλεψης. Η ανάλυση και η επεξεργασία με τη χρήση του στατιστικού πακέτου της R για την κάθε χρονοσειρά που μελετήθηκε βρίσκεται στο παράρτημα στο τέλος.

### 5.1 Δεδομένα τουριστών<sup>20</sup>

Η μεταβλητή που θα εξετάσουμε περιγράφει τον αριθμό των τουριστών που επισκέπτονται τις τουριστικές μονάδες της Ελλάδας το χρονικό διάστημα Ιαν2001-Μαρ2016. Το δείγμα αποτελείται από 183 μηνιαίες παρατηρήσεις<sup>21</sup>.

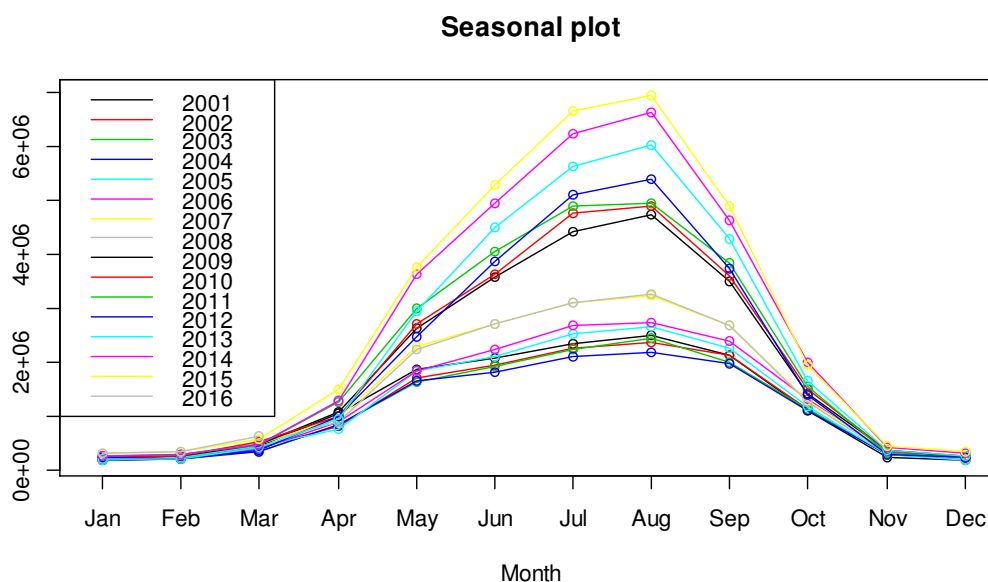


Εικόνα 5.1-1: Δεδομένα τουριστών στην Ελλάδα το χρονικό διάστημα 2001-2016

<sup>20</sup> Η ανάλυση και οι κώδικες παρατίθενται στο παράρτημα, στο αρχείο `tourists_final.r`

<sup>21</sup> Πηγή: Eurostat

Η γραφική τους παράσταση φαίνεται στην εικόνα 5.1-1. Παρατηρούμε ότι όσο κινούμαστε από τα αριστερά προς τα δεξιά οι διακυμάνσεις στα δεδομένα μεγαλώνουν. Η χρονοσειρά δηλαδή παρουσιάζει τάση στη διακύμανση. Επίσης παρουσιάζει και εποχικότητα η οποία ξεχωρίζει στην εικόνα 5.1-2 από το διάγραμμα εποχικότητας.



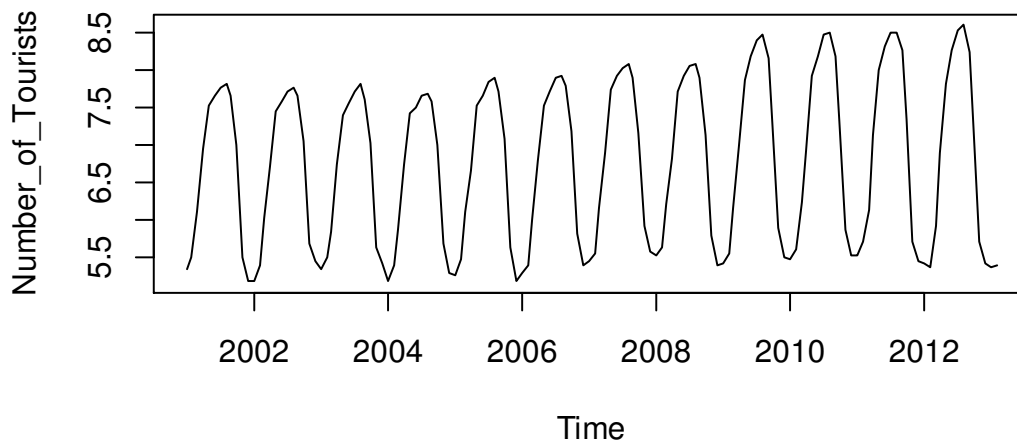
Εικόνα 5.1-2: Διάγραμμα εποχικότητας δεδομένων τουριστών

Το πρώτο πράγμα που κάνουμε είναι να χωρίσουμε τα δεδομένα σε 2 κομμάτια. Το πρώτο κομμάτι καλύπτει το χρονικό διάστημα Ιαν2001-Φεβ2013 και το δεύτερο το χρονικό διάστημα Μαρ2013-Μαρ2016. Η εφαρμογή της μεθόδου και η εκτίμηση του μοντέλου γίνεται στο πρώτο κομμάτι. Το δεύτερο κομμάτι θα μας βοηθήσει να συγκρίνουμε στο τέλος πόσο καλές προβλέψεις δίνει το μοντέλο που θα εκτιμήσουμε σε σχέση με τις υπόλοιπες μεθόδους που θα εφαρμόσουμε σε αυτά τα δεδομένα.

### ΠΡΟΒΛΕΨΕΙΣ ΜΕ ΜΟΝΤΕΛΛΑ ARIMA

Για να εφαρμόσουμε τη μέθοδο ARIMA πρέπει να μετασχηματίσουμε τη σειρά αφού δεν είναι στάσιμη. Μέσω του μετασχηματισμού BoxCox ( $\lambda=0$ -Ισοδυναμεί με λογαρίθμηση των τιμών) προκύπτει η χρονοσειρά που φαίνεται στην εικόνα 5.1-3.

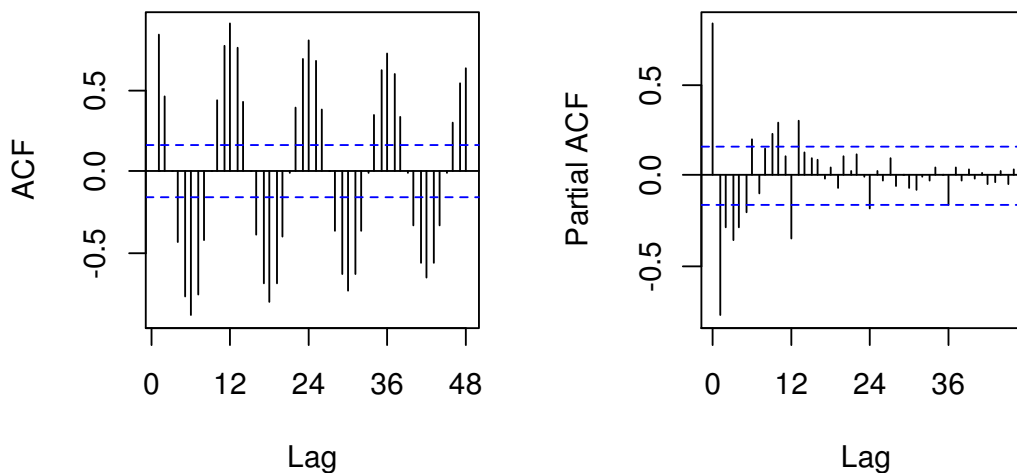
Στα δεδομένα αυτά έχουμε μετατρέψει τα κανονικά για να εκφράζονται τα νούμερα σε χιλιάδες και να είναι ευανάγνωστα στις γραφικές παραστάσεις.



Εικόνα 5.1-3: Δεδομένα τουριστών με μετασχηματισμό BoxCox ( $\lambda=0$ )

Συνεχίζοντας μελετάμε τα διαγράμματα ACF και PACF (εικόνα 5.1-4).

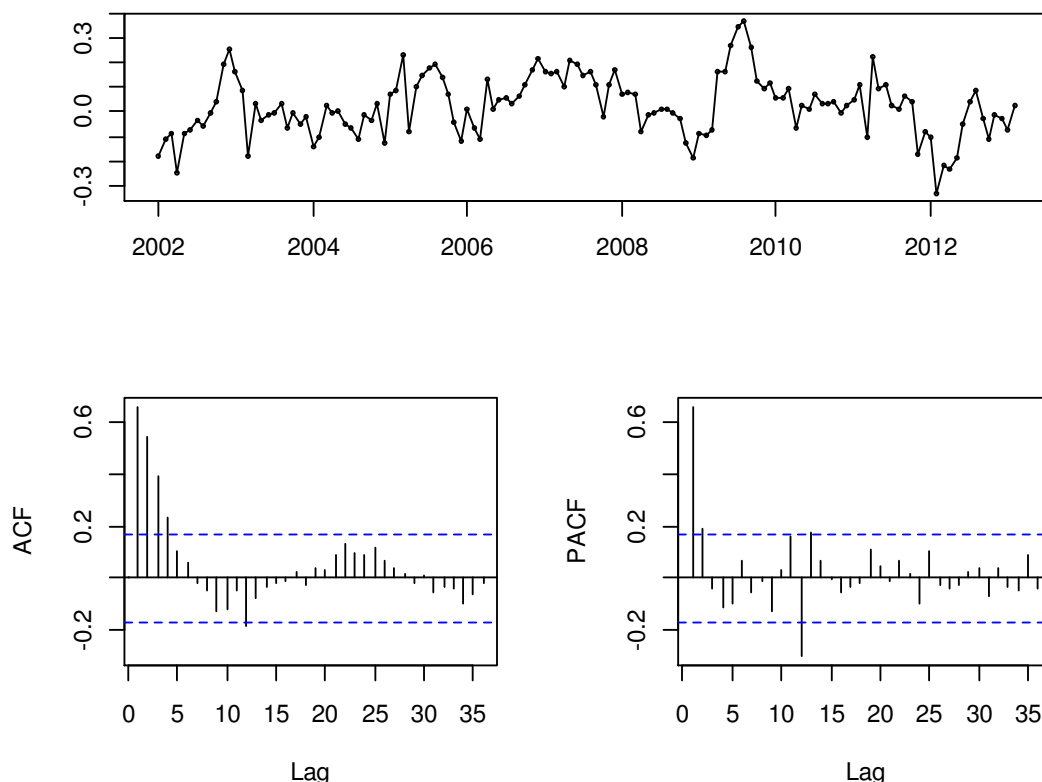
Και από το διάγραμμα των δεδομένων αλλά και από τον αυξημένο και στατιστικά σημαντικό συντελεστή μερικής αυτοσυσχέτισης για χρονική υστέρηση 12 περιόδων, παρατηρούμε ότι η χρονοσειρά παρουσιάζει εποχικότητα.



Εικόνα 5.1-4: Συναρτήσεις αυτοσυσχέτισης και μερικής αυτοσυσχέτισης σε μετασχηματισμένα δεδομένα τουριστών

Θα εφαρμόσουμε τη μέθοδο εποχικής διαφορίσης. Τα αποτελέσματα φαίνονται παρακάτω:

### Data after a seasonal difference



Εικόνα 5.1-5: Δεδομένα τουριστών μετά από εφαρμογή εποχιακών διαφορών

Η σειρά φαίνεται στάσιμη πλέον και το υπόδειγμα που φαίνεται να την περιγράφει μέχρι στιγμής είναι το  $ARIMA(p, 0, q)(P, 1, Q)_s$ . Μένει να υπολογίσουμε τα  $p, q, P, Q$ . Από τη συνάρτηση αυτοσυσχέτισης ACF παρατηρούμε μια εκθετική μείωση των πρώτων συντελεστών υποδεικνύοντας ένα μη εποχικό AR υπόδειγμα. Η τάξη του καθορίζεται από το γράφημα της PACF όπου παρατηρείται να υπάρχουν στατιστικά σημαντικοί συντελεστές στους χρόνους υστέρησης 1 και 2. Οπότε υπάρχει ένδειξη ενός μη εποχικού υποδείματος AR(2). Όσο για το εποχικό κομμάτι, από την PACF φαίνεται να υπάρχουν στατιστικά σημαντικοί συντελεστές μόνο στην χρονική υστέρηση 12. Οπότε υπάρχει η ένδειξη ενός εποχικού υποδείματος AR(1). Αυτή είναι η πρώτη εκτίμηση που φαίνεται από τα γραφήματα. Οπότε καταλήγουμε στο υπόδειγμα  $ARIMA(2,0,0)(1,1,0)_{12}$ .

Προχωράμε στην εκτίμηση του υποδείματος που επιλέξαμε. Εφαρμόζουμε τη μέθοδο ARIMA στην R και παρατηρούμε ότι η παράμετρος για το εποχικό υπόδειγμα AR(1) δεν είναι στατιστικά σημαντική (πίνακας 5.1-1). Αυτό μας οδηγεί να επιλέξουμε κάποιο άλλο υπόδειγμα. Θα συγκρίνουμε πολλά με βάση το  $ARIMA(p,0,q)(P, 1, Q)_{12}$  για τα διάφορα  $p, q, P, Q$ . Το βασικό κριτήριο θα είναι η τιμή του AIC. Καλύτερο υπόδειγμα θα θεωρηθεί αυτό με το μικρότερο AIC. Τα αποτελέσματα δείχνουν ως καλύτερο το  $ARIMA(1,0,0)(2,1,1)_{12}$ .

ARIMA(2,0,0)(1,1,0)[12]

BoxCoxtransformation: lambda= 0

Coefficients:

ar1 ar2 sar1

0.6084 0.1653 -0.3370

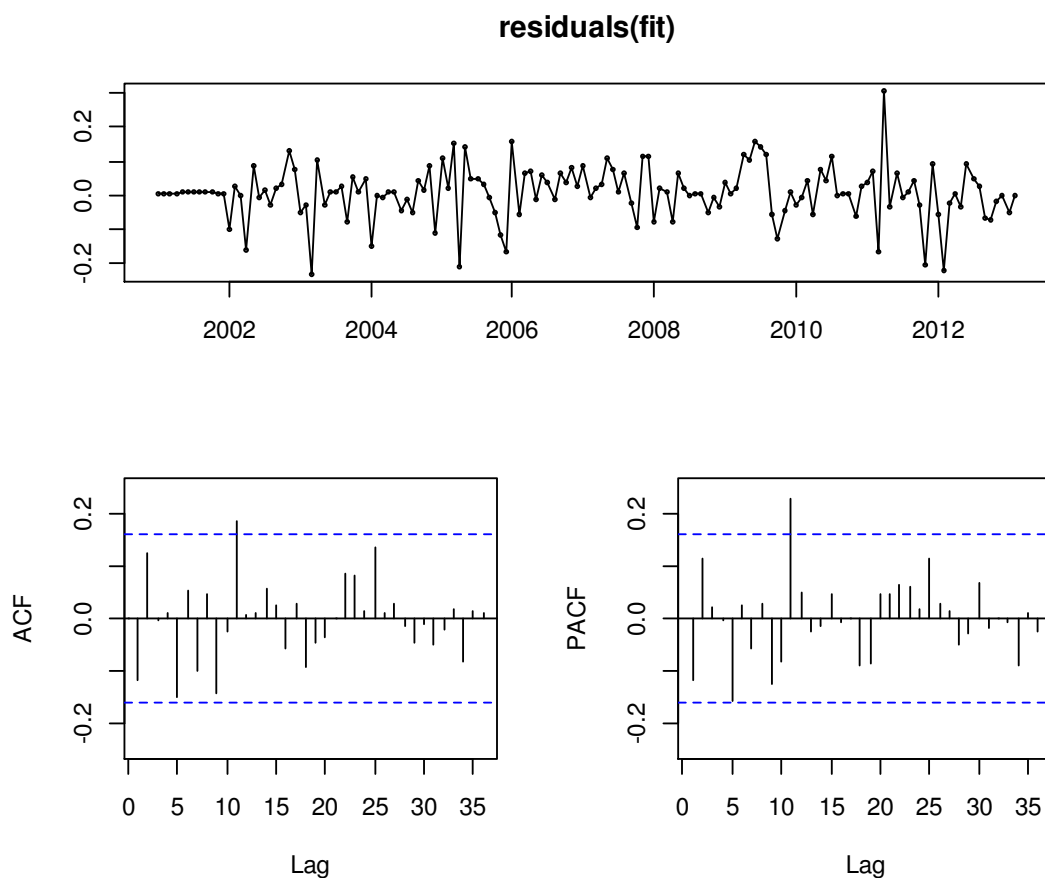
s.e. 0.0864 0.0857 0.0842

sigma^2 estimated as 0.007342: log likelihood=139.56

AIC=-271.12 AICc=-270.81 BIC=-259.53

Πίνακας 5.1-1: Εφαρμογή υποδείγματος ARIMA(2,0,0)(1,1,0)<sub>12</sub> στα δεδομένα τουριστών

Σε επόμενο βήμα θα ελέγξουμε πόσο καλά προσαρμόζεται το υπόδειγμα που διαλέξαμε στα δεδομένα. Θα μελετήσουμε τους συντελεστές αυτοσυσχέτισης και μερικής αυτοσυσχέτισης των υπολοίπων. Το αποτέλεσμα φαίνεται στην εικόνα



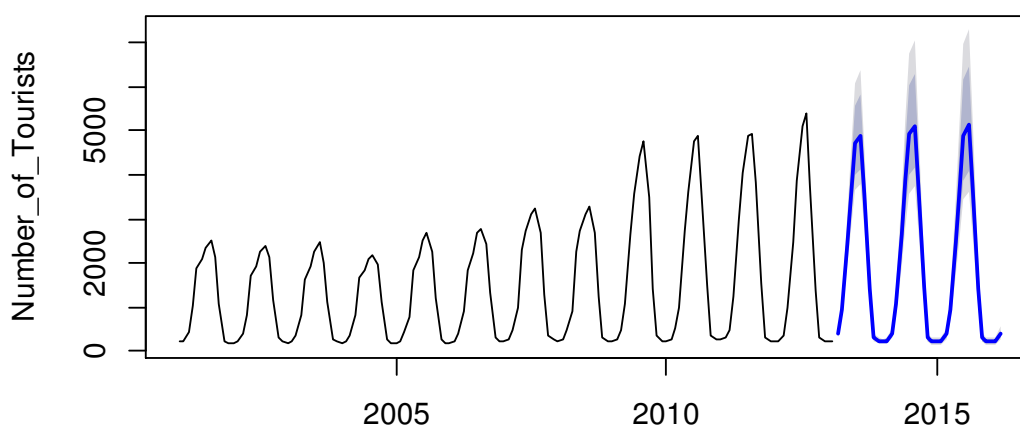
Εικόνα 5.1-6: Γράφημα υπολοίπων για δεδομένα τουριστών με εφαρμογή του υποδείγματος ARIMA(1,0,0)(2,1,1)<sub>12</sub>.



Υπάρχουν στατιστικά σημαντικοί συντελεστές τόσο στην ACF όσο και στην PACF που μπορεί να οδηγήσουν σε λάθος διαστήματα εμπιστοσύνης της πρόβλεψης λόγω της αυτοσυσχέτισης που παρουσιάζουν τα υπόλοιπα. Όμως μπορούμε να πούμε ότι ακολουθούν σχεδόν καλά το πρότυπο του λευκού θορύβου. Και το τεστ των Box-Pierce ενισχύει αυτή την άποψη ( $X\text{-squared} = 27.899$ ,  $p\text{-value} = 0.6744$ ).

Το επόμενο βήμα είναι οι προβλέψεις. Θα προβλέψουμε για τόσες χρονικές περιόδους όσες και το test set μας. Στη συγκεκριμένη περίπτωση είναι 37.

### Forecasts from ARIMA(1,0,0)(2,1,1)[12]



Εικόνα 5.1-7: Προβλέψεις για δεδομένα τουριστών (σε χιλιάδες) με το υπόδειγμα  $ARIMA(1,0,0)(2,1,1)_{12}$

Στο σημείο αυτό θα υπολογιστούν τα σφάλματα πρόβλεψης. Παρακάτω δίνεται ο πίνακας με κάποιες ενδεικτικές τιμές της χρονοσειράς που συμπεριλάβαμε στην εκτίμηση του μοντέλου και οι προβλέψεις για τις μελλοντικές 37 περιόδους.

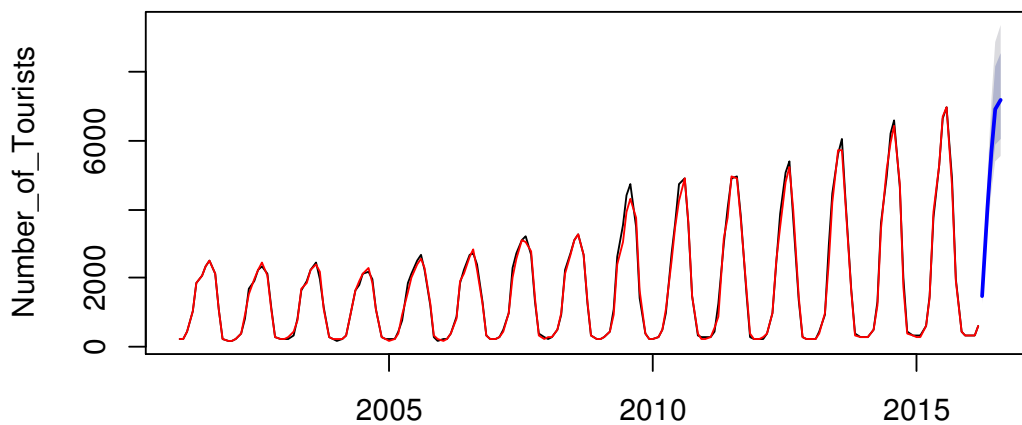
Year	2011	2012	2013	2014	2015	2016
Jan	253.314	227.762	211.598	221.5937	217.8246	216.8127
Feb	301.648	215.909	222.145	237.4247	218.0132	230.4154
Mar	462.902	371.926	392.1149	392.0623	381.2401	392.2425
Apr	1257.512	998.431	952.7183	1071.639	969.3338	1500.967
May	3000.032	2477.023	2451.252	2629.224	2455.041	4078.348
Jun	4056.852	3860.455	3591.804	3861.085	3705.33	5697.459
Jul	4892.728	5098.168	4714.594	4910.228	4882.447	6939.873
Aug	4948.3	5405.678	4905.838	5113.918	5127.166	7219.927
Sep	3838.208	3732.457	3571.005	3739.046	3640.295	

Oct	1551.962	1383.517	1408.148	1446.079	1394.797	
Nov	301.794	297.717	303.0652	297.2028	300.4976	
Dec	232.37	226.455	227.1066	227.811	226.6961	

Πίνακας 5.1-2: Προβλέψεις για τα δεδομένα τουριστών με το υπόδειγμα  $ARIMA(1,0,0)(2,1,1)_{12}$

Στον πίνακα 5.1-2 βλέπουμε 3 κατηγορίες τιμών. Με τα μαύρα γράμματα είναι οι πραγματικές παρατηρήσεις τις χρονοσειράς. Με τα κόκκινα γράμματα είναι οι προβλέψεις του υποδείματος  $ARIMA(1,0,0)(2,1,1)_{12}$  για το test set. Τέλος με τα μπλε γράμματα είναι οι τελικές προβλέψεις του επιλεγμένου υποδείματος για 5 μελλοντικές περιόδους. Την αντιστοιχία μπορούμε να τη δούμε στην εικόνα 5.1-8

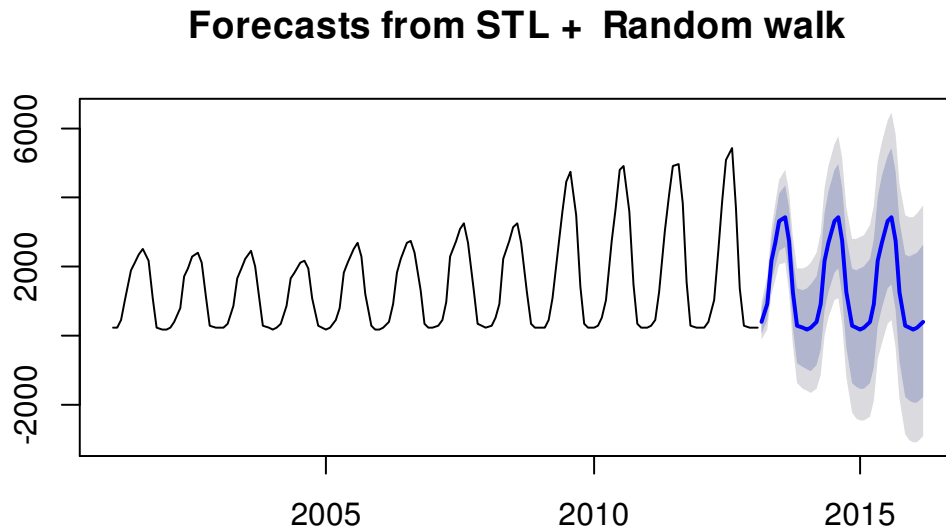
### Forecasts from $ARIMA(1,0,0)(2,1,1)[12]$



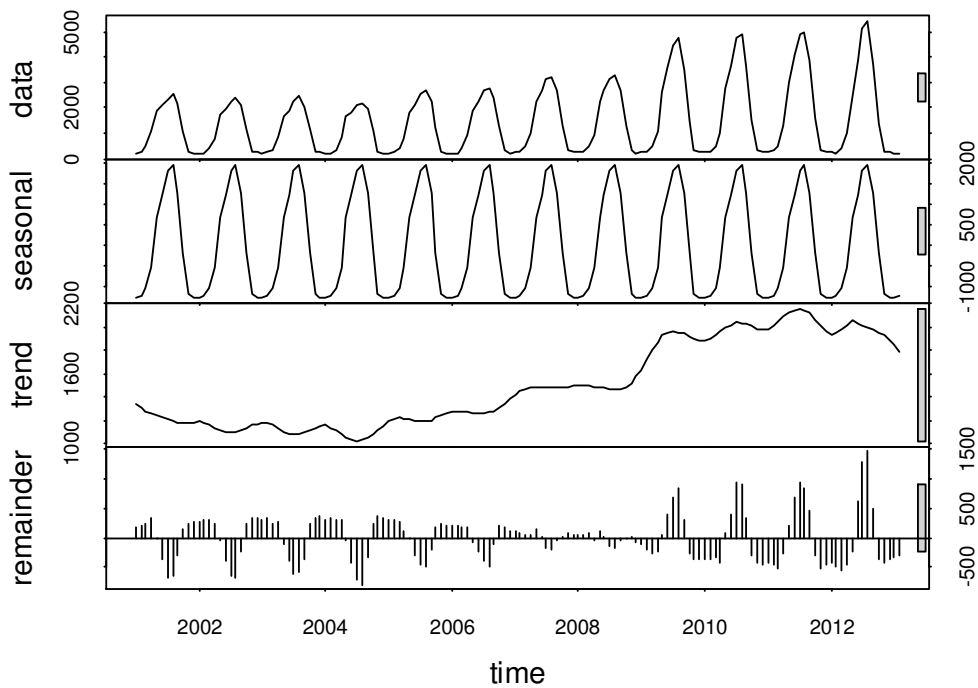
Εικόνα 5.1-8: Γράφημα των δεδομένων τουριστών και των προβλεπόμενων τιμών του υποδείματος  $ARIMA(1,0,0)(2,1,1)_{12}$

### ΜΕΘΟΔΟΣ ΑΠΟΣΥΝΘΕΣΗΣ

Η αποσύνθεση της χρονοσειράς των τουριστών και οι προβλέψεις για το test set φαίνονται στα παρακάτω διαγράμματα:



Εικόνα 5.1-9: Προβλέψεις δεδομένων τουριστών με τη μέθοδο STL χρησιμοποιώντας τη μέθοδο naive



Εικόνα 5.1-10: Αποσύνθεση δεδομένων τουριστών με χρήση μεθόδου STL

Η διασπορά των τιμών σύμφωνα με τη μέθοδο αποσύνθεσης φαίνεται να εξηγείται κατά 83% από την εποχικότητα, 7% από την τάση και 10% από τις τυχαίες διακυμάνσεις.

#### ΜΕΘΟΔΟΣ HOLT-WINTERS

Τα δεδομένα χαρακτηρίζονται από μεταβολή στη διασπορά των τιμών με το χρόνο. Οπότε το κατάλληλο μοντέλο πρόβλεψης θα θεωρήσουμε ότι είναι το πολλαπλασιαστικό.

Θα εξετάσουμε 2 περιπτώσεις. Στην πρώτη το μοντέλο υπολογίζεται με τη βοήθεια της μεθόδου Holt Winters του στατιστικού πακέτου R και δίνει τις εκτιμήσεις μέσω της μεθόδου χωρίς να του προσδιορίσουμε κάποιες αρχικές τιμές. Στη δεύτερη περίπτωση οι τιμές για να ξεκινήσει η μέθοδος προκύπτουν σύμφωνα με τις εξισώσεις (4.18), (4.19) και (4.20). Η επιλογή των παραμέτρων και στις 2 περιπτώσεις γίνεται ελαχιστοποιώντας το MSE.

Περίπτωση 1: Η επιλογή της μεθόδου φαίνεται στον παρακάτω πίνακα.

Holt-Winters exponential smoothing with trend and additive seasonal component.

Smoothing parameters:

alpha: 0.1475818

beta : 0.02084195

gamma: 1

Coefficients:

[,1]

a 1783.362215

b 5.002323

s1 -1385.734508

s2 -721.409426

s3 826.809414

s4 2217.237932

s5 3402.133351

s6 3627.235620

s7 1961.020286

s8 -373.345020

s9 -1469.027507

s10 -1547.941376

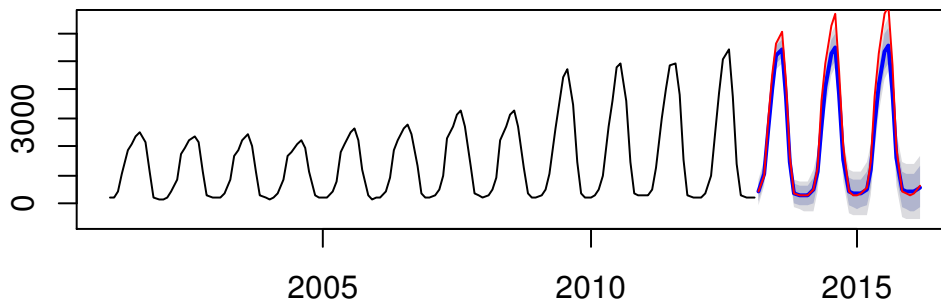
s11 -1567.058825

s12 -1561.217215

Πίνακας 5.1-3: Μοντέλο περίπτωσης 1 για δεδομένα τουριστών με εφαρμογή της μεθόδου Holt-Winters

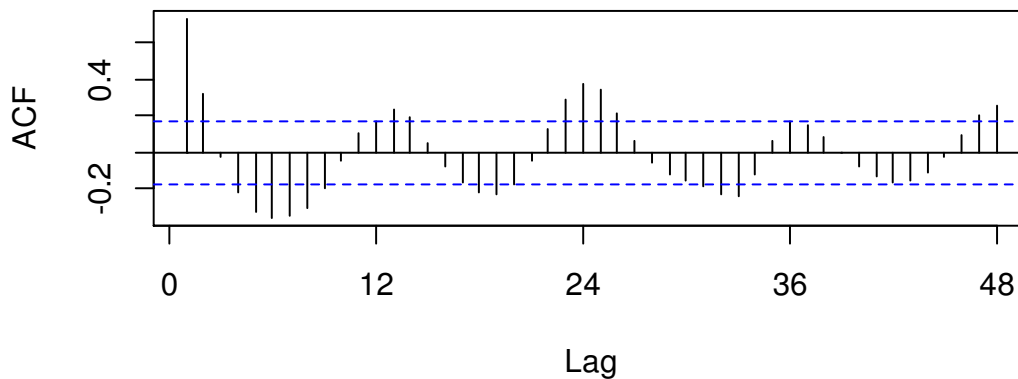
Θα χρησιμοποιήσουμε το μοντέλο για να προβλέψουμε 37 μελλοντικές τιμές που αντιστοιχούν στο test set που επιλέξαμε στην αρχή (εικόνα 5.1-11) και στη συνέχεια θα εξετάσουμε τα υπόλοιπα (residuals) (εικόνα 5.1-12).

### Forecasts from HoltWinters



Εικόνα 5.1-11: Προβλέψεις με τη μέθοδο Holt-Winters σε δεδομένα τουριστών (περίπτωση 1)

### Series hwforecast\$residuals



Εικόνα 5.1-12: Συνάρτηση αυτοσυσχέτισης υπολοίπων για πρόβλεψη περίπτωσης 1 με εφαρμογή της Holt-Winters σε δεδομένα τουριστών

Στην εικόνα 5.1-11 βλέπουμε τις προβλέψεις της μεθόδου με τη μπλε γραμμή και τις πραγματικές τιμές για το ίδιο χρονικό διάστημα με την κόκκινη γραμμή. Δε φαίνεται να είναι τόσο κοντά. Επίσης από τη συνάρτηση αυτοσυσχέτισης βλέπουμε ότι τα υπόλοιπα έχουν μεγάλη εξάρτηση μεταξύ τους. Οπότε και οι προβλέψεις δε θα είναι αξιόπιστες.

Πάνω κάτω η ίδια εικόνα υπάρχει και στη δεύτερη περίπτωση.

Holt-Winters exponential smoothing with trend and additive seasonal component.

Smoothing parameters:

alpha: 0.1388274

beta : 0.002558621

gamma: 1

Coefficients:

[,1]

a 1796.620281

b 1.465084

s1 -1410.730464

s2 -747.835875

s3 796.661920

s4 2186.743449

s5 3374.785862

s6 3605.218614

s7 1939.956511

s8 -393.819750

s9 -1487.560199

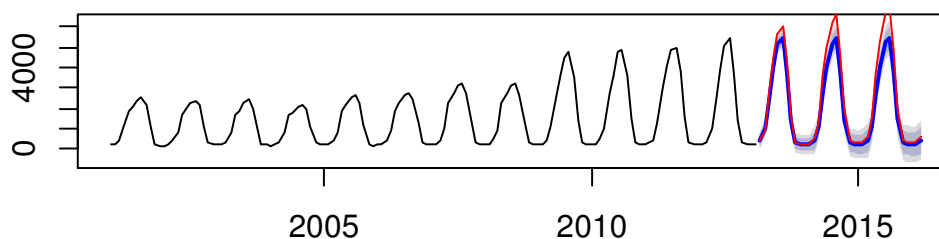
s10 -1564.558903

s11 -1581.940724

s12 -1574.475281

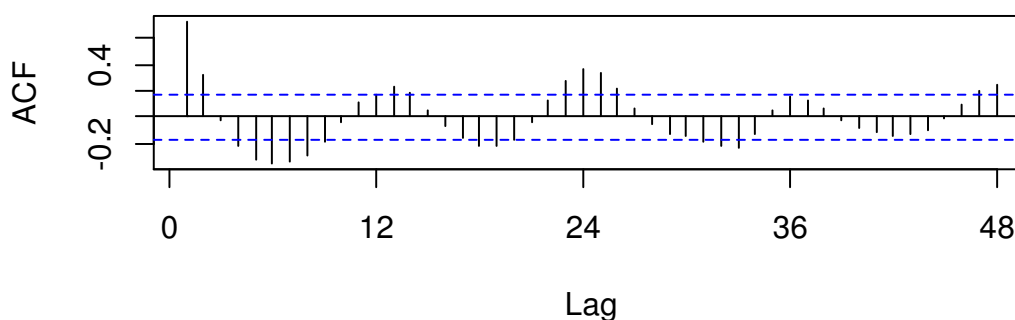
Πίνακας 5.1-4: Μοντέλο περίπτωσης 2 για δεδομένα τουριστών με εφαρμογή της μεθόδου Holt-Winters

### Forecasts from HoltWinters



Εικόνα 5.1-13: Προβλέψεις με τη μέθοδο Holt-Winters σε δεδομένα τουριστών (περίπτωση 2)

### Series hwforecast1\$residuals



Εικόνα 5.1-14: Συνάρτηση αυτοσυσχέτισης υπολοίπων για πρόβλεψη περίπτωσης 2 με εφαρμογή της Holt-Winters σε δεδομένα τουριστών

Και τα 2 μοντέλα παρουσιάζουν συσχετίσεις στα υπόλοιπα (residuals). Αυτό μας οδηγεί στο συμπέρασμα ότι υπάρχει πληροφορία που δεν έχουμε εκμεταλλευτεί και το μοντέλο ενδεχομένως να επιφέρει βελτίωση. Εξετάζοντας και τα στατιστικά μέτρα για τα σφάλματα με κριτήριο το μικρότερο RMSE καλύτερο μοντέλο δείχνει να είναι το πρώτο.

	ME	RMSE	MAE	MPE	MAPE	ACF1	Theil's U
Model_1	384.61	600.1	409.67	-7.03	13.41	0.80	0.31
Model_2	-483.97	681.6	487.15	-19.18	19.52	0.81	0.42

Πίνακας 5.1-5: Στατιστικά μέτρα σφαλμάτων πρόβλεψης για τις 2 περιπτώσεις της μεθόδου Holt-Winters με εφαρμογή σε δεδομένα κατανάλωσης ηλεκτρικής ενέργειας

### 5.1.1 Σύγκριση μεθόδων και επιλογή μοντέλου

Συγκρίναμε 4 μεθόδους. Arima, Holt-Winters, STL Decomposition και Naive. Τα αποτελέσματα των στατιστικών μέτρων για τα σφάλματα πρόβλεψης φαίνονται στον παρακάτω πίνακα:

Μέθοδος	RMSE	MAE	MPE	MAPE	ACF1	Theil.s.U
HW	600.1026	409.6678	7.028708	13.40972	0.799956	0.310394
ARIMA	811.724	582.3987	23.83548	23.83548	0.806287	0.47348
STL	1542.955	1043.343	33.98893	33.98893	0.796846	0.648559
Naive	3269.683	2279.797	68.8545	68.8545	0.811917	1.521384

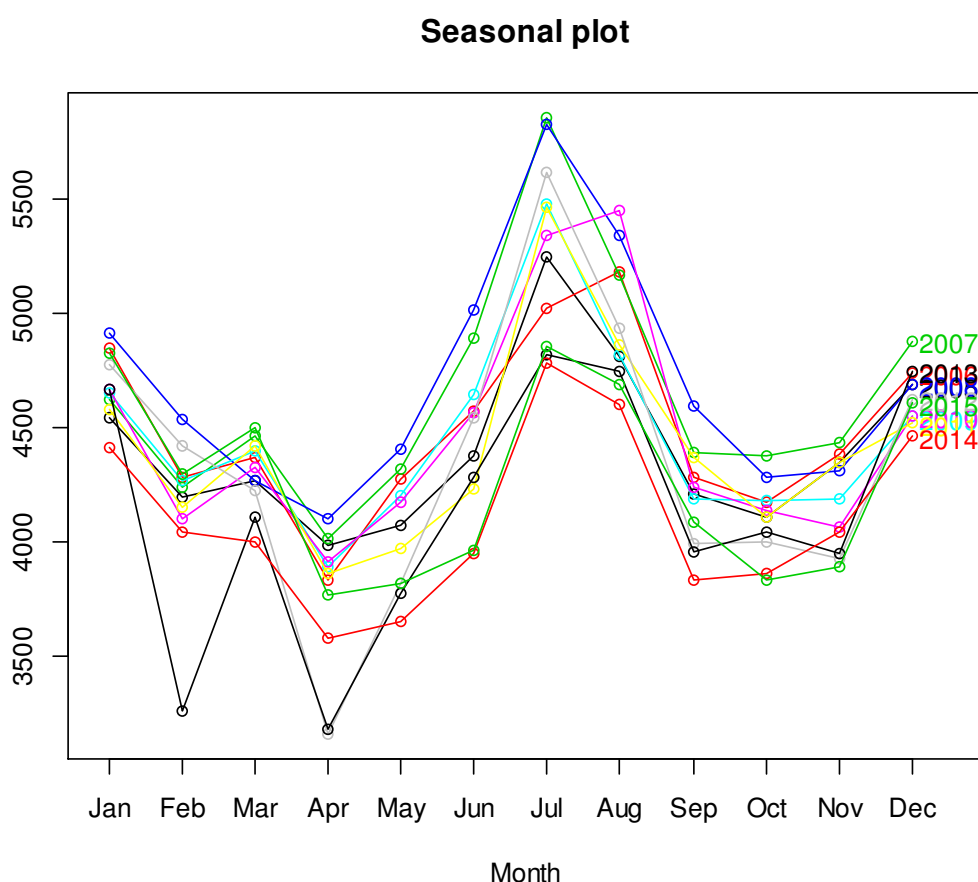
Το καλύτερο μοντέλο φαίνεται να είναι της μεθόδου Holt-Winters.



## 5.2 Δεδομένα κατανάλωσης ηλεκτρικής ενέργειας<sup>22</sup>

Η μεταβλητή που θα εξετάσουμε περιγράφει την κατανάλωση ηλεκτρικής ενέργειας στην Ελλάδα το χρονικό διάστημα Ιαν2005-Δεκ2015. Το δείγμα αποτελείται από 132 μηνιαίες παρατηρήσεις.

Μια πρώτη επαφή με τα δεδομένα έγινε στην παράγραφο 4.1. Στην εικόνα 4.1-1 μπορούμε να δούμε τη γραφική παράσταση των δεδομένων. Φαίνεται να παρουσιάζουν εποχικότητα τόσο από το γράφημα αυτό όσο και από το εποχικό γράφημα της εικόνας 5.2-1



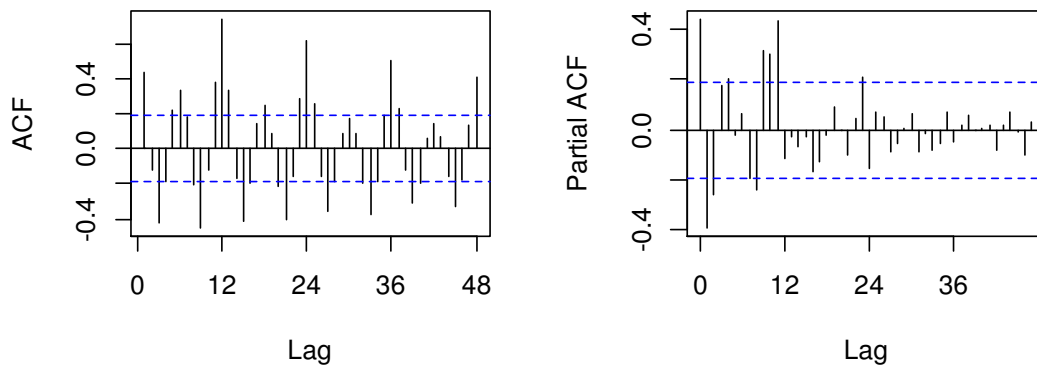
Εικόνα 5.2-1: Εποχικό γράφημα δεδομένων κατανάλωσης ηλεκτρικής ενέργειας

Χωρίσαμε τα δεδομένα σε 2 κομμάτια. Το πρώτο κομμάτι καλύπτει το χρονικό διάστημα Ιαν2005-Οκτ2013 και το δεύτερο το χρονικό διάστημα Νοε2013-Δεκ2015. Η εφαρμογή της μεθόδου και η εκτίμηση του μοντέλου γίνεται στο πρώτο κομμάτι. Το δεύτερο κομμάτι θα μας βοηθήσει στη σύγκριση με άλλες μεθόδους και στην αξιολόγηση του μοντέλου.

<sup>22</sup> Ο κώδικας με τις εντολές στην R παρατίθεται στο παράρτημα, στο αρχείο elec\_final.r

### ΠΡΟΒΛΕΨΕΙΣ ΜΕ ΜΟΝΤΕΛΑ ARIMA

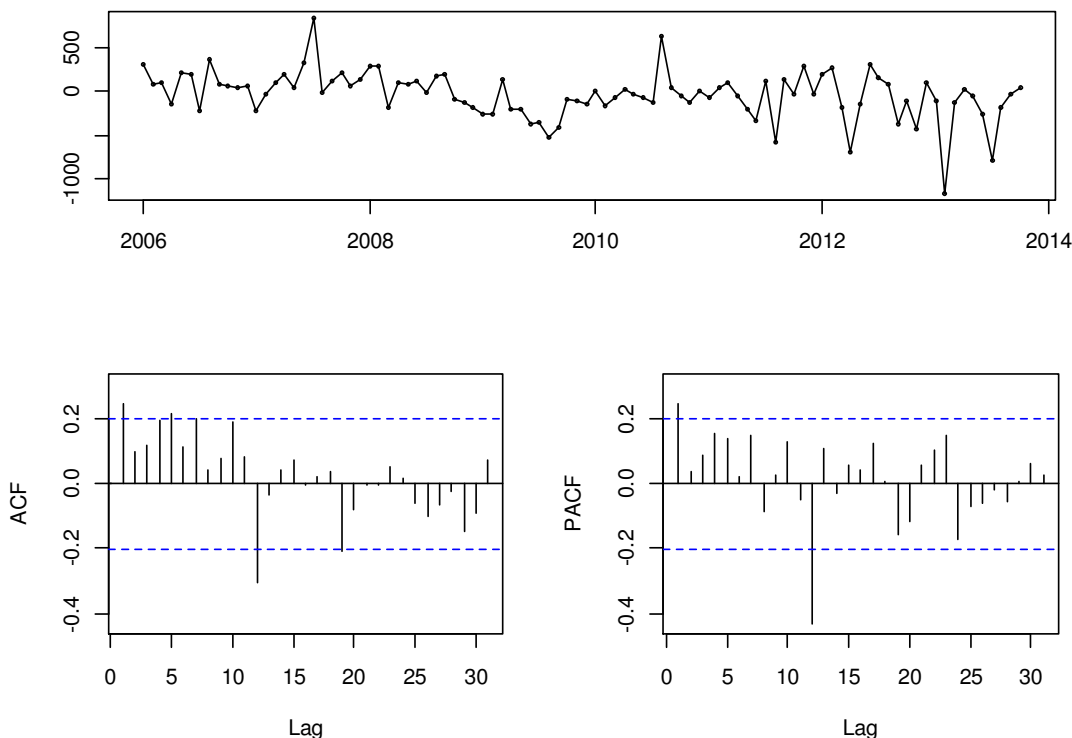
Για να εφαρμόσουμε τη μέθοδο ARIMA πρέπει να εξετάσουμε αν η σειρά είναι στάσιμη. Με μια πρώτη ματιά από το γράφημα των δεδομένων δεν παρατηρείται κάποια μεταβολή ως προς το μέσο. Ίσως μόνο στη διασπορά.. Μια πιο ξεκάθαρη εικόνα θα μας δώσουν τα διαγράμματα ACF και PACF της εικόνας. 5.2-2.



Εικόνα 5.2-2: Διαγράμματα ACF και PACF των δεδομένων κατανάλωσης ηλεκτρικής ενέργειας το χρονικό διάστημα Ιαν2005-Οκτ2013

Η χρονοσειρά φαίνεται να μην είναι στάσιμη. Και επίσης στο διάγραμμα ACF φαίνεται και το μοτίβο της εποχικότητας. Επειδή τα δεδομένα είναι εποχικά, θα εφαρμόσουμε τη μέθοδο εποχικής διαφορίσης. Τα αποτελέσματα φαίνονται παρακάτω:

**Data after a seasonal difference**



Εικόνα 5.2-3: Δεδομένα ηλεκτρικής κατανάλωσης μετά από εφαρμογή εποχικών διαφορών

Η σειρά φαίνεται στάσιμη και το υπόδειγμα που φαίνεται να την περιγράφει μέχρι στιγμής είναι το  $ARIMA(p, 0, q)(P, 1, Q)_s$ . Μένει να υπολογίσουμε τα  $p, q, P, Q$ . Από τα διαγράμματα και των 2 συναρτήσεων αυτοσυσχέτισης μπορούμε να συμπεράνουμε ότι υπάρχουν ενδείξεις μη εποχικών υποδειγμάτων AR και MA. Από τη συνάρτηση αυτοσυσχέτισης ACF παρατηρούμε ότι υπάρχει ένας στατιστικά σημαντικός συντελεστής στο χρόνο υστέρησης 1 και αντίστοιχα στο διάγραμμα PACF υπάρχει ένας στατιστικά σημαντικός συντελεστής στον ίδιο χρόνο υστέρησης. Οπότε υπάρχει ένδειξη ενός μη εποχικού  $AR(1)$  και ενός μη εποχικού  $MA(1)$ . Όσο για το εποχικό κομμάτι, από την PACF φαίνεται να υπάρχει στατιστικά σημαντικός συντελεστής μόνο στην χρονική υστέρηση 12. Οπότε υπάρχει η ένδειξη ενός εποχικού υποδείγματος  $AR(1)$ . Και το ίδιο ισχύει και στο διάγραμμα ACF οπότε υπάρχει η ένδειξη ενός εποχικού υποδείγματος  $MA(1)$ . Αυτή είναι η πρώτη εκτίμηση που φαίνεται από τα γραφήματα. Οπότε καταλήγουμε στο υπόδειγμα  $ARIMA(1,0,1)(1,1,1)_{12}$ .

Προχωράμε στην εκτίμηση του υποδείγματος που επιλέξαμε. Εφαρμόζουμε τη μέθοδο ARIMA στην R και παρατηρούμε ότι η παράμετρος για το εποχικό υπόδειγμα  $AR(1)$  δεν είναι στατιστικά σημαντική (πίνακας 5.2-4). Αυτό μας οδηγεί να επιλέξουμε κάποιο άλλο υπόδειγμα. Θα συγκρίνουμε πολλά με βάση το  $ARIMA(p,0,q)(P, 1, Q)_{12}$  για τα διάφορα  $p, q, P, Q$ . Το βασικό κριτήριο θα είναι η τιμή του AIC. Καλύτερο υπόδειγμα θα θεωρηθεί αυτό με το μικρότερο AIC. Τα αποτελέσματα δείχνουν ως καλύτερο το  $ARIMA(1,0,1)(0,1,1)_{12}$ .

ARIMA(1,0,1)(1,1,1)[12]

Coefficients:

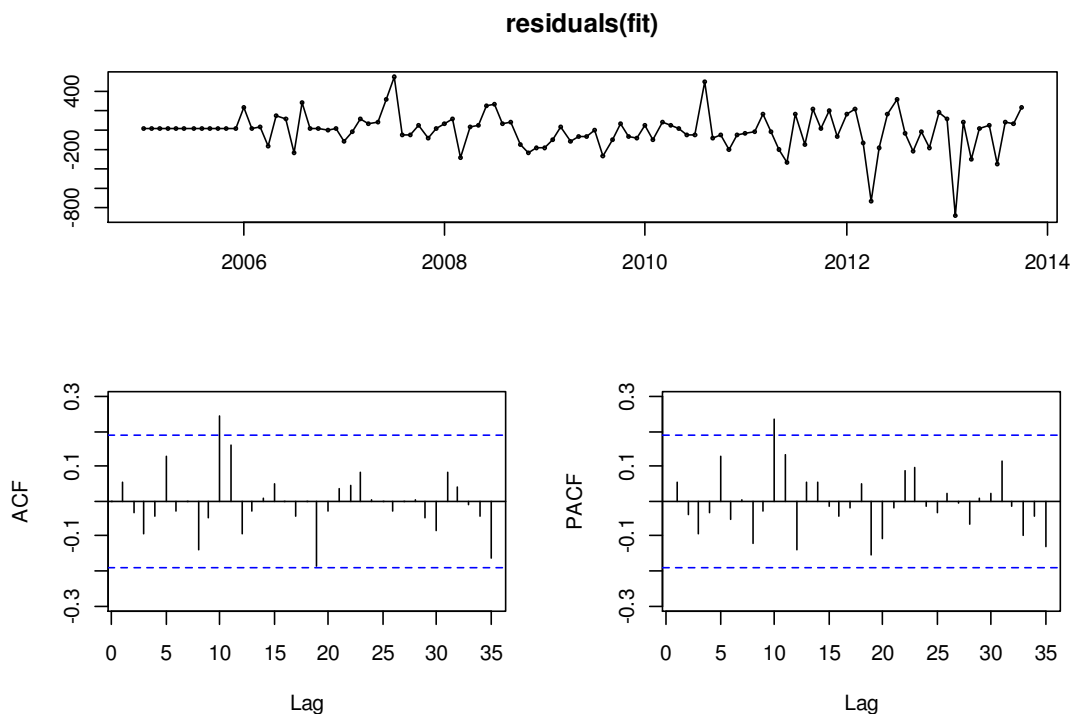
ar1	ma1	sar1	sma1
0.9759	-0.7622	-0.1357	-0.7106
s.e. 0.0326	0.0857	0.1675	0.1846

sigma^2 estimated as 43823: log likelihood=-638.86

AIC=1287.72 AICc=1288.4 BIC=1300.44

Εικόνα 5.2-4: Εφαρμογή υποδείγματος ARIMA(1,0,1)(1,1,1)<sub>12</sub> στα δεδομένα κατανάλωσης ηλεκτρικής ενέργειας

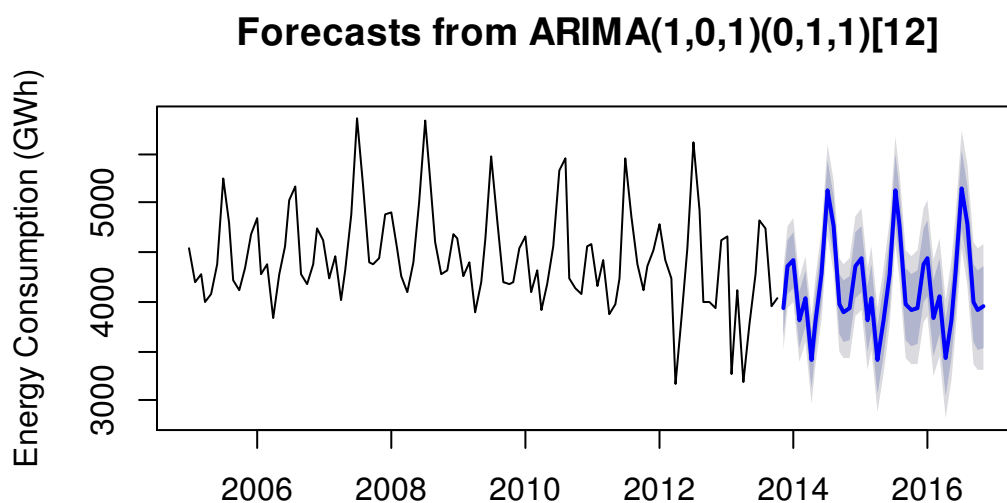
Σε επόμενο βήμα θα ελέγξουμε πόσο καλά προσαρμόζεται το υπόδειγμα που διαλέξαμε στα δεδομένα. Θα μελετήσουμε τους συντελεστές αυτοσυσχέτισης και μερικής αυτοσυσχέτισης των υπολοίπων. Το αποτέλεσμα φαίνεται στην εικόνα



Εικόνα 5.2-5: Γράφημα υπολοίπων για δεδομένα κατανάλωσης ηλεκτρικής ενέργειας με εφαρμογή του υποδείγματος ARIMA(1,0,1)(0,1,1)<sub>12</sub>.

Υπάρχουν στατιστικά σημαντικοί συντελεστές τόσο στην ACF όσο και στην PACF που μπορεί να οδηγήσουν σε λάθος διαστήματα εμπιστοσύνης της πρόβλεψης λόγω της αυτοσυσχέτισης που παρουσιάζουν τα υπόλοιπα. Όμως μπορούμε να πούμε ότι ακολουθούν σχεδόν καλά το πρότυπο του λευκού θορύβου. Και το τεστ των Box-Pierce ενισχύει αυτή την άποψη ( $X\text{-squared} = 21.264$ ,  $p\text{-value} = 0.5649$ ).

Το επόμενο βήμα είναι οι προβλέψεις. Θα προβλέψουμε για τόσες χρονικές περιόδους όσες και το test set μας. Στη συγκεκριμένη περίπτωση είναι 26.



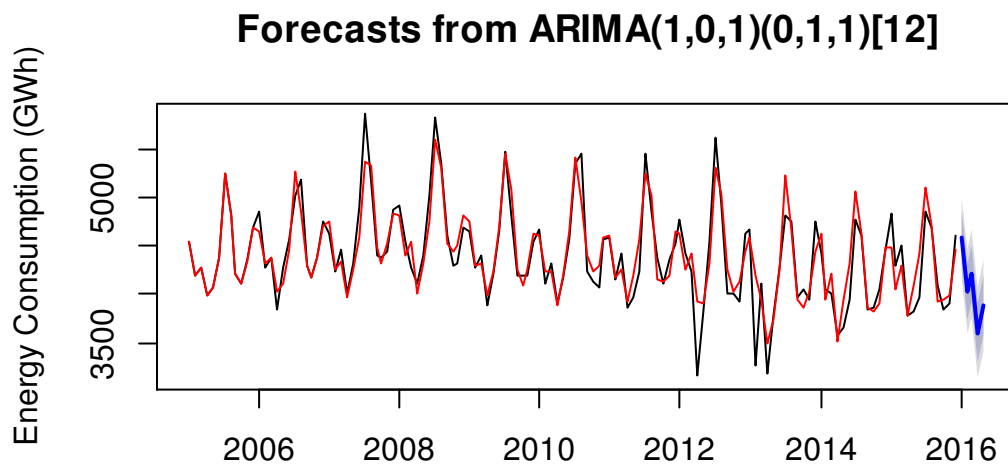
Εικόνα 5.2-6: Προβλέψεις για δεδομένα τουριστών (σε χιλιάδες) με το υπόδειγμα  
ARIMA(1,0,1)(0,1,1)<sub>12</sub>

Στο σημείο αυτό θα υπολογιστούν τα σφάλματα πρόβλεψης. Παρακάτω δίνεται ο πίνακας με κάποιες ενδεικτικές τιμές της χρονοσειράς που συμπεριλάβαμε στην εκτίμηση του μοντέλου και οι προβλέψεις για τις μελλοντικές 26 περιόδους.

Year	2011	2012	2013	2014	2015	2016
Jan	4311	4192	4070	4425.66	4435.337	4583.099
Feb	4691	4550	4556	3815.324	3824.8	4030.916
Mar	4653	4665	4583	4034.106	4043.384	4212.655
Apr	4270	4107	4157	3409.917	3419.002	3601.949
May	4399	4325	4422	3798.209	3807.104	3880.268
Jun	3893	3913	3866	4277.412	4286.121	
Jul	4206	4178	3974	5125.778	5134.307	
Aug	4649	4569	4235	4762.241	4770.592	
Sep	5479	5342	5462	3973.307	3981.483	
Oct	4822	5447	4862	3905.584	3913.59	
Nov	4193	4239	3927.075	3937.168	3945.008	
Dec	4186	4140	4355.082	4364.965	4372.641	

Πίνακας 5.2-1: Προβλέψεις για τα δεδομένα ηλεκτρικής κατανάλωσης με το υπόδειγμα  
ARIMA(1,0,1)(0,1,1)<sub>12</sub>

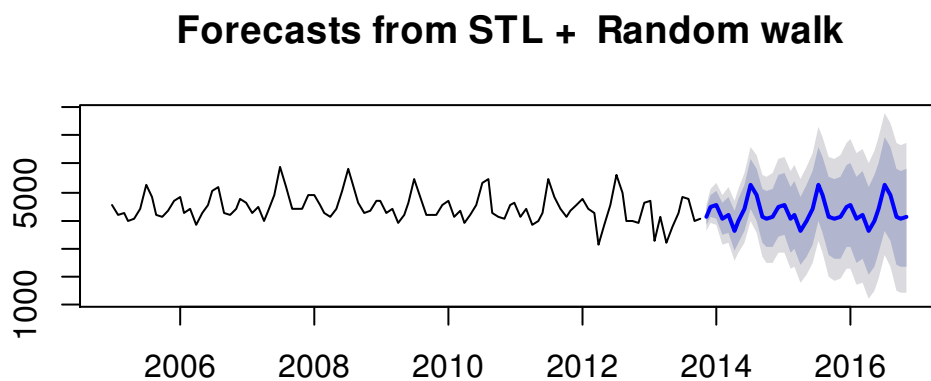
Στον πίνακα 5.2-1 βλέπουμε 3 κατηγορίες τιμών. Με τα μαύρα γράμματα είναι οι πραγματικές παρατηρήσεις τις χρονοσειράς. Με τα κόκκινα γράμματα είναι οι προβλέψεις του υποδείγματος  $ARIMA(1,0,1)(0,1,1)_{12}$  για το testset. Τέλος με τα μπλε γράμματα είναι οι τελικές προβλέψεις του επιλεγμένου υποδείγματος για 5 μελλοντικές περιόδους. Την αντιστοιχία μπορούμε να τη δούμε στην εικόνα 5.2-7



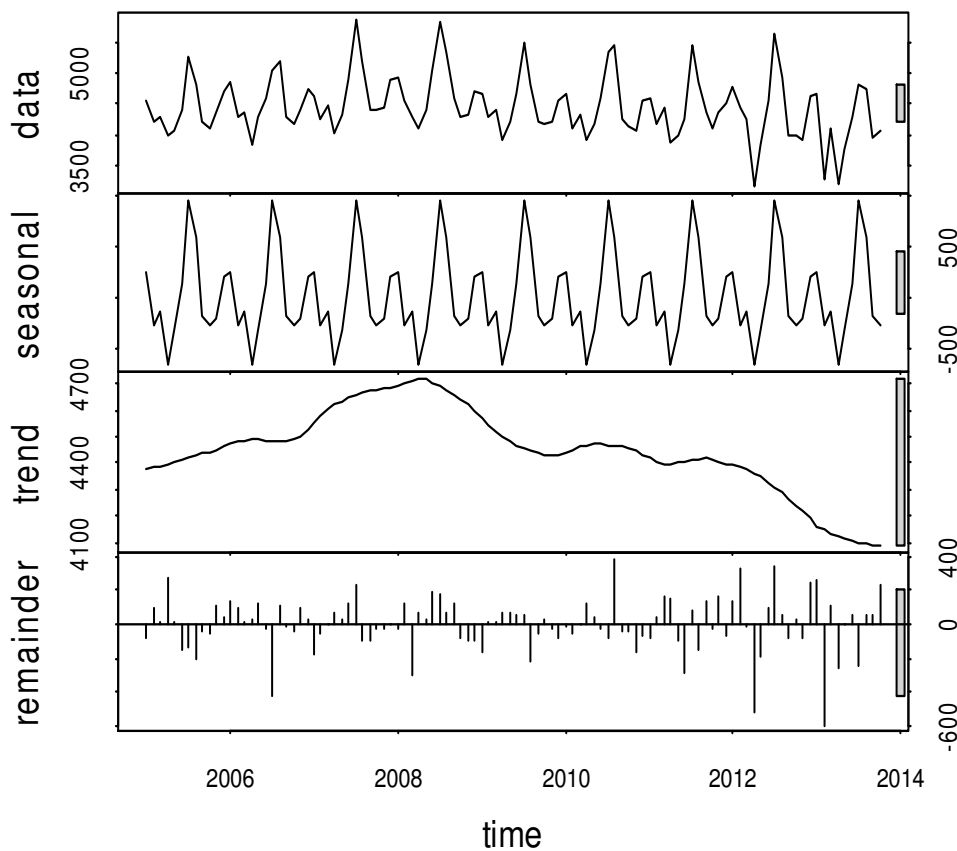
Εικόνα 5.2-7: Γράφημα των δεδομένων ηλεκτρικής κατανάλωσης και των προβλεπόμενων τιμών του υποδείγματος  $ARIMA(1,0,1)(0,1,1)_{12}$

#### ΜΕΘΟΔΟΣ ΑΠΟΣΥΝΘΕΣΗΣ

Οι προβλέψεις καθώς και η αποσύνθεση φαίνονται στα παρακάτω διαγράμματα:



Εικόνα 5.2-8: Προβλέψεις στα δεδομένα τουριστών με τη μέθοδο αποσύνθεσης STL κάνοντας χρήση της μεθόδου naïve.



Εικόνα 5.2-9: Αποσύνθεση χρονοσειράς τουριστών

Εδώ έχουμε μια περίπτωση χρονοσειράς που δε φαίνεται να επηρεάζεται από την τάση αλλά κυρίως από την εποχικότητα. Για να μιλήσουμε και με νούμερα, η διασπορά των τιμών φαίνεται να εξηγείται κατά 80% από την εποχικότητα, 10% από την τάση και 10% από τις τυχαίες διακυμάνσεις.

#### ΜΕΘΟΔΟΣ HOLT-WINTERS

Επειδή δε φαίνεται κάποια έντονη μεταβολή της διακύμανσης των παρατηρήσεων με το χρόνο θα υποθέσουμε ότι το μοντέλο που θα εκτιμήσουμε έχει προσθετική εποχικότητα. Με το ίδιο τρόπο που δουλέψαμε στην ενότητα 4.2.2.4 θα δουλέψουμε και εδώ. Θα υπολογίσουμε 2 μοντέλα και θα εκτιμήσουμε το καλύτερο. Για το πρώτο μοντέλο η εκτίμηση έγινε με τη βοήθεια της μεθόδου Holt-Winters στην R. Η μέθοδος εφαρμόζει αυτόματα τη μέθοδο αποσύνθεσης για να διασπάσει τη χρονοσειρά στις συνιστώσες. Μετά οι αρχικές τιμές του επιπέδου και της τάσης εκτιμούνται με εφαρμογή απλή γραμμικής παλινδρόμησης πάνω στη συνιστώσα της τάσης που απομονώθηκε. Αντίστοιχα για την εποχικότητα εφαρμόζεται κάποιος κινητός μέσος στη συνιστώσα της εποχικότητας. Έπειτα η επιλογή των παραμέτρων γίνεται ελαχιστοποιώντας το MSE. Το μοντέλο που υπολογίστηκε φαίνεται στον πίνακα 5.2-2.

Holt-Winters exponential smoothing with trend and additive seasonal component.

Smoothing parameters:

alpha: 0.1469264

beta : 0.04461461

gamma: 0.2685157

Coefficients:

[,1]

a 4141.13903

b -9.49252

s1 -284.42368

s2 149.43765

s3 243.22124

s4 -421.64209

s5 -155.96597

s6 -832.82095

s7 -380.20024

s8 105.07489

s9 943.87410

s10 569.30116

s11 -221.55616

s12 -285.71566

Πίνακας 5.2-2: Μέθοδος Holt-Winters για δεδομένα κατανάλωσης ηλεκτρικής ενέργειας(περίπτωση 1)

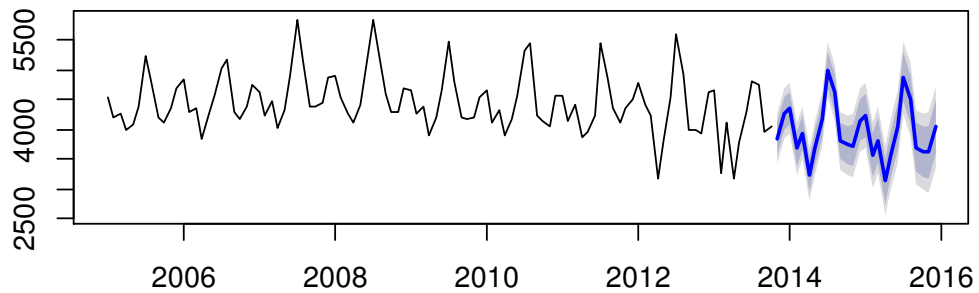
Θα εφαρμόσουμε τη μέθοδο για να προβλέψουμε τις 26 τιμές που βρίσκονται στο test set που έχουμε επιλέξει. Τα αποτελέσματα φαίνονται στην εικόνα 5.2-10 μαζί με τα 80% και 95% διαστήματα εμπιστοσύνης. Κάνοντας έναν έλεγχο των υπολοίπων διακρίνουμε κάποιες αυτοσυσχετίσεις σε 2 χρονικές υστερήσεις που είναι στατιστικά σημαντικές (εικόνα 5.2-11) . Αυτό είναι ένδειξη ότι το μοντέλο μπορεί να βελτιωθεί.

Θα δούμε τώρα το δεύτερο μοντέλο. Σε αυτό οι αρχικές τιμές προήλθαν όπως υποδεικνύουν οι εξισώσεις (4.18), (4.19), (4.25).

Τα αποτελέσματα και οι εκτιμήσεις των παραμέτρων φαίνονται στον πίνακα 5.2-3.

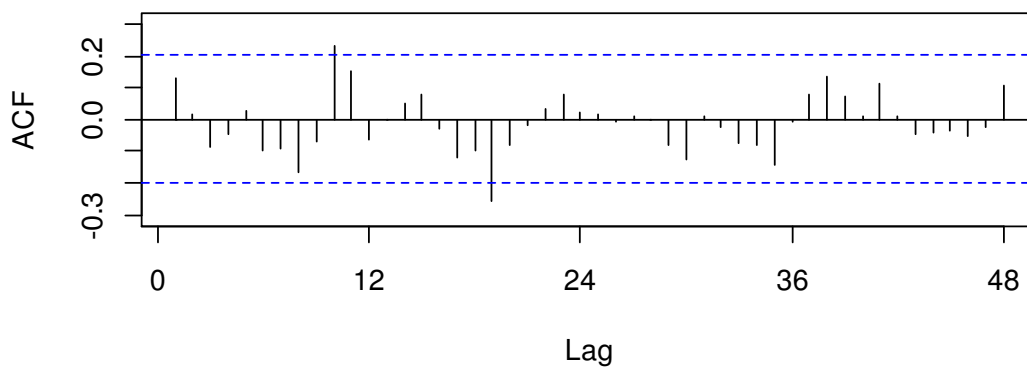


### Forecasts from HoltWinters



Εικόνα 5.2-10: Προβλέψεις με τη μέθοδο Holt-Winters για δεδομένα κατανάλωσης ηλεκτρικής ενέργειας.

### Series hwforecast\$residuals



Εικόνα 5.2-11: Συνάρτηση αυτοσυσχέτισης των υπολοίπων με εφαρμογή της μεθόδου Holt-Winters σε δεδομένα κατανάλωσης ηλεκτρικής ενέργειας για την περίπτωση 1

Holt-Winters exponential smoothing with trend and additive seasonal component.

Smoothing parameters:

alpha: 0.1664458

beta : 0

gamma: 0.3211827

Coefficients:

[,1]

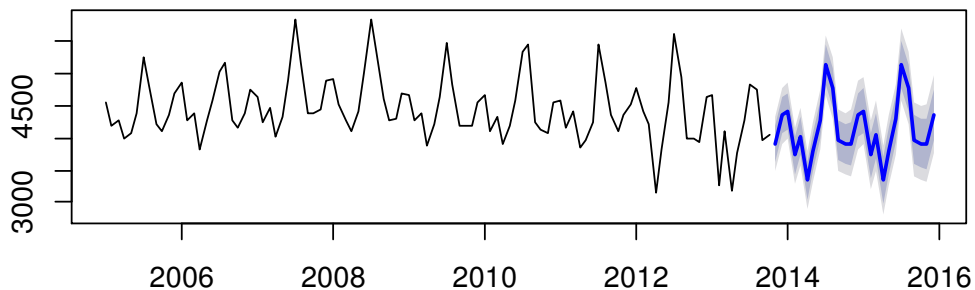
a 4172.1983053

b	0.6666667
s1	-270.7324959
s2	178.4105423
s3	244.8250417
s4	-444.2364739
s5	-145.1238022
s6	-831.3196432
s7	-384.3966022
s8	108.7048231
s9	952.7733093
s10	589.0104137
s11	-209.7129469
s12	-265.5988056

Πίνακας 5.2-3: Μέθοδος Holt-Winters για δεδομένα κατανάλωσης ηλεκτρικής ενέργειας(περίπτωση 2)

Η εφαρμογή της μεθόδου και η πρόβλεψη 26 μελλοντικών τιμών αποτυπώνονται στο γράφημα της εικόνας 5.2-12.

### Forecasts from HoltWinters

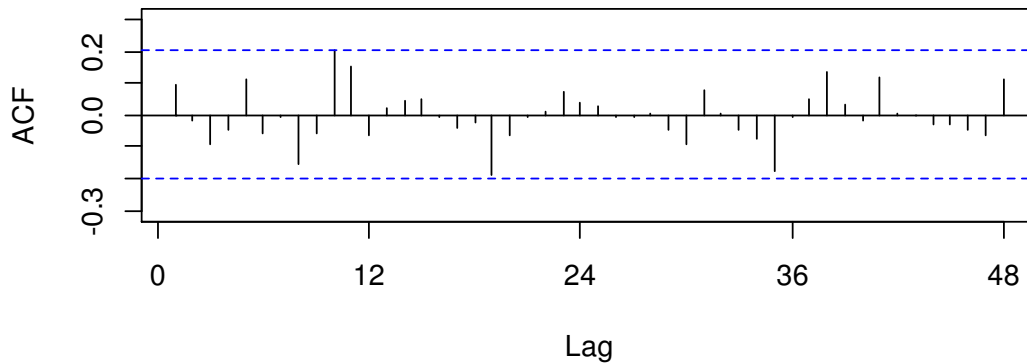


Εικόνα 5.2-12: Πρόβλεψη μεθόδου Holt-Winters στα δεδομένα κατανάλωσης ηλεκτρικής ενέργειας(περίπτωση 2)

Στη συνέχεια θα εξετάσουμε τα υπόλοιπα(residuals) για να δούμε πόσο ανεξάρτητα είναι (εικόνα 5.2-13).

Το μοντέλο αυτό συγκριτικά με αυτό της πρώτης περίπτωσης φαίνεται να δουλεύει καλύτερα. Στη συνέχεια θα συγκρίνουμε τα 2 μοντέλα να δούμε ποιο δίνει καλύτερες προβλέψεις με βάση και το RMSE (πίνακας 5.2-4) .

### Series hwforecast1\$residuals

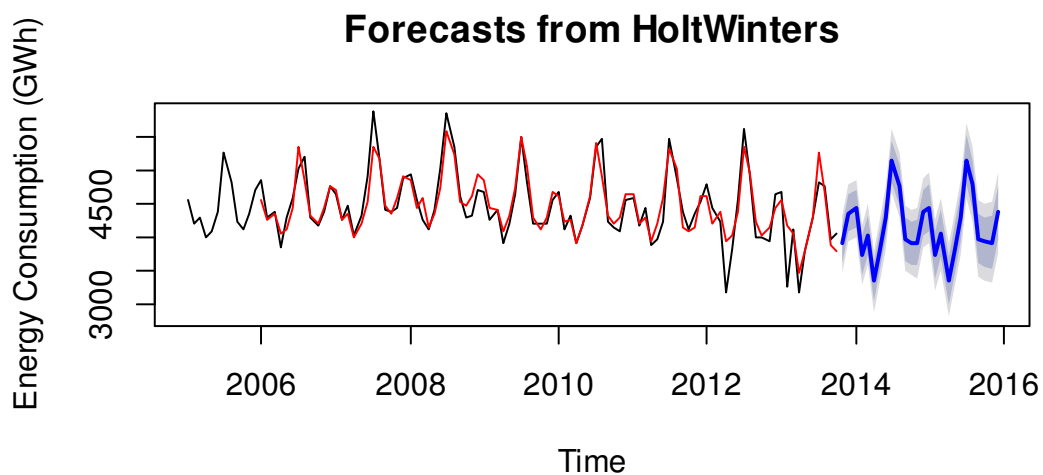


Εικόνα 5.2-13: Συνάρτηση αυτοσυσχέτισης των υπολοίπων για την εφαρμογή της μεθόδου Holt-Winters σε δεδομένα κατανάλωσης ηλεκτρικής ενέργειας (περίπτωση 2)

	ME	RMSE	MAE	MPE	MAPE	ACF1	Theil's U
Model_1	233.16	352.57	279.99	-5.53	6.63	0.58	0.75
Model_2	-56.08	262.44	279.99	-1.30	4.94	0.57	0.57

Πίνακας 5.2-4: Στατιστικά μέτρα σφαλμάτων πρόβλεψης για τις 2 περιπτώσεις της μεθόδου Holt-Winters με εφαρμογή σε δεδομένα κατανάλωσης ηλεκτρικής ενέργειας

Μετά από όλη την παραπάνω ανάλυση μπορούμε να συμπεράνουμε ότι το μοντέλο που δουλεύει καλύτερα είναι αυτό της δεύτερης περίπτωσης.



Εικόνα 5.2-14: Προβλέψεις επιλεγμένου μοντέλου με εφαρμογή της μεθόδου Holt-Winters σε δεδομένα κατανάλωσης ηλεκτρικής ενέργειας.

### 5.2.1 Σύγκριση μεθόδων και επιλογή μοντέλου

Συγκρίναμε 4 μεθόδους. Τα αποτελέσματα των σφαλμάτων πρόβλεψης φαίνονται στον παρακάτω πίνακα:

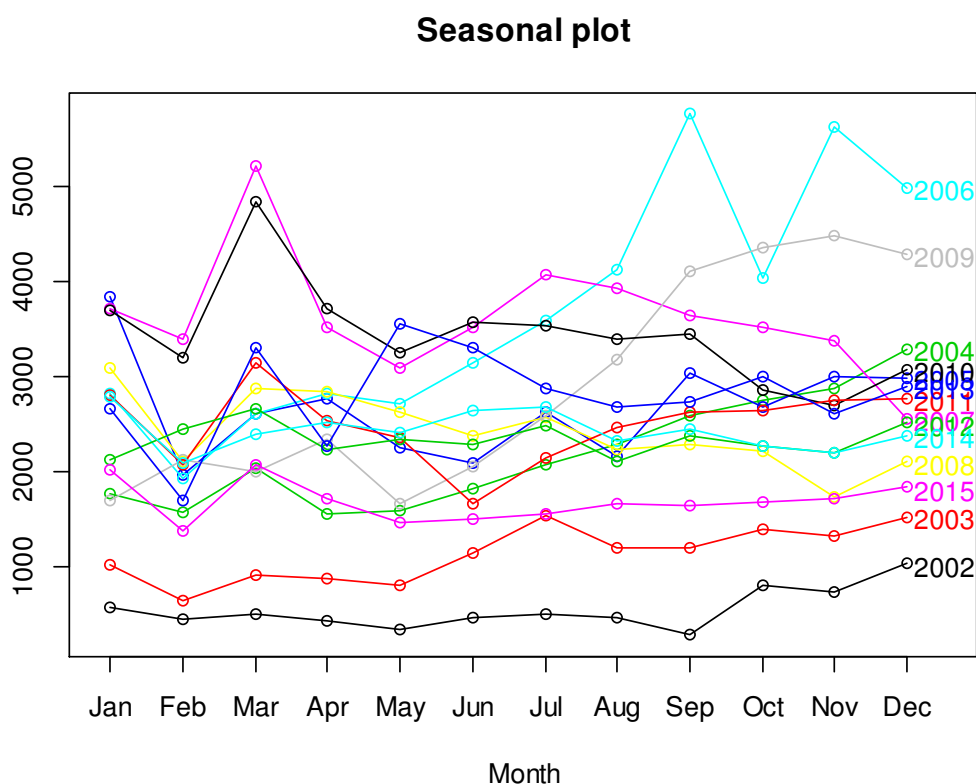
Μέθοδος	RMSE	MAE	MPE	MAPE	ACF1	Theil.s.U
ARIMA	242.8856	194.8705	0.927318	4.556473	0.5878	0.529336
HW	262.4446	210.2459	1.297991	4.939103	0.574222	0.567196
STL	262.5204	223.6127	-2.99647	5.298453	0.607761	0.593261
Naive	420.6225	336.3077	2.792547	7.684958	0.289912	0.920501

Το καλύτερο μοντέλο από αυτά που μελετήθηκαν φαίνεται να είναι το υπόδειγμα  $ARIMA(1,0,1)(0,1,1)_{12}$ .

### 5.3 Δεδομένα πωλήσεων<sup>23</sup>

Η μεταβλητή που θα εξετάσουμε περιγράφει τις μηνιαίες πωλήσεις μιας εταιρίας το χρονικό διάστημα Ιαν2002-Δεκ2015. Το δείγμα αποτελείται από 168 μηνιαίες παρατηρήσεις.

Μια πρώτη επαφή με τα δεδομένα έγινε στην παράγραφο 4.2. Στην εικόνα 4.2-2 μπορούμε να δούμε τη γραφική παράσταση των δεδομένων. Δε φαίνεται καθαρά η εποχικότητα. Θα κάνουμε ένα εποχικό διάγραμμα να δούμε αν βγάζουμε από εκεί κάποιο συμπέρασμα (εικόνα 5.3-1). Ούτε από αυτό φαίνεται κάτι ξεκάθαρα.



Εικόνα 5.3-1: Εποχικό γράφημα δεδομένων πωλήσεων

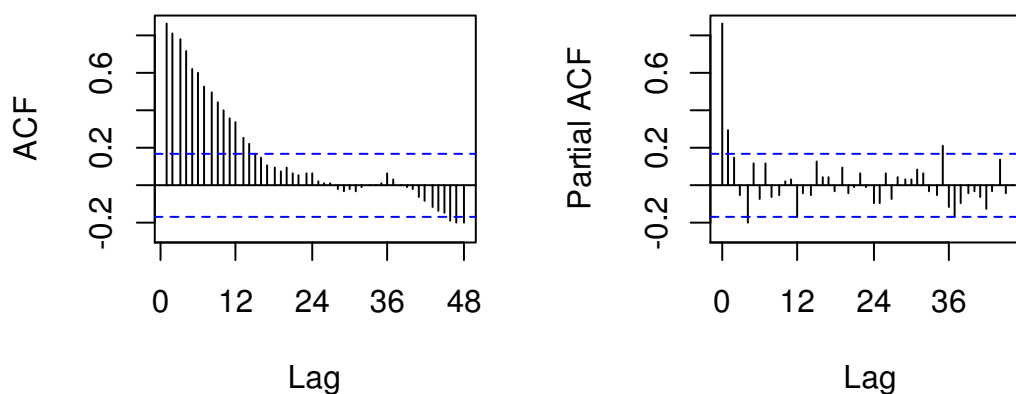
Τα δεδομένα χωρίστηκαν σε 2 κομμάτια. Το πρώτο κομμάτι καλύπτει το χρονικό διάστημα Ιαν2002-Φεβ2013 και το δεύτερο το χρονικό διάστημα Μαρ2013-Δεκ2015. Η εφαρμογή της μεθόδου και η εκτίμηση του μοντέλου γίνεται στο πρώτο κομμάτι. Το δεύτερο κομμάτι θα μας βοηθήσει να συγκρίνουμε στο τέλος τις μεθόδους.

#### ΠΡΟΒΛΕΨΕΙΣ ΜΕ ΜΟΝΤΕΛΑ ARIMA

Για να εφαρμόσουμε τη μέθοδο ARIMA πρέπει να εξετάσουμε αν η σειρά είναι στάσιμη. Με μια πρώτη ματιά από το γράφημα των δεδομένων παρατηρείται μεταβολή του μέσου της χρονοσειράς και μεταβολή των διακυμάνσεων. Θα μπορούσαμε να πούμε ότι έχουμε και εμφάνιση κυκλικής συνιστώσας. Θα δούμε και

<sup>23</sup> Ο κώδικας με τις εντολές στην R παρατίθεται στο παράρτημα, στο αρχείο sales\_final.r

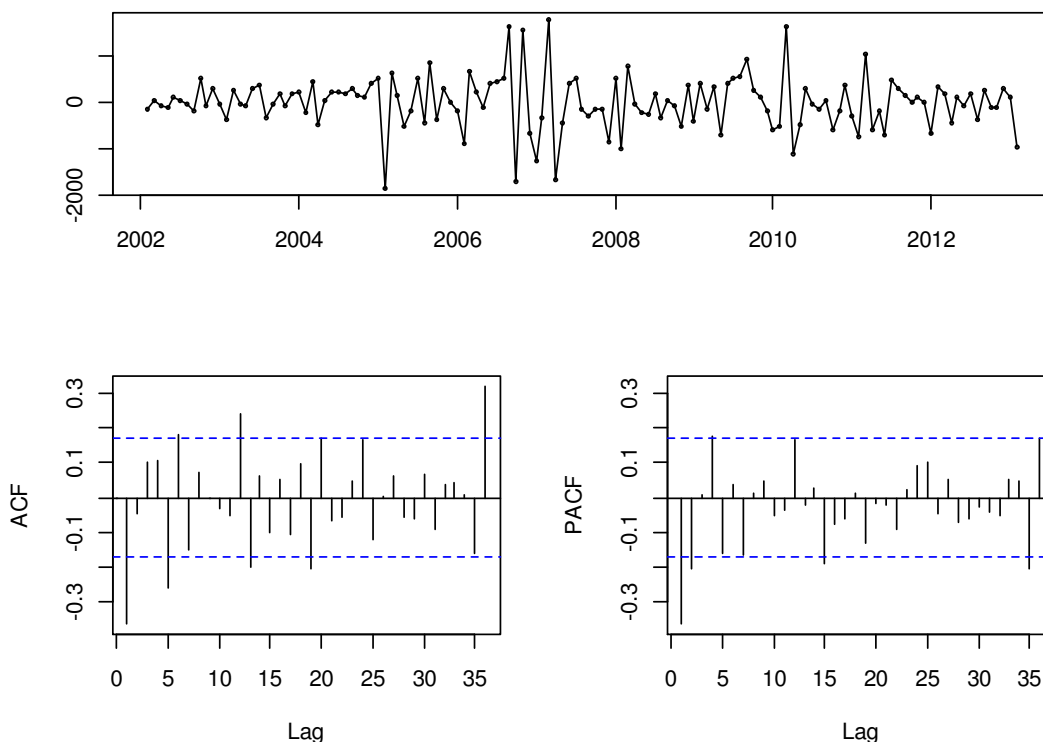
από τα γραφήματα των ACF και PACF αν μπορούμε να βγάλουμε κάποιο άλλο συμπέρασμα (Εικόνα 5.3-2).



Εικόνα 5.3-2: Διαγράμματα ACF και PACF των δεδομένων πωλήσεων το χρονικό διάστημα Ιαν2002-Φεβ2013

Η χρονοσειρά φαίνεται να μην είναι στάσιμη. Και επίσης το διάγραμμα ACF είναι χαρακτηριστικό για εμφάνιση τάσης στα δεδομένα. Δε φαίνεται κάτι για την εποχικότητα. Θα εφαρμόσουμε τη μέθοδο διαφόρισης πρώτης τάξης. Τα αποτελέσματα φαίνονται παρακάτω:

### Data after first differencing



Εικόνα 5.3-3: Δεδομένα πωλήσεων μετά από εφαρμογή διαφώρισης πρώτης τάξης

Η σειρά φαίνεται στάσιμη και το υπόδειγμα που φαίνεται να την περιγράφει μέχρι στιγμής είναι το  $ARIMA(p, 1, q)(P, 0, Q)_s$ . Μένει να υπολογίσουμε τα  $p, q, P, Q$ . Από τα διαγράμματα και των 2 συναρτήσεων αυτοσυσχέτισης μπορούμε να συμπεράνουμε ότι υπάρχουν ενδείξεις μη εποχικού υποδείγματος AR. Στο διάγραμμα PACF παρατηρούμε την ύπαρξη δύο στατιστικά σημαντικών συντελεστών στους χρόνους υστέρησης 1 και 2. Οπότε υπάρχει ένδειξη ενός μη εποχικού  $AR(2)$ . Όσο για το εποχικό κομμάτι, από την PACF δε φαίνεται να υπάρχει κάτι. Από την ACF όμως θα μπορούσαμε να πούμε ότι υπάρχει ένδειξη ενός εποχικού υποδείγματος  $MA(3)$ . Οπότε με τις ενδείξεις που έχουμε καταλήγουμε στο υπόδειγμα  $ARIMA(2,1,0)(0,0,3)_{12}$ .

Προχωράμε στην εκτίμηση του υποδείγματος που επιλέξαμε. Εφαρμόζουμε τη μέθοδο ARIMA στην R και παρατηρούμε ότι η παράμετρος για το εποχικό υπόδειγμα  $MA(2)$  δεν είναι στατιστικά σημαντική (πίνακας 5.3-4). Αυτό μας οδηγεί να επιλέξουμε κάποιο άλλο υπόδειγμα. Θα συγκρίνουμε πολλά με βάση το  $ARIMA(p,1,q)(P, 0, Q)_{12}$  για τα διάφορα  $p, q, P, Q$ . Το βασικό κριτήριο θα είναι η τιμή του AIC. Καλύτερο υπόδειγμα θα θεωρηθεί αυτό με το μικρότερο AIC. Αυτό είναι ισοδύναμο με το μοντέλο που παρουσιάζει το μικρότερο μέσο τετραγωνικό σφάλμα MSE. Τα αποτελέσματα δείχνουν ως καλύτερο το  $ARIMA(0,1,5)(1,0,2)_{12}$ .

$ARIMA(2,1,0)(0,0,3)[12]$

Coefficients:

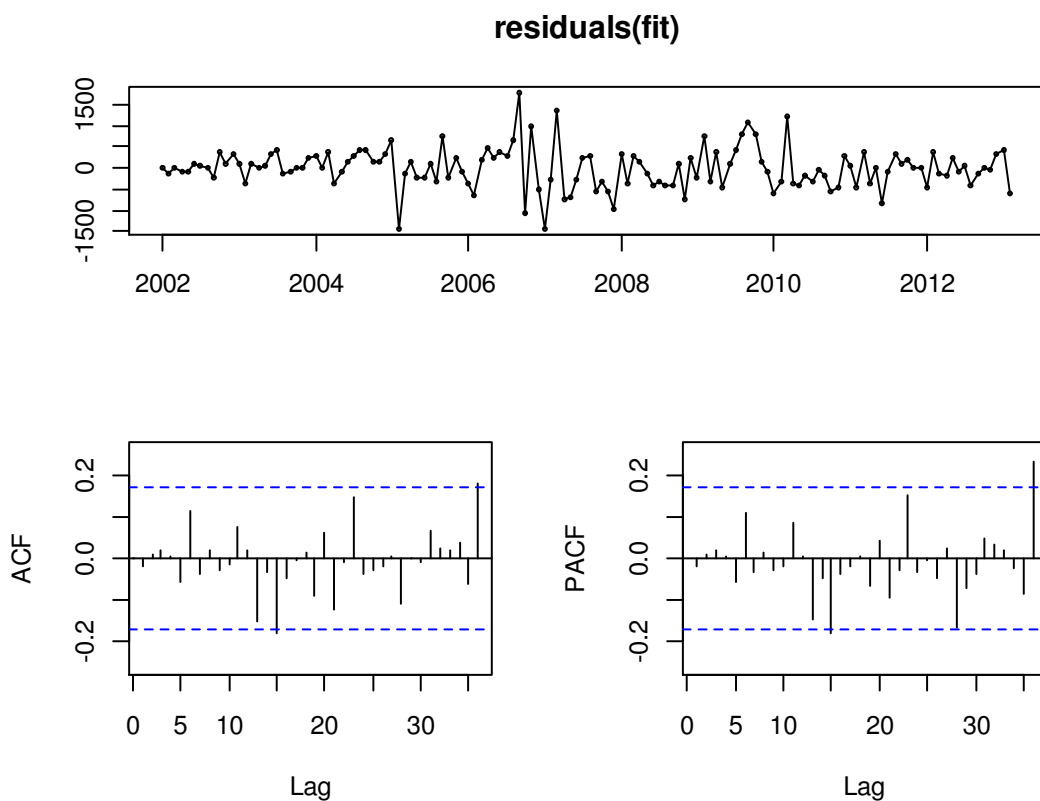
ar1	ar2	sma1	sma2	sma3	
-0.4392	-0.2144	0.1773	0.0614	0.3345	
s.e.	0.0864	0.0853	0.0919	0.0929	0.0971

sigma^2 estimated as 237468: log likelihood=-1011.73

AIC=2035.47 AICc=2036.14 BIC=2052.81

Εικόνα 5.3-4: Εφαρμογή υποδείγματος ARIMA(2,1,0)(0,0,3)<sub>12</sub> στα δεδομένα κατανάλωσης ηλεκτρικής ενέργειας

Σε επόμενο βήμα θα μελετήσουμε τους συντελεστές αυτοσυσχέτισης και μερικής αυτοσυσχέτισης των υπολοίπων (εικόνα 5.3-5).



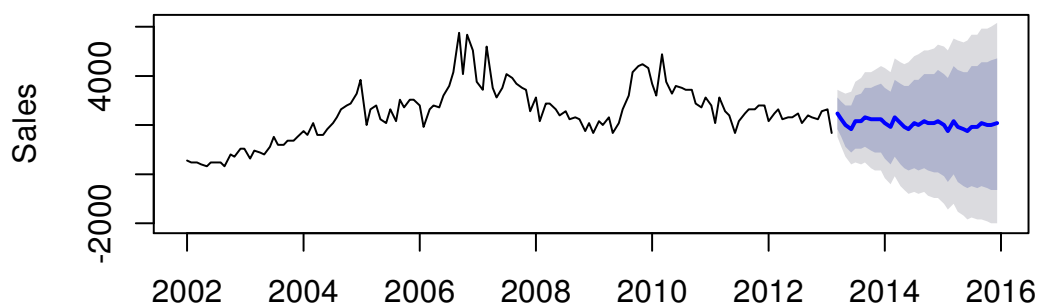
Εικόνα 5.3-5: Γράφημα υπολοίπων για δεδομένα πωλήσεων με εφαρμογή του υποδείγματος ARIMA(0,1,5)(1,0,2)<sub>12</sub>.

Εκτός από έναν με δύο στατιστικά σημαντικούς συντελεστές που εμφανίζονται μπορούμε να πούμε ότι τα διαγράμματα είναι ικανοποιητικά ως προς την προσαρμογή της μεθόδου. Επίσης και το τεστ των Box-Pierce δείχνει καλή προσαρμογή (X-squared = 18.536, p-value = 0.4209).



Το επόμενο βήμα είναι να υπολογίσουμε τις 34 προβλέψεις που αντιστοιχούν στα δεδομένα του test set.

### Forecasts from ARIMA(0,1,5)(1,0,2)[12]



Εικόνα 5.3-6: Προβλέψεις για δεδομένα πωλήσεων με το υπόδειγμα  $ARIMA(0,1,5)(1,0,2)_{12}$

Στην εικόνα 5.3-6 αποτυπώνεται γραφικά το αποτέλεσμα του υποδείγματος που επιλέξαμε για την πρόβλεψη μελλοντικών παρατηρήσεων. Στον πίνακα που ακολουθεί περιέχονται οι εκτιμήσεις που παρατηρούμε στο γράφημα.

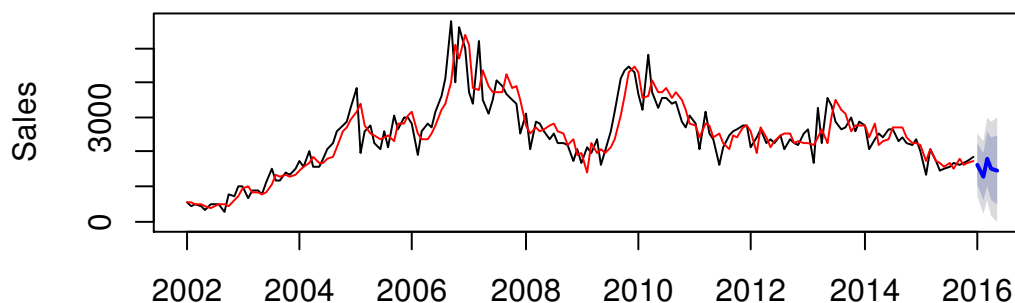
Year	2011	2012	2013	2014	2015	2016
Jan	2804	2113	2649	2067.887	2010.961	1637.842
Feb	2067	2446	1687	1871.885	1774.397	1274.259
Mar	3138	2659	2472.845	2286.348	2157.388	1813.770
Apr	2535	2226	2232.54	2027.469	1939.871	1531.281
May	2348	2342	1964.809	1915.663	1833.584	1477.767
Jun	1657	2277	1846.946	1862.268	1780.873	
Jul	2146	2470	2107.872	2022.526	1926.153	
Aug	2466	2101	2136.617	2009.162	1914.689	
Sep	2614	2368	2295.311	2152.595	2037.737	
Oct	2639	2272	2203.928	2076.732	1972.656	
Nov	2744	2189	2208.027	2072.227	1968.791	
Dec	2771	2512	2259.854	2144.044	2030.402	

Πίνακας 5.3-1: Προβλέψεις για τα δεδομένα ηλεκτρικής κατανάλωσης με το υπόδειγμα  $ARIMA(0,1,5)(1,0,2)_{12}$

Ο πίνακας 5.3-1 περιλαμβάνει 3 κατηγορίες τιμών. Με τα μαύρα γράμματα είναι οι πραγματικές παρατηρήσεις τις χρονοσειράς. Με τα κόκκινα γράμματα είναι οι προβλέψεις του υποδείγματος  $ARIMA(0,1,5)(1,0,2)_{12}$  για το testset. Τέλος με τα

μπλε γράμματα είναι οι τελικές προβλέψεις του επιλεγμένου υποδείγματος για 5 μελλοντικές περιόδους. Την αντιστοιχία γραφικά μπορούμε να τη δούμε στην εικόνα 5.2-7

### Forecasts from ARIMA(0,1,5)(1,0,2)[12]

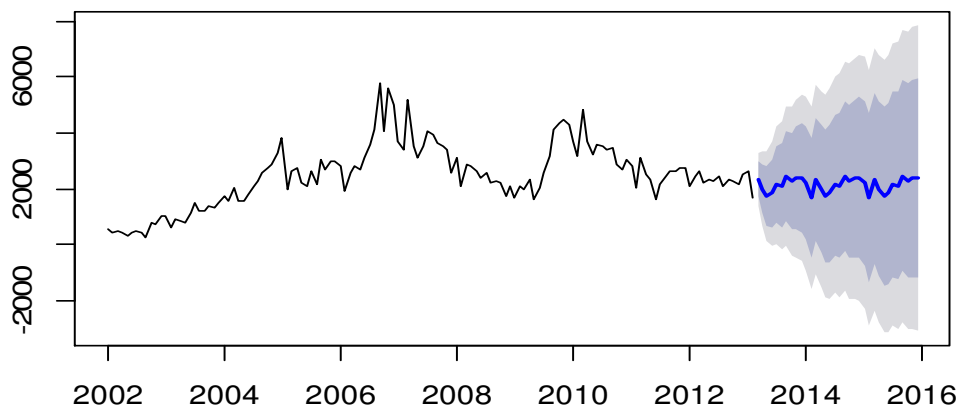


Εικόνα 5.3-7: Γράφημα των δεδομένων ηλεκτρικής κατανάλωσης και των προβλεπόμενων τιμών του υποδείγματος ARIMA(0,1,5)(1,0,2)<sub>12</sub>

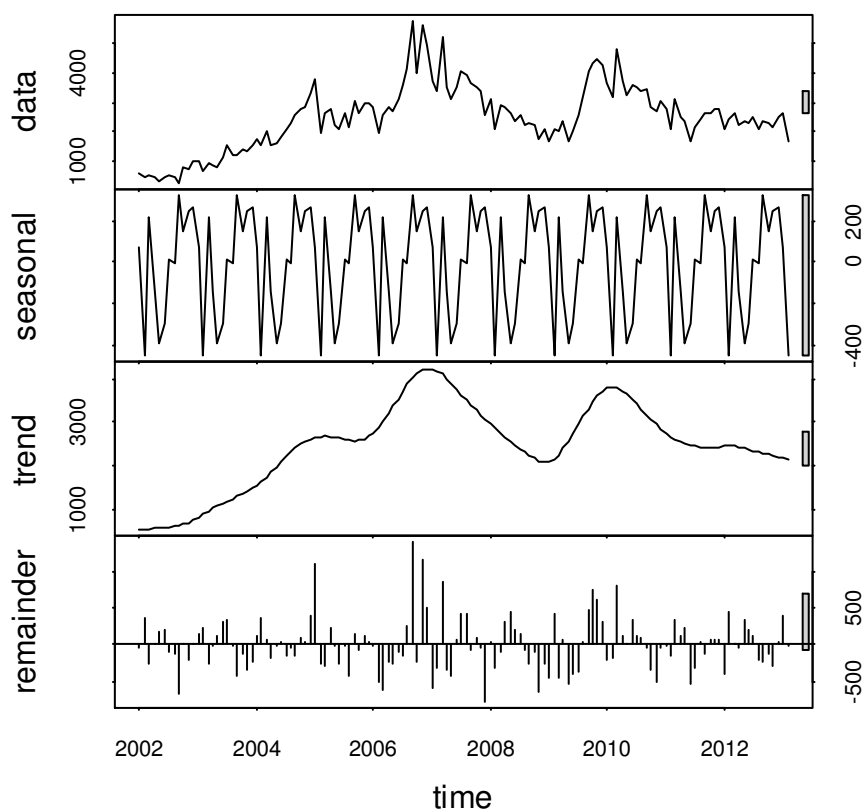
#### ΜΕΘΟΔΟΣ ΑΠΟΣΥΝΘΕΣΗΣ

Οι συνιστώσες της χρονοσειράς όπως προέκυψαν από τη μέθοδο STL φαίνονται στην εικόνα 5.3-9. Η εκτίμηση των προβλέψεων έγινε με εφαρμογή της απλής μεθόδου naïve.

### Forecasts from STL + Random walk



Εικόνα 5.3-8: Προβλέψεις των πωλήσεων με εφαρμογή της μεθόδου αποσύνθεσης STL



Εικόνα 5.3-9: Συνιστώσες των δεδομένων πωλήσεων με τη μέθοδο αποσύνθεσης STL

Η διασπορά των τιμών σύμφωνα με τη μέθοδο αποσύνθεσης φαίνεται να εξηγείται κατά 5% από την εποχικότητα, 75% από την τάση και 10 % από τις τυχαίες διακυμάνσεις.

#### ΜΕΘΟΔΟΣ HOLT-WINTERS

Η ανάλυση βρίσκεται στην ενότητα 4.2.2-4

### 5.3.1 Σύγκριση μεθόδων και επιλογή μοντέλου

Τα σφάλματα πρόβλεψης για την κάθε μέθοδο που μελετήθηκε περιέχονται στον πίνακα που ακολουθεί:

Μέθοδος	RMSE	MAE	MPE	MAPE	ACF1	Theil.s.U
ARIMA	567.2625	447.0572	6.622299	18.32243	0.675377	1.243469
STL	621.5039	494.9558	1.253883	21.41385	0.741419	1.508634
HW	765.1015	627.9107	-18.639	31.33061	0.807335	2.347191
Naive	839.5401	677.1765	22.13108	25.97584	0.691105	1.71884

Το καλύτερο μοντέλο δείχνει να είναι το υπόδειγμα  $ARIMA(0,1,5)(1,0,2)_{12}$ .

## 5.4 Γενικά συμπεράσματα

Από τις 3 παραπάνω εφαρμογές μπορούμε να διαπιστώσουμε ότι σε γενικές γραμμές τα υποδείγματα ARIMA είχαν καλές επιδόσεις σε σχέση με τα υπόλοιπα μοντέλα. Φυσικά η εύρεση του καταλληλότερου μοντέλου δεν είναι εύκολη διαδικασία. Υπάρχουν πολλοί παράμετροι που παίζουν ρόλο και επηρεάζουν τις προβλέψεις και φαίνεται να υπάρχει περιθώριο βελτίωσης των μοντέλων που εκτιμήσαμε, αν κρίνουμε και από τα σφάλματα πρόβλεψης.

Κάτι που θα μπορούσε να γίνει θα ήταν να εφαρμόσουμε και άλλες μεθόδους, ώστε να δούμε αν τα υποδείγματα ARIMA έχουν πράγματι πολύ καλές επιδόσεις ως προς τα δεδομένα που επιλέξαμε να μελετήσουμε. Επίσης για τις μεθόδους που ήδη έχουμε επιλέξει, θα μπορούσαμε ενδεχομένως να πειραματιστούμε περισσότερο με τις αρχικές συνθήκες για την Holt-Winters και με τις τάξεις των υποδειγμάτων για τα μοντέλα ARIMA.

## Παράρτημα

### Ενότητα 4.1

#### Δεδομένα κατανάλωσης ηλεκτρικής ενέργειας

Αρχείο elec\_ma.r

Αναφέρεται στη μέθοδο απλού κινητού μέσου.

```
library(forecast)
el <- read.csv("el_con.csv", header=TRUE, stringsAsFactors=FALSE)
str(el)

#create time series
y=ts(el[,2],frequency=12, start=2005)

#Figure 4.1-1
plot(y,ylab="Energy Consumption (GWh)")

# Calculate moving averages
ma3 <- ma(y, order=3, centre=FALSE)
ma5 <- ma(y, order=5, centre=FALSE)
ma7 <- ma(y, order=7, centre=FALSE)
ma9 <- ma(y, order=9, centre=FALSE)

#extract data
ma=list(ma3,ma5,ma7,ma9)
MA=Reduce(rbind,ma)
row.names(MA)=c("ma3","ma5","ma7","ma9")
library(xlsx)
write.xlsx(MA,"elecma.xlsx")

#Figure 4.1-2
par(mfrow=c(2,2))
plot(ma3, ylab="Energy Consumption (GWh)", main="3MA", col=2, lwd=3)
lines(y, ylab="Energy Consumption (GWh)")
plot(ma5, ylab="Energy Consumption (GWh)", main="5MA", col=3, lwd=3)
lines(y, ylab="Energy Consumption (GWh)")
plot(ma7, ylab="Energy Consumption (GWh)", main="7MA", col=4, lwd=3)
lines(y, ylab="Energy Consumption (GWh)")
plot(ma9, ylab="Energy Consumption (GWh)", main="9MA", col=5, lwd=3)
lines(y, ylab="Energy Consumption (GWh)")
```

Αρχείο elec\_wma.r

Αναφέρεται στη μέθοδο σταθμισμένου κινητού μέσου

```
library(forecast)
el <- read.csv("el_con.csv", header=TRUE, stringsAsFactors=FALSE)
str(el)
y=ts(el[,2],frequency=12, start=2005)
```

```

#Create function for weighted moving average
wmv = function (s, weights, order){
  ## default order
  if(is.na(order)==T){
    order =2
  }
  ## check if weight vector and order are same
  if(length(weights)!=order){
    stop("Different order and weight counts")
  }
  wm.out = numeric()
  for (i in 1:(length(s)-order+1)){

    w.sum = sum(weights)# weight sum
    w = weights/w.sum # standardize weights
    s1 = s[i:(i+order-1)]
    wm = as.vector(weighted.mean(s1,w) )
    wm.out[i] = wm
  }
  # add fist elements as missing..
  m = numeric()
  for(p in 1:(order-1)){
    m[p] = NA
  }
  wm.out = append(m,wm.out)
  wm.out
}

# Weighted Moving Average with order 5
v5 = c(1,2,3,2,1)
wma.5 = wmv(y,v5,5)
# Weighted Moving Average with order 3
v3 = c(1,2,1)
wma.3 = wmv(y,v3,3)

#extract data
wma=list(wma.3,wma.5)
WMA=Reduce(rbind,wma)
row.names(WMA)=c("wma3","wma5")
library(xlsx)
write.xlsx(WMA,"elecwma.xlsx")

#turns wma to time series objects
wma.5=ts(wma.5,frequency=12, start=2005)
wma.3=ts(wma.3,frequency=12, start=2005)

#plots wma.3 and wma.5(Figure 4.1-3)
plot(y, ylab="Energy Consumption (GWh)")
lines(wma.3, main="3MA", col=2, lwd=2, lty=1)
lines(wma.5, main="5MA", col=4, lwd=2, lty=1)
legend(x="topright", legend=c('data','3WMA','5WMA'), col=c(1,2,4), lty=c(1,1,1))

```

Αρχείο elec\_Loess.r

Εφαρμογή μεθόδου τοπικής παλινδρόμησης Loess.

```
library(forecast)
el <- read.csv("el_con.csv", header=TRUE, stringsAsFactors=FALSE)
str(el)
y=ts(el[,2],frequency=12, start=2005)

#creates a vector with elements the time of the times series
x=as.numeric(time(y))

#loess method
y.loess=loess(y~x, span=3)
y.loess1=loess(y~x, span=10)
y.loess2=loess(y~x, span=22)
y.predict=predict(y.loess)
y.predict1=predict(y.loess1)
y.predict2=predict(y.loess2)

#Figure 4.1-4
plot(x,y, xlab="time", ylab= "Energy Consumption (GWh)",main="Loess Local
Regression")
lines(x,y.predict, col=2)
lines(x,y.predict1, col=3)
lines(x,y.predict2, col=4)
legend(x="topright", c("data",expression(k == 3, k == 10, k == 22)), col=c(1,2,3,4),lty=1,
lwd=2)
```

Αρχείο elec\_classical.r

Περιέχει την εφαρμογή της κλασικής μεθόδου αποσύνθεσης.

```
--Trend--
trend_consumption = ma(y, order = 12, centre = T)
#Figure 4.1-5
plot(as.ts(y), ylab="Energy Consumption (GWh)", main= "trend for electricity
consumption")
lines(trend_consumption)
--Detrend data--
detrend_consumption = y - trend_consumption
#Figure 4.1-6
plot(as.ts(detrend_consumption), ylab="Energy Consumption(GWh)", main= "de-
trended data for electricity consumption")
--Seasonality--
m_consumption = t(matrix(data = detrend_consumption, nrow = 12))
seasonal_consumption = colMeans(m_consumption, na.rm = T)
#Figure 4.1-6
plot(as.ts(rep(seasonal_consumption,11)),ylab="Energy Consumption (GWh)",
```

```

main="seasonal component")
--Irregular series--
random_consumption = y - trend_consumption - seasonal_consumption
#Figure 4.1-7
plot(as.ts(random_consumption))
#final(extra):
recomposed_consumption                                     =
trend_consumption+seasonal_consumption+random_consumption
plot(as.ts(recomposed_consumption))

```

## Ενότητα 4.2

### Δεδομένα πωλήσεων

Αρχείο saleses.r

Αναφέρεται στη μέθοδο απλής εκθετικής εξομάλυνσης. Περιέχει το βέλτιστο αποτέλεσμα που δίνει η μέθοδος καθώς και έναν αλγόριθμο που εξετάζει πιθανές τιμές της  $\alpha$  και επιστρέφει αυτή που ελαχιστοποιεί τα μέτρα εκτίμησης σφαλμάτων MSE και MAPE. Παρατηρείται ότι οι τιμές της βέλτιστης σταθεράς εξομάλυνσης  $\alpha$  για το κάθε μέτρο διαφέρουν. Τέλος, απεικονίζει γραφικά τα δεδομένα και τις προβλέψεις για διάφορες τιμές της παραμέτρου εξομάλυνσης  $\alpha$ . Στα σχήματα διακρίνονται οι προβλέψεις χωριστά.

```

library(forecast)

getwd()
sales <- read.csv("sales.csv", header=TRUE, stringsAsFactors=FALSE)
str(sales)

#Create time series
salests <- ts(sales,frequency=12, start=2002)
str(salests)
n=nrow(salests)

#Simple exponential smoothing for  $\alpha=0,1, 0,9$ , optimal
fit <- ses(salests[,2], initial="optimal", h=1)
summary(fit)
fit1 <- ses(salests[,2], alpha=0.1, initial="simple", h=1)
summary(fit1)
fit2 <- ses(salests[,2], alpha=0.9, initial="simple", h=1)
summary(fit2)

library(Metrics)
y=salests[2:n,2]
y_hat=fitted(fit)[2:n]
y1_hat=fitted(fit1)[2:n]
y2_hat=fitted(fit2)[2:n]

#Compute meansquared error for the optimal solution

```



```

mse(salests[2:n,2],fitted(fit)[2:n])

#Create function that computes the mean absolute percent error
mape <- function(y,y_hat){
+ result <- mean(abs(((y - y_hat)/y)*100))
+ return(result)}

#Compute the mean absolute percent error for the optimal solution
mape( y,y_hat)

# Forecasts for different values of a between (0,1) with one step ahead (computes forecasts
for the next period only(h=1))
library(Metrics)
fit <- c()
i=0
while(i < 1){
f <- ses(salests[,2], alpha=i , initial="simple", h=1)
fi <- c(fi,f$mean)
i=i+0.01
}
> str(fi)

#Finds MSE for different values of a between(0,1)with step 0.01 and returns the minimum
#(result shows: a=0.56)

MSE <- c()
i=0
while(i < 1){
f <- ses(salests[,2], alpha=i , initial="simple", h=1)
MSE <- c(MSE,mse(salests[2:n,2],fitted(f)[2:n]))
i=i+0.01
}
a=seq(0,1,0.01)
str(a)
names(MSE)=a[1:100]
number<-which.min(MSE)
number
MSE[number]

#Finds MAPE for different values of a between (0,1) with step 0.01 and returns the
minimum
#(result shows: a=0.53)

MAPE <- c()
i=0
while(i < 1){
f <- ses(salests[,2], alpha=i , initial="simple", h=1)
MAPE <- c(MAPE,mape(salests[2:n,2],fitted(f)[2:n]))
i=i+0.01
}
a=seq(0,1,0.01)
str(MAPE)

```

```

names(MAPE)= a[1:100]
num <- which.min(MAPE)
MAPE[num]

#Plots
#Picture 4.2-6
plot.ts(salests[,2], ylab="Sales", type="o")
lines(fitted(fit3p),col=4,type="o")
lines(fitted(fit1p),col=3,type="o")
legend("topleft",lty=1, col=c(1,"green","blue"),
+ c("data", expression(alpha == 0.1), expression(alpha == 0.9)),pch=1)

#Picture 4.2-7
plot.ts(salests[,2],xlim=c(2012,2016),ylab="Sales", type="b")
lines(fit2$mean, col="blue", type="b", xlim=c(2012,2016))
lines(fit1$mean, col="green", type="b", xlim=c(2012,2016))
lines(fit$mean, col="red", type="b",xlim=c(2012,2016))
legend("topleft",lty=1, col=c(1,"red","green","blue"), c("data",expression(alpha
==0.5558), expression(alpha == 0.1), expression(alpha == 0.9)),pch=1)

```

#### Αρχείο salesARRSES.r

Αναφέρεται στη μέθοδο απλής εκθετικής εξομάλυνσης με αναπροσαρμοζόμενο ρυθμό ανταπόκρισης (ARRSES /Adaptive-Response-Rate SES). Η εφαρμογή έγινε σε δεδομένα πωλήσεων μιας εταιρίας το χρονικό διάστημα 2002-2015. Εκτός από τις παρατηρήσεις, περιέχει τον κώδικα για τις προβλέψεις που δίνει η μέθοδος μέσα από έναν αλγόριθμο όπου υπολογίζει και τη σταθερά εξομάλυνσης για κάθε παρατήρηση. Περιλαμβάνονται και οι εντολές για τα στατιστικά μέτρα εκτίμησης σφαλμάτων MSE και MAPE. Τέλος, απεικονίζει γραφικά τα δεδομένα και τις προβλέψεις.

```

library(forecast)

getwd()
sales <- read.csv("sales.csv", header=TRUE, stringsAsFactors=FALSE)
str(sales)

#Create time series
salests <- ts(sales, frequency=12, start=2002)
str(salests)

#Calculate ARRSES for fixed value of beta (0.2)
y=salests[,2]
beta=0.2
n=length(y)
alpha=rep(0,length(y))
f=rep(0,length(y))
a=rep(0,length(y))

```

```

m=rep(o,length(y))
e=rep(o,length(y))
f[1]=y[1]
a[1]=o
m[1]=o
e[1]=o
alpha[1]=0.2
alpha[2]=0.2

for(j in 2:length(y)){
  f[j]=f[(j-1)]+alpha[(j-1)]*e[(j-1)]
  e[j]=y[j]-f[j]
  a[j]=a[(j-1)]+beta*(e[(j)]-a[(j-1)])
  m[j]=m[(j-1)]+beta*(abs(e[j])-m[(j-1)])
  alpha[(j+1)]=abs(a[(j)]/m[(j)])
}
#Return Forecast
forecast=f[n]+alpha[(n-1)]*e[(n-1)]
forecast

#Calculate MSE
library(Metrics)
mse(y[2:n],f[2:n])

#Calculate MAPE
mape <- function(y,y_hat){
+ result <- mean(abs(((y - y_hat)/y)*100))
+ return(result)}
mape(y[2:n],f[2:n])

#Finds ARRSES for different values of beta between (0,0.5) and returns the value that
minimizes MSE

Library(Metrics)
y=salests[,2]
beta=c()
n=length(y)
alpha=rep(o,length(y))
f=rep(o,length(y))
a=rep(o,length(y))
m=rep(o,length(y))
e=rep(o,length(y))
f[1]=y[1]
a[1]=o
m[1]=o
e[1]=o
alpha[1]=0.2
alpha[2]=0.2
MSE=c()
i=o
while (i < 0.5){

```

```

for(j in 2:length(y)){
f[j]=f[(j-1)]+alpha[(j-1)]*e[(j-1)]
e[j]=y[j]-f[j]
a[j]=a[(j-1)]+i*(e[(j)]-a[(j-1)])
m[j]=m[(j-1)]+i*(abs(e[j])-m[(j-1)])]
alpha[(j+1)]=abs(a[(j)]/m[(j)])}
MSE=c(MSE,mse(y[2:n],f[2:n]))
beta=c(beta,i)
i=i+0.01
}
r=seq(0,0.5,0.01)
str(r)
names(MSE)=a[1:length(r)]
number<-which.min(MSE)
number

#Plots
#Picture 4.2-8
fi <- ts(f[2:n], start=c(2002, 2), end=c(2015, 12), frequency=12)
plot.ts(y, ylab="Sales", type="b")
lines(fi,col=4,type="b")
legend(x='topright', c("data",expression(beta==0.27)), lwd=2, lty=1, col=c(1,'blue'))
>
#Pictyre 4.2-9
alphat <- ts(alpha, start=c(2002, 2), end=c(2016, 1), frequency=12)
plot.ts(alphat, main="Values of alpha(t)", ylab="alpha(t)")

```

#### Αρχείο salesHolt.r

Αναφέρεται στη μέθοδο του Holt (εκθετική εξομάλυνση με προσαρμογή στην τάση). Η εφαρμογή έγινε σε δεδομένα πωλήσεων μιας εταιρίας το χρονικό διάστημα 2002-2015. Περιέχει τον κώδικα για την επεξεργασία των παρατηρήσεων και τις προβλέψεις που δίνει η μέθοδος. Η τελική πρόβλεψη γίνεται για 5 μελλοντικές περιόδους. Γίνεται εφαρμογή τριών διαφορετικών μοντέλων με επιλογή αυτού που ελαχιστοποιεί τα σφάλματα. Περιλαμβάνονται και οι εντολές για διάφορα στατιστικά μέτρα εκτίμησης σφαλμάτων. Τέλος, απεικονίζει γραφικά τα δεδομένα και τις προβλέψεις .

```

library(forecast)
library(Metrics)

getwd()
sales <- read.csv("sales.csv", header=TRUE, stringsAsFactors=FALSE)
str(sales)

#Create time series
salests <- ts(sales,frequency=12, start=2002)
str(salests)
# Parameters for Option 2

```

```

y=salests[,2]
n=length(y)

#Computes values of alpha and beta that minimize MSE
alpha=c()
beta=c()
f=rep(0,n)
l=rep(0,n)
b=rep(0,n)
l[1]=y[1]
b[1]=y[2]-y[1]
f[1]=l[1]+b[1]
MSE=c()
for (i in 1:99){
  i=i/100
  for (j in 1:99){
    j=j/100
    for(k in 2:n){
      l[k]=i*y[k]+(1-i)*(l[k-1]+b[k-1])
      b[k]=j*(l[k]-l[k-1])+(1-j)*b[k-1]
      f[k]=l[k]+b[k]
    }
    MSE=c(MSE,mse(y[2:n],f[1:n-1]))
    beta=c(beta,j)
    alpha=c(alpha,i)
  } }

w=seq(1,length(MSE))
names(MSE)=w[1:length(w)]
number=which.min(MSE)
number
alpha[number]
beta[number]
rm(y)

#Divide data into two sets
y <-window(salests[,2],end=c(2013,2))
y_test<-window(salests[,2],start=c(2013,3))

#parameters for option 3
y1=window(y,start=c(2002,1), end=c(2003,12))
> fit1=tslm(y1~trend)
> summary(fit1)

#Fit models
models<-list(
  mod_1=holt(y,initial="optimal", h=34),
  mod_2=holt(y,alpha=0.58,beta=0.04,initial="simple", h=34),
  mod_3=HoltWinters(y,gamma=FALSE, l.start=232.569, b.start=48.148))

#Create forecasts

```

```

forecasts <- lapply(models, forecast, 34)

#Check Errors
acc <- lapply(forecasts, function(f){
  accuracy(f, y_test)[2,,drop=FALSE]
})

#Plots
#Picture 4.2-10
par(mfrow=c(3,1))
for (f in forecasts){
  plot(f)
  lines(y_test, col='red')
}
#Picture 4.2-11
fit=HoltWinters(y,alpha=0.5807, beta=1e-04 gamma=FALSE, l.start=477.234,
b.start=11.5558 )
plot(forecast(fit,h=5), xlab="Time", ylab="Sales")
lines(fit$fitted[,2], col=4, lty=5, lwd=2)
legend(x="topright", c("data",expression(alpha==0.58, beta == 1e-04)),
col=c(1,"blue","blue"), lty=c(1,5,5), lwd=2)

#Picture 4.2-12
plot(fit$fitted, main="Results using Holts Method on sales data")

```

#### Αρχείο salesHoltWinters.r

Αναφέρεται στη μέθοδο Holt-Winters. Η εφαρμογή έγινε σε δεδομένα πωλήσεων μιας εταιρίας το χρονικό διάστημα 2002-2015. Περιέχει τον κώδικα για την επεξεργασία των παρατηρήσεων και τις προβλέψεις που δίνει η μέθοδος. Η τελική πρόβλεψη είναι για 5 μελλοντικές περιόδους. Γίνεται εφαρμογή δύο διαφορετικών μοντέλων με επιλογή αυτού που ελαχιστοποιεί τα σφάλματα. Περιλαμβάνονται και οι εντολές για διάφορα στατιστικά μέτρα εκτίμησης σφαλμάτων. Τέλος, απεικονίζει γραφικά τα δεδομένα και τις προβλέψεις .

```

library(forecast)
library(Metrics)
sales <- read.csv("sales.csv", header=TRUE, stringsAsFactors=FALSE)
str(sales)
salests <- ts(sales,frequency=12, start=2002)
y=salests[,2]

#Divide data into two sets(initial and test set)
y_i <-window(y,end=c(2013,2))
y_test<-window(y,start=c(2013,3))
n=length(y_i)

f=rep(o,n)

```

```

l=rep(o,n)
b=rep(o,n)
S=rep(o,n)

#length of seasonality
s=12

#Initial values
l[s]=mean(y[1:s])

e=rep(o,2*s)
for (i in 1:2*s){
  e[i]=(y[s+i]-y[i])/s
}
b[s]=mean(e)

for(i in 1:s){
  S[i]=y[i]-l[12]
}

#Function for computing MSE using the Holt-Winters method
MSE=function(i,j,m){
  for(k in 13:n){
    l[k]=i*(y[k]/S[k-s])+(1-i)*(l[k-1]+b[k-1])
    b[k]=j*(l[k]-l[k-1])+(1-j)*b[k-1]
    S[k]=m*(y[k]/l[k])+(1-m)*S[k-s]
    f[k]=(l[k]+b[k])*S[k]}
  mse(y[s:n],f[s:n])
}

#return best values for parameters alpha,beta,gamma by minimizing MSE
best_alpha=-1
best_beta=-1
best_gamma=-1
best_MSE=10000000000
current_MSE=700000
for (i in 1:99){
  i=i/100
  for (j in 1:99){
    j=j/100
    for (m in 1:99){
      m=m/100
      current_MSE=MSE(i,j,m)
      if( current_MSE< best_MSE){
        best_MSE=current_MSE
        best_alpha=i
        best_beta=j
        best_gamma=m
      }
    }
  }
}

```

```

}

#Fit models for additive seasonality
models<-list(
  mod_1=HoltWinters(y_i),
  mod_2=HoltWinters(y_i, l.start=l[s], b.start=b[s], s.start=S[1:s]))
#---best model fitted over all observations
fit1=HoltWinters(y, alpha=0.5665786,beta=0,gamma=0.3859994, l.start=l[12],
b.start=b[12], s.start=S[1:12])

#Create forecasts
forecasts <- lapply(models, forecast, 34)

#Check Errors
acc <- lapply(forecasts, function(f){
  accuracy(f, y_test)[2,,drop=FALSE]
})

#Plots

#Picture 4.2-13
par(mfrow=c(2,1))
plot(forecasts$mod_1, main="alpha=0.58,beta=0.01,gamma=0.39")
lines(y_test, col='red')
plot(forecasts$mod_2, main="alpha=0.567,beta=0,gamma=0.386")
lines(y_test, col='red')

#Picture 4.2-14
plot(forecast(fit1,h=5), xlab="Time", ylab="Sales")
lines(fit1$fitted[,1], col="red", lty=1)
legend(x="topright", c("data",expression(alpha==0.57, beta == 0, gamma == 0.36)),
col=c(1,"red","red","red"), lty=c(1,1,1,1), lwd=2)
>
#Fit models for multiplicative seasonality
rm(S)
for(i in 1:s){
  S[i]=y[i]/l[12]
}

models1<-list(
  model_1=HoltWinters(y_i, seasonal="mult"),
  model_2=HoltWinters(y_i, l.start=l[12], b.start=b[12], s.start=S[1:12], seasonal="mult"))
#---best model fitted over all observations
fit2=HoltWinters(y, alpha=0.3914628,beta=0,gamma=0.8528836, l.start=l[12],
b.start=b[12], s.start=S[1:12])

#Create forecasts
forecasts1 <- lapply(models1, forecast, 34)

#Check Errors
acc1 <- lapply(forecasts1, function(f){

```



```

accuracy(f, y_test)[2,,drop=FALSE]
})

#Plots
#Picture 4.2-15
par(mfrow=c(2,1))
plot(forecasts$model_1, main="alpha=0.42,beta=0.005,gamma=0.87")
lines(y_test, col='red')
plot(forecasts$model_2, main="alpha=0.39,beta=0,gamma=0.85")
lines(y_test, col='red')

#Picture 4.2-16
plot(forecast(fit2,h=5), xlab="Time", ylab="Sales")
lines(fit2$fitted[,1], col="red", lty=1)
legend(x="topright", c("data",expression(alpha==0.39, beta == 0, gamma == 0.85)),
col=c(1,"red","red","red"), lty=c(1,1,1,1), lwd=2)

#Picture 4.2-17
plot(fit2$fitted, main="Estimation of time series components" )

```

## Ενότητα 5.1

### Αρχείο tourists\_final.r

Περιέχει τις μεθοδολογίες ARIMA, STL decomposition και Holt-Winters. Η εφαρμογή έγινε σε δεδομένα τουριστών που επισκέπτονται την Ελλάδα το χρονικό διάστημα 2001-2016. Ο κώδικας αφορά την επεξεργασία των παρατηρήσεων, τις εκτιμήσεις των παραμέτρων των μοντέλων, τις προβλέψεις που δίνουν οι μέθοδοι για το test set και τις τελικές προβλέψεις για 5 μελλοντικές περιόδους. Απεικονίζει γραφικά τα δεδομένα και τις προβλέψεις για κάθε μέθοδο και στο τέλος γίνεται η σύγκριση των σφαλμάτων πρόβλεψης.

```

library(forecast)
tour=read.csv("non_tour.csv", header=TRUE, stringsAsFactors=FALSE)
str(tour)
y=ts(tour[,2],frequency=12,start=2001)

#Seasonplot
seasonplot(y,12, col=c(1,2,3,4,5,6,7,8,9,10,11,12,13,14,15,16))
legend("topleft",c("2001","2002","2003","2004","2005","2006","2007","2008","2009","2010",
,"2011","2012","2013","2014","2015","2016"),col=c(1,2,3,4,5,6,7,8,9,10,11,12,13,14,15,16),lty=1)

#divide data to 2 sets
y_i <-window(y,end=c(2013,2))
y_test<-window(y,start=c(2013,3))
n=length(y_i)

----ARIMA method----

#Takes the log of the observations

```

```

x=BoxCox(y_i,o)

#Picture 5.1-3
plot(x,ylab="Number_of_Tourists")

#Picture 5.1-4
par(mfcol=c(1,2))
Acf(x, lag.max=48, main="")
Pacf(x, lag.max=48, main="")

#Picture 5.1-5
tsdisplay(diff(x,12),main="Data after a seasonal difference")

#Hypothetical model
tour1=Arima(y_i, order=c(2,o,o), seasonal=c(1,1,o),lambda=o)
tour1
#Best model
fit <- auto.arima(y_i, lambda=o, d=o, D=1, max.order=6, stepwise=FALSE,
approximation=FALSE)

#Picture 5.1-6
tsdisplay(residuals(fit))

#Box Pierce test
Box.test(residuals(fit), lag=36, fitdf=4, type="Box-Pierce")

#Picture 5.1-7
plot(forecast(fit,h=37,lambda=o), ylab="Number_of_Tourists")

fit_total=Arima(y, order=c(1,o,o), seasonal=c(2,1,1),lambda=o)
fi=forecast(fit_total,h=5,lambda=o)
#Picture 5.1-8
plot(fi, ylab="Number_of_Tourists")
lines(fi$fitted, col="blue", lwd=2)

---STL decomposition---

d <- stl(y_i,s.window="periodic")

#Figure 5.1-9
plot(d)
fcast=forecast(d,method="naive", h=37)
#Figure 5.1-10
plot(fcast)

#the fraction of the variance which corresponds to each component
apply(d$time.series, 2, var) / var(y_i)

---Holt-Winters method---

#set initial level
l=rep(o,n)

```

```

b=rep(o,n)
S=rep(o,n)
s=12
l[12]=mean(y[1:s])

e=rep(o,2*s)
for (i in 1:2*s){
  e[i]=(y[s+i]-y[i])/s
}
b[12]=mean(e)
for(i in 1:s){
  S[i]=y[i]/l[12]
}

#method for 2 cases
mod_1=HoltWinters(y_i)
mod_2=HoltWinters(y_i, l.start=l[12], b.start=b[12], s.start=S[1:12])
hwforecast=forecast(mod_1, h=37)
hwforecast1=forecast(mod_2, h=37)

#picture 5.1-11
plot(hwforecast)
lines(y_test, col='red')

#picture 5.1-12
Acf(hwforecast$residuals, lag.max=48)

#picture 5.1-13
plot(hwforecast1)
lines(y_test, col='red')

#picture 5.1-14
Acf(hwforecast1$residuals, lag.max=48)

#comparison of the 2 cases
models<-list(
mod_1=HoltWinters(y_i),
mod_2=HoltWinters(y_i, l.start=l[12], b.start=b[12], s.start=S[1:12]))

forecasts <- lapply(models, forecast, 37)

acc <- lapply(forecasts, function(f){
  accuracy(f, y_test)[2,,drop=FALSE]
})
acc <- Reduce(rbind, acc)
row.names(acc) <- names(forecasts)
acc <- acc[order(acc[, 'RMSE']),]
round(acc, 2)

----Final comparison concerning the forecast errors of each method----

```

```

models1<-list(
model_1=Arima(y_i, order=c(1,0,0), seasonal=c(2,1,1),lambda=0),
model_2=stl(y_i,s.window="periodic"),
model_3=HoltWinters(y_i))

forecasts1 <- list(
f1=forecast(models1$model_1, h=37),
f2=forecast(models1$model_2,method="naive", h=37),
f3=forecast(models1$model_3, h=37),
f4=naive(y_i, h=37))

acc <- lapply(forecasts1, function(f){
  accuracy(f, y_test)[2,,drop=FALSE]
})
acc <- Reduce(rbind, acc)
row.names(acc) <- names(forecasts1)
acc <- acc[order(acc[, 'RMSE']),]
round(acc, 2)

```

Αρχείο elec\_final.r

Περιέχει τις μεθοδολογίες ARIMA, STL decomposition και Holt-Winters που εφαρμόστηκαν σε δεδομένα (μηνιαία) κατανάλωσης ηλεκτρικής ενέργειας της Ελλάδας το χρονικό διάστημα Ιαν2005-Δεκ2015. Ο κώδικας αφορά την επεξεργασία των παρατηρήσεων, τις εκτιμήσεις των παραμέτρων των μοντέλων, τις προβλέψεις που δίνουν οι μέθοδοι για το test set και τις τελικές προβλέψεις για 5 μελλοντικές περιόδους. Απεικονίζει γραφικά τα δεδομένα και τις προβλέψεις για κάθε μέθοδο και στο τέλος γίνεται η σύγκριση των σφαλμάτων πρόβλεψης.

```

library(forecast)
el <- read.csv("el_con.csv", header=TRUE, stringsAsFactors=FALSE)
str(el)
y=ts(el[,2],frequency=12, start=2005)
y_i <-window(y,end=c(2013,10))
y_test<-window(y,start=c(2013,11))
plot(y, ylab="Energy Consumption (GWh)", main="Electricity Consumption")

#Seasonplot-Figure 5.2-1
seasonplot(y,12, col=c(1,2,3,4,5,6,7,8,9,10,11),year.labels=TRUE, lty=1)

---ARIMA method---

#Figure 5.2-2
par(mfcol=c(1,2))
Acf(y_i, lag.max=48, main="")
Pacf(y_i, lag.max=48, main="")

#plot seasonal difference- Figure 5.2-3
tsdisplay(diff(y_i,12),main="Data after a seasonal difference")

```

```

#Hypothetical model
id=Arima(y_i, order=c(1,0,1), seasonal=c(1,1,1))

#Best model
fit <- auto.arima(y_i, d=0, D=1, max.order=8,stepwise=FALSE, approximation=FALSE)
---Results--
>fit
Series: y_i
ARIMA(1,0,1)(0,1,1)[12]

Coefficients:
      ar1      ma1      sma1
      0.9792 -0.7737 -0.8139
s.e. 0.0321 0.0796 0.1620

sigma^2 estimated as 42858: log likelihood=-639.17
AIC=1286.35 AICc=1286.8 BIC=1296.52
-----
#Figure 5.2-5
tsdisplay(residuals(fit))

#Box Pierce test
Box.test(residuals(fit), lag=26, fitdf=3, type="Box-Pierce")

#Figure 5.2-6
plot(forecast(fit,h=37), ylab="Energy Consumption (GWh)")

#table 5.2-1
fo=forecast(fit,h=26)
library(xlsx)
write.xlsx(fo$mean,"consumption.xlsx")
write.xlsx(y_i,"consumption.xlsx")

#Figure 5.2-7
fit_total=Arima(y, order=c(1,0,1), seasonal=c(0,1,1))
fi=forecast(fit_total,h=5)
plot(fi, ylab="Energy Consumption (GWh)")
lines(fi$fitted, col="red", lwd=1)
--extra--
#future values
fi$mean

----STL decomposition ---

d <- stl(y_i,s.window="periodic")
#Figure 5.2-9
plot(d)
fcast=forecast(d,method="naive", h=37)
#Figure 5.2-8
plot(fcast)

```

```

#the fraction of the variance which corresponds to each component
apply(d$time.series, 2, var) / var(y_i)

----Holt-Winters ----
--Case 1--
mod_1=HoltWinters(y_i)
hwforecast=forecast(mod_1, h=26)

#Figure 5.2-10
plot(hwforecast)

#Figure 5.2-11
Acf(hwforecast$residuals, lag.max=48)

--Case 2--
#Initial level
l=rep(o,n)
b=rep(o,n)
S=rep(o,n)
s=12
l[12]=mean(y[1:s])

e=rep(o,2*s)
for (i in 1:2*s){
  e[i]=(y[s+i]-y[i])/s
}
b[12]=mean(e)
for(i in 1:s){
  S[i]=y[i]-l[12]
}
mod_2=HoltWinters(y_i, l.start=l[12], b.start=b[12], s.start=S[1:12])
hwforecast1=forecast(mod_2, h=26)

#Figure 5.2-12
plot(hwforecast1)

#Figure 5.2-13
Acf(hwforecast1$residuals, lag.max=48)

#Comparison for the forecast errors of each case

models<-list(mod_1,mod_2)
forecasts <- lapply(models, forecast, 26)
acc <- lapply(forecasts, function(f){
  accuracy(f, y_test)[2,,drop=FALSE]
})
acc <- Reduce(rbind, acc)
row.names(acc) <- names(forecasts)
acc <- acc[order(acc[, 'MASE'],)]
round(acc, 2)

#Picture 5.2-14

```

```
plot(hwforecast1, xlab="Time", ylab="Energy Consumption (GWh)")
lines(mod_2$fitted[,1], col="red", lty=1)
```

--- Final comparison concerning the forecast errors of each method ---

```
models<-list(
model_1=Arima(y_i, order=c(1,0,1), seasonal=c(0,1,1)),
model_2=stl(y_i,s.window="periodic"),
model_3=HoltWinters(y_i, l.start=l[12], b.start=b[12], s.start=S[1:12]))
```

```
forecasts <- list(
f1=forecast(models$model_1, h=26),
f2=forecast(models$model_2,method="naive", h=26),
f3=forecast(models$model_3, h=26),
f4=naive(y_i, h=37))
```

```
acc <- lapply(forecasts, function(f){
  accuracy(f, y_test)[2,,drop=FALSE]
})
acc <- Reduce(rbind, acc)
row.names(acc) <- names(forecasts)
acc <- acc[order(acc[, 'RMSE']),]
round(acc, 2)
```

#### Αρχείο sales\_final.r

Περιέχει τις μεθοδολογίες ARIMA και STL decomposition που εφαρμόστηκαν σε δεδομένα (μηνιαία) πωλήσεων μιας εταιρίας το χρονικό διάστημα Ιαν2002-Δεκ2015 (τα αποτελέσματα της μεθόδου Holt-Winters βρίσκονται στο αρχείο salesHoltWinters.r). Ο κώδικας αφορά την επεξεργασία των παρατηρήσεων, τις εκτιμήσεις των παραμέτρων των μοντέλων, τις προβλέψεις που δίνουν οι μέθοδοι για το test set και τις τελικές προβλέψεις για 5 μελλοντικές περιόδους. Απεικονίζει γραφικά τα δεδομένα και τις προβλέψεις για κάθε μέθοδο και στο τέλος γίνεται η σύγκριση των σφαλμάτων πρόβλεψης.

```
library(forecast)
sales <- read.csv("sales.csv", header=TRUE, stringsAsFactors=FALSE)
str(sales)
salests <- ts(sales,frequency=12, start=2002)
y=salests[,2]
y_i <-window(y,end=c(2013,2))
y_test<-window(y,start=c(2013,3))

#Seasonal plot - Figure 5.3-1
seasonplot(y, 12, col=c(1,2,3,4,5,6,7,8,9,10,11,12,13,14),year.labels=TRUE, lty=1,
main="Seasonal plot")
```

--- **ARIMA method**---

```
#Figure 5.3-2
par(mfcol=c(1,2))
Acf(y_i, lag.max=48, main="")
Pacf(y_i, lag.max=48, main="")

#apply first differences
y1=diff(y_i,1)

#Figure 5.3-3
tsdisplay(diff(y_i,1),main="Data after first differencing")

#Hypothetical model
id=Arima(y_i, order=c(2,1,0), seasonal=c(0,0,3))

#Best model
fit <- auto.arima(y_i, d=1, D=0, max.order=9, stepwise=FALSE, approximation=FALSE)

#Figure 5.3-5
tsdisplay(residuals(fit))

#Box Pierce test
Box.test(residuals(fit), lag=26, fitdf=8, type="Box-Pierce")

#Figure 5.3-6
plot(forecast(fit,h=34), ylab="Sales")
#table 5.3-1
fo=forecast(fit,h=34)
library(xlsx)
write.xlsx(fo$mean,"salesarima.xlsx")
write.xlsx(y_i,"salearima.xlsx")

fit_total=Arima(y, order=c(0,1,5), seasonal=c(1,0,2))
fi=forecast(fit_total,h=5)
#Figure 5.3-7
plot(fi, ylab="Sales")
lines(fi$fitted, col="red", lwd=1)

---STL decomposition---

```
d <- stl(y_i,s.window="periodic")
#Figure 5.3-9
plot(d)
fcast=forecast(d, method="naive")
#Figure 5.3-8
plot(fcast)

# the fraction of the variance which corresponds to each component
apply(d$time.series, 2, var) / var(y_i)
```


```



### ---Holt-Winters---

Check salesHoltWinters.r file above

--- Final comparison concerning the forecast errors of each method---

```
models<-list(  
model_1=Arima(y_i, order=c(0,1,5), seasonal=c(1,0,2)),  
model_2=stl(y_i,s.window="periodic"),  
model_3=HoltWinters(y_i))
```

```
forecasts <- list(  
f1=forecast(models$model_1, h=34),  
f2=forecast(models$model_2,method="naive", h=34),  
f3=forecast(models$model_3, h=34),  
f4=naive(y_i, h=34))
```

```
acc <- lapply(forecasts, function(f){  
  accuracy(f, y_test)[2,,drop=FALSE]  
})  
acc <- Reduce(rbind, acc)  
row.names(acc) <- names(forecasts)  
acc <- acc[order(acc[, 'RMSE']),]  
round(acc, 2)
```

# Βιβλιογραφία

## Ξένα

Diebold, F.X. Elements of forecasting.-4<sup>th</sup> Edition. United States of America: Thomson South-Western

Dr. Mahendra, R.G. Industrials statistics and operational management  
6: Forecasting Techniques.  
<http://docplayer.net/7341442-Industrial-statistics-and-operational-management.html>

Hillier F.S. & Lieberman G.J. Introduction to Operations Research, 7<sup>th</sup> Edition.  
United States of America McGraw Hill

Hyndman, R.J. & Athanasopoulos, G. Forecasting: principles and practice  
<https://www.otexts.org/book/fpp>

Makridakis, S. & Hyndman, R.J. Wheelwright, S.C. (1998). Forecasting: Methods and applications.-3<sup>rd</sup> Edition. United States of America: John Wiley & Sons.

Makridakis, S & Wheelwright, S.C. Forecasting methods for management. United States of America: John Wiley & Sons.

R.B. Cleveland, W.S. Cleveland, J.E. McRae, and I. Terpenning, "STL: A Seasonal-Trend Decomposition Procedure Based on Loess," Journal of Official Statistics, Vol. 6, No.1, pp. 3-73, 1990.

## Ελληνική

Καρκάζης, Ι. (1988). Εισαγωγή στην επιχειρησιακή έρευνα. Πανεπιστήμιο Αιγαίου. Αθήνα: Το Οικονομικόν.

Κολέτσος, Ι. & Στογιάννης, Δ. (2015). Εισαγωγή Στην Επιχειρησιακή Έρευνα (2η έκδοση). Αθήνα: Εκδόσεις Συμμεών.

Ξηροκόστα, Α.Δ. (1999). Επιχειρησιακή Έρευνα. Αθήνα: Εκδόσεις Συμμετρία.

Πετρόπουλος, Φ. (2011). Επιχειρησιακές προβλέψεις. Αθήνα: Εκδόσεις Συμμετρία.

Πραστάκος, Γ. (1994). Επιχειρησιακή έρευνα: για τη λήψη επιχειρηματικών αποφάσεων. Αθήνα: Εκδόσεις Σταμούλης.

Υψηλάντης, Π.Γ. (2006). Επιχειρησιακή Έρευνα: εφαρμογές στη σημερινή επιχείρηση. Αθήνα : Εκδόσεις Προπομπός.

Jarrett, J. (1996). Μέθοδοι προβλέψεων. Αθήνα: Εκδόσεις Gutenberg.

## **Ιστοσελίδες**

<http://research.stowers-institute.org/efg/R/Statistics/loess.htm>

<http://robjhyndman.com/>

[http://users.uom.gr/~drits/lessons/Lesson\\_4\\_eidika.pdf](http://users.uom.gr/~drits/lessons/Lesson_4_eidika.pdf)

<http://www.quantlego.com/howto/holt-winters-smoothing-and-forecast/>

<https://cran.r-project.org/web/packages/Metrics/Metrics.pdf>

<http://dni-institute.in/blogs/forecasting-moving-average-and-weighted-moving-average/>

<http://a-little-book-of-r-for-time-series.readthedocs.io/en/latest/src/timeseries.html>

<http://www.itl.nist.gov/div898/handbook/pmd/section1/pmd144.htm>

## **Δεδομένα**

[https://www.quandl.com/data/ENTSOE/GR\\_CONS-Electricity-Consumption-Greece](https://www.quandl.com/data/ENTSOE/GR_CONS-Electricity-Consumption-Greece)

<http://www.adata.com/en/about/content-sales-1/>

<http://ec.europa.eu/eurostat/data/database>