

ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΕΙΡΑΙΩΣ

Τμήμα Ψηφιακών Συστημάτων



**ΠΡΟΒΛΕΨΗ ΤΕΧΝΙΚΩΝ ΧΡΟΝΟΣΕΙΡΩΝ ΜΕ  
ΟΙΚΟΝΟΜΕΤΡΙΚΕΣ ΜΕΘΟΔΟΥΣ ΚΑΙ ΝΕΥΡΩΝΙΚΑ  
ΔΙΚΤΥΑ**

**ΕΦΑΡΜΟΓΗ ΜΟΝΤΕΛΩΝ ΣΕ ΕΚΠΑΙΔΕΥΤΙΚΑ PORTAL**

Καχριμάνη Ελευθερία

Μεταπτυχιακή Διπλωματική Εργασία

Η εργασία υποβάλλεται για την μερική κάλυψη των απαιτήσεων με στόχο την απόκτηση του Μεταπτυχιακού Διπλώματος Σπουδών στην Διδακτική της Τεχνολογίας και τα Ψηφιακά Συστήματα

Επιβλέπων: Νικήτας-Μαρίνος Σγούρος, Καθηγητής

Φεβρουάριος 2012

ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΕΡΡΑΙΑ

Στη μικρή μου Νάσια

## Περίληψη

Η παρούσα μεταπτυχιακή διπλωματική εργασία μελετά την πρόβλεψη χρονοσειρών με οικονομετρικές μεθόδους και νευρωνικά δίκτυα, με στόχο την εφαρμογή των διαφόρων μοντέλων στην πρόβλεψη μελλοντικών τιμών της επισκεψιμότητας σε μια ηλεκτρονική εκπαιδευτική πύλη (educational portal).

Στην κατεύθυνση αυτή, μελετώνται αρχικά οι βασικές έννοιες που σχετίζονται με τη διαδικασία της πρόβλεψης και την αναγκαιότητά της, τις διαφορετικές μεθόδους πρόβλεψης ανάλογα με τον αριθμό και την ποιότητα των διαθέσιμων δεδομένων καθώς επίσης και με τη δυνατότητα μετρησιμότητας δεικτών ενός εκπαιδευτικού portal.

Στη συνέχεια ιδιαίτερη βαρύτητα δίνεται στο θεωρητικό πλαίσιο γύρω από την κυριότερη μέθοδο πρόβλεψης, που είναι η ανάλυση χρονοσειρών, λόγω της ευρείας χρήσης της στην πράξη. Παρουσιάζονται και αναλύονται εκτενώς τα πλεονεκτήματα από τη χρήση των χρονοσειρών για τη διεξαγωγή προβλέψεων καθώς επίσης και όλα τα μοντέλα ανάλυσης χρονοσειρών από το πιο απλό στο πιο περίπλοκο. Επίσης, γίνεται επισκόπηση της χρήσης των νευρωνικών δικτύων στην διαδικασία της πρόβλεψης μέσα από την παρουσίαση και ανάλυση των κυριότερων αρχιτεκτονικών των δικτύων αυτών.

Τέλος, γίνεται εφαρμογή των μοντέλων χρονοσειρών και νευρωνικών δικτύων, σε μια υπάρχουσα χρονοσειρά που σχετίζεται με ένα εκπαιδευτικό portal, με σκοπό την πρόβλεψη της επισκεψιμότητας στο portal αυτό. Μέσα από την εφαρμογή αυτή, προκύπτουν ενδιαφέροντα συμπεράσματα σχετικά με τη χρήση της ανάλυσης χρονοσειρών για την αποτίμηση και την περαιτέρω εξέλιξη των εκπαιδευτικών portals.

## Ευχαριστίες

Θα ήθελα να εκφράσω τις θερμότερες ευχαριστίες μου στον Καθηγητή του Πανεπιστημίου Πειραιώς, Νικήτα – Μαρίνο Σγούρο και την Επίκουρο Καθηγήτρια του Πανεπιστημίου Πειραιώς Φωτεινή Παρασκευά, για την επίβλεψη αυτής της διπλωματικής εργασίας και ιδιαίτερες ευχαριστίες στον Επίκουρο Καθηγητή του Πανεπιστημίου Πειραιώς Ιωάννη Αθ. Παραβάνη για τη βοήθεια, καθώς και την ηθική στήριξη που μου παρείχε καθ' όλη τη διάρκεια της εκπόνησης της παρούσας μεταπτυχιακής διπλωματικής εργασίας.

Τέλος, εκφράζω την ευγνωμοσύνη και τις ευχαριστίες μου στην οικογένειά μου και τους φίλους μου για την υποστήριξη, την υπομονή και την αμέριστη συμπαράστασή τους κατά τη διάρκεια των μεταπτυχιακών μου σπουδών.

## Περιεχόμενα

Περίληψη .....	i
Ευχαριστίες .....	ii
Περιεχόμενα .....	iii
Κατάλογος Σχημάτων .....	vi
Κατάλογος Πινάκων .....	viii
Κατάλογος Εικόνων .....	ix
Συνοτομογραφίες .....	x
ΚΕΦΑΛΑΙΟ 1 Εισαγωγή.....	1
1.1 Περιγραφή του αντικειμένου μελέτης της εργασίας.....	1
1.2 Κύριοι στόχοι της εργασίας .....	2
1.3 Περιγραφή του τρόπου διερεύνησης του αντικειμένου της εργασίας .....	2
1.4 Σύνοψη .....	3
ΚΕΦΑΛΑΙΟ 2 Βιβλιογραφική επισκόπηση .....	4
2.1 Εισαγωγή.....	4
2.2 Η έννοια της πρόβλεψης .....	5
2.2.1 Ορισμός πρόβλεψης .....	5
2.2.2 Ρόλος και αρχές της πρόβλεψης.....	5
2.2.3 Κατηγορίες μεθόδων πρόβλεψης .....	6
2.2.3.1 Παράγοντες επιλογής της κατάλληλης μεθόδου πρόβλεψης.....	8
2.2.4 Βασικά στάδια στη διαδικασία πρόβλεψης.....	9
2.3 Η έννοια των δεικτών (Indicators).....	10
2.3.1 Ορισμοί και χαρακτηριστικά δεικτών.....	10
2.3.2 Οι δείκτες στον Παγκόσμιο Ιστό .....	12
2.3.3 Μετρικές χρήσης portal.....	13
2.3.3.1 Τι είναι portal .....	13
2.3.3.2 Σημαντικές μετρικές portal .....	14
2.4 Η έννοια της χρονοσειράς.....	17
2.4.1 Ορισμός χρονοσειράς.....	17

2.4.2	Πρόβλεψη χρονοσειράς .....	21
2.4.3	Πλεονεκτήματα μοντέλου χρονοσειρών για πρόβλεψη.....	22
2.4.4	Ανάλυση χρονοσειρών (Time series analysis).....	23
2.4.4.1	Μέθοδοι εξομάλυνσης (smoothing methods).....	23
2.4.4.2	Διάσπαση χρονοσειρών.....	24
2.4.4.3	Ανάλυση ARIMA-Box Jenkins .....	27
2.4.4.4	Βασικές έννοιες προχωρημένης ανάλυσης.....	28
2.5	Τεχνητά Νευρωνικά δίκτυα .....	34
2.5.1	Νευρωνική δομή .....	36
2.5.2	Τύποι συναρτήσεων ενεργοποίησης .....	38
2.5.3	Αρχιτεκτονικές δικτύων .....	39
ΚΕΦΑΛΑΙΟ 3	Μεθοδολογία.....	42
3.1	Εισαγωγή.....	42
3.2	Συλλογή στοιχείων.....	42
3.3	Μέθοδοι εξομάλυνσης .....	45
3.3.1	Η μέθοδος του κινητού μέσου (Moving average).....	45
3.3.2	Απλή εκθετική εξομάλυνση (Simple exponential smoothing) .....	47
3.3.3	Διπλός κινητός μέσος (double moving average).....	49
3.3.4	Διπλή εκθετική εξομάλυνση (Double exponential smoothing) .....	50
3.3.5	Εκθετική εξομάλυνση με προσαρμογή στην τάση (Exponential smoothing adjusted for trend) .....	52
3.3.6	Εκθετική εξομάλυνση με προσαρμογή στην τάση και στην εποχικότητα (Exponential smoothing adjusted for trend and seasonality) .....	53
3.4	Διάσπαση χρονοσειρών.....	54
3.4.1	Ανάλυση και διαχωρισμός της εποχικότητας .....	55
3.4.2	Προσαρμογή και Ανάλυση της Τάσης.....	56
3.4.3	Το Μοντέλο Γραμμικής Τάσης.....	57
3.4.4	Απομόνωση της κυκλικότητας.....	58
3.5	Ανάλυση ARIMA .....	60
3.5.1	Μοντέλα χρονοσειρών ARIMA.....	60

3.5.2	Επιλογή του καλύτερου μοντέλου ARIMA .....	68
3.5.3	Διαγνωστικός έλεγχος .....	69
3.5.4	Πρόβλεψη με τη χρήση μοντέλων ARIMA .....	70
3.6	Ανάλυση Νευρωνικού Δικτύου .....	71
3.6.1	Μονοστρωματικά Νευρωνικά Δίκτυα εμπρόσθιας τροφοδότησης ....	71
	Το πολυεπίπεδο perceptron – MLP .....	71
ΚΕΦΑΛΑΙΟ 4	Αποτελέσματα .....	78
4.1	Εισαγωγή .....	78
4.2	Αποτελέσματα μεθόδων εξομάλυνσης .....	78
4.3	Αποτελέσματα διάσπασης χρονοσειρών .....	81
4.4	Αποτελέσματα μοντέλου ARIMA .....	83
4.5	Αποτελέσματα Νευρωνικού δικτύου – MLP .....	86
ΚΕΦΑΛΑΙΟ 5	Συμπεράσματα .....	90
5.1	Εισαγωγή .....	90
5.2	Συμπεράσματα .....	90
5.3	Προτάσεις για περαιτέρω μελέτη .....	93
Βιβλιογραφία .....		95
ΠΑΡΑΡΤΗΜΑ Α .....		98
ΠΑΡΑΡΤΗΜΑ Β .....		113

## Κατάλογος Σχημάτων

Σχήμα 2.1: Στασιμότητα .....	17
Σχήμα 2.2: Τάση .....	18
Σχήμα 2.3: Εποχικότητα .....	18
Σχήμα 2.4: Κυκλικότητα.....	19
Σχήμα 2.5: Συνδυασμός διαφόρων στοιχείων χρονοσειράς .....	19
Σχήμα 2.6: ACF-PACF.....	31
Σχήμα 2.7: Δομή βιολογικού νευρώνα .....	36
Σχήμα 2.8: Δομή τεχνητού νευρώνα.....	37
Σχήμα 2.9: Η βηματική συνάρτηση .....	38
Σχήμα 2.10: Η λογιστική συνάρτηση.....	39
Σχήμα 3.1: Ιστόγραμμα χρονοσειράς.....	44
Σχήμα 3.2: Boxplot χρονοσειράς.....	44
Σχήμα 3.3: Χρονοσειρά ημερησίων τιμών για 730 ημέρες .....	45
Σχήμα 3.4: Πολυεπίπεδο perceptron με δύο κρυμμένα επίπεδα.....	72
Σχήμα 3.5: Η κατεύθυνση ροής των δύο σημάτων κατά την εκπαίδευση του MLP ..	73
Σχήμα 4.1: Διπλή εκθετική εξομάλυνση χρονοσειράς με $\alpha=0,005$ και $\beta=0,005$ (TestSet).....	79
Σχήμα 4.2: Μέθοδος Winters (πολ/κο μοντέλο) στη χρονοσειρά με $\alpha=0,1$ , $\beta=0,1$ και $\gamma=0,1$ (TestSet).....	79
Σχήμα 4.3: Μέθοδος Winters(προσθετικό μοντέλο) στη χρονοσειρά με $\alpha=0,1$ , $\beta=0,1$ και $\gamma=0,1$ (TestSet).....	80
Σχήμα 4.4: Μέθοδος Διάσπασης Χρονοσειράς (πολλαπλασιαστικό μοντέλο) .....	81
Σχήμα 4.5: Συνιστώσες Χρονοσειράς (πολλαπλασιαστικό μοντέλο).....	81



Σχήμα 4.6: Ανάλυση εποχικότητας Χρονοσειράς (πολλαπλασιαστικό μοντέλο) .....	82
Σχήμα 4.7: Διαγράμματα υπολοίπων αυτοσυσχέτισης (ACF) και μερικής αυτοσυσχέτισης (PACF) .....	85
Σχήμα 4.8: Μοντέλο ARIMA(1,1,1)(1,1,1) <sub>7</sub> .....	85
Σχήμα 4.9: Διάγραμμα πρόβλεψης επισκεπτων εβδομάδας (training set).....	88
Σχήμα 4.10: Διάγραμμα πρόβλεψης επισκεπτων εβδομάδας (test set).....	88
Σχήμα 4.11: Διάγραμμα μελλοντικών προβλέψεων επισκεπτών (training set) .....	89
Σχήμα 4.12: Διάγραμμα μελλοντικών προβλέψεων επισκεπτών (test set).....	89

## Κατάλογος Πινάκων

Πίνακας 2.1: Βασικοί στατιστικοί δείκτες .....	20
Πίνακας 2.2: Κατηγορίες σφαλμάτων και αποκλίσεων .....	21
Πίνακας 3.1: Συγκεντρωτικός πίνακας στατιστικής ανάλυσης χρονοσειράς .....	44
Πίνακας 3.2: Εκτίμηση της τάξης των AR και MA με τη βοήθεια των ACF-PACF .	65
Πίνακας 4.1: Σύνοψη αποτελεσμάτων των μεθόδων εξομάλυνσης .....	80
Πίνακας 4.2: Σύνοψη αποτελεσμάτων μεθόδου διάσπασης .....	82
Πίνακας 4.3: Πίνακας τιμών σφαλμάτων μοντέλου ARIMA(1,1,1)(1,1,1).....	83
Πίνακας 4.4: Πίνακας στατιστικών αποτελεσμάτων μοντέλου ARIMA(1,1,1)(1,1,1) .....	84
Πίνακας 4.5: Πίνακας παραμέτρων μοντέλου ARIMA(1,1,1)(1,1,1).....	84
Πίνακας 4.6: Πίνακας υπολοίπων αυτοσυσχέτισης .....	84
Πίνακας 4.7: Πίνακας υπολοίπων μερικής αυτοσυσχέτισης .....	84
Πίνακας 4.8: Παράμετροι αλγορίθμου εκπαίδευσης νευρωνικού δικτύου .....	86
Πίνακας 4.9: Αποτελέσματα εκπαίδευσης νευρωνικού δικτύου .....	87
Πίνακας 4.10: Αποτελέσματα εκπαίδευσης νευρωνικού δικτύου .....	87
Πίνακας 5.1: Σύνοψη αποτελεσμάτων μεθόδων πρόβλεψης .....	91

## Κατάλογος Εικόνων

Εικόνα 2.1: Διάγραμμα μοντέλου χρονοσειρών .....	22
Εικόνα 2.2: Αρχιτεκτονικές νευρωνικών δικτύων .....	41

## Συντομογραφίες

### Λατινικές

ACF	AutoCorrelation function
AIC	Akaike's Information Criterion
AR	AutoRegression
ARIMA	AutoRegressive Integrated-Moving Average
ARMA	AutoRegressive Moving Average
BIC	Bayesian Information Criterion
MA	Moving Average
MAD	Mean Absolute Deviation
MAE	Mean Absolute Error
MAPE	Mean Absolute Percentage Error
ME	Mean Error
MLP	Multilayer Perceptron
MPE	Mean Percentage Error
MSD	Mean Square Deviation
MSE	Mean Square Error
OECD/DAC	Organisation for Economic Co-Operation and Development / Development Assistance Committee
PACF	Partial Autocorrelation Coefficient
USAID	U.S. Agency for International Development

## ΚΕΦΑΛΑΙΟ 1

### Εισαγωγή

#### 1.1 Περιγραφή του αντικειμένου μελέτης της εργασίας

Η παρούσα μεταπτυχιακή διπλωματική εργασία επικεντρώνεται στη μελέτη της πρόβλεψης χρονοσειρών με οικονομετρικές μεθόδους και νευρωνικά δίκτυα, με στόχο την εφαρμογή των διαφόρων μοντέλων στην πρόβλεψη μελλοντικών τιμών της επισκεψιμότητας σε μια ηλεκτρονική εκπαιδευτική πύλη (educational portal).

Η πρόβλεψη μελλοντικών καταστάσεων αποτελεί πλέον ένα πολύ σημαντικό εργαλείο στα χέρια των υπευθύνων του επιχειρηματικού σχεδιασμού, καθώς δίνει τη δυνατότητα δημιουργίας μιας εικόνας του μελλοντικού περιβάλλοντος μέσα στο οποίο θα λειτουργήσει μια επιχείρηση ή ένας οργανισμός. Έτσι, οι υπεύθυνοι είναι σε θέση να πάρουν αποφάσεις σχετικά με τις αλλαγές που ίσως θελήσουν να κάνουν ή για την κατασκευή καινοτόμων και ανταγωνιστικών λύσεων για την επιχείρησή τους. Η ανάλυση χρονοσειρών και τα νευρωνικά δίκτυα είναι δυο επιστημονικά πεδία που βοηθούν καταλυτικά στην κατεύθυνση αυτή αφού μπορούν να δώσουν σχετικά ακριβείς προβλέψεις με βάση ιστορικά στοιχεία του παρελθόντος.

Από την άλλη πλευρά η ανάπτυξη του Παγκόσμιου Ιστού (Web) και η τεχνολογική εξέλιξη των τελευταίων ετών, έχει οδηγήσει τις επιχειρήσεις, τους οργανισμούς ή ακόμα και μεμονωμένα άτομα στην ηλεκτρονική παρουσία τους στο Internet, μέσω της δημιουργίας διαφόρων ειδών ιστοσελίδων, όπως απλές ιστοσελίδες (sites), πύλες ενημέρωσης (portals), ηλεκτρονικά καταστήματα (eshops) κλπ. Για να είναι όλες αυτές οι σελίδες ανταγωνιστικές και αποτελεσματικές, είναι απαραίτητο να γίνονται μετρήσεις και προβλέψεις σχετικά με τη χρήση τους και την αποδοχή τους από τους χρήστες.

Στο πλαίσιο αυτό, η παρούσα εργασία επικεντρώνεται στον τρόπο με τον οποίο μπορεί να γίνει ένας συνδυασμός όλων των παραπάνω, δηλαδή στον τρόπο με τον

οποίο μπορεί η ανάλυση χρονοσειρών και νευρωνικών δικτύων να χρησιμοποιηθεί αποτελεσματικά από τους εμπλεκόμενους στο σχεδιασμό και την ανάπτυξη εκπαιδευτικών portals, με σκοπό τη βελτίωση της λειτουργίας τους και την περαιτέρω ανάπτυξη τους ή την κατασκευή νέων καινοτόμων λύσεων. Το θέμα αυτό κρίνεται ιδιαίτερα σημαντικό καθώς προσπαθεί να συνδυάσει τις διάφορες τάσεις των τριών αυτών σημαντικών γνωστικών πεδίων.

Στις επόμενες ενότητες παρουσιάζονται οι κύριοι στόχοι της εργασίας αυτής καθώς επίσης και η μεθοδολογία ανάπτυξής της.

## **1.2 Κύριοι στόχοι της εργασίας**

Οι κύριοι στόχοι της παρούσας μεταπτυχιακής διπλωματικής εργασίας παρουσιάζονται παρακάτω:

- Η επισκόπηση των βασικών εννοιών που αφορούν στην πρόβλεψη και τους δείκτες. Μέσα από την επισκόπηση αυτή θα προκύψουν τα δομικά στοιχεία για την ανάλυση των χρονοσειρών και των νευρωνικών δικτύων.
- Η επισκόπηση των βασικών εννοιών που σχετίζονται με τις χρονοσειρές και τα μοντέλα ανάλυσής τους καθώς επίσης και έννοιες που σχετίζονται με τα νευρωνικά δίκτυα και τις αρχιτεκτονικές τους.
- Η εφαρμογή των διαφόρων μοντέλων χρονοσειρών και νευρωνικών δικτύων στην πράξη, αναλύοντας την επισκεψιμότητα σε μια εκπαιδευτική πύλη (portal).

## **1.3 Περιγραφή του τρόπου διερεύνησης του αντικειμένου της εργασίας**

Τα βασικά βήματα που ακολουθούνται για την προσέγγιση του θέματος της παρούσας εργασίας είναι τα παρακάτω:

- Επισκόπηση των βασικών εννοιών που σχετίζονται με την πρόβλεψη και τους δείκτες που χρησιμοποιούνται κατά τη διαδικασία πρόβλεψης. Στην κατεύθυνση αυτή, αναλύεται η έννοια της πρόβλεψης, η αναγκαιότητα για πρόβλεψη, τα βασικά στάδια της διαδικασίας αυτής και οι διάφορες μέθοδοι πρόβλεψης που συναντώνται στη βιβλιογραφία. Επίσης, παρουσιάζεται η έννοια των δεικτών και τα χαρακτηριστικά τους και η έννοια αυτή εξειδικεύεται στο πεδίο του Παγκόσμιου Ιστού με σκοπό να αναλυθούν οι μετρήσιμοι δείκτες μιας εκπαιδευτικής πύλης (portal), ώστε στη συνέχεια να εφαρμοστούν σε αυτό οι διάφορες τεχνικές πρόβλεψης.
- Επισκόπηση των βασικών εννοιών που σχετίζονται με την ανάλυση χρονοσειρών. Στο πλαίσιο αυτό περιγράφεται η έννοια της χρονοσειράς, τα διάφορα μοντέλα ανάλυσης χρονοσειρών και οι βασικότερες έννοιες της προχωρημένης οικονομετρικής ανάλυσης.
- Επισκόπηση της χρήσης των νευρωνικών δικτύων για τη διεξαγωγή προβλέψεων ως μια εναλλακτική στα διάφορα μοντέλα χρονοσειρών. Στην κατεύθυνση αυτή αναλύεται η έννοια του νευρωνικού δικτύου, παρουσιάζεται η βασικότερη αρχιτεκτονική των δικτύων αυτών και αναλύεται ο τρόπος εκπαίδευσής τους, με σκοπό την εφαρμογή τους στην πρόβλεψη μελλοντικών τιμών της επισκεψιμότητας ενός εκπαιδευτικού portal.
- Εφαρμογή όλων των μοντέλων που παρουσιάζονται θεωρητικά στην πράξη, με ανάλυση και επεξήγηση των αποτελεσμάτων, καθώς και εξαγωγή συμπερασμάτων σχετικά με τη δυνατότητα εφαρμογής τους για την διεξαγωγή ασφαλών προβλέψεων για εκπαιδευτικά portals.

#### **1.4 Σύνοψη**

Στο πρώτο αυτό κεφάλαιο της παρούσας εργασίας, παρουσιάστηκαν το αντικείμενο και οι στόχοι της μεταπτυχιακής αυτής διπλωματικής εργασίας, οι λόγοι που κάνουν το ζήτημα που εξετάζεται να είναι σημαντικό καθώς επίσης και η μεθοδολογία που ακολουθείται για την προσέγγιση του θέματος που πραγματεύεται η εργασία.

## ΚΕΦΑΛΑΙΟ 2

### Βιβλιογραφική επισκόπηση

#### 2.1 Εισαγωγή

Με δεδομένο ότι η παρούσα μεταπτυχιακή διπλωματική εργασία επικεντρώνεται στη μελέτη της χρήσης των χρονοσειρών και των νευρωνικών δικτύων για τη διεξαγωγή προβλέψεων, οι βασικές τις συνιστώσες είναι η έννοια της πρόβλεψης, της χρονοσειράς, των νευρωνικών δικτύων καθώς επίσης και του σημαντικού εργαλείου για να επιτευχθούν τα παραπάνω το οποίο είναι οι δείκτες (indicators). Όλα αυτά αναλύονται θεωρητικά και γίνεται μια προσπάθεια εφαρμογής τους στην πράξη προβλέποντας τη μελλοντική επισκεψιμότητα ενός εκπαιδευτικού portal.

Στο κεφάλαιο αυτό λοιπόν γίνεται μια σύντομη και περιεκτική βιβλιογραφική επισκόπηση στις παραπάνω έννοιες, το ρόλο και τη χρησιμότητά τους με σκοπό να δοθεί το απαραίτητο θεωρητικό υπόβαθρο για την περαιτέρω ανάλυση και εφαρμογή τους που θα γίνει στα επόμενα κεφάλαια.

Πιο συγκεκριμένα στην ενότητα 2.2 παρουσιάζεται η έννοια της πρόβλεψης, ο ρόλος της, οι αρχές που τη διέπουν, οι διάφοροι τύποι μεθόδων πρόβλεψης καθώς επίσης και τα βασικά στάδια στη διαδικασία της πρόβλεψης.

Στην ενότητα 2.3 παρουσιάζονται οι δείκτες (indicators), τα χαρακτηριστικά τους και οι συγκεκριμένες κατηγορίες τους που αφορούν την εργασία αυτή. Επίσης γίνεται αναφορά στην έννοια του portal, που αποτελεί το πεδίο εφαρμογής της ανάλυσης που πραγματοποιείται στην παρούσα εργασία.

Στην ενότητα 2.4 γίνεται μια εννοιολογική προσέγγιση των χρονοσειρών, παρουσιάζονται τα πλεονεκτήματα της χρήσης τους για τη διεξαγωγή προβλέψεων καθώς επίσης αναφέρονται συνοπτικά οι κύριες μέθοδοι της ανάλυσης χρονοσειρών.

Τέλος, στην ενότητα 2.5 παρουσιάζονται οι βασικές έννοιες των τεχνητών νευρωνικών δικτύων, όπως η δομή τους, οι συναρτήσεις που τα αποτελούν καθώς επίσης και οι διάφορες αρχιτεκτονικές των δικτύων αυτών.



## 2.2 Η έννοια της πρόβλεψης

### 2.2.1 Ορισμός πρόβλεψης

Στη βιβλιογραφία συναντώνται διάφοροι ορισμοί για την έννοια της πρόβλεψης (forecast).

Σύμφωνα με τον Armstrong, 2001, *πρόβλεψη* είναι «η εκτίμηση πραγματικής τιμής ή κατάστασης, αναφερόμενη σε μελλοντική χρονική περίοδο».

Το Business Dictionary ορίζει την πρόβλεψη ως «ένα εργαλείο σχεδιασμού που βοηθά στην αντιμετώπιση της αβεβαιότητας του μέλλοντος, χρησιμοποιώντας δεδομένα από το παρελθόν και το παρόν και κάνοντας ανάλυση των τάσεων που εμφανίζονται».

### 2.2.2 Ρόλος και αρχές της πρόβλεψης

Η πρόβλεψη είναι μια σημαντική εισαγωγή στον επιχειρηματικό σχεδιασμό και την δημιουργία πολιτικής. Πράγματι, η πρόβλεψη και ο σχεδιασμός είναι συσχετιζόμενες έννοιες. Μια επιχείρηση δε μπορεί να κάνει σχέδια χωρίς να κάνει προβλέψεις, και δε μπορεί να κάνει προβλέψεις χωρίς να έχει κάποιου τύπου πλάνο σαν πλαίσιο για την προβλεπτική διαδικασία. Η διαδικασία της πρόβλεψης προσπαθεί να δημιουργήσει μια εικόνα του είδους του μελλοντικού περιβάλλοντος στο οποίο θα πραγματοποιηθούν τα σχέδια και οι δραστηριότητες της επιχείρησης (Wacker and Sprague, 1998).

Η πρόβλεψη αποτελεί ένα πολύ σημαντικό βήμα στον αποτελεσματικό σχεδιασμό καθώς βοηθά να εξασφαλιστεί η σωστή χρήση των πόρων μιας επιχείρησης ή ενός οργανισμού. (Waddell and Sohal, 1994; Newbold and Bos, 1994; Klassen and Flores Benito, 2001).

Παρακάτω συνοψίζονται κάποιες βασικές αρχές των προβλέψεων.

1. Καμία πρόβλεψη δεν είναι τέλεια: καθώς περιλαμβάνει το στοιχείο της αβεβαιότητας, η πρόβλεψη θα περιέχει κάποιο σφάλμα (δηλ, τη διαφορά μεταξύ της πρόβλεψης και της πραγματικότητας). Με βάση αυτό, στόχος της διαδικασίας πρόβλεψης είναι η ελαχιστοποίηση του σφάλματος για την όσο το δυνατόν ακριβέστερη προσέγγιση της πραγματικότητας.

2. Μια πρόβλεψη είναι περισσότερο ακριβής για ομάδες στοιχείων παρά για μεμονωμένα στοιχεία: Αυτό συμβαίνει γιατί οι μέγιστες και ελάχιστες τιμές των διαφόρων στοιχείων αλληλοεξουδετερώνονται με αποτέλεσμα η ομάδα των στοιχείων να έχει σταθερή συμπεριφορά ακόμα και αν τα μεμονωμένα στοιχεία συμπεριφέρονται με ασταθή τρόπο.

3. Η πρόβλεψη είναι περισσότερο ακριβής όταν είναι βραχυπρόθεσμη παρά όταν είναι μακροπρόθεσμη: όσο κοντινότερος είναι ο χρονικός ορίζοντας της πρόγνωσης τόσο μικρότερος είναι ο βαθμός αβεβαιότητας και άρα τόσο μικρότερο το σφάλμα που θα περιέχει. Ένα κλασικό παράδειγμα αφορά στην πρόβλεψη του καιρού: ένα μετεωρολογικό δελτίο για τις επόμενες δύο ή τρεις μέρες είναι πάρα πολύ πιθανό να βγει αληθινό. Αντίθετα, η πρόγνωση για τον καιρό του επόμενου μήνα έχει μεγάλες πιθανότητες να αποδειχτεί λανθασμένη.

### **2.2.3 Κατηγορίες μεθόδων πρόβλεψης**

Οι μέθοδοι πρόβλεψης συμβάλλουν στη σωστή και έγκαιρη λήψη αποφάσεων και στον σωστό σχεδιασμό. Αν η διαδικασία της πρόβλεψης χρησιμοποιεί υψηλής ποιότητας δεδομένα, οι προβλέψεις που προκύπτουν μπορεί να είναι αρκετά ακριβείς (Wacker & Sprague, 1998). Στην πράξη όμως, οι προβλέψεις που επιχειρούμε, επηρεάζονται από διάφορους παράγοντες. Για παράδειγμα, τα δεδομένα δεν είναι πάντα ακριβή και αξιόπιστα. Μπορεί, επίσης, να μην είναι και αρκετά για τη διενέργεια μιας σωστής πρόβλεψης-εκτίμησης. Ακόμη, είναι γνωστό, πως το παρελθόν δεν είναι πάντα ο σωστός οδηγός εκτίμησης για το μέλλον. Επίσης το υπό μελέτη φαινόμενο μπορεί να δέχεται σημαντική επιρροή από απρόβλεπτους εξωτερικούς παράγοντες.

Όλα τα παραπάνω οδηγούν στον διαχωρισμό των μεθόδων πρόβλεψης σε δύο κύριες κατηγορίες, οι οποίες είναι:

*Ποιοτικές ή κριτικές προβλέψεις (qualitative ή judgmental forecasting):*

Οι μέθοδοι αυτές χρησιμοποιούνται, όταν οι διαθέσιμες πληροφορίες για το παρελθόν είναι λίγες ή ανύπαρκτες. Σε αυτές τις περιπτώσεις απαιτείται εμπειρία, γνώση και

κριτική ικανότητα. Χρησιμοποιούνται, κυρίως, σε συνδυασμό με τις ποσοτικές μεθόδους, ενώ δεν απαντούν μόνες σε αριθμητικές προβλέψεις. Αυτού του είδους η πρόβλεψη είναι υποκειμενική και περιλαμβάνει το στοιχείο της προκατάληψης (bias). (Σούκος & Κάντζιος, 2006)

*Ποσοτικές προβλέψεις (quantitative forecasting):*

Οι μέθοδοι αυτές βασίζονται στη μαθηματική μοντελοποίηση άρα είναι αντικειμενικές και επαναλήψιμες (δηλ, παράγουν το ίδιο αποτέλεσμα κάθε φορά που εισάγουμε τα ίδια δεδομένα). Οι ποσοτικές μέθοδοι πρόβλεψης χρησιμοποιούνται, όταν υπάρχει διαθέσιμη πληροφορία για το παρελθόν (επαρκή δεδομένα), όταν μπορεί να γίνει ποσοτικοποίηση της πληροφορίας (αριθμητικά δεδομένα) και τέλος όταν εικάζεται ότι το πρότυπο συμπεριφοράς του παρελθόντος θα διατηρηθεί και στο μέλλον.

Οι ποσοτικές μέθοδοι μπορούν να διακριθούν σε αυτές που βασίζονται σε:

- *μοντέλα χρονοσειρών (time series models)*. Με τα μοντέλα αυτά η πρόβλεψη γίνεται με χρήση ιστορικών δεδομένων. Πιο συγκεκριμένα τα μοντέλα αυτά προϋποθέτουν ότι η απαραίτητη πληροφορία για την πρόβλεψη περιέχεται στη χρονοσειρά των στοιχείων. Χρονοσειρά είναι μια σειρά παρατηρήσεων που λαμβάνονται σε κανονικά διαστήματα μέσα σε ένα καθορισμένο χρονικό διάστημα. Η έννοια της χρονοσειράς περιγράφεται αναλυτικά στην ενότητα 2.4.

$\text{Είσοδοι}(x_i) \rightarrow \text{Σύστημα}(f) \rightarrow \text{Εξοδος}(y)$
--

- *αιτιοκρατικά ή επεξηγηματικά μοντέλα (causal models)*. Με τα μοντέλα αυτά η πρόβλεψη γίνεται με χρήση ανεξάρτητων παραμέτρων (επεξηγηματικές μεταβλητές) οι οποίες επηρεάζουν το σύστημα (π.χ. η θερμοκρασία επηρεάζει την ταχύτητα του ανέμου). Τα μοντέλα αυτά χρησιμοποιούνται με επιτυχία για χάραξη τακτικής και λήψης αποφάσεων.

$\text{Ανεξάρτητες μεταβλητές}(x_i) \rightarrow \text{Σύστημα}(f) \rightarrow \text{Εξαρτημένη μεταβλητή}(y)$
---

Η δυσκολία στην κατηγορία αυτή έγκειται στην εύρεση της μαθηματικής σχέσης με την οποία επηρεάζεται η ζητούμενη μεταβλητή από τις παραμέτρους αυτές.

Η διαμόρφωση των προβλέψεων στηρίζεται κατά κανόνα σε δύο βασικές προϋποθέσεις. Η πρώτη αφορά στην ποσοτική μέθοδο που θα χρησιμοποιηθεί και η δεύτερη στον τρόπο συμπεριφοράς της μεταβλητής. Με άλλα λόγια, η μέθοδος που επιλέγεται θα πρέπει να προσδιορίζει με τον καλύτερο τρόπο, τη συμπεριφορά των τιμών της μεταβλητής. Όσο καλύτερα αναγνωρίζεται ο τρόπος δημιουργίας των τιμών της μεταβλητής, τόσο καλύτερες προβλέψεις αναμένεται να διαμορφώνονται. Για να ισχύσει, όμως, κάτι τέτοιο, θα πρέπει, επιπροσθέτως, η υπάρχουσα δομή του τρόπου συμπεριφοράς των τιμών της μεταβλητής να παραμένει σταθερή ή περίπου σταθερή και στο άμεσο μέλλον (Αγιακόγλου & Οικονόμου, 2004).

Με βάση τη στατιστική ανάλυση, η συμπεριφορά των τιμών μιας μεταβλητής προσδιορίζεται από δύο μέρη, που είναι το **συστηματικό** (systematic) και το **τυχαίο** (random). Το συστηματικό μέρος αναφέρεται σε όλους εκείνους τους παράγοντες που μπορούν να καθοριστούν και οι οποίοι ερμηνεύουν τον τρόπο συμπεριφοράς της μεταβλητής. Αντίθετα, το τυχαίο μέρος αφορά όλους εκείνους τους τυχαίους και απρόσμενους παράγοντες οι οποίοι, αν και επηρεάζουν τις τιμές της μεταβλητής, δεν μπορούν, ωστόσο, να προσδιοριστούν. Αυτό σημαίνει πως, όταν αναφερόμαστε στον προσδιορισμό του τρόπου συμπεριφοράς των τιμών μιας μεταβλητής, εννοούμε αποκλειστικά τον καθορισμό του συστηματικού μέρους. Έτσι, οι προβλέψεις που προέρχονται από τη χρησιμοποίηση των ποσοτικών μεθόδων, διαμορφώνονται με βάση το συστηματικό μέρος και όχι το τυχαίο, η συμπεριφορά του οποίου δε μπορεί να προσδιοριστεί, αφού δεν παρατηρείται. Η διαδικασία διαμόρφωσης των προβλέψεων αποτελείται από κάποια στάδια, που θα αναλυθούν στην επόμενη ενότητα (Αγιακόγλου & Οικονόμου, 2004).

#### *2.2.3.1 Παράγοντες επιλογής της κατάλληλης μεθόδου πρόβλεψης*

Οι κύριοι παράγοντες επιλογής της καταλληλότερης μεθόδου πρόβλεψης συνοψίζονται παρακάτω:

- Σημασία της απόφασης που θα ληφθεί

- Διαθεσιμότητα σχετικών δεδομένων
- Χρονικός ορίζοντας της πρόβλεψης
- Το κόστος προετοιμασίας της πρόβλεψης
- Ο χρόνος για τον οποίο απαιτείται η ύπαρξη πρόβλεψης
- Ο αριθμός των φορών για τις οποίες θα χρειαστεί να πραγματοποιηθεί παρόμοιου τύπου πρόβλεψη
- Σταθερότητα του περιβάλλοντος

#### **2.2.4 Βασικά στάδια στη διαδικασία πρόβλεψης**

Τα βασικά στάδια σε μια διαδικασία πρόβλεψης είναι τα εξής (Αγιακόγλου & Οικονόμου, 2004):

##### *I. Καθορισμός Προβλήματος (Problem Definition)*

Αποτελεί το πιο σημαντικό αλλά και το πιο δύσκολο μέρος στη διαδικασία πρόβλεψης. Αυτό συμβαίνει διότι στο στάδιο αυτό θα πρέπει να καταστούν σαφή και κατανοητά ορισμένα θέματα, όπως το πώς θα χρησιμοποιηθούν οι προβλέψεις και από ποιους.

##### *II. Συγκέντρωση Πληροφοριών (Gathering Information)*

Σε αυτό το στάδιο απαιτούνται τουλάχιστον δύο είδη πληροφοριών. Το πρώτο είναι τα στατιστικά (συνήθως αριθμητικά) δεδομένα και το δεύτερο η κρίση, η πείρα και η εμπειρία του προσωπικού που ασχολούνταν με αυτή τη συλλογή για αυτό το χρονικό διάστημα. Επίσης οι παραπάνω πληροφορίες πρέπει να συλλεχθούν πριν ξεκινήσει η διαδικασία της πρόβλεψης.

##### *III. Προκαταρκτική Ανάλυση (Exploratory Analysis)*

Στο στάδιο αυτό γίνεται μια πρώτη ανάλυση του είδους της πληροφορίας που αποκομίζεται από τα ακατέργαστα ιστορικά δεδομένα. Αρχικά, αναπαριστώνται γραφικά τα δεδομένα και στη συνέχεια, υπολογίζονται κάποιοι βασικοί στατιστικοί δείκτες, όπως η μέση τιμή, η τυπική απόκλιση, ελάχιστο, μέγιστο και γραμμική τάση. Οι παραπάνω δείκτες αναδεικνύουν κάποια δευτερεύοντα χαρακτηριστικά της χρονοσειράς. Ο σκοπός στο βήμα αυτό είναι να δοθούν απαντήσεις σε ερωτήματα

όπως αν υπάρχουν λανθασμένα πρότυπα, αν υπάρχει σημαντική τάση ή εποχικότητα και τέλος, αν υπάρχουν ασυνήθιστες τιμές (outliers). Η ανάλυση αυτή έχει ως στόχο την επιλογή του καταλληλότερου μοντέλου πρόβλεψης που λογικά αναμένεται να δώσει ικανοποιητικές προβλέψεις.

#### *IV. Επιλογή και Προσαρμογή Μοντέλου (Choosing & Fitting models).*

Στη φάση αυτή, γίνεται η επιλογή και καθορισμός των παραμέτρων διάφορων ποσοτικών μοντέλων πρόβλεψης που έχουν επιλεγεί στο προηγούμενο βήμα.

#### *V. Χρήση και αποτίμηση του μοντέλου πρόβλεψης (Using and Evaluating a forecasting model).*

Στο τελικό στάδιο, αφού έχει γίνει η επιλογή του μοντέλου και έχουν καθοριστεί οι παράμετροί του, το μοντέλο αυτό χρησιμοποιείται ώστε να παραχθούν προβλέψεις. Κατά την εξέλιξη της διαδικασίας, γίνεται αποτίμηση των πλεονεκτημάτων και μειονεκτημάτων του μοντέλου και, εφόσον κριθεί απαραίτητο, επαναλαμβάνονται κάποια βήματα στη διαδικασία.

### **2.3 Η έννοια των δεικτών (Indicators)**

Η αναζήτηση στην βιβλιογραφία πάνω στην έννοια των «δεικτών» (indicators), μας παρέχει μια τεράστια ποσότητα αποτελεσμάτων σχετικά με διάφορες κατηγορίες δεικτών όπως: επιχειρηματικούς δείκτες, οικονομικούς δείκτες, κοινωνικούς δείκτες, περιβαλλοντικούς δείκτες, εκπαιδευτικούς δείκτες, δείκτες ποιότητας ζωής, κλπ.

Το κοινό στοιχείο όλων αυτών των κατηγοριών είναι το ότι αναφέρονται σε συγκεκριμένη πληροφορία. Έτσι, οι δείκτες παίζουν ένα πολύ σημαντικό ρόλο στην αποτίμηση ενός οργανισμού ή μιας επιχείρησης (κερδοσκοπικού ή μη) ή ακόμα στην αποτίμηση της λειτουργίας του ίδιου του κράτους και των επιμέρους τομέων του, καθώς η πληροφορία αυτή δίνει τη δυνατότητα να παρθούν σωστές αποφάσεις για το μέλλον.

#### **2.3.1 Ορισμοί και χαρακτηριστικά δεικτών**

Σύμφωνα με τον OECD/DAC, 2002, ένας δείκτης είναι:

«Ένας ποσοτικός ή ποιοτικός παράγοντας – μεταβλητή, ο οποίος παρέχει ένα απλό και αξιόπιστο μέσο για τη μέτρηση της αποτελεσματικότητας, για τον εντοπισμό της ανάγκης για παρεμβάσεις και αλλαγές ή για τη αξιολόγηση της προόδου ενός τομέα ανάπτυξης».

Σύμφωνα με το USAID, «Δείκτης είναι μια μεταβλητή σκοπός της οποίας είναι η μέτρηση της αλλαγής σε φαινόμενα ή διαδικασίες».

Ακόμη πιο συγκεκριμένα το USAID ορίζει την έννοια του δείκτη αποδοτικότητας ως «Οι δείκτες αποδοτικότητας είναι μεταβλητές με ένα συγκεκριμένο χαρακτηριστικό ή μια συγκεκριμένη διάσταση, που χρησιμοποιούνται για την επίτευξη στόχων και επιμέρους αποτελεσμάτων. Οι δείκτες αυτοί υποδηλώνουν κατά πόσο υπάρχει πρόοδος προς τα επιθυμητά αποτελέσματα ή ακόμα δείχνουν το λόγο για τον οποίο δεν υπάρχει κάποια πρόοδος.»

Μέσα λοιπόν από τους διάφορους ορισμούς που συναντώνται στη βιβλιογραφία μπορούμε να συνοψίσουμε τα παρακάτω γενικά χαρακτηριστικά των δεικτών (Johnston, 1981; Land & Spillerman, 1975; Greaney & Kellaghan, 1996; Owen, Hodgkinson & Tuijnman, 1995; Shavelson et al. 1997; Cave, Hanney, Henkel & Kogan, 1997):

- ✓ Η τιμή ενός δείκτη εκφράζει μια ποσότητα
- ✓ Οι δείκτες περιγράφουν συνθήκες οι οποίες μπορούν να βελτιωθούν
- ✓ Οι τιμές των δεικτών εξαρτώνται από το χρόνο
- ✓ Οι δείκτες θεωρούνται συχνά βασικό στοιχείο στην ανάπτυξη θεωριών για κάποιο συγκεκριμένο θέμα

Συνοψίζοντας όλα τα παραπάνω, για να είναι ένας δείκτης αποτελεσματικός στη χρήση του πρέπει να είναι σαφώς και λεπτομερώς ορισμένος. Επιπλέον είναι σημαντικό στην περιγραφή του δείκτη που χρησιμοποιείται, να αναφέρεται ποια είναι η μονάδα μέτρησης που χρησιμοποιείται. Όλα τα τεχνικά στοιχεία και οποιαδήποτε άλλη λεπτομέρεια πρέπει να ορίζεται εξ'αρχής, έτσι ώστε να διασφαλίζεται το ότι διαφορετικοί άνθρωποι σε διαφορετικές χρονικές στιγμές, θα μπορούν να συλλέξουν τα ίδια ακριβώς δεδομένα για ένα συγκεκριμένο δείκτη.

### 2.3.2 Οι δείκτες στον Παγκόσμιο Ιστό

Όπως έχει ήδη γίνει φανερό από τα παραπάνω, οι δείκτες είναι το απαραίτητο στοιχείο για τη διεξαγωγή προβλέψεων. Η εργασία αυτή επικεντρώνεται στην διεξαγωγή πρόβλεψης που αφορά σε ένα portal, έτσι είναι απαραίτητο να γίνει αναφορά στους δείκτες σε επίπεδο παγκόσμιου ιστού (web).

Η ανάπτυξη του Παγκόσμιου Ιστού (Web) και η τεχνολογική εξέλιξη των τελευταίων ετών, έχει οδηγήσει τις επιχειρήσεις, τους οργανισμούς ή ακόμα και μεμονωμένα άτομα στην ηλεκτρονική παρουσία τους στο Internet, μέσω της δημιουργίας διαφόρων ειδών ιστοσελίδων, όπως απλές ιστοσελίδες (sites), πύλες ενημέρωσης (portals), ηλεκτρονικά καταστήματα (eshops) κλπ. Για να είναι όλες αυτές οι σελίδες ανταγωνιστικές και αποτελεσματικές, είναι απαραίτητο να γίνονται μετρήσεις και προβλέψεις σχετικά με τη χρήση τους, την αποδοχή τους από τους συμμετέχοντες-πελάτες και την ευκολία που έχουν στην πλοήγησή τους.

Έτσι στον χώρο του Παγκόσμιου Ιστού, η έννοια των δεικτών (indicators) ισοδυναμεί με την έννοια *μετρικές ιστού (web metrics ή web analytics)*.

Οι *μετρικές ιστού (Web metrics-Web analytics)* είναι το μέσο που απεικονίζει τον τρόπο με τον οποίο οι χρήστες αλληλεπιδρούν με μια ιστοσελίδα (Burby & Brown, 2007).

Με άλλα λόγια οι *μετρικές ιστού* είναι η διαδικασία ανάλυσης της συμπεριφοράς των επισκεπτών-χρηστών μιας ιστοσελίδας. Η χρήση αυτών των μετρικών δίνει τη δυνατότητα σε μια επιχείρηση ή έναν οργανισμό να προσελκύσει περισσότερους επισκέπτες, να διατηρήσει τους ήδη υπάρχοντες χρήστες, να προσφέρει νέες ή βελτιστοποιημένες υπηρεσίες στους χρήστες της και τέλος να αυξήσει ακόμα και τα έσοδά της μέσα από τη χρήση της ιστοσελίδας της.

Συνοπτικά λοιπόν η στατιστική ανάλυση των μετρικών ιστού έχει ως στόχο:

- Τη βελτιστοποίηση της λειτουργίας της ιστοσελίδας και του περιεχομένου της
- Τη βελτίωση της εμπειρίας του χρήστη κατά την πλοήγησή του στην ιστοσελίδα
- Την εύρεση των δυνατών και αδύνατων σημείων της ιστοσελίδας



Συνεπώς η πληροφορία που αντλείται από την ανάλυση των δεδομένων αυτών, είναι σημαντική καθώς αποκαλύπτει το τι πρέπει να βελτιωθεί, τι προτιμούν οι χρήστες και τι δεν είναι τόσο σημαντικό για αυτούς.

### **Σημαντικές μετρικές ιστού**

Κάποιοι από τους πιο σημαντικούς δείκτες που μπορούν να χρησιμοποιηθούν για την αξιολόγηση και την πρόβλεψη της πορείας μιας ιστοσελίδας είναι (Burby & Brown, 2007):

- Ο αριθμός των επισκέψεων στην ιστοσελίδα.
- Ο αριθμός των σελίδων που «βλέπει» ένας χρήστης.
- Ο αριθμός των νέων επισκεπτών μιας ιστοσελίδας.
- Χρόνος παραμονής στην ιστοσελίδα.

### **2.3.3 Μετρικές χρήσης portal**

Στην ενότητα αυτή θα εξειδικευτεί ακόμα περισσότερο η έννοια του δείκτη – μετρικής πάνω στο κυρίως αντικείμενο της εργασίας αυτής που είναι οι πύλες (portals). Έτσι θα παρουσιαστεί αρχικά ο ορισμός του portal, ποιες είναι οι διαφορές του σε σχέση με μια απλή ιστοσελίδα και ποιοι δείκτες μέτρησης είναι σημαντικοί για την απεικόνιση της χρήσης του.

#### **2.3.3.1 Τι είναι portal**

Η ανάπτυξη του Internet και της τεχνολογίας έχει ως αποτέλεσμα ότι οι χρήστες μπορούν να ζητούν την πρόσβαση σε πληροφορίες από όπου κι αν βρίσκονται και στο χρόνο που αυτοί θέλουν. Δεδομένου ότι η δομή του διαδικτύου είναι αχανής και η πληροφορία βρίσκεται διεσπαρμένη σε πολλές ιστοσελίδες, προέκυψε η ανάγκη δημιουργίας ιστοσελίδων που απευθύνονται σε συγκεκριμένες ομάδες ανθρώπων με συγκεκριμένα ενδιαφέροντα ή συγκεκριμένο επάγγελμα ή ένα συγκεκριμένο hobby, με σκοπό να τους καθοδηγούν στην γρήγορη ανεύρεση των πληροφοριών που αναζητούν. Οι ιστοσελίδες αυτές ονομάζονται πύλες – *Portals*.

Έτσι, το *Portal* είναι μια ιστοσελίδα που χρησιμεύει ως πύλη σε άλλους προορισμούς ή δραστηριότητες στο Διαδίκτυο. Ο σκοπός του είναι να καθοδηγήσει τους επισκέπτες σε οτιδήποτε θέλουν να βρουν on-line και να παρέχει όσο το δυνατόν περισσότερες υπηρεσίες όπως : δωρεάν υπηρεσίες ηλεκτρονικού ταχυδρομείου, online chat forums, εξατομικευμένες ειδήσεις, αγορές, υπηρεσίες αναζήτησης και πολλά άλλα.

Παρακάτω παρουσιάζονται κάποια παραδείγματα portals από τον ελληνικό χώρο που απευθύνονται σε όλους τους συμμετέχοντες στην εκπαιδευτική διαδικασία και κυρίως τους εκπαιδευτικούς.

- Πύλη ενημέρωσης για Πρωτοβάθμια και Δευτεροβάθμια Εκπαίδευση : <http://www.pde.gr>
- Πύλη ενημέρωσης για εκπαιδευτικούς: [www.alfavita.gr/index.php](http://www.alfavita.gr/index.php)
- Έδρα Εκπαίδευσης: [www.edra.gr](http://www.edra.gr)
- Πανελλήνιο Σχολικό Δίκτυο : [www.sch.gr](http://www.sch.gr)
- Πύλη ενημέρωσης για εκπαιδευτικούς : <http://www.ekp.gr/>
- Πύλη ενημέρωσης για εκπαιδευτικούς : <http://www.tetradio.gr/>

Η παρούσα εργασία μελετά και εξετάζει το πρώτο αναφερόμενο portal, ως ένα ενδεικτικό παράδειγμα.

#### 2.3.3.2 Σημαντικές μετρικές portal

Η δημιουργία ενός portal από μια επιχείρηση-οργανισμό απαιτεί σημαντικές επενδύσεις σε χρόνο, προσπάθεια και χρήματα. Με την εξέλιξη και την ευρεία υιοθέτηση των portals τα τελευταία χρόνια, η μέτρηση της χρήσης τους είναι πλέον αναγκαία για να είναι ανταγωνιστικά. Παρόλο που η χρήση ενός portal δεν είναι από μόνη της ένας επαρκής δείκτης για την ανάλυση ανταποδοτικότητας (return on investment (ROI) analysis), δεν παύει να αποτελεί έναν από τους κύριους παράγοντες. Η γνώση του πως χρησιμοποιείται το portal από τους χρήστες και οι διάφορες εκδοχές της χρήσης του μέσα στο χρόνο μπορεί να αποδείξει ή όχι αν ένα portal είναι επιτυχημένο (Haddad,2004).

Σε αντιστοιχία με την ανάλυση της χρήσης μιας ιστοσελίδας, η καταγραφή του αριθμού των επισκέψεων σε ένα portal είναι σημαντική για την κατανόηση του ποια μέρη του portal του προσδίδουν αξία και ποια ίσως χρίζουν βελτίωσης. Παρόλα αυτά σε αντίθεση με τις παραδοσιακές ιστοσελίδες, τα portal έχουν κάποια μοναδικά χαρακτηριστικά τα οποία αποτελούν ξεχωριστές ευκαιρίες να μετρηθεί η χρήση τους και η συμπεριφορά των χρηστών.

Η βασική διαφορά των portal από μια τυπική ιστοσελίδα είναι το ότι τα portal απευθύνονται σε συγκεκριμένες κατηγορίες χρηστών. Αυτό σημαίνει ότι οι χρήστες ενός portal έχουν κάποιο κοινό χαρακτηριστικό (π.χ κοινό επάγγελμα ή κοινό ενδιαφέρον) και δεν είναι ανώνυμοι. Για να εισέλθει ένας χρήστης σε ένα portal και να μπορεί να δει όλες τις πληροφορίες που παρέχονται, πρέπει πρώτα να εγγραφεί σε αυτό δίνοντας ένα όνομα χρήστη (username) και έναν κωδικό (password). Αυτό είναι το στοιχείο που διαφοροποιεί σημαντικά την διαδικασία επίσκεψης σε ένα portal από την απλή πλοήγηση σε μια τυπική ιστοσελίδα (Haddad,2004).

Έτσι ενώ σε μια ιστοσελίδα η μέτρηση επισκεψιμότητας είναι απλά καταγραφή των ανώνυμων διαφορετικών IP διευθύνσεων που έχουν πρόσβαση στη σελίδα και πλοηγούνται σε αυτή, στα portals καταγράφεται ο συγκεκριμένος χρήστης με βάση τον κωδικό του και όλες οι ενέργειες που κάνει συσχετίζονται αυτόματα με αυτό το μοναδικό κωδικό (user ID). Τέτοιες ενέργειες είναι για παράδειγμα: ποιες σελίδες του portal έχει επισκεφθεί, ποια έγγραφα έχει «κατεβάσει»(download), τι αναζητήσεις περιεχομένου έκανε, αν επισκέφθηκε ή όχι το forum, τι συζητήσεις έκανε κλπ. Συνεπώς, η μέτρηση της χρήσης ενός portal δίνει τη δυνατότητα στους διαχειριστές του να μάθουν περισσότερα σχετικά με τη συμπεριφορά των συγκεκριμένων χρηστών καθώς επίσης να γίνεται σύγκριση μεταξύ των χρηστών και αυτοί να ομαδοποιούνται με βάση τη χρήση που κάνουν (π.χ οι 10 πιο ενεργοί χρήστες).

Συνοπτικά λοιπόν μπορούμε να πούμε ότι κάποιες βασικές μετρικές των portal είναι οι παρακάτω:

- ✓ Σύνολο μελών
- ✓ Σύνολο μηνυμάτων
- ✓ Σύνολο θεμάτων
- ✓ Σύνολο κατηγοριών

- ✓ Συνδεδεμένοι χρήστες
- ✓ Περισσότεροι συνδεδεμένοι
- ✓ Συνολικές εμφανίσεις σελίδων
- ✓ Μέσος όρος εγγραφών ανά ημέρα
- ✓ Μέσος όρος μηνυμάτων ανά ημέρα
- ✓ Μέσος όρος θεμάτων ανά ημέρα
- ✓ Σύνολο πινάκων
- ✓ Μέσος όρος συνδεδεμένων ανά ημέρα
- ✓ Λόγος ανδρών/γυναικών
- ✓ Μέσος όρος εμφανίσεων σελίδων ανά ημέρα

Ο αριθμός των επισκεπτών σε ένα portal ή αλλιώς σύμφωνα με τα παραπάνω οι περισσότεροι συνδεδεμένοι χρήστες ανά ημέρα, είναι ο δείκτης που μας απασχόλησε στην εργασία αυτή και πάνω στη χρονοσειρά του οποίου έγινε η πρακτική εφαρμογή των διαφόρων μοντέλων που παρουσιάζονται. Έτσι θεωρείται χρήσιμο να αναλυθεί λίγο περισσότερο, προκειμένου να γίνει κατανοητός ο λόγος για τον οποίο επιλέχθηκε ο συγκεκριμένος δείκτης έναντι των άλλων που αναφέρθηκαν παραπάνω.

Συνοπτικά λοιπόν ο δείκτης αυτός αναφέρεται στον αριθμό των μεμονωμένων ατόμων – χρηστών του portal σε μια συγκεκριμένη χρονική περίοδο ημέρα, μήνας, έτος τα οποία είχαν κάποια δραστηριότητα μέσα στο portal (π.χ επίσκεψη στα διάφορα θέματά του).

Στο συγκεκριμένο εκπαιδευτικό portal, το οποίο επιλέχθηκε για την εργασία αυτή, και είναι η πύλη ενημέρωσης για Πρωτοβάθμια και Δευτεροβάθμια Εκπαίδευση, οι κυριότεροι χρήστες είναι οι εκπαιδευτικοί και οι μαθητές. Έτσι ο δείκτης αυτός μπορεί να δώσει στοιχεία για το ποιες συγκεκριμένες ημέρες ο αριθμός των συνδεδεμένων χρηστών ήταν μεγαλύτερος (π.χ σε περίοδο μεταθέσεων εκπαιδευτικών πόσοι εκπαιδευτικοί αναζητούσαν πληροφορίες στο συγκεκριμένο portal) ή ποιες ημέρες ο αριθμός των χρηστών μειωνόταν σημαντικά (π.χ σε περίοδο διακοπών, πόσοι χρήστες ασχολούνταν με τα εκπαιδευτικά θέματα).

## 2.4 Η έννοια της χρονοσειράς

### 2.4.1 Ορισμός χρονοσειράς

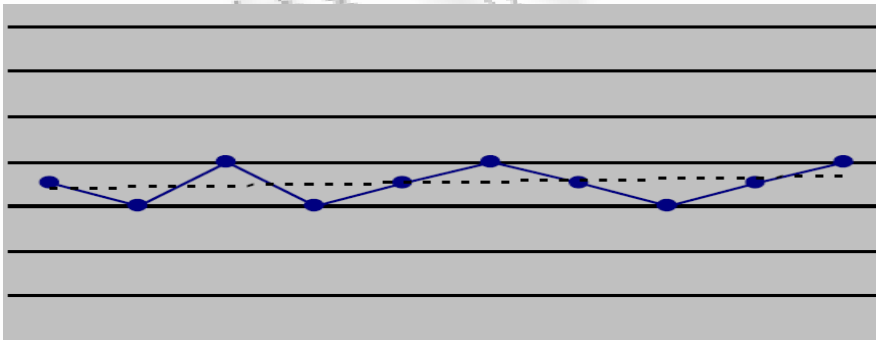
Με τον όρο χρονοσειρά εννοούμε μια σειρά από παρατηρήσεις που παίρνονται σε ορισμένες χρονικές στιγμές ή περιόδους που ισαπέχουν μεταξύ τους.

Τα τρία κυριότερα στοιχεία της ανάλυσης χρονοσειρών είναι η περιγραφή, η επεξήγηση και η πρόβλεψη των εξαρτημένων δεδομένων. Η περιγραφή επιτυγχάνεται με τη βοήθεια διαφόρων γραφημάτων, η επεξήγηση χρησιμοποιώντας κάποια μορφή μοντέλα για να εξερευνηθούν οι μηχανισμοί δημιουργίας της χρονοσειράς και η πρόβλεψη περιλαμβάνει τη χρησιμοποίηση ενός μοντέλου για να προβλεφθούν μελλοντικές τιμές της σειράς.

#### *Ποιοτικά χαρακτηριστικά χρονοσειρών*

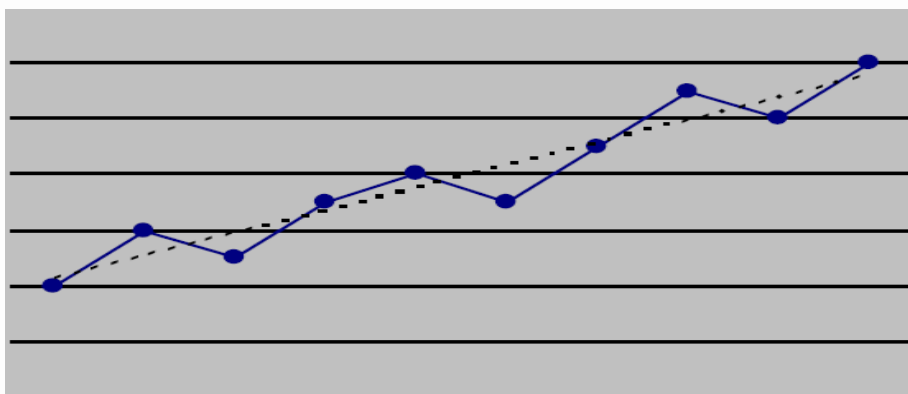
Οι χρονοσειρές είναι ιστορικά δεδομένα που απαρτίζονται από διαδοχικές παρατηρήσεις μέσα σε ένα χρονικό διάστημα. Οι παρατηρήσεις γίνονται ανά σταθερό χρονικό βήμα και μπορούν να είναι ετήσιες, τριμηνιαίες, μηνιαίες, εβδομαδιαίες, ημερήσιες κτλ. Τα ποιοτικά χαρακτηριστικά αυτών των παρατηρήσεων είναι τα εξής (Σούκος & Κάντζιος, 2006):

- ✓ *Στασιμότητα* (Stationary), όταν οι τιμές κυμαίνονται γύρω από μία μέση τιμή.



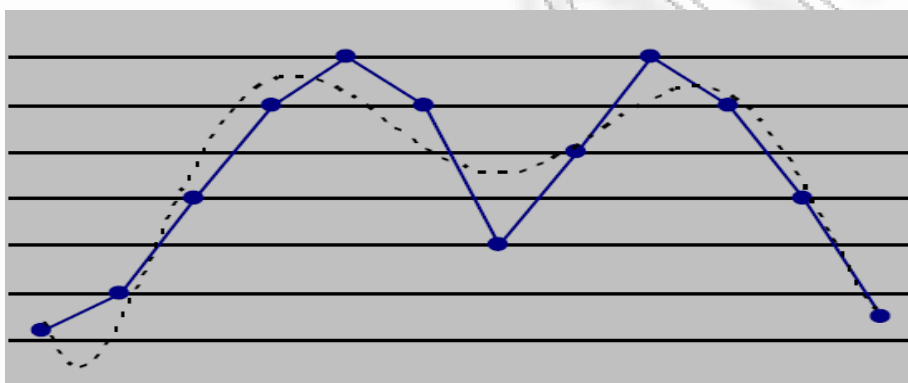
**Σχήμα 2.1:** Στασιμότητα

- ✓ *Τάση* (Trend) όταν υπάρχει μια μακροπρόθεσμη αύξηση ή μείωση του επιπέδου των τιμών.



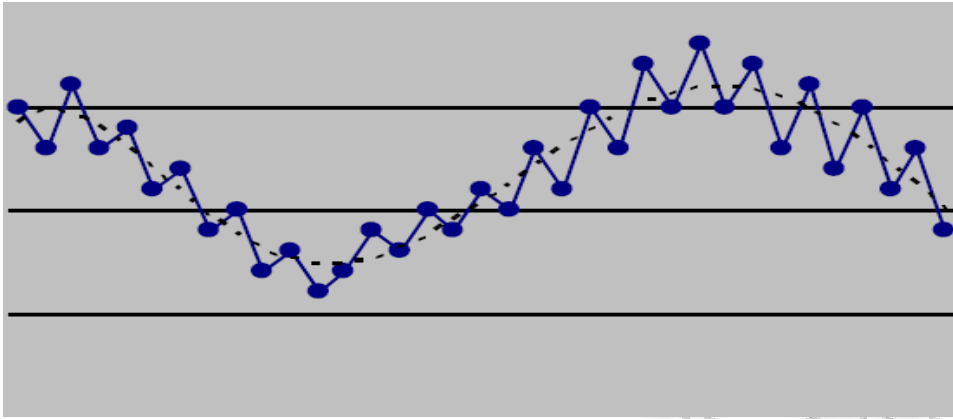
Σχήμα 2.2: Τάση

- ✓ *Εποχικότητα* (Seasonal) όταν η χρονοσειρά επηρεάζεται από εποχιακούς παράγοντες.



Σχήμα 2.3: Εποχικότητα

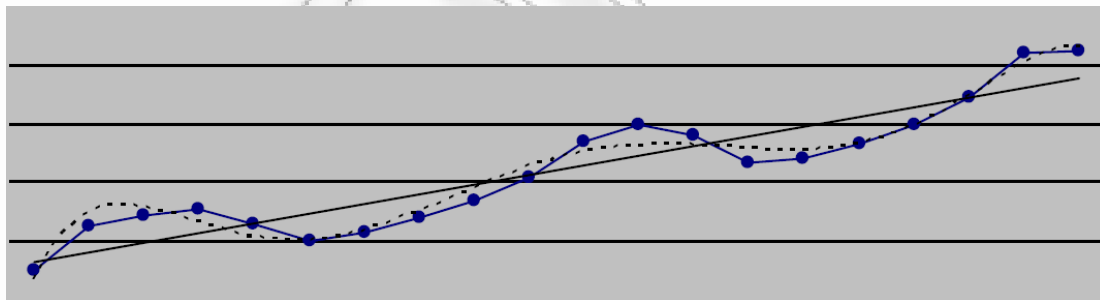
- ✓ *Κυκλικότητα* (Cyclical) όταν οι τιμές αυξομειώνονται, αλλά όχι σε σταθερές περιόδους. Ο χρονικός ορίζοντας είναι συνήθως μεγαλύτερος (5-10 χρόνια) και είναι συνήθως συνδεδεμένη με τις διακυμάνσεις στο επίπεδο της συνολικής οικονομίας (γνωστές ως περίοδοι οικονομικής ύφεσης και οικονομικής ανάπτυξης)



**Σχήμα 2.4:** Κυκλικότητα

- ✓ *Ασυνέχειες* (outliers), οι οποίες οφείλονται είτε σε ασυνήθιστες τιμές, είτε σε αλλαγή επιπέδου (level shift).
- ✓ *Τυχειότητα* (Irregular-Random), όταν έχουμε διακυμάνσεις λόγω τυχαίων γεγονότων.

Αν εξετάσουμε μια οποιαδήποτε χρονοσειρά, διαπιστώνουμε ότι αποτελεί συνδυασμό ενός ή περισσότερων από τα παραπάνω στοιχεία. Στο παρακάτω διάγραμμα φαίνεται μια χρονοσειρά με εμφανή την ύπαρξη αυξητικής τάσης και, ταυτόχρονα, εποχικότητας.



— γραμμική αυξητική τάση - - - - - εποχικότητα

**Σχήμα 2.5:** Συνδυασμός διαφόρων στοιχείων χρονοσειράς

Όπως έχουμε ήδη αναφέρει, οι περισσότερες μέθοδοι προβλέψεων στηρίζονται, σε κάποιο βαθμό, στις βασικές έννοιες της στατιστικής ανάλυσης. Ο πίνακας 2.1, παρουσιάζει συνοπτικά τους πιο βασικούς στατιστικούς δείκτες:



**Πίνακας 2.1:** Βασικοί στατιστικοί δείκτες

Δείκτης	Τύπος
Μέση τιμή (mean)	$Mean = \frac{\sum_{i=1}^n x_i}{n}$
Διάμεσος (median)	Η μεσαία παρατήρηση αν (n) περιττός Ο μέσος των δύο «μεσαίων» παρατηρήσεων αν (n) άρτιος
Μέση Απόλυτη Απόκλιση (Mean absolute deviation)	$MAD = \frac{1}{n} \sum  Y_i - \bar{Y} $
Μέση Τετραγωνική Απόκλιση (Mean Square deviation)	$MSD = \frac{1}{n} \sum (Y_i - \bar{Y})^2$
Διακύμανση (Variance)	$S^2 = \frac{1}{n-1} \sum (Y_i - \bar{Y})^2$
Τυπική απόκλιση (Standard Deviation)	$S = \sqrt{S^2} = \sqrt{\frac{1}{n-1} \sum (Y_i - \bar{Y})^2}$
Συνδιακύμανση (Covariance)	$COV_{xy} = \frac{1}{n-1} \sum (X_i - \bar{X})(Y_i - \bar{Y})$
Συσχέτιση (Correlation)	$r_{xy} = \frac{COV_{xy}}{S_x S_y} = \frac{\sum (X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum (X_i - \bar{X})^2} \sqrt{\sum_{i=1}^n (Y_i - \bar{Y})^2}}, +1 \geq r_{xy} \geq -1$
Αυτοδιακύμανση (Autocovariance)	$C_k = \frac{1}{n} \sum_{t=k+1}^n (Y_t - \bar{Y})(Y_{t-k} - \bar{Y})$
Αυτοσυσχέτιση (Autocorrelation lagged k)	$r_k = \frac{\sum_{t=k+1}^n (Y_t - \bar{Y})(Y_{t-k} - \bar{Y})}{\sum_{i=1}^n (Y_i - \bar{Y})^2}$



Επίσης, σκόπιμο είναι να παρουσιάσουμε τα κριτήρια αξιολόγησης των μεθόδων πρόβλεψης τα οποία χρησιμοποιούνται για την επιλογή της «κατάλληλης» μεθόδου. Τα κριτήρια αυτά βασίζονται στις τιμές των αποκλίσεων των προβλεπόμενων τιμών από τις αντίστοιχες πραγματικές τιμές της χρονοσειράς. Όσο πιο μικρές είναι οι αποκλίσεις αυτές, τόσο πιο καλή θεωρείται η μέθοδος πρόβλεψης που χρησιμοποιήθηκε, δεδομένου ότι είναι σε θέση να προσδιορίζει, με σχετικά μεγάλη ακρίβεια, τις τιμές της χρονοσειράς. Αντίθετα, όσο περισσότερο απέχουν οι πραγματικές τιμές από τις προβλεπόμενες, τόσο λιγότερο αξιόλογη είναι η χρησιμοποιούμενη μέθοδος πρόβλεψης.

**Πίνακας 2.2:** Κατηγορίες σφαλμάτων και αποκλίσεων

Δείκτης	Τύπος
Σφάλμα	$e_t = Act - F$
Μέσο σφάλμα	$ME = \frac{1}{n} \sum_{t=1}^n e_t$
Μέσο απόλυτο σφάλμα	$MAE = \frac{1}{n} \sum_{t=1}^n  e_t $
Μέσο τετραγωνικό σφάλμα	$MSE = \frac{1}{n} \sum_{t=1}^n e_t^2$
Μέσο ποσοστιαίο σφάλμα	$MPE = \frac{1}{n} \sum_{t=1}^n PE_t = \frac{1}{n} \sum_{t=1}^n \left( \frac{Act - F}{Act} \right) 100$
Μέσο απόλυτο ποσοστιαίο σφάλμα	$MAPE = \frac{\sum_{i=1}^n \left  \frac{F_i - Act_i}{Act_i} \right }{n} \cdot 100$

#### 2.4.2 Πρόβλεψη χρονοσειράς

Η πρόβλεψη χρονοσειράς (timeseries prediction) είναι μια μέθοδος απεικόνισης παρελθοντικών σημείων της χρονοσειράς σε μελλοντικά.

Η μέθοδος βασίζεται στην υπόθεση ότι η μεταβολή της τιμής του μεγέθους ακολουθεί ένα συγκεκριμένο πρότυπο («λανθάνον πρότυπο») που επαναλαμβάνεται στο χρόνο

και παραμένει σταθερό. Βασικός στόχος αυτής της μεθόδου είναι η αναγνώριση του ακολουθούμενου προτύπου των ιστορικών δεδομένων και η προέκταση του στο μέλλον.

Με μια έννοια, το μοντέλο χρονοσειρών θεωρείται σαν ένα «μαύρο κουτί» (black box) που δεν κάνει καμία προσπάθεια να ανακαλύψει τους συντελεστές που επηρεάζουν τη συμπεριφορά του.

Η παραστατική μορφή του μοντέλου χρονοσειρών δίνεται στην παρακάτω εικόνα:



**Εικόνα 2.1:** Διάγραμμα μοντέλου χρονοσειρών

Είσοδος του μοντέλου είναι παρελθοντικές τιμές  $X_i$  μέχρι τη χρονική στιγμή  $x = t$  και έξοδος  $Y$  είναι η πρόβλεψη της μελλοντικής τιμής τη χρονική στιγμή  $x = t+P$ . Η βασική μέθοδος για τον τύπο αυτό πρόβλεψης είναι η δημιουργία μιας απεικόνισης από  $D$  σημεία της χρονοσειράς που απέχουν κατά  $\Delta$  μονάδες μεταξύ τους, π.χ.  $x(t-(D-1)\Delta), \dots, x(t-\Delta), x(t)$  προκειμένου να επιτευχθεί πρόβλεψη της μελλοντικής τιμής  $x(t+P)$ .

#### 2.4.3 Πλεονεκτήματα μοντέλου χρονοσειρών για πρόβλεψη

Υπάρχουν τρία βασικά πλεονεκτήματα της θεώρησης του «μαύρου κουτιού» που αποτελούν και τους βασικούς λόγους της συχνότατης επιλογής του μοντέλου των χρονοσειρών:

1. Δεν υπάρχει πάντα η δυνατότητα να συσχετίσουμε ένα μεταβαλλόμενο μέγεθος με κάποιους παράγοντες και πολύ περισσότερο να προσδιορίσουμε τον τρόπο αλληλεπίδρασής τους.

2. Σε πολλές περιπτώσεις ενδιαφερόμαστε να προσδιορίσουμε μόνο το τι θα συμβεί και όχι το γιατί.
3. Το κόστος που απαιτείται στην περίπτωση αυτή είναι πολύ μικρότερο σε σχέση με άλλες κατηγορίες μοντέλων (όπως το επεξηγηματικό).

#### **2.4.4 Ανάλυση χρονοσειρών (Time series analysis)**

Για τη διενέργεια προβλέψεων υπάρχει ένας μεγάλος αριθμός μεθόδων που μπορούν να χρησιμοποιηθούν ανάλογα με την εξεταζόμενη περίπτωση. Οι περισσότερες από αυτές αναπτύχθηκαν πριν από μερικές δεκαετίες και εφαρμόστηκαν σε ευρεία κλίμακα λόγω της ραγδαίας ανάπτυξης κατάλληλων υπολογιστικών προγραμμάτων. Στόχος μας είναι η χρησιμοποίηση εκείνων των μεθόδων που είναι εύχρηστες, εφαρμόζονται συχνά και έχουν αποδειχθεί στην πράξη ότι δίνουν καλά αποτελέσματα. Μάλιστα, αν ληφθεί υπόψη η ευκολία με την οποία προσδιορίζονται οι προβλέψεις με τη χρήση των H/Y, τότε το κύριο βάρος της έρευνας θα πρέπει να εστιάζεται αφενός στην αξιολόγηση των αποτελεσμάτων και αφετέρου στην επιλογή της καταλληλότερης μεθόδου πρόβλεψης.

Η *ανάλυση χρονοσειρών (time series analysis)* ασχολείται αποκλειστικά με τη διερεύνηση της διαχρονικής συμπεριφοράς των τιμών μιας μεταβλητής, οι παρατηρήσεις της οποίας προέρχονται από χρονοσειρά. Η πρόβλεψη των μελλοντικών τιμών της μεταβλητής σύμφωνα με την γραμμική ανάλυση χρονοσειρών μπορεί να προέλθει από τις ακόλουθες τρεις κατηγορίες μεθόδων προβλέψεων (Αγιακόγλου & Οικονόμου, 2004):

- ✓ **Μέθοδοι εξομάλυνσης (smoothing methods)**
- ✓ **Διάσπαση Χρονοσειρών (time series decomposition)**
- ✓ **Ανάλυση ARIMA (Autoregressive Intergated Moving Average Analysis)**

##### **2.4.4.1 Μέθοδοι εξομάλυνσης (smoothing methods)**

Οι μέθοδοι εξομάλυνσης είναι τεχνικές με τις οποίες προσδιορίζονται οι μελλοντικές τιμές μιας μεταβλητής με βάση τον τρόπο εφαρμογής τους. Οι τεχνικές αυτές

ονομάζονται μέθοδοι εξομάλυνσης, διότι η δημιουργία των προβλέψεων προέρχεται από την εξομάλυνση της διαχρονικής εξέλιξης των τιμών της μεταβλητής, ώστε να αναγνωριστεί καλύτερα ο τρόπος συμπεριφοράς της. Ορισμένες από αυτές τις μεθόδους μπορούν να εφαρμοστούν και σε περιπτώσεις μικρού αριθμού παρατηρήσεων της μεταβλητής.

Οι μέθοδοι αυτές θα περιγραφούν και θα αναλυθούν εκτενέστερα στην ενότητα 3.3 καθώς επίσης θα παρουσιαστούν και τα παραδείγματα εφαρμογής τους στο αντικείμενο της εργασίας αυτής.

Συνοπτικά, οι μέθοδοι εξομάλυνσης που θα παρουσιαστούν είναι:

- Η μέθοδος του κινητού μέσου (Moving average)
- Απλή εκθετική εξομάλυνση (Simple exponential smoothing)
- Διπλός κινητός μέσος (double moving average)
- Διπλή εκθετική εξομάλυνση (Double exponential smoothing)
- Εκθετική εξομάλυνση με προσαρμογή στην τάση (Exponential smoothing adjusted for trend)
- Εκθετική εξομάλυνση με προσαρμογή στην τάση και στην εποχικότητα (Exponential smoothing adjusted for trend and seasonality)

#### 2.4.4.2 Διάσπαση χρονοσειρών

Οι μέθοδοι διάσπασης χρονοσειρών στηρίζονται στην υπόθεση ότι οι τιμές μιας χρονοσειράς σχηματίζονται από τα στοιχεία που τη συνθέτουν και τα οποία είναι η *τάση*, η *κυκλικότητα*, η *εποχικότητα* και η *μη-κανονικότητα*. Για τη δημιουργία των προβλέψεων με τη μέθοδο αυτή, η χρονοσειρά διασπάται στα τέσσερα παραπάνω συνθετικά της στοιχεία και προσδιορίζεται η επιρροή που έχει καθένα από αυτά στη διαμόρφωση των τιμών της μεταβλητής.

Οι χρονοσειρές μελετώνται κάτω από την υπόθεση ότι με βάση τη συμπεριφορά στο παρελθόν μπορεί να προβλεφθεί η μελλοντική πορεία της διακύμανσης της παρατηρούμενης μεταβλητής. Μολονότι η πρόβλεψη μελλοντικών καταστάσεων δεν

γίνεται χωρίς σφάλμα, η ανάλυση και ο διαχωρισμός (decomposition) των χρονοσειρών έχει αξία στη διαδικασία της πρόβλεψης γιατί έτσι μπορεί να μειωθεί το σφάλμα πρόβλεψης. *Επισημαίνεται ότι είναι αδύνατο να προβλεφθεί το μέλλον με απόλυτη ακρίβεια και ότι πάντοτε θα υπάρχει, ως ένα βαθμό, σφάλμα στην πρόβλεψη.* Οι μέθοδοι για την ανάλυση και το διαχωρισμό χρονοσειρών είναι στατιστικές τεχνικές βελτίωσης των προβλέψεων μέσω της διάσπασης του προτύπου της χρονοσειράς σε υπο-πρότυπα που ονομάζονται συνιστώσες. Για το διαχωρισμό των χρονοσειρών δεχόμαστε ότι τα δεδομένα εμφανίζουν την εξής εικόνα:

**Δεδομένα = πρότυπο + σφάλμα =  $f$  ( συστηματικές συνιστώσες) + σφάλμα**

Το σφάλμα αυτό ή *υπολειπόμενη (residual) απόκλιση*, υποθέτουμε ότι μπορεί να εκφραστεί ως η διαφορά ανάμεσα στο συνδυασμό αποτελεσμάτων υπομοντέλων της χρονοσειράς (συστηματικές συνιστώσες) και των παρατηρηθεισών τιμών της. Σ' αυτήν την παράγραφο θα μελετήσουμε τη φύση και την έκταση (scope) αυτών των συνιστωσών. Οι τεχνικές διαχωρισμού που χρησιμοποιούνται για τη μελέτη αυτών των συνιστωσών δεν αποτελούν ενιαίο σώμα της στατιστικής θεωρίας. Επίσης, δεν πληρούνται συνήθως όλες οι προϋποθέσεις για επιστημονική στατιστική διερεύνηση, όπως, για παράδειγμα, η ύπαρξη τυχαίας δειγματοληψίας. Επομένως, οι διαδικασίες διάσπασης των χρονοσειρών συχνά αξιολογούνται με κριτήρια που επηρεάζονται από την κρίση εκείνου που διενεργεί την πρόβλεψη. Το υπόλοιπο της παραγράφου αφιερώνεται σε συζητήσεις γύρω από την ανάλυση χρονοσειρών και τις εφαρμογές της στη διαδικασία της πρόβλεψης. Σε αντίθεση με τις μεθόδους εξομάλυνσης, δεν υπάρχει σώμα στατιστικής θεωρίας που να έχει αναπτυχθεί γι' αυτό ακριβώς το ζήτημα. Επομένως, απαιτείται καλή αξιολογική κρίση προκειμένου να χρησιμοποιηθούν οι μέθοδοι αυτές σε προβλήματα πρόβλεψης. Εν τούτοις, επισημαίνοντας τους στατιστικούς περιορισμούς αυτών των τεχνικών, μπορούμε να μειώσουμε το διάστημα κατά το οποίο θα χρειαστεί να λειτουργήσει η υποκειμενική κρίση κι επομένως να βελτιώσουμε τη διαδικασία πρόβλεψης. (Σούκος & Κάντζιος, 2006)

Οι υπεύθυνοι για την πρόβλεψη αναλύουν αυτές τις συνιστώσες για να καθορίσουν την επίδρασή τους πάνω στην εξέλιξη της χρονοσειράς. Οι τέσσερις συνιστώσες αλληλεπιδρούν για να εξηγήσουν τη διακύμανση στις τιμές των δεδομένων

διαχρονικά. Οι επιδράσεις αυτές είναι υπεύθυνες για την συχνά ασταθή συμπεριφορά των χρονοσειρών.

### *Τάση*

Τάση είναι η συνεχής κίνηση της χρονοσειράς κατά μήκος μιας χρονικής περιόδου. Η τάση μπορεί να εμφανιστεί με γραμμική ή καμπυλόγραμμη μορφή. Συνήθως, οι δυνάμεις που επηρεάζουν ή παράγουν αυτές τις διαχρονικές κινήσεις περιλαμβάνουν μεταβολές στην τεχνολογία, στην παραγωγικότητα, στον εξωτερικό ανταγωνισμό και άλλα (οικονομικές μεταβλητές).

### *Εποχικότητα*

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

### *Κυκλικότητα*

Η κυκλικότητα αναφέρεται σε κυματοειδείς διακυμάνσεις (συχνά μη- κανονικές) που καλύπτουν χρονικό διάστημα μεγαλύτερο του έτους και προέρχονται από μεταβολές της γενικής δραστηριότητας. Αποτελούν τη διαφορά μεταξύ της αναμενόμενης τιμής της χρονοσειράς που προέρχεται από την τάση και την πραγματική τιμή της χρονοσειράς. Επομένως, η κυκλική κίνηση μπορεί να θεωρηθεί ως υπολειπόμενη απόκλιση κυμαινόμενη γύρω από την τάση που σχετίζεται με τη μεταβαλλόμενη γενική δραστηριότητα.

### *Μη κανονικότητα*

Η μη-συστηματική συνιστώσα, ή μη-κανονικές αποκλίσεις, προκαλείται από μη-περιοδικά ή απρόβλεπτα γεγονότα. Τα γεγονότα αυτά περιλαμβάνουν απεργίες, σεισμούς, πολέμους, καιρικές μεταβολές και αλλαγές νομοθεσίας που έχουν επιδράσεις στις τιμές της χρονοσειράς (οικονομικές μεταβλητές).

#### 2.4.4.3 Ανάλυση ARIMA-Box Jenkins

Τα παραπάνω μοντέλα πρόβλεψης αποτελούν τις βασικές τεχνικές χρονοσειρών. Στην υποενότητα αυτή γίνεται μια μικρή αναφορά σε πιο προχωρημένες προσεγγίσεις της ανάλυσης χρονοσειρών. Παρουσιάζονται οι μέθοδοι αυτοπαλινδρόμησης κινητού μέσου (ARMA) χρονοσειρών. Αυτές οι μέθοδοι πρόβλεψης περιλαμβάνουν γενικευμένες μεθόδους αναπροσαρμοζόμενου φιλτραρίσματος και τεχνικές Box-Jenkins.

Η ανάλυση ARIMA, γνωστή και ως Box-Jenkins ανάλυση, είναι μια περισσότερο σύνθετη στατιστική μέθοδος από τις δύο προηγούμενες, με την οποία αναγνωρίζεται ο τρόπος δημιουργίας των τιμών της μεταβλητής. Με τη μέθοδο αυτή επιλέγεται από μια μεγάλη κατηγορία υποδειγμάτων, εκείνο που θεωρείται ως το πιο κατάλληλο για τη διαμόρφωση των προβλέψεων.

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

Τα στοχαστικά μοντέλα περιέχουν τον τυχαίο παράγοντα (τυχαίο σφάλμα ή σφάλμα πρόβλεψης), τις τιμές του μεγέθους οι οποίες εμφανίστηκαν σε προηγούμενες χρονικές στιγμές και ίσως κάποιους άλλους στοχαστικούς παράγοντες. Το μοντέλο που προκύπτει είναι ένας γραμμικός συνδυασμός των παραπάνω ποσοτήτων. Γενικά, έχουν αναπτυχθεί πολλά και ποικίλα τέτοια μοντέλα για την περιγραφή των διακυμάνσεων κάποιου μεγέθους μέσα στο χρόνο. Τα μοντέλα ARIMA χρησιμοποιούνται ευρύτατα γιατί βρίσκουν εφαρμογή στη μελέτη πολλών μεγεθών και φαίνεται να δίνουν μια "καλή" εικόνα της διαχρονικής τους συμπεριφοράς, καθώς και ικανοποιητικά αποτελέσματα στην πρόβλεψη των μελλοντικών τιμών του μεγέθους.

Τα μοντέλα ARIMA έχουν μελετηθεί εκτεταμένα από τους Box και Jenkins, σε βαθμό που τα ονόματα των παραπάνω να είναι σχεδόν συνώνυμα με τις ARIMA διαδικασίες και τις εφαρμογές τους στην ανάλυση και την πρόβλεψη χρονοσειρών. Οι Box-Jenkins πρότειναν μια οικογένεια αλγεβρικών μοντέλων πρόβλεψης, από τα οποία μπορεί κάποιος να διαλέξει το "καταλληλότερο" για την πρόβλεψη μιας δεδομένης χρονοσειράς. Στα μοντέλα αυτά οι προβλέψεις βασίζονται αποκλειστικά στις παρελθούσες τιμές και τα εμφανισθέντα πρότυπα συμπεριφοράς της χρονοσειράς που εξετάζεται (Κουγιουμτζής, 2005).

Επειδή το θεωρητικό υπόβαθρο των μοντέλων ARIMA είναι αρκετά σύνθετο, είναι προτιμότερο θα δοθεί έμφαση στην προσπάθεια κατανόησης της ουσίας της μεθοδολογίας.

#### 2.4.4.4 Βασικές έννοιες προχωρημένης ανάλυσης

##### **Η συνάρτηση αυτοσυσχέτισης (autocorrelation function)**

Ο πλέον χρήσιμος στατιστικός δείκτης στην ανάλυση χρονοσειρών είναι ο συντελεστής αυτοσυσχέτισης (ή η συσχέτιση της χρονοσειράς με τον εαυτό της, για παρατηρήσεις που απέχουν μεταξύ τους 1,2 ή περισσότερες περιόδους). Ο συντελεστής αυτοσυσχέτισης (autocorrelation coefficient) ορίζεται:

$$r_k = \frac{\sum_{t=k+1}^n (Y_t - \bar{Y})(Y_{t-k} - \bar{Y})}{\sum_{t=1}^n (Y_t - \bar{Y})^2} \quad (2.1)$$

Ο  $r_1$  υποδηλώνει πώς οι διαδοχικές παρατηρήσεις (καθυστέρηση 1) της χρονοσειράς σχετίζονται μεταξύ τους, ο  $r_2$  πώς σχετίζονται οι παρατηρήσεις που απέχουν δύο χρονικές περιόδους (καθυστέρηση 2) κλπ. Οι συντελεστές αυτοσυσχέτισης για καθυστερήσεις 1,2,..., δημιουργούν τη συνάρτηση αυτοσυσχέτισης (autocorrelation function) ή ACF.

##### **Μοντέλο "λευκού θορύβου" (white noise model)**

Η σχέση  $Y_t = c + e_t$  περιγράφει ένα απλό τυχαίο μοντέλο, όπου η παρατήρηση  $Y_t$  εκφράζεται από δύο μέρη: ένα συνολικό επίπεδο  $c$  και μια συνιστώσα τυχαίου



σφάλματος  $e_t$ . Ένα τέτοιο μοντέλο, το οποίο συχνά καλείται μοντέλο "λευκού θορύβου" και αποτελεί θεμελιώδες μοντέλο σε πολλές τεχνικές ανάλυσης χρονοσειρών.

### Δειγματική κατανομή των αυτοσυσχετίσεων

Για ένα μοντέλο "λευκού θορύβου" η δειγματική θεωρία του  $r_k$  είναι γνωστή και συνεπώς μπορούν να μελετηθούν οι ιδιότητες της ACF. Γενικότερα, ένας τρόπος προσέγγισης του προβλήματος είναι να εξετασθεί καθεμιά από τις τιμές του  $r_k$  και με βάση το τυπικό σφάλμα (standard error) να ελεγχθεί εάν η τιμή αυτή είναι σημαντικά διάφορη του μηδενός.

Θεωρητικά, όλοι οι συντελεστές αυτοσυσχέτισης μια σειράς τυχαίων αριθμών πρέπει να είναι ίσοι με το μηδέν, αλλά στην πράξη, οι συντελεστές αυτοσυσχέτισης του δείγματος δεν είναι ακριβώς μηδέν επειδή τα δείγματα είναι πεπερασμένα. Έχει δειχθεί, ότι για ένα μοντέλο λευκού θορύβου η κατανομή των συντελεστών αυτοσυσχέτισης μπορεί να προσεγγισθεί από την καμπύλη της κανονικής κατανομής με μέση τιμή 0 και τυπικό σφάλμα  $1/n$  όπου  $n$  ο αριθμός των παρατηρήσεων. Η πληροφορία αυτή μπορεί να χρησιμοποιηθεί για την ανάπτυξη ελέγχων υποθέσεων. Για παράδειγμα, 95% των συντελεστών αυτοσυσχέτισης του δείγματος πρέπει να ανήκουν στο διάστημα που ορίζεται από τη μέση τιμή συν ή πλην 1.96 standard errors, δηλαδή για ένα μοντέλο λευκού θορύβου στο  $\pm 1.96/n$ . Εάν αυτό δεν συμβαίνει τότε πιθανώς δεν πρόκειται για μοντέλο λευκού θορύβου.

### Στατιστικοί δείκτες Q, Q\* (Box-Pierce, Ljung-Box)

Ο δείκτης Q εξετάζει συνολικά τους συντελεστές αυτοσυσχέτισης και ελέγχει εάν το σύνολο είναι σημαντικά διάφορο του μηδενός.

$$Q = n \sum_{k=1}^h r_k^2 \quad (2.2)$$

όπου  $h$  η τάξη των μεταβλητών καθυστέρησης και  $n$  το πλήθος των παρατηρήσεων. Συνήθως  $h \approx 20$  ή  $h \approx 36$ . Ο Q ακολουθεί την κατανομή  $\chi^2$  με  $h-m$  βαθμούς ελευθερίας όπου  $m$  οι παράμετροι του μοντέλου, και η τιμή του συγκρίνεται με τις

τιμές του πίνακα  $x_2$  για να εκτιμηθεί το επίπεδο της σημαντικότητάς του. Για ένα μοντέλο λευκού θορύβου  $m=0$ , έτσι σαν εναλλακτικός δείκτης χρησιμοποιείται ο δείκτης  $Q^*$  ο οποίος επίσης εξετάζει συνολικά τους συντελεστές αυτοσυσχέτισης.

$$Q^* = n(n+2) \sum_{k=1}^h (n-k)^{-1} r_k^2 \quad (2.3)$$

Ο  $Q^*$  ακολουθεί την ίδια κατανομή με τον  $Q$ .

### Συντελεστής μερικής αυτοσυσχέτισης (Partial Autocorrelation Coefficient-PACF)

Ο συντελεστής μερικής αυτοσυσχέτισης χρησιμοποιείται σαν μέτρο του βαθμού της σχέσης ανάμεσα στην  $Y_t$  και την  $Y_{t-k}$ , όταν οι επιδράσεις όλων των άλλων μεταβλητών καθυστέρησης 1,2,3 , ... ,  $k-1$  έχουν αφαιρεθεί. Ο συντελεστής μερικής αυτοσυσχέτισης τάξης  $k$  (για καθυστέρηση  $k$ ), συμβολίζεται με  $a_k$  και μπορεί να υπολογισθεί εφαρμόζοντας τη μέθοδο της πολλαπλής γραμμικής παλινδρόμησης με εξαρτημένη μεταβλητή την  $Y_t$  και ανεξάρτητες μεταβλητές τις  $Y_{t-1}$  , ... ,  $Y_{t-k}$  :

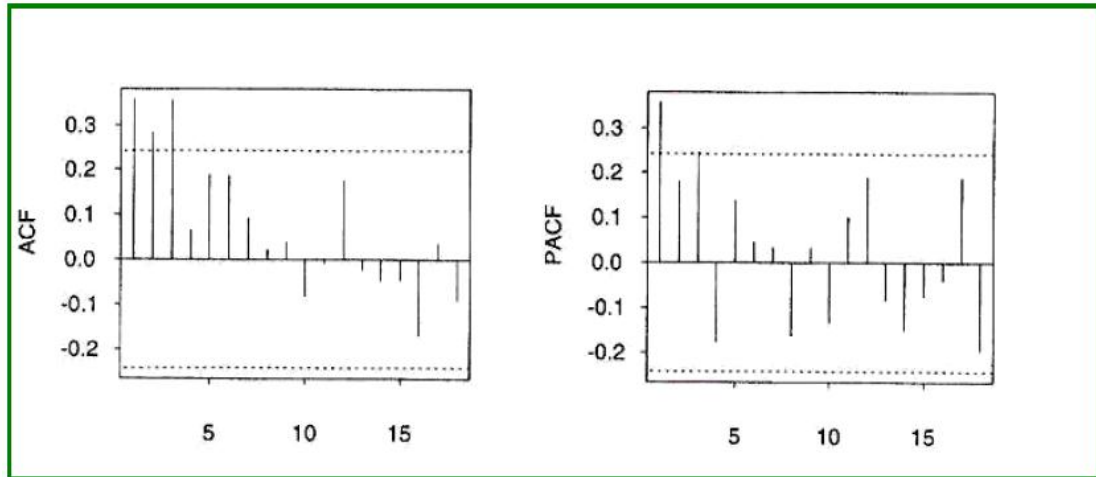
$$Y_t = b_0 + b_1 Y_{t-1} + \dots + b_k Y_{t-k} \quad (2.4)$$

Ο συντελεστής  $a_k$  ισούται με τον συντελεστή  $b_k$ . Πρέπει να σημειωθεί ότι ο πρώτος συντελεστής μερικής αυτοσυσχέτισης  $a_1$  είναι πάντα ίσος με τον πρώτο συντελεστή αυτοσυσχέτισης  $r_1$ .

### Αναγνώριση Εποχικότητας

Η ύπαρξη εποχικότητας σε μια χρονοσειρά, μπορεί να αναγνωρισθεί από τις υψηλές τιμές των συντελεστών αυτοσυσχέτισης ή μερικής αυτοσυσχέτισης για τις αντίστοιχες εποχιακές καθυστερήσεις. Συχνά σημαντικές είναι και οι αυτοσυσχετίσεις για καθυστερήσεις πολλαπλάσιες των εποχιακών καθυστερήσεων. Έτσι, για μηνιαία δεδομένα που παρουσιάζουν εποχικότητα (εποχικότητα 12), υψηλές αυτοσυσχετίσεις μπορούν να παρατηρηθούν για καθυστέρηση 12 αλλά και για καθυστερήσεις 24 και 36.

Στο παρακάτω σχήμα, στην ACF όμως παρατηρούνται υψηλές αυτοσυσχετίσεις για καθυστερήσεις 1,2 και 3 και στην PACF σημαντικές μερικές αυτοσυσχετίσεις για καθυστερήσεις 1 και 3. Αυτό αποτελεί ένδειξη ότι η χρονοσειρά δεν είναι λευκού θορύβου. Επίσης η αυτοσυσχέτιση για καθυστέρηση 12 είναι σχετικά υψηλή αν και δεν είναι στατιστικά σημαντική. Αυτό ίσως υποδηλώνει την ύπαρξη κάποιας εποχικότητας στη χρονοσειρά αν και δεν μπορεί να εξαχθεί κάποιο ασφαλές συμπέρασμα.



Σχήμα 2.6: ACF-PACF

### Έλεγχος της στασιμότητας της χρονοσειράς

Εάν η χρονοσειρά είναι στάσιμη, τότε τα δεδομένα κυμαίνονται γύρω από ένα σταθερό μέσο, ανεξάρτητα του χρόνου, και η διακύμανση παραμένει σταθερή. Συνήθως είναι δυνατόν να ελεγχθεί η στασιμότητα με τη χρήση της γραφικής παράστασης της χρονοσειράς :

- ✓ Εάν δεν παρατηρείται αλλαγή της μέσης τιμής κατά μήκος του χρόνου, τότε η χρονοσειρά είναι στάσιμη ως προς τη μέση τιμή.
- ✓ Εάν δεν παρατηρείται αλλαγή της διακύμανσης κατά μήκος του χρόνου, τότε η χρονοσειρά είναι στάσιμη ως προς τη διακύμανση.

Το διάγραμμα των αυτοσυσχετίσεων μπορεί επίσης να χρησιμοποιηθεί για τον έλεγχο της στασιμότητας ως προς τη μέση τιμή. Οι αυτοσυσχετίσεις στάσιμων χρονοσειρών φθίνουν στο μηδέν με γρήγορο ρυθμό, ενώ για μη στάσιμες χρονοσειρές φθίνουν με αργό ρυθμό καθώς αυξάνει ο αριθμός των καθυστερήσεων.

### Μετατροπή μη στάσιμης χρονοσειράς σε στάσιμη

Τάσεις ή άλλα μη στάσιμα πρότυπα στο επίπεδο της χρονοσειράς, έχουν σαν αποτέλεσμα θετικές αυτοσυσχετίσεις οι οποίες επικρατούν στο διάγραμμα των αυτοσυσχετίσεων. Για το λόγο αυτό, είναι σημαντικό να αφαιρεθούν τα μη στάσιμα πρότυπα ώστε να εμφανισθεί η πραγματική δομή των συσχετίσεων και να αναπτυχθεί το κατάλληλο μοντέλο. Ένας τρόπος να επιτευχθεί αυτό είναι με τη χρήση της μεθόδου της διαφοράς.

Οι σειρές διαφορών πρώτης τάξης προκύπτουν από τις διαφορές των διαδοχικών παρατηρήσεων της αρχικής χρονοσειράς :

$$Y'_t = Y_t - Y_{t-1} \quad (2.5)$$

Μερικές φορές η διαφοράση πρώτης τάξης δεν αρκεί για τη μετατροπή μιας μη στάσιμης χρονοσειράς σε στάσιμη και είναι αναγκαίο να διαφοριστούν τα δεδομένα για δεύτερη φορά :

$$Y''_t = Y'_t - Y'_{t-1} = (Y_t - Y_{t-1}) - (Y_{t-1} - Y_{t-2}) = Y_t - 2Y_{t-1} + Y_{t-2} \quad (2.6)$$

Η σειρά διαφορών δεύτερης τάξης έχει n-2 δεδομένα.

### Μοντέλο τυχαίου περιπάτου (random walk)

Ένα μοντέλο μπορεί να γραφεί σαν :

$$Y_t = Y_{t-1} + e_t \quad (2.7)$$

όπου  $e_t$  είναι λευκός θόρυβος.

Το μοντέλο είναι ευρέως γνωστό σαν μοντέλο τυχαίου περιπάτου και χρησιμοποιείται για μη στάσιμα δεδομένα. Τα μοντέλα τυχαίου περιπάτου παρουσιάζουν μεγάλες περιόδους εμφανών τάσεων, οι οποίες μπορούν να αλλάξουν κατεύθυνση με απρόβλεπτο τρόπο.

### Εποχιακή διαφόριση

Για μη στάσιμα εποχιακά δεδομένα, η στάσιμη χρονοσειρά προκύπτει με τη χρήση των εποχιακών διαφορών. Σαν εποχιακή διαφορά ορίζεται η διαφορά μιας παρατήρησης και της αντίστοιχης παρατήρησης του προηγούμενου έτους. Έτσι για μηνιαία δεδομένα που παρουσιάζουν εποχικότητα, η εποχιακή διαφορά ορίζεται :

$$Y'_t = Y_t - Y_{t-12} \quad (2.8)$$

Γενικά, οι σειρές εποχιακών διαφορών προκύπτουν από τις διαφορές μεταξύ δεδομένων που απέχουν μεταξύ τους  $s$  χρονικές περιόδους, όπου  $s$  η εποχικότητα. Για μηνιαία δεδομένα  $s = 12$ , για τριμηνιαία  $s = 4$  κλπ. Όμοια με τις διαφορές πρώτης τάξης, η διαφóριση μπορεί να επαναληφθεί ώστε να προκύψουν δεύτερης τάξης εποχιακές διαφορές, αν και αυτό σπάνια κρίνεται αναγκαίο.

### Τελεστής Ολίσθησης B

Μια πολύ χρήσιμη σημειογραφική επινόηση είναι ο τελεστής ολίσθησης B ο οποίος ορίζεται ως :

$$BY_t = Y_{t-1} \quad (2.9)$$

Όπως προκύπτει από τον ορισμό, ο χρονικός δείκτης κάθε μεταβλητής η οποία πολλαπλασιάζεται με τον τελεστή, "ολισθαίνει" τόσες περιόδους προς τα πίσω, όσος είναι ο εκθέτης  $k$  του τελεστή. Ισχύει γενικότερα

$$B^k Y_t = Y_{t-k} \quad (2.10)$$

Ας σημειωθεί ότι ο τελεστής B δεν έχει καμιά επίδραση όταν εφαρμόζεται σε σταθερές ποσότητες. Με τη χρήση του τελεστή ολίσθησης είναι εύκολο να περιγραφεί η διαδικασία της διαφóρισης.

Για διαφóριση πρώτης τάξης είναι :

$$Y'_t = Y_t - Y_{t-1} = Y_t - BY_t = (1 - B)Y_t \quad (2.11)$$

Όμοια για διαφóριση δεύτερης τάξης :

$$Y''_t = Y'_t - Y'_{t-1} = (Y_t - Y_{t-1}) - (Y_{t-1} - Y_{t-2}) = Y_t - 2Y_{t-1} + Y_{t-2} = (1 - 2B + B^2)Y_t = (1 - B)^2 Y_t \quad (2.12)$$

Γενικότερα η διαφορά d τάξης μπορεί να περιγραφεί ως  $(1-B)^d Y_t$ . Για εποχιακή διαφορά πρώτης τάξης δεδομένων με μηνιαία εποχικότητα είναι:

$$Y'_t = Y_t - Y_{t-12} = Y_t - B^{12}Y_t = (1 - B^{12})Y_t \quad (2.13)$$

και γενικότερα για εποχιακή διαφορά πρώτης τάξης δεδομένων μη εποχικότητα s είναι:

$$Y'_t = Y_t - Y_{t-s} = Y_t - B^s Y_t = (1 - B^s)Y_t \quad (2.14)$$

Η εποχιακή διαφορά η οποία ακολουθείται από διαφορά πρώτης τάξης γράφεται ως:

$$(1 - B)(1 - B^s)Y_t \quad (2.15)$$

## 2.5 Τεχνητά Νευρωνικά δίκτυα

Η προσομοίωση της ανθρώπινης νοημοσύνης και κατ' επέκταση των ευφυών συστημάτων αποτελεί θεμελιώδη προϋπόθεση της Υπολογιστικής Νοημοσύνης. Ο χώρος αυτός έχει μεγάλη εξέλιξη τις τελευταίες δεκαετίες με αποτέλεσμα να δημοσιευτούν πολλές εργασίες και να εμφανιστούν πολλές πρακτικές εφαρμογές.

Ο άνθρωπος δεν είναι γρήγορος ούτε ακριβής στους υπολογισμούς του, όπως ένας σύγχρονος υπολογιστής. Μάλιστα, ακριβώς το αντίθετο ισχύει και ο υπολογιστής είναι πολύ πιο ικανός να εκτελέσει πολύπλοκες πράξεις. Σε πολλούς τομείς όμως, η ανθρώπινη ικανότητα υπερτερεί κατά πολύ των ικανοτήτων ενός υπολογιστή. Για παράδειγμα, ο άνθρωπος αναγνωρίζει πολύ εύκολα αντικείμενα και κατανοεί τις σχέσεις μεταξύ τους μέσα στο φυσικό περιβάλλον έστω και αν αυτά είναι παραμορφωμένα ή δεν είναι εξ' ολοκλήρου ορατά.

Η ικανότητα μάθησης από εμπειρία είναι από τα κύρια χαρακτηριστικά της ανθρώπινης ευφυΐας. Η ανθρώπινη μνήμη έχει την ικανότητα να αποθηκεύει μεγάλη ποσότητα και ποικιλία γνώσης και να συσχετίζει πληροφορίες πολύ γρήγορα και χωρίς ιδιαίτερη προσπάθεια. Σε αντίθεση, ο υπολογιστής έχει την ικανότητα να αποθηκεύει τεράστιες ποσότητες πληροφοριών αλλά δυσκολεύεται να τις εκμεταλλευτεί. Μια εξήγηση για την ανωτερότητα της ανθρώπινης ικανότητας σε σχέση με αυτήν ενός υπολογιστικού συστήματος πιθανολογείται ότι είναι η έλλειψη

κατάλληλου λογισμικού, γεγονός που εμποδίζει την αναπαραγωγή της ανθρώπινης ικανότητας με μηχανιστικές μεθόδους. Μια δεύτερη εξήγηση είναι το γεγονός ότι ο ανθρώπινος εγκέφαλος και οι υπολογιστές λειτουργούν εντελώς διαφορετικά. Ο άνθρωπος είναι πιο έξυπνος από τον υπολογιστή επειδή ο εγκέφαλος χρησιμοποιεί μια αρχιτεκτονική που είναι πολύ πιο κατάλληλη να αντιμετωπίσει τη φυσική επεξεργασία πληροφοριών. Ο εγκέφαλος είναι ένας ισχυρότατος υπολογιστής, που αποτελείται από έναν τεράστιο αριθμό (10<sup>10</sup>) στοιχείων ή νευρώνων, καθένας από τους οποίους επικοινωνεί με χιλιάδες άλλους. Από κάθε νευρώνα ξεκινούν χιλιάδες ίνες, οι οποίες συναντούν ίνες από άλλους νευρώνες σε σημεία που αποκαλούνται συνάψεις. (Κίνγκ,2004)

Τα Τεχνητά Νευρωνικά Δίκτυα (ΤΝΔ) προέκυψαν από τη διαπίστωση ότι ο τρόπος λειτουργίας του ανθρώπινου εγκεφάλου είναι εντελώς διαφορετικός από τον τρόπο λειτουργίας ενός τυπικού ηλεκτρονικού υπολογιστή. Ο εγκέφαλος είναι ένα αρκετά πολύπλοκο, μη γραμμικό παράλληλο υπολογιστικό σύστημα και έχει την ικανότητα να οργανώνει τα στοιχειώδη δομικά του συστατικά, γνωστά ως νευρώνες (neurons), με τέτοιο τρόπο ώστε να επιτυγχάνεται η εκτέλεση πολύπλοκων υπολογισμών και λειτουργιών (αναγνώριση προτύπων, έλεγχος κίνησης κ.ά.) σε ταχύτητες πολύ πιο γρήγορες από αυτές που μπορούν να επιτευχθούν ακόμα και με τον γρηγορότερο σημερινό υπολογιστή. (Haykin,2009)

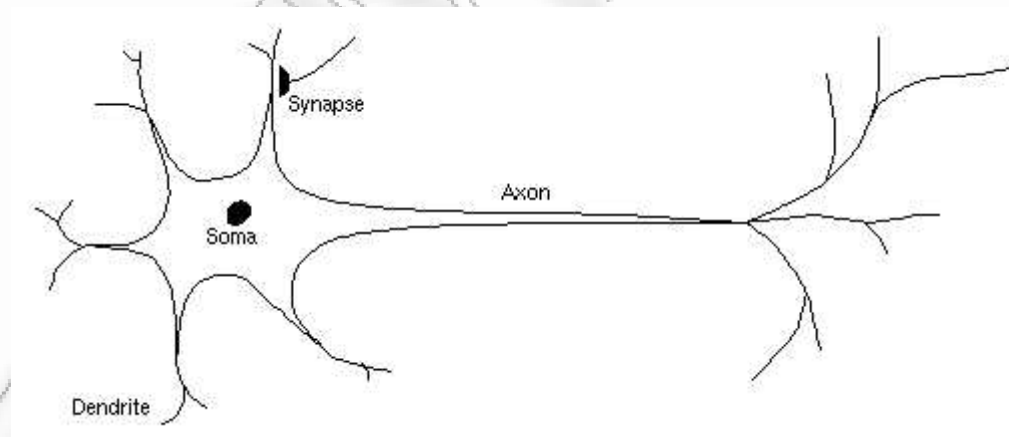
Ο ορισμός των ΤΝΔ απορρέει από το βιολογικό ανάλογο, που είναι το εξής: Ο εγκέφαλος αποτελείται από ένα μεγάλο αριθμό διασυνδεδεμένων νευρώνων, δηλαδή νευρικών κυττάρων. Καθένας νευρώνας δέχεται ερεθίσματα (εισόδους) από άλλους νευρώνες, μέσω συνδέσεων, τα οποία επηρεάζουν την κατάστασή του, και, ανάλογα με αυτή τη κατάσταση, στέλνει με τη σειρά του ερεθίσματα (εξόδους) σε άλλους νευρώνες. Κάθε σύνδεση μεταξύ δύο νευρώνων χαρακτηρίζεται από μία τιμή ισχύος η οποία υποδηλώνει πόσο ισχυρή είναι η μεταξύ τους αλληλεπίδραση. (Λύκας,2008) Με βάση τα παραπάνω, μπορούμε να ορίσουμε ένα Τεχνητό Νευρωνικό Δίκτυο ως εξής:

«Τεχνητό Νευρωνικό Δίκτυο είναι μια αρχιτεκτονική δομή (δίκτυο) αποτελούμενη από ένα πλήθος διασυνδεδεμένων μονάδων επεξεργασίας (τεχνητοί νευρώνες). Κάθε μονάδα επεξεργασίας χαρακτηρίζεται από εισόδους και εξόδους. Υλοποιεί τοπικά ένα υπολογισμό με βάση τις εισόδους που δέχεται και μεταδίδει το αποτέλεσμα (έξοδος)

σε άλλες μονάδες επεξεργασίας με τις οποίες συνδέεται. Οι τιμές των βαρών των συνδέσεων αποτελούν τη γνώση που είναι αποθηκευμένη στο ΤΝΔ και καθορίζουν τη λειτουργικότητά του. Συνήθως ένα ΤΝΔ αναπτύσσει μια συνολική λειτουργικότητα μέσω μιας μορφής εκπαίδευσης (μάθησης)». (Λύκας,2008)

### 2.5.1 Νευρωνική δομή

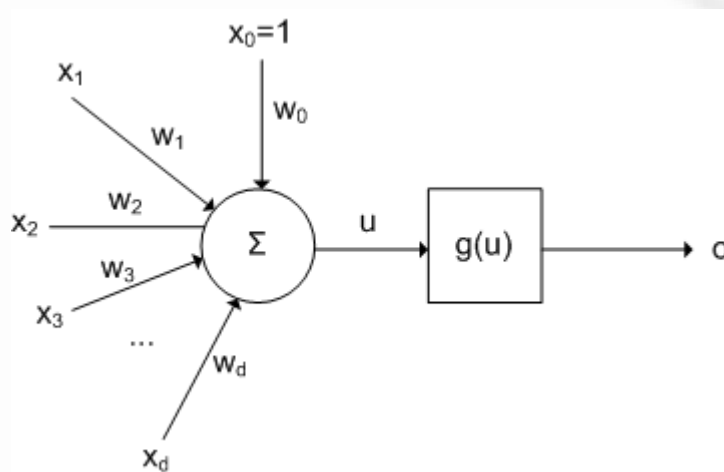
Όπως ήδη αναφέραμε και είναι γνωστό από τη βιολογία, η δομική μονάδα του εγκεφάλου είναι ο νευρώνας. Ένας τυπικός βιολογικός νευρώνας αποτελείται, όπως φαίνεται και από το παρακάτω σχήμα, από το *σώμα*, που περιλαμβάνει τον *πυρήνα* του, τους *δενδρίτες*, μέσω των οποίων λαμβάνει σήματα από γειτονικούς νευρώνες (σημεία εισόδου) και τον *άξονα* που είναι η έξοδος του νευρώνα και το μέσο σύνδεσής του με τους άλλους νευρώνες. Σε κάθε δενδρίτη υπάρχει ένα απειροελάχιστο κενό που ονομάζεται *σύναψη*. Οι συνάψεις μέσω χημικών διαδικασιών επιταχύνουν ή επιβραδύνουν τη ροή ηλεκτρονίων προς το σώμα του νευρώνα. Η ικανότητα μάθησης και μνήμης που παρουσιάζει ο εγκεφαλος οφείλεται στην ικανότητά των συνάψεων να μεταβάλλουν την αγωγιμότητά τους. Τα ηλεκτρικά σήματα που εισέρχονται στο σώμα του νευρώνα μέσω των δενδριτών συνδυάζονται και αν το αποτέλεσμα ξεπερνά κάποια τιμή κατωφλίου το σήμα διαδίδεται με τη βοήθεια του άξονα και στους άλλους νευρώνες.



**Σχήμα 2.7:** Δομή βιολογικού νευρώνα



Τα τεχνητά νευρωνικά δίκτυα είναι εμπνευσμένα από το βιολογικό νευρωνικό σύστημα. Τα στοιχεία που αποτελούν ένα τεχνητό νευρωνικό δίκτυο είναι οι νευρώνες οι οποίοι συνδέονται μεταξύ τους μέσω βαρών. Τα βάρη αυτά μπορούν να προσαρμόζονται κατά τη διαδικασία της μάθησης. Μπορούμε να διαχωρίσουμε τα ΤΝΔ σε δύο βασικές κατηγορίες. Στην *πρώτη* έχουμε τα **στατικά νευρωνικά δίκτυα** τα οποία δεν περιέχουν στοιχεία με μνήμη αλλά μπορούν να έχουν ως εισόδους προηγούμενες τιμές των εισόδων και στη *δεύτερη* έχουμε τα **δυναμικά νευρωνικά δίκτυα** τα οποία διαθέτουν στοιχεία μνήμης και είναι κατάλληλα για την προτυποποίηση μη γραμμικών δυναμικών συστημάτων. Στο παρακάτω σχήμα βλέπουμε τη μορφή ενός τεχνητού νευρώνα:



**Σχήμα 2.8:** Δομή τεχνητού νευρώνα

Ο νευρώνας του σχήματος αποτελείται από  $d$  συνδέσεις εισόδου, κάθε μία από τις οποίες δέχεται ένα σήμα εισόδου  $x_i$  ( $i = 1, 2, 3, \dots, d$ ) και χαρακτηρίζεται από μια τιμή βάρους  $w_i$  ( $i = 1, 2, 3, \dots, d$ ). Η είσοδος  $x_0$  είναι μια σύνδεση σταθερής διέγερσης με τιμή που ισούται μόνιμως με 1 και βάρος  $w_0$  που ονομάζεται πόλωση (bias) του νευρώνα. Η εισαγωγή της πόλωσης προσδίδει στο νευρώνα επιπλέον υπολογιστικές δυνατότητες μάθησης και προσαρμογής στα δεδομένα εκπαίδευσης. (Λύκας,2008)

Ο υπολογισμός που επιτελεί ένας νευρώνας διακρίνεται σε δύο στάδια:

- 1) Υπολογισμός της συνολικής εισόδου

$$u(x) = \sum_{i=1}^d w_i x_i + w_0 \tag{2.16}$$

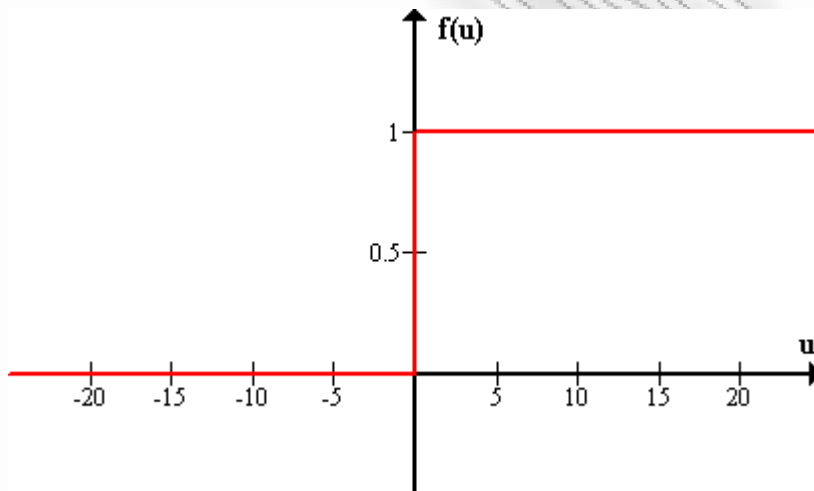
- 2) Υπολογισμός της εξόδου  $o(\mathbf{x})$  του νευρώνα περνώντας τη συνολική είσοδο  $u(\mathbf{x})$  από μια συνάρτηση ενεργοποίησης (activation function)  $g: o = g(u)$ .

### 2.5.2 Τύποι συναρτήσεων ενεργοποίησης

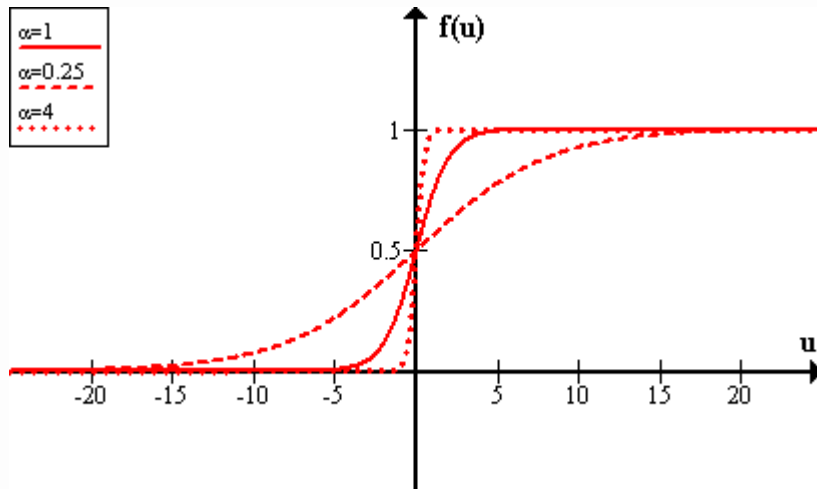
Η συνάρτηση ενεργοποίησης στον βιολογικό νευρώνα θεωρείται πως είναι η βηματική συνάρτηση ή συνάρτηση κατωφλιού (threshold function) η γραφική παράσταση της οποίας φαίνεται στο Σχήμα 2-9

$$f(u) = \begin{cases} 1, & u \geq 0 \\ 0, & u < 0 \end{cases} \quad (2.16)$$

Στην κατασκευή των ΤΝΔ για λόγους που θα φανούν στη συνέχεια, όταν θα αναφερθούμε στην εκπαίδευση των ΤΝΔ, χρησιμοποιούμε συναρτήσεις που το γράφημά τους μοιάζει με τη βηματική, αλλά σε αντίθεση με αυτή είναι συνεχείς και



Σχήμα 2.9: Η βηματική συνάρτηση



Σχήμα 2.10: Η λογιστική συνάρτηση

παραγωγίσιμες. Οι συναρτήσεις αυτής της κατηγορίας ονομάζονται σιγμοειδής (sigmoids), αφού το γράφημά τους μοιάζει με τελικό σίγμα ( $\zeta$ ). Ο πιο γνωστός και ευρέως χρησιμοποιούμενος τύπος σιγμοειδούς συνάρτησης είναι η λογιστική (logistic), η γραφική παράσταση της οποίας φαίνεται στο Σχήμα 2.10 και ορίζεται από τη σχέση

$$f(u) = \frac{1}{1 + e^{-au}} \quad (2.17)$$

### 2.5.3 Αρχιτεκτονικές δικτύων

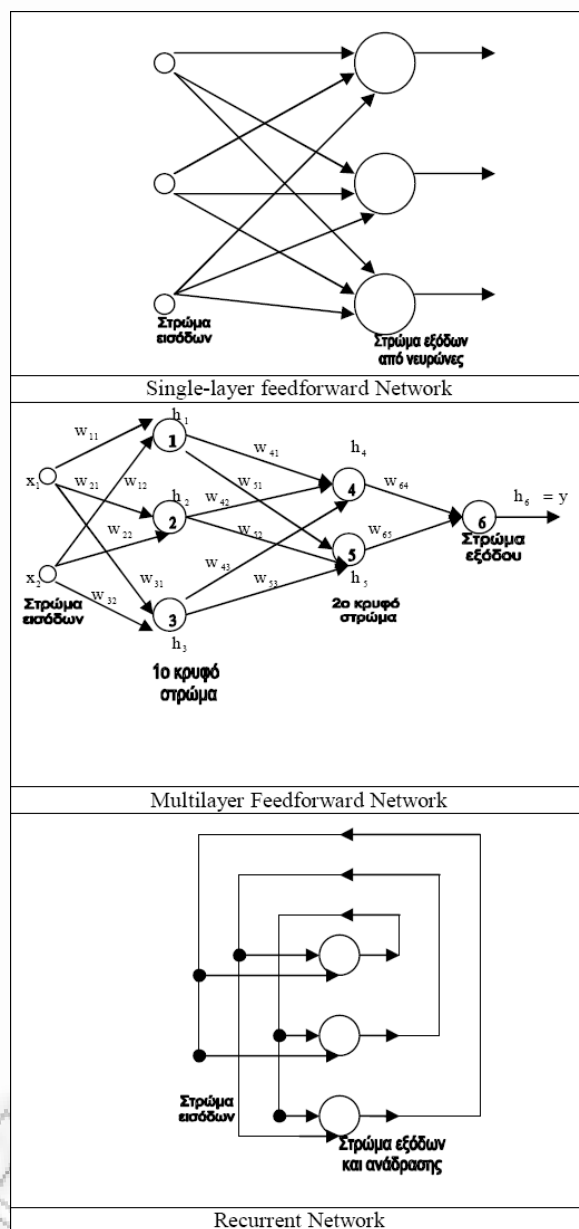
Οι αρχιτεκτονικές δικτύων είναι αλληλένδετες με τους αλγόριθμους εκμάθησης που χρησιμοποιούνται για τα δίκτυα.

Οι αρχιτεκτονικές των νευρωνικών δικτύων μπορούν να διαχωριστούν σε τρεις κύριες κατηγορίες (Εικόνα 2.2):

1. Τα *Μονοστρωματικά Νευρωνικά Δίκτυα Εμπρόσθιας Τροφοδότησης* (Single-layer feedforward Networks). Πρόκειται για την απλούστερη περίπτωση Τεχνητών Νευρωνικών Δικτύων που αποτελούν και την ευρύτερη χρησιμοποιούμενη κατηγορία. Στη περίπτωση αυτή έχουμε ένα στρώμα εισόδου που αποτελείται από τις εισόδους του νευρωνικού και συνδέεται με ένα στρώμα εξόδου αποτελούμενο από νευρώνες (υπολογιστικοί κόμβοι). Η

μετακίνηση δεδομένων γίνεται μόνο προς τα εμπρός και δεν υφίσταται ανάδραση.

2. Τα *Πολυστρωματικά Νευρωνικά Δίκτυα Εμπρόσθιας Τροφοδότησης* (Multilayer Feedforward Networks). Στην περίπτωση αυτή μεταξύ του στρώματος εισόδων και του στρώματος εξόδων παρεμβάλεται ένα ή περισσότερα ενδιάμεσα στρώματα, τα επονομαζόμενα κρυφά στρώματα (hidden layers), τα οποία αποτελούνται από νευρώνες. Με τη χρήση των κρυφών στρωμάτων επιτυγχάνεται μεγαλύτερος βαθμός επεξεργασίας των δεδομένων. Συνήθως κάθε κόμβος ενός στρώματος είναι συνδεδεμένος με καθέναν από τους κόμβους του στρώματος που ακολουθεί. Σε αυτή την περίπτωση το νευρωνικό δίκτυο ονομάζεται πλήρως συνδεδεμένο (fully connected). Υπάρχουν περιπτώσεις δικτύων όπου λείπουν κάποιες συναπτικές ενώσεις, οπότε το δίκτυο ονομάζεται μερικώς συνδεδεμένο (partially connected).
3. *Επαναλαμβανόμενα Νευρωνικά Δίκτυα* (Recurrent Networks). Τα νευρωνικά δίκτυα που ανήκουν στη συγκεκριμένη κατηγορία διαφέρουν από τα νευρωνικά δίκτυα εμπρόσθιας μετάδοσης από το γεγονός ότι έχουν τουλάχιστον μια ανάδραση.



Εικόνα 2.2: Αρχιτεκτονικές νευρωνικών δικτύων

Στην ενότητα 3.7.1 γίνεται εκτενής ανάλυση και εφαρμογή ενός μονοστρωματικού νευρωνικού δικτύου.

## ΚΕΦΑΛΑΙΟ 3

### Μεθοδολογία

#### 3.1 Εισαγωγή

Στο κεφάλαιο αυτό παρουσιάζεται αναλυτικά η μεθοδολογία που ακολουθήθηκε για την υλοποίηση της παρούσας εργασίας.

Πιο συγκεκριμένα στην ενότητα 3.2, γίνεται μια πρώτη αναφορά στα δεδομένα που χρησιμοποιήθηκαν για την εφαρμογή των μοντέλων πρόβλεψης χρονοσειρών, στην ενότητα 3.3 παρουσιάζονται και αναλύονται οι μέθοδοι εξομάλυνσης χρονοσειρών, όπως η μέθοδος του κινητού μέσου, η απλή εκθετική εξομάλυνση, ο διπλός κινητός μέσος, η διπλή εκθετική εξομάλυνση, η εκθετική εξομάλυνση με προσαρμογή στην τάση και τέλος η εκθετική εξομάλυνση με προσαρμογή στην τάση και στην εποχικότητα.

Στην ενότητα 3.4 παρουσιάζονται οι τεχνικές διάσπασης χρονοσειρών και στην ενότητα 3.5 παρουσιάζεται εκτενώς η ανάλυση ARIMA η οποία χρησιμοποιείται κατά κόρον στην εργασία αυτή.

Τέλος η ενότητα 3.6 πραγματεύεται τα τεχνητά νευρωνικά δίκτυα και τις αρχιτεκτονικές τους και παρουσιάζεται ο τρόπος εκπαίδευσης ενός μονοστρωματικού δικτύου εμπρόσθιας τροφοδότησης.

#### 3.2 Συλλογή στοιχείων

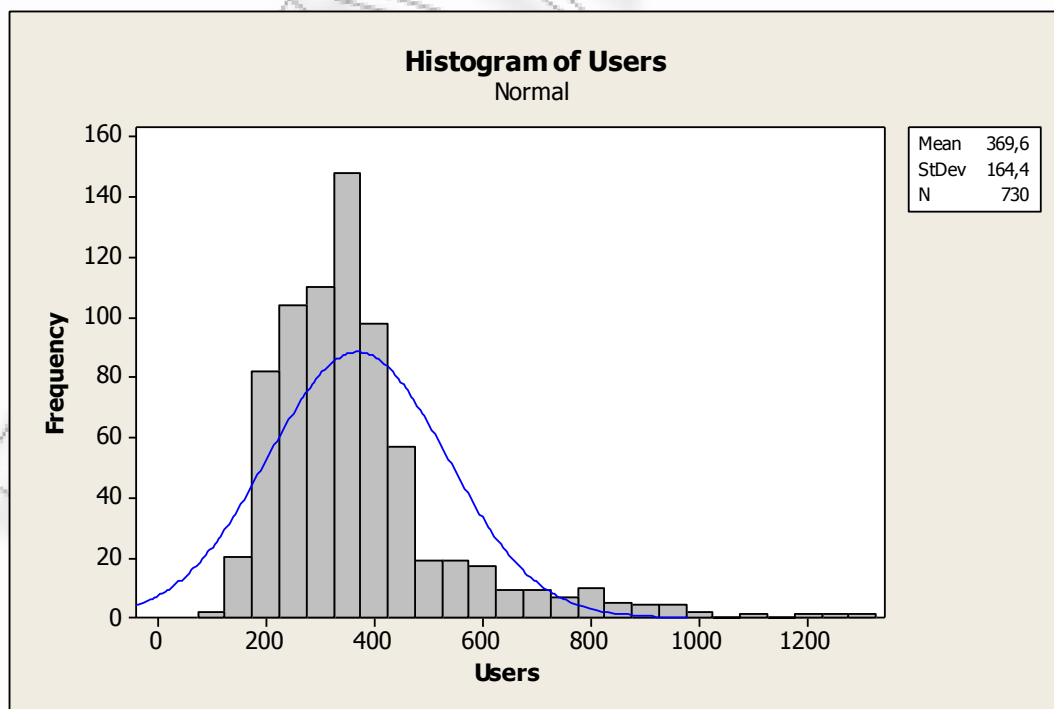
Σύμφωνα με το σκοπό της εργασίας θέλουμε να προσεγγίσουμε το πρόβλημα της πρόβλεψης των επισκεπτών σε ένα εκπαιδευτικό portal, έτσι έχουμε στη διάθεσή μας τα ημερήσια δεδομένα που συλλέχθηκαν από τον ιστότοπο ενός εκπαιδευτικού portal (PDE - Η κοινότητα των εκπαιδευτικών). Τα δεδομένα που χρησιμοποιούνται για την ανάλυση αυτή περιέχουν τον μέγιστο αριθμό συνδεδεμένων χρηστών για τα έτη 2010 και 2011 που όπως θα δούμε παρακάτω παρατηρείται κάποιου βαθμού εποχικότητα.

Για την αξιολόγηση της απόδοσης των μοντέλων πρόβλεψης, η κάθε χρονοσειρά χωρίζεται σε δύο τμήματα:

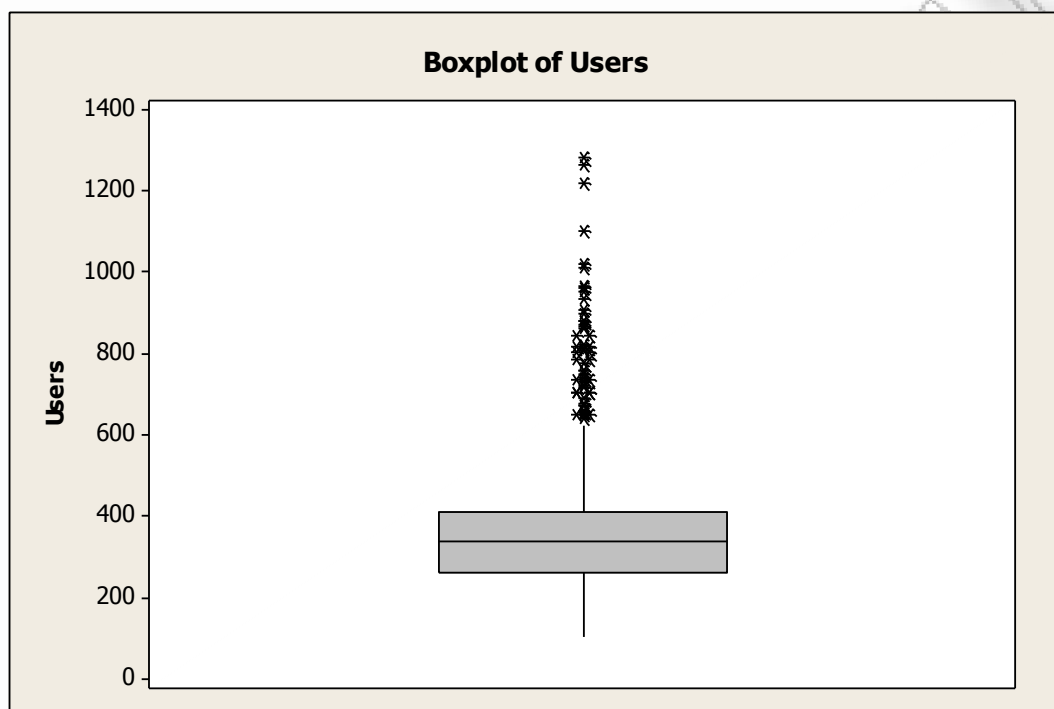
- Το πρώτο σύνολο σημείων για την εκμάθηση του μοντέλου (training set). Για τα διάφορα μοντέλα αυτό το σύνολο σημείων δίνει τις απαραίτητες παραμέτρους που θα χρησιμοποιηθούν στην εφαρμογή.
- Το δεύτερο σύνολο σημείων χρησιμοποιείται στην εκτίμηση του λάθους πρόβλεψης (test set). Αυτά είναι τα σημεία αναφοράς για τα οποία γίνεται πρόβλεψη και συγκρίνεται με τις πραγματικές παρατηρήσεις.

Στην στατιστική ανάλυση των χρονοσειρών συνήθως χρησιμοποιούμε το 80% των δεδομένων για την προσαρμογή ή την εκμάθηση του μοντέλου (training set) και το 20% για την αξιολόγηση της ικανότητας πρόβλεψης (test set). Στην εργασία αυτή θα χρησιμοποιήσουμε 694 ημερήσιες τιμές για το training set και 36 ημερήσιες τιμές για το test set. Το δείγμα που παίρνουμε για το test set είναι κατά πολύ μικρότερο απ' τις βιβλιογραφικές αναφορές, κάτι το οποίο έχει γίνει αυθαίρετα, άλλωστε σε βάθος χρόνου η απόκλιση των τιμών πρόβλεψης από τις πραγματικές αυξάνετε.

Στη συνέχεια παραθέτουμε τα διαγράμματα της χρονοσειράς μας έτσι όπως προκύπτουν από τα δεδομένα μας. Επίσης, βλέπουμε και μερικούς βασικούς στατιστικούς δείκτες.



**Σχήμα 3.1:** Ιστόγραμμα χρονοσειράς



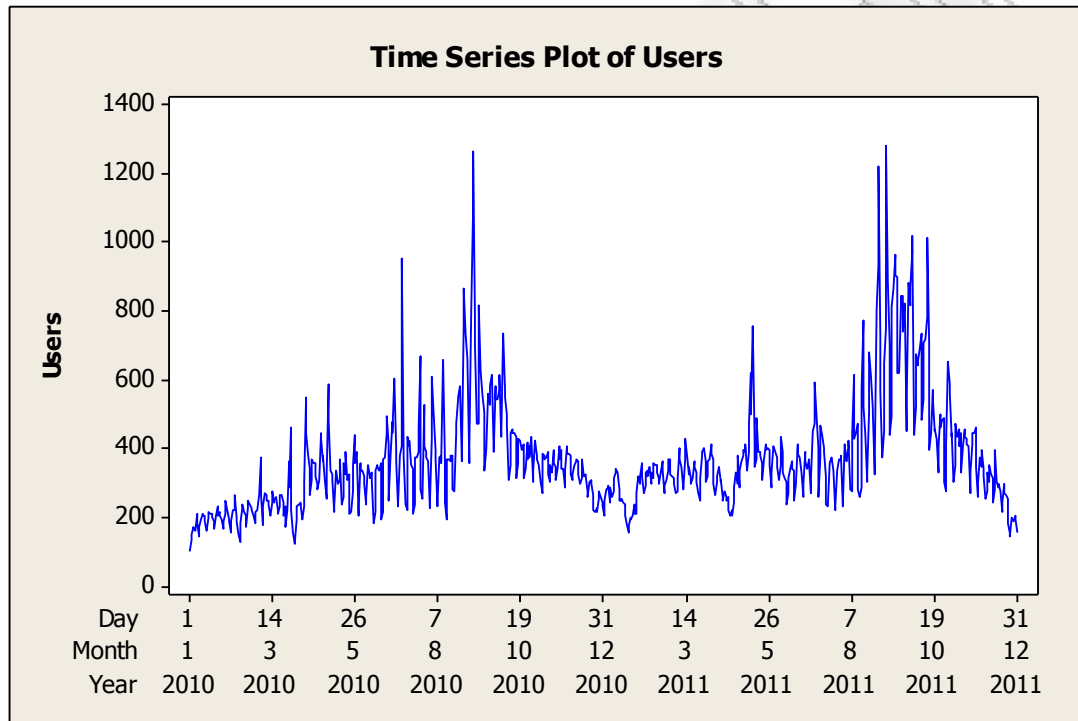
**Σχήμα 3.2:** Boxplot χρονοσειράς

**Πίνακας 3.1:** Συγκεντρωτικός πίνακας στατιστικής ανάλυσης χρονοσειράς

Μεταβλητή (Variable)	Χρήστες (Users)
Variance	27027,37
CoefVar	44,48
Minimum	101,00
Q1	262,00
Median	339,50
Q3	410,00
Maximum	1284,00
Skewness	1,95
Kurtosis	5,43



Παρατηρώντας την χρονοσειρά μας στο παρακάτω διάγραμμα, είναι εμφανές ότι υπάρχει μία αυξητική τάση στους επισκέπτες του portal με ιδιαίτερες διακυμάνσεις κυρίως το 3<sup>ο</sup> τρίμηνο του έτους. Η διακύμανση αυτή και η έκρηξη στους επισκέπτες είναι συνδεδεμένες με την αρχή του ακαδημαϊκού έτους. Η εποχικότητα του φαινομένου και η αύξηση των χρηστών είναι δύο θέματα που θα μας απασχολήσουν ιδιαίτερα στην ανάλυση των μοντέλων που ακολουθεί.



Σχήμα 3.3: Χρονοσειρά ημερησίων τιμών για 730 ημέρες

### 3.3 Μέθοδοι εξομάλυνσης

#### 3.3.1 Η μέθοδος του κινητού μέσου (Moving average)

Η πρόβλεψη χρονοσειρών με τη μέθοδο των κινητών μέσων εμπεριέχει τον υπολογισμό του μέσου όρου του δείγματος παρατηρήσεων, καθώς και τη χρησιμοποίηση αυτού του μέσου σαν πρόβλεψη για την επόμενη περίοδο. Ο αριθμός των παρατηρήσεων του δείγματος που συμπεριλαμβάνονται στον υπολογισμό του

μέσου προσδιορίζεται στην αρχή της διαδικασίας πρόβλεψης. Χρησιμοποιούμε τον όρο κινητός μέσος γιατί, καθώς κάθε νέα παρατήρηση γίνεται διαθέσιμη στο δείγμα, υπολογίζεται ένας νέος μέσος απορρίπτοντας την παλαιότερη παρατήρηση του δείγματος από τον μέσο και συμπεριλαμβάνοντας τη νεότερη. Καθώς υπολογίζεται κάθε νέος μέσος, γίνεται πρόβλεψη για την επόμενη περίοδο. Συνεπώς, κάθε πρόβλεψη χρησιμοποιεί τον ίδιο αριθμό παρατηρήσεων δείγματος από τη χρονοσειρά και περιλαμβάνει μόνο την πλέον πρόσφατη παρατήρηση.

Η μέθοδος του κινητού μέσου όρου, σε αλγεβρική μορφή, διατυπώνεται ως εξής:

$$F_{t+1} = \frac{X_t + X_{t+1} + \dots + X_{t-n+1}}{n} = \frac{1}{n} \left( \sum_{i=t-n+1}^t X_i \right) \quad (3.1)$$

όπου  $t$  είναι η πιο πρόσφατη παρατήρηση και  $t+1$  είναι η επόμενη περίοδος. Ο τύπος αυτός απαιτεί ο προβλέπων να διαθέτει τις τιμές των παρατηρήσεων του παρελθόντος. Με την προσθήκη μιας νέας παρατήρησης και την εξάλειψη της παλαιότερης, μπορούμε να επαναδιατυπώσουμε τον τύπο ως εξής:

$$F_{t+1} = \frac{1}{n} \left( \sum_{i=t-n}^{t-1} X_i \right) + \frac{1}{n} (X_t - X_{t-n}) = F_t + \frac{X_t}{n} - \frac{X_{t-n}}{n} \quad (3.2)$$

Ο τύπος του κινητού μέσου όρου εκφράζει απλώς μια προσαρμογή στην πρόβλεψη ( $F_t$ ) που έγινε την προηγούμενη περίοδο. Είναι φανερό πως αν αυξηθεί το  $n$ , το εξομαλυντικό αποτέλεσμα είναι μεγαλύτερο γιατί γίνεται μια πολύ μικρότερη προσαρμογή για κάθε νέα χρονική περίοδο. Εάν τα δεδομένα είναι πολύ άστατα, το γεγονός αυτό αποτελεί ένα πολύ ελκυστικό χαρακτηριστικό.

Η μέθοδος του κινητού μέσου όρου είναι μια πολύ εύκολη διαδικασία πρόβλεψης, αφού οι απαιτήσεις σε στατιστικά στοιχεία είναι μικρές. Εν τούτοις, η τεχνική αυτή δε χρησιμοποιείται ευρέως, δεδομένου ότι η προϋπόθεση να είναι τα δεδομένα *στάσιμα* είναι περιοριστική. Πάντως, η μέθοδος του κινητού μέσου όρου είναι πιο αποτελεσματική από τη χρησιμοποίηση του απλού αριθμητικού μέσου μιας ολόκληρης χρονοσειράς. Είναι πιο εύκαμπτη γιατί το  $n$  μπορεί να μεταβληθεί ώστε να ανταποκρίνεται στα παρατηρούμενα πρότυπα των δεδομένων. (Σούκος & Κάντζιος, 2006)

### 3.3.2 Απλή εκθετική εξομάλυνση (Simple exponential smoothing)

Υπάρχουν σοβαροί περιορισμοί στη χρήση του κινητού μέσου όρου. *Πρώτον*, για τους σχετικούς υπολογισμούς απαιτούνται οι παρελθούσες  $n$  παρατηρήσεις του δείγματος. Αν πρόκειται να προβλεφθεί ένας μεγάλος αριθμός μεγεθών (μεταβλητών), τα δεδομένα απαιτούν μεγάλο αποθηκευτικό χώρο. *Δεύτερον*, δίνεται ίση στάθμιση σε όλες τις παρατηρήσεις που αφορούν το παρελθόν, χωρίς να δίνεται καμία στάθμιση στις παρατηρήσεις που είναι πριν από την περίοδο  $(t-n+1)$ . Οι πρόσφατες παρατηρήσεις, εν τούτοις, μπορεί να περιέχουν περισσότερες πληροφορίες από τις παλαιότερες, πράγμα που έχει σημασία για τις μελλοντικές προβλέψεις. Έτσι, κάποιος μπορεί να θέλει να αποδώσει μεγαλύτερη στάθμιση στην πρόβλεψη η οποία βασίζεται στις πιο πρόσφατες παρατηρήσεις παρά στις παλαιότερες. Μια μέθοδος που απλοποιεί τους υπολογισμούς της πρόβλεψης και έχει μικρές απαιτήσεις σε δεδομένα, ονομάζεται *εκθετική εξομάλυνση*. Επίσης, η μέθοδος αυτή δίνει αυτοδιορθούμενες προβλέψεις αφού διαθέτει μηχανισμό ενσωματωμένων προσαρμογών, ο οποίος ρυθμίζει τις τιμές αλλάζοντάς αυτές, προς την αντίθετη κατεύθυνση από εκείνη που κινήθηκαν τα λάθη προηγούμενων περιόδων. Για να αναπτύξουμε αυτή τη μέθοδο αρχίζουμε με την εξίσωση (3.2) και την ξαναγράφουμε ως εξής:

$$F_{t+1} = \frac{X_t}{n} - \frac{F_t}{n} + F_t \quad (3.3)$$

εάν κάνουμε την αντικατάσταση  $F_t = X_{t-n}^2$ . Παραπέρα, η εξίσωση αυτή μπορεί να γραφεί ως:

$$F_{t+1} = \frac{1}{n} X_t + \left(1 - \frac{1}{n}\right) F_t \quad (3.4)$$

Η εξίσωση (3.4) δίνει μια πρόβλεψη βασισμένη στη στάθμιση της πιο πρόσφατης παρατήρησης με συντελεστή  $1/n$  και της πιο πρόσφατης πρόβλεψης με συντελεστή στάθμισης  $[1-(1/n)]$ . Από τη στιγμή που ο αριθμός των περιόδων,  $n$ , είναι σταθερός, το κλάσμα  $1/n$  πρέπει να είναι μεγαλύτερο από το μηδέν και μικρότερο από τη μονάδα. Αν αντικαταστήσουμε το  $1/n$  με  $a$ , το βασικό μοντέλο γράφεται ως εξής:

$$F_{t+1} = aX_t + (1-a)F_t \quad (3.5)$$

όπου  $t$  είναι η τρέχουσα περίοδος, τα  $F_{t+1}$  και  $F_t$ , είναι τιμές πρόβλεψης για την επόμενη και την τρέχουσα περίοδο και  $X_t$  είναι η τιμή που παρατηρήθηκε την τρέχουσα περίοδο. Το  $a$  ονομάζεται σταθερά εξομάλυνσης και παίρνει τιμές από 0 έως 1. Αφού η παραπάνω εξίσωση περιλαμβάνει μόνο μία σταθερά, το μοντέλο αυτό είναι μοντέλο εκθετικής εξομάλυνσης μιας παραμέτρου.

Η εκθετική εξομάλυνση μιας παραμέτρου είναι πολύ απλή μέθοδος, αφού μόνο μια τιμή, η πρόβλεψη της τελευταίας περιόδου, είναι αυτή που πρέπει να διασωθεί. Στην ουσία, ολόκληρη η χρονοσειρά εμπεριέχεται σ' αυτή την πρόβλεψη. Εάν εκφράσουμε το  $F_t$  σε όρους της προηγούμενης παρατήρησης  $X_{t-1}$  και των τιμών της πρόβλεψης  $F_{t-1}$ , τότε το ισοδύναμο για την πρόβλεψη της επόμενης περιόδου γίνεται:

$$F_{t+1} = aX_t + (1-a)[aX_{t-1} + (1-a)F_{t-1}] \quad (3.6)$$

ή πιο απλά

$$F_{t+1} = aX_t + a(1-a)X_{t-1} + (1-a)^2 F_{t-1} \quad (3.7)$$

Η νέα αυτή εξίσωση είναι μοντέλο δευτεροβάθμιας εκθετικής εξομάλυνσης μιας παραμέτρου. Μπορούμε να συνεχίσουμε έτσι για έναν αριθμό προηγούμενων περιόδων, πράγμα που δείχνει ότι όλες οι προηγούμενες τιμές του  $X$  αντανακλώνται στην τρέχουσα πρόβλεψη. Έτσι, το όνομα αυτής της διαδικασίας προέρχεται από τις διαδοχικές σταθμίσεις  $a, a(1-a), a(1-a)^2, a(1-a)^3, \dots$ , οι οποίες μειώνονται εκθετικά. Οι πιο πρόσφατες περίοδοι στη χρονοσειρά λαμβάνουν μεγαλύτερη στάθμιση στον υπολογισμό της πρόβλεψης. Πρακτικά, οι αρκετά παλιές τιμές της  $X$  εξαιρούνται. Η διαδικασία πρόβλεψης μπορεί να τροποποιηθεί οποιαδήποτε στιγμή με τη μεταβολή της τιμής της  $a$ . Μπορούμε να ξαναγράψουμε την εξίσωση (3.5) ως εξής:

$$F_{t+1} = aF_t + a(X_t - F_t) \quad (3.8)$$

ή πιο απλά

$$F_{t+1} = F_t + ae_t \quad (3.9)$$

όπου  $ae_t$ , το σφάλμα πρόβλεψης για την περίοδο  $t$ , είναι η πραγματική τιμή μείον την τιμή της πρόβλεψης. Επομένως, βλέπουμε ότι η πρόβλεψη που δίνεται από την εκθετική εξομάλυνση είναι η παλαιά πρόβλεψη συν μια προσαρμογή για το σφάλμα που έγινε στην τελευταία πρόβλεψη.

Όταν το  $\alpha$  βρίσκεται πλησίον του 1, η νέα πρόβλεψη περιέχει μια ουσιώδη προσαρμογή για το σφάλμα της προηγούμενης πρόβλεψης. Αντίθετα, εάν το  $\alpha$  βρίσκεται πολύ κοντά στο 0, η νέα πρόβλεψη θα περιέχει μικρή μόνο προσαρμογή για το σφάλμα. Επομένως, το αποτέλεσμα του μεγέθους του  $\alpha$  είναι όμοιο με τα αποτελέσματα των διαφόρων τιμών για τον αριθμό των παρατηρήσεων του δείγματος όταν υπολογίζουμε τον κινητό μέσο όρο. Τέλος, η εκθετική εξομάλυνση μιας παραμέτρου δίνει προβλέψεις που ακολουθούν το πρότυπο στα δεδομένα ενός δείγματος. Αυτό συμβαίνει γιατί η διαδικασία προσαρμόζει μόνο την επόμενη πρόβλεψη ως προς κάποιο ποσοστό του πιο πρόσφατου σφάλματος πρόβλεψης, και δεν μπορεί να προβλέψει μεταβολές στην κατεύθυνση της χρονοσειράς.

Για να λύσουμε το πρόβλημα της επιλογής εφαρμόζουμε την ανάλυση ευαισθησίας στις ιστορικές χρονοσειρές, χρησιμοποιώντας διαφορετικές τιμές για τη σταθερά της εξομάλυνσης. Για κάθε τιμή πάνω από ένα εύρος τιμών, προετοιμάζεται ένα πρότυπο πρόβλεψης με τη μέθοδο της εκθετικής εξομάλυνσης και υπολογίζεται το κατάλληλο μέτρο της ακριβείας της πρόβλεψης. Στην πράξη, μελέτες που έγιναν, δείχνουν ότι οι τιμές από 0.05 έως 0.30 ταιριάζουν πολύ καλά στα πρότυπα της εκθετικής εξομάλυνσης. Τιμές της  $\alpha$  που είναι μεγαλύτερες από 0.30 συνήθως δείχνουν ότι κάποιο εναλλακτικό μοντέλο πρόβλεψης θα είναι περισσότερο κατάλληλο. (Σούκος & Κάντζιος, 2006)

### 3.3.3 Διπλός κινητός μέσος (*double moving average*)

Η μέθοδος αυτή μπορεί να χρησιμοποιηθεί για την πρόβλεψη των τιμών μιας χρονοσειράς, οι παρατηρήσεις της οποίας παρουσιάζουν ανοδική ή πτωτική πορεία που εκφράζεται από κάποια γραμμική τάση. Για τη διαμόρφωση των προβλέψεων με τη μέθοδο αυτή υπολογίζεται ένας δεύτερος κινητός από τον απλό κινητό μέσο, ενώ στη συνέχεια λαμβάνεται υπ' όψιν και η γραμμική τάση των παρατηρήσεων της χρονοσειράς. Για το λόγο αυτό η μέθοδος ονομάζεται πολύ συχνά και μέθοδος γραμμικού κινητού μέσου (*linear moving average*). Η εφαρμογή της μεθόδου του διπλού κινητού μέσου βασίζεται στην ακόλουθη διαδικασία (Αγιακόγλου & Οικονόμου, 2004):

- Υπολογίζεται ο απλός κινητός μέσος  $m$ -περιόδων,  $M_t$ , ως:

$$M_{t+1} = \frac{1}{m} \sum_{j=1}^m X_{t-j+1} \quad (3.10)$$

➤ Υπολογίζεται ο διπλός κινητός μέσος m-περιόδων,  $M'_t$ , ως:

$$M'_{t+1} = \frac{1}{m} \sum_{j=1}^m X_{t-j+1} \quad (3.11)$$

➤ Υπολογίζεται η διαφορά  $a_t$  ως:

$$a_t = 2M_t - M'_t \quad (3.12)$$

➤ Υπολογίζεται ο παράγοντας προσαρμογής για τη τάση,  $b_t$ , ως:

$$b_t = \frac{2}{m-1} (M_t - M'_t) \quad (3.13)$$

➤ Υπολογίζεται η πρόβλεψη  $F_{t+h}$ , για την h μελλοντική περίοδο ως:

$$F_{t+h} = a_t + hb_t \quad (3.14)$$

Η μέθοδος αυτή, σε αντίθεση με τις δύο προηγούμενες μεθόδους προβλέψεων που παρουσιάσαμε, μπορεί για  $h > 1$  να χρησιμοποιηθεί για τη διενέργεια προβλέψεων για περισσότερες από μια μελλοντικές περιόδους, ενώ για  $h=1$  δίνει την πρόβλεψη για την επόμενη περίοδο. Βέβαια, η χρήση της προϋποθέτει την ύπαρξη μεγάλου αριθμού παρατηρήσεων, ιδιαίτερα μάλιστα όταν η τιμή του m είναι σχετικά μεγάλη. Όπως και στη μέθοδο του απλού κινητού μέσου, όταν η τιμή του m δεν είναι γνωστή, επιλέγουμε εκείνη την τιμή που ελαχιστοποιεί την τιμή του κριτηρίου MSE ή κάποιου άλλου κριτηρίου στα δεδομένα της χρονοσειράς.

### 3.3.4 Διπλή εκθετική εξομάλυνση (Double exponential smoothing)

Ονομάζεται επίσης μέθοδος Brown. Χρησιμοποιείται κυρίως σε σειρές που παρουσιάζουν τάση. Η βασική φιλοσοφία της μεθόδου είναι σχεδόν ίδια με αυτήν του διπλού κινητού μέσου, δηλαδή η εξομάλυνση των παρατηρήσεων της χρονοσειράς γίνεται δύο φορές, ενώ στη διαμόρφωση των προβλέψεων λαμβάνεται υπόψη και η τάση. Η εφαρμογή της μεθόδου στηρίζεται στην ακόλουθη διαδικασία (Αγιακόγλου & Οικονόμου, 2004):

- Εξομαλύνονται οι αρχικές παρατηρήσεις της χρονοσειράς με τη μέθοδο της απλής εκθετικής εξομάλυνσης ως ακολούθως:

$$A_t = aX_t + (1-a)A_{t-1} \quad (3.15)$$

όπου  $a$  είναι η σταθερά εξομάλυνσης, για  $0 \leq a \leq 1$ ,  $A_t$  οι εξομαλυνθείσες τιμές της χρονοσειράς που προκύπτουν από την εξομάλυνση, για  $t=2, 3, \dots, n$  ενώ για  $t=1$  ορίζεται ως αρχική συνθήκη  $A_1 = X_1$ .

- Εξομαλύνονται οι εξομαλυνθείσες τιμές  $A_t$  της χρονοσειράς με τη μέθοδο της απλής εκθετικής εξομάλυνσης ως ακολούθως:

$$A'_t = aA_t + (1-a)A'_{t-1} \quad (3.16)$$

όπου  $A'_t$  οι εξομαλυνθείσες τιμές της χρονοσειράς που προκύπτουν από την δεύτερη εξομάλυνση, για  $t=2, 3, \dots, n$  ενώ για  $t=1$ ,  $A'_1 = A_1$ .

- Υπολογίζεται η διαφορά  $a_t$  ως:

$$a_t = aA_t - A'_t \quad (3.17)$$

- Υπολογίζεται ο παράγοντας προσαρμογής για τη τάση,  $b_t$ , ως:

$$b_t = \frac{a}{1-a} (A_t - A'_t) \quad (3.18)$$

- Υπολογίζεται η πρόβλεψη  $F_{t+h}$ , για την  $h$  μελλοντική περίοδο ως:

$$F_{t+h} = a_t + hb_t \quad (3.19)$$

Σε αντίθεση με την απλή εκθετική εξομάλυνση, μπορούμε να εφαρμόσουμε αυτή τη μέθοδο για περισσότερες από μια μελλοντικές περιόδους. Επίσης, αν η τιμή της σταθεράς εξομάλυνσης  $a$  δεν είναι γνωστή, κάτι που συμβαίνει όταν εφαρμόζουμε τη μέθοδο για πρώτη φορά στα δεδομένα μιας χρονοσειράς, επιλέγουμε κατά τα γνωστά εκείνη την τιμή του  $a$  που ελαχιστοποιεί την τιμή του κριτηρίου MSE ή κάποιου άλλου κριτηρίου στα δεδομένα της χρονοσειράς. Σημειώνεται επίσης ότι ο αριθμός των παρατηρήσεων που απαιτούνται για την εφαρμογή της μεθόδου αυτής είναι αρκετά μικρότερος από τον αντίστοιχο αριθμό της μεθόδου του διπλού κινητού μέσου.

### 3.3.5 Εκθετική εξομάλυνση με προσαρμογή στην τάση (*Exponential smoothing adjusted for trend*)

Ονομάζεται επίσης μέθοδος Holt. Χρησιμοποιείται επίσης όταν υπάρχει τάση στις παρατηρήσεις της χρονοσειράς. Η μέθοδος Holt έχει δύο παραμέτρους εξομάλυνσης, την παράμετρο  $\alpha$  για την εξομάλυνση των τιμών της χρονοσειράς και την παράμετρο  $\beta$  για την εξομάλυνση της τάσης, σε αντίθεση με τη μέθοδο της διπλής εκθετικής εξομάλυνσης που έχει μόνο μια. Η εφαρμογή της μεθόδου βασίζεται στην ακόλουθη διαδικασία (Αγιακόγλου & Οικονόμου, 2004):

- Η εξομάλυνση των τιμών της χρονοσειράς γίνεται με την ακόλουθη σχέση:

$$A_t = \alpha X_t + (1 - \alpha)(A_{t-1} + T_{t-1}) \quad (3.20)$$

όπου  $\alpha$  είναι η σταθερά για την εξομάλυνση για  $0 \leq \alpha \leq 1$ ,  $A_t$  οι εξομαλυνθείσες τιμές της χρονοσειράς που προκύπτουν από την εξομάλυνση, για  $t=2, 3, \dots, n$  ενώ για  $t=1$  ορίζεται ως αρχική συνθήκη  $A_1=X_1$ .

- Η εξομάλυνση της τάσης γίνεται ως εξής:

$$T_t = \beta(A_t - A_{t-1}) + (1 - \beta)T_{t-1} \quad (3.21)$$

όπου  $\beta$ , για  $0 \leq \beta \leq 1$ , είναι η σταθερά για την εξομάλυνση της τάσης,  $T_t$  οι εξομαλυνθείσες τιμές της τάσης, για  $t=2, 3, \dots, n$  ενώ για  $t=1$  ορίζεται ως αρχική συνθήκη η  $T_1=0$ .

- Υπολογίζεται η πρόβλεψη  $F_{t+h}$ , για την  $h$  μελλοντική περίοδο ως:

$$F_{t+h} = A_t + hT_t \quad (3.22)$$

όπου  $h=1, 2, 3, \dots$ .

Οι τιμές των παραμέτρων  $\alpha$  και  $\beta$  για μια συγκεκριμένη χρονοσειρά είναι αυτές που ελαχιστοποιούν την τιμή του κριτηρίου MSE ή κάποιου άλλου κριτηρίου στα δεδομένα της χρονοσειράς. Αξίζει να σημειωθεί ότι με τη μέθοδο Holt οι τιμές της τάσης εξομαλύνονται απευθείας, ενώ με τη μέθοδο Brown η τάση προσδιορίζεται από τον παράγοντα  $bt$ , αφού προηγουμένως εξομαλυνθούν δύο φορές οι τιμές της χρονοσειράς. Αυτό σημαίνει ότι με τη μέθοδο Holt γίνεται καλύτερη εκτίμηση των τιμών της τάσης, απ' ό,τι με τη μέθοδο Brown που είναι πολύ ευαίσθητη στις τυχαίες



διακυμάνσεις της χρονοσειράς. Για αυτό το λόγο, στην πράξη, η μέθοδος Holt εφαρμόζεται πιο συχνά.

### 3.3.6 Εκθετική εξομάλυνση με προσαρμογή στην τάση και στην εποχικότητα (Exponential smoothing adjusted for trend and seasonality)

Ο Winters ανέπτυξε μια μέθοδο για την προσαρμογή της εποχικής ή περιοδικής κίνησης μέσα στο πλαίσιο της γραμμικής εκθετικής εξομάλυνσης. Επομένως, η διαδικασία του Winters μπορεί να εφαρμοστεί για προβλέψεις με βάση μια χρονοσειρά που εμφανίζει και τάση και εποχικό πρότυπο. Η επέκταση συνίσταται στην ύπαρξη επιπλέον εξίσωσης για τον υπολογισμό της εποχικής συνιστώσας της χρονοσειράς. Προκειμένου να κάνουμε πρόβλεψη με αυτό το μοντέλο χρειαζόμαστε τέσσερις εξισώσεις (Αγιακόγλου & Οικονόμου, 2004):

- Κάνουμε επίκαιρη την εκθετικά εξομαλυνθείσα σειρά

$$S_t = a \frac{X_t}{I_t - L} + (1 - a)(S_{t-1} + b_{t-1}) \quad (3.23)$$

όπου L είναι το εύρος της εποχικότητας, δηλαδή ο αριθμός των τριμήνων ή μηνών του έτους.

- Κάνουμε επίκαιρη την εκτίμηση της εποχικότητας

$$I_t = \beta \frac{X_t}{S_t} + (1 - \beta)I_{t-L} \quad (3.24)$$

όπου I είναι ο παράγοντας προσαρμογής της εποχικότητας.

- Κάνουμε επίκαιρο τον εκτιμητή της τάσης

$$b_t = \nu(S_t + b_t \times m)I_{t-L+m} \quad (3.25)$$

- Προβλέπουμε m περιόδους στο μέλλον

$$F_{t+m} = (S_t + b_t \times m)I_{t-L+m} \quad (3.26)$$

Με την εξίσωση (3.23) γίνονται επίκαιρες οι εξομαλυνθείσες τιμές της σειράς. Στην εξίσωση (3.23), το  $X_t$  διαιρείται δια του  $I_t - L$  που προσαρμόζει τις αρχικές παρατηρήσεις,  $X_t$  για εποχικότητα και στη συνέχεια αναιρεί τις επιδράσεις της

εποχικότητας όσο καλύτερα αυτές μπορεί να μετρηθούν από τη χρονολογική σειρά. Η εξίσωση (3.24) δίνει την εκτίμηση της εποχικής συνιστώσας,  $X_t / S_t$  πολλαπλασιασμένη επί τη σταθερά  $\beta$  συν την παλαιά εποχική εκτίμηση,  $I_t - L$  πολλαπλασιασμένη επί  $(1-\beta)$ . Επομένως, η επικαιροποίηση των εποχικών εκτιμήσεων είναι από μόνη της μια διαδικασία εκθετικής εξομάλυνσης. Επίσης, ο  $X_t$  διαιρείται με  $S_t$  προκειμένου να εκφραστεί η τιμή ως δείκτης παρά ως απόλυτο μέγεθος. Αυτό επιτρέπει την εύρεση του μέσου όρου των νέων εποχικών εκτιμήσεων με βάση τον εποχικό δείκτη της προηγούμενης περιόδου. Η εξίσωση (3.25) εκφράζει τη σύγχρονη τιμή της συνιστώσας της τάσης που επιτυγχάνεται με τη συνηθισμένη διαδικασία εκθετικής εξομάλυνσης. Τέλος, μετά την εξίσωση (3.25), λαμβάνουμε την εξίσωση (3.26) για τις μελλοντικές περιόδους. Η διαφορά είναι ότι αυτή η εκτίμηση για τη μελλοντική περίοδο,  $t+m$ , πολλαπλασιάζεται επί  $I_t - L + m$ . Αυτός είναι ο τελικός διαθέσιμος εποχικός δείκτης και αποτελεί την προσαρμογή της πρόβλεψης για εποχικότητα. (Σούκος & Κάντζιος, 2006)

### 3.4 Διάσπαση χρονοσειρών

Στην ανάλυση χρονοσειρών (ιδιαίτερα οικονομικών), υποτίθεται ότι οι τρεις συστηματικές συνιστώσες και η τυχαία συνιστώσα είναι αποτέλεσμα διαφορετικών αιτιών. Διαχωρίζοντας τρεις συστηματικές συνιστώσες, ρίχνουμε κάποιο φως στη φύση των παραγόντων που τις προκαλούν. Η γνώση αυτή μας επιτρέπει να προβάλλουμε την επίδραση της συστηματικής συνιστώσας στο μέλλον. Στη συνέχεια, κάνοντας εκτιμήσεις ή υποθέσεις για τη μη-κανονική συνιστώσα μπορούμε να προχωρήσουμε στην πρόβλεψη.

Εάν το  $Y_t$  συμβολίζει την πραγματική παρατήρηση σε χρόνο  $t$ , εκφράζουμε το  $Y_t$  σε όρους της τάσης της σειράς,  $T_t$  σε χρόνο  $t$ , τον εποχικό παράγοντα,  $S_t$  σε χρόνο  $t$ , την κυκλική συνιστώσα,  $C_t$  και τους μη-κανονικούς παράγοντες με  $I_t$ . Στη συνέχεια υποθέτουμε ότι η σχέση παίρνει την παρακάτω μορφή:

$$Y_t = S_t * T_t * C_t * I_t \quad (3.27)$$

Η ισότητα αυτή υποδηλώνει ότι η τιμή της μεταβλητής σε χρόνο  $t$  είναι το γινόμενο των τεσσάρων συνιστωσών που αναφέρθηκαν. Ακόμη, η εξίσωση δείχνει ότι τα  $T$ ,  $S$ ,

C και αλληλοσχετίζονται. Εναλλακτικά, η σχέση μπορεί να υποθεθεί ότι είναι προσθετική ως εξής:

$$Y_t = S_t + T_t + C_t + I_t \quad (3.28)$$

Σε αυτή την εξίσωση οι συνιστώσες δεν αλληλοσχετίζονται. Το μέγεθος της μιας συνιστώσας είναι ασυσχέτιστο με το μέγεθος οποιασδήποτε άλλης. Από τα δυο μοντέλα διαχωρισμού, το προσθετικό είναι πιο δύσκολο για παραπέρα ανάλυση. Επίσης, το προσθετικό μοντέλο υποθέτει ότι οι συντελεστές μεταξύ τους είναι ανεξάρτητοι. Για παράδειγμα, αν χρησιμοποιήσουμε το προσθετικό μοντέλο, δεχόμαστε ότι η τάση δεν επηρεάζει τον εποχικό συντελεστή όσο υψηλός ή χαμηλός κι αν γίνεται. Κατά κανόνα, η παραδοχή αυτή δεν είναι ισχυρή, εκτός από την περίπτωση της πολύ βραχυχρόνιας πρόβλεψης. Επομένως, θα δώσουμε έμφαση στο πολλαπλασιαστικό μοντέλο, εφόσον στην πραγματικότητα όλοι οι χρήστες της διαχωριστικής (decomposition) ανάλυσης το βρίσκουν πιο ευλογοφανές.

#### **3.4.1 Ανάλυση και διαχωρισμός της εποχικότητας**

Η γνώση των εποχικών προτύπων είναι ζωτικός παράγοντας για τον σωστό και έγκυρο σχεδιασμό. Εκτός από την κατάστρωση ετήσιου επιχειρηματικού προγράμματος, η γνώση των εποχικών κινήσεων μας επιτρέπει να αναλύουμε τις επιδόσεις του παρελθόντος, να αλλάζουμε το διαχρονικό πρότυπο παραγωγής και διάθεσης κι ίσως ακόμα να μεταβάλλουμε και το ίδιο το εποχικό πρότυπο.

Αν και υπάρχουν διάφορες μέθοδοι για τη μέτρηση της εποχικής διακύμανσης, θα εστιάσουμε την προσοχή μας σε μία, στη μέθοδο του λόγου της πραγματικής τιμής προς τον κινητό μέσο (ratio-to-moving average). Η λογική που βρίσκεται πίσω από αυτή και άλλες μεθόδους είναι η εξάλειψη της επίδρασης της τάσης, των κυκλικών και μη- κανονικών στοιχείων από τα αρχικά δεδομένα των χρονοσειρών, προκειμένου να διασφαλιστεί ότι παραμένει μόνο η επίδραση του εποχικού παράγοντα. Με δεδομένα σε τρίμηνη βάση, κατασκευάζουμε έναν κινητό μέσο όρο τεσσάρων τριμήνων για να εκτιμήσουμε τους εποχικούς παράγοντες από τις παρατηρήσεις της χρονοσειράς. Αν χρησιμοποιηθούν μηνιαία δεδομένα, μπορεί να κατασκευαστεί ένας κινητός μέσος όρος 12 μηνών.

Στόχος είναι η εξάλειψη των άλλων συνιστωσών έτσι ώστε να απομείνουν μόνο οι εποχικές διακυμάνσεις. Αυτοί οι κινητοί μέσοι όροι θα μπορούσαν να εξαλείψουν τις

κυμάνσεις που οφείλονται σε εποχικότητα και συνυπάρχουν στα δεδομένα. Η βασική παραδοχή που βρίσκεται πίσω από τη μέθοδο του κινητού μέσου όρου είναι ότι το εποχικό πρότυπο επανεμφανίζεται σε κανονική βάση κάθε επόμενο έτος. Μολονότι το εποχικό πρότυπο δεν είναι απόλυτα ακριβές κατά τα διαδοχικά έτη, ωστόσο υπάρχει μέσα σε λογικά όρια και ο κινητός μέσος όρος ελαχιστοποιεί την επίδραση σε έναν τουλάχιστον βαθμό. Για να εξηγήσουμε τη θεωρία, θα υποθέσουμε ότι οι συνιστώσες της χρονοσειράς είναι πολλαπλασιαστικής μορφής (εξίσωση 3.27). Ο όρος ratio-to-moving-average χρησιμοποιείται επειδή παίρνουμε τη μεταβλητή  $Y_t$  της χρονοσειράς και τη διαιρούμε με τον κινητό μέσο όρο ως εξής:

$$\frac{Y_t}{MA} = \frac{S_t \times T_t \times C_t \times I_t}{T_t \times C_t \times I_t} = S_t \quad (3.29)$$

Η ποσότητα που βρίσκουμε, η  $S_t$ , είναι η καλύτερη δυνατή μέτρηση της εποχικής συνιστώσας στα δεδομένα της χρονοσειράς. (Σούκος & Κάντζιος, 2006)

### 3.4.2 Προσαρμογή και Ανάλυση της Τάσης

Μελετούμε τις τάσεις για να προσδιορίσουμε τη μακρόχρονη κατεύθυνση της χρονοσειράς. Γενικά, αυτές οι εξελίξεις προέρχονται από μια ποικιλία αιτιών π.χ. στις οικονομικές χρονοσειρές τέτοιες αιτίες είναι οι μεταβολές της τεχνολογίας, των τιμών, του πληθυσμού και της παραγωγικότητας. Για παράδειγμα, το σύνολο των αποταμιευτικών καταθέσεων σε μια τράπεζα ταμιευτηρίου μπορεί να αυξάνεται από έτος σε έτος για ένα χρονικό διάστημα, επειδή ο αριθμός των καταθετών συνεχίζει να αυξάνεται. Παραπέρα, ο κάθε καταθέτης τείνει να έχει μεγαλύτερη αποταμίευση σε χρηματικές μονάδες, πράγμα που σχετίζεται με μια γενική αύξηση στα εισοδήματα των καταθετών. Εν μέρει, αυτό προέρχεται από μια γενική αύξηση των τιμών που ονομάζεται πληθωρισμός (τάση). Επίσης, το γενικό ανοδικό πρότυπο των εσόδων ενός νοσοκομείου σχετίζεται με τη γενική αύξηση (τάση) των τιμών ή των τιμών των νοσοκομειακών υπηρεσιών. Στην περίπτωση του ανέμου, μια ένδειξη τάσης, θα μπορούσε να παρατηρηθεί π.χ. όταν για κάποιες ημέρες ξεσπούν έντονα καιρικά φαινόμενα, στη υπό μελέτη περιοχή, και συνοδεύονται από ισχυρούς ανέμους.

Ένα πρώτο βήμα στην ανάλυση της συνιστώσας της τάσης είναι ο καθορισμός του λόγου για τον οποίο γίνεται η ανάλυσή της. Μπορεί κάποιος να θέλει να προβάλει την

τάση στο μέλλον ή να την εξαλείψει από την αρχική χρονοσειρά. Φυσικά, ίσως κάποιος να επιθυμεί να πράξει και τα δυο μαζί. Η γνώση του σκοπού βοηθάει τον ασχολούμενο με τις προβλέψεις στην επιλογή της μεθόδου και του εύρους της χρονοσειράς που θα αναλύσει. Εάν ο σκοπός είναι η προβολή της τάσης στο μέλλον, είναι σημαντικό να γνωρίζει το πόσο μακριά στο μέλλον πρόκειται να γίνει αυτή η προβολή. Μια προβολή για 15 χρόνια κατά κανόνα απαιτεί τη χρησιμοποίηση χρονοσειράς μακρύτερης διάρκειας από ό,τι μια προβολή τριών μόνο χρόνων.

Το πρώτο βήμα για την ανάλυση της τάσης είναι η γραφική παράσταση της χρονοσειράς, τόσο σε αριθμητική όσο και σε ημι-λογαριθμική κλίμακα πριν την επιλογή της μεθόδου μέτρησης. Αν και υπάρχουν στατιστικά κριτήρια για τον καθορισμό της «σπουδαιότητας» της τάσης, μια εξέταση της γραφικής παράστασης ως προς τον χρόνο θα είναι συνήθως επαρκής. (Σούκος & Κάντζιος, 2006)

### 3.4.3 Το Μοντέλο Γραμμικής Τάσης

Η αξιολογική κρίση (judgement) χρησιμοποιείται πολύ συχνά ως κριτήριο για τον προσδιορισμό του μοντέλου με την καλύτερη προσαρμογή. Μερικές φορές ο διενεργών προβλέψεις χαράζει απλώς μια γραμμή διαμέσου της σειράς των παρατηρήσεων, προκειμένου να αποκαλύψει το σχήμα και την κατεύθυνση της τάσης της χρονοσειράς. Ο προβλέπων, χαράζοντας τη γραμμή της τάσης πρέπει να έχει την ικανότητα να αναγνωρίζει την επαναλαμβανόμενη εποχική διακύμανση και τους κύκλους μέσω των οποίων διέρχεται η τάση. Με άλλα λόγια, θα πρέπει ο προβλέπων να έχει εξοικειωθεί σημαντικά με αυτές τις συγκεκριμένες σειρές, διαφορετικά η διαδικασία αυτή μπορεί να αποδειχθεί εξαιρετικά δύσκολη. Γι' αυτό οι διενεργούντες προβλέψεις συνήθως εμπιστεύονται την αντικειμενική μέθοδο των ελαχίστων τετραγώνων για να αποφύγουν την υποκειμενικότητα που υπεισέρχεται στη μέθοδο που βασίζεται στην αξιολογική κρίση. Για να αρχίσουμε, θα ορίσουμε την εκτιμηθείσα γραμμική εξίσωση ως εξής:

$$\hat{Y}_t = a + bX_t \quad (3.30)$$

$Y_t$  είναι η προβλεφθείσα τιμή τάσης για τη μεταβλητή  $Y$  για μια επιλεγείσα κωδικοποιημένη χρονική περίοδο  $X$ .  $a$  είναι η τιμή της τάσης όταν  $X=0$  (η διαφορά ύψους).  $b$  είναι η κλίση της εξίσωσης (η μεταβολή στο  $Y^{\wedge}$  που σχετίζεται με μια

μοναδιαία μεταβολή στη X). Το X παριστάνει την παρατήρηση για την κωδικοποιημένη χρονική περίοδο. Κατά την εκτίμηση της εξίσωσης της τάσης, ο τύπος για την κλίση είναι:

$$b = \frac{n \sum XY - \sum X \sum Y}{n \sum X^2 - (\sum X)^2} \quad (3.31)$$

Αντικαθιστούμε την κλίση στον τύπο, για να βρούμε τη διαφορά ύψους:

$$a = \frac{\sum Y}{n} - b \frac{\sum X}{n} \quad (3.32)$$

Αν και η εξίσωση της ευθύγραμμης τάσης εφαρμόζεται πιο συχνά στις χρονοσειρές από οποιοδήποτε άλλο μοντέλο για να περιγράψει μακροχρόνιες κινήσεις, η χρήση άλλων μοντέλων για την περιγραφή μεταβολών συχνά είναι πιο ενδεδειγμένες.

#### 3.4.4 Απομόνωση της κυκλικότητας

Στη παράγραφο αυτή θα ασχοληθούμε με μια εφαρμογή της μεθόδου του κινητού μέσου όρου προκειμένου να απομονωθεί η κυκλική συνιστώσα. Αυτή η διαδικασία απομόνωσης είναι παρόμοια με τη θεωρία απομόνωσης της τάσης και της εποχικής συνιστώσας, με τις οποίες ασχοληθήκαμε νωρίτερα. Η απομόνωση μιας κυκλικής συνιστώσας επιτυγχάνεται με την απάλειψη της εποχικότητας από τα αρχικά δεδομένα ως εξής:

$$\frac{Y_t}{\text{εποχικός δείκτης}} = \frac{S_t \times T_t \times C_t \times I_t}{S_t} = T_t \times C_t \times I_t \quad (3.33)$$

Τα δεδομένα αυτά, τα οποία είναι απαλλαγμένα από την εποχικότητα, μπορούν να απαλλαγούν από την τάση ως εξής:

$$\frac{T_t \times C_t \times I_t}{T_t} = C_t \times I_t \quad (3.34)$$

Οι σειρές που προκύπτουν περιλαμβάνουν μόνο την επίδραση που σχετίζεται με την κυκλική συνιστώσα και τη μη-κανονική διακύμανση. Η μη-κανονική διακύμανση μπορεί να εξαλειφθεί από τη χρονοσειρά μέσω υπολογισμού του κινητού μέσου όρου,

που εξομαλύνει αυτές τις μη-κανονικές διακυμάνσεις από τα δεδομένα. Ένα απλό μοντέλο κινητού μέσου,  $M_t$ , παρουσιάζεται στην ακόλουθη εξίσωση:

$$M_t = \frac{Z_t + Z_{t-1} + \dots + Z_{t-n+1}}{n} \quad (3.35)$$

όπου  $M_t$  είναι ο κινητός μέσος όρος τον χρόνο  $t$ .  $Z_t$  είναι η τιμή στην περίοδο  $t$ , η απαλλαγμένη από την εποχικότητα και την τάση, δηλαδή  $Z_t = C_t * I_t$  και  $n$  είναι ο αριθμός των όρων στον κινητό μέσο όρο. Για ευκολία στους υπολογισμούς, η εξίσωση μπορεί να ξαναγραφεί ως:

$$M_t = M_{t-1} + \frac{Z_t - Z_{t-n}}{n} \quad (3.36)$$

Επομένως, για τη χρονική περίοδο  $t$ , ο κινητός μέσος όρος είναι ο αριθμητικός μέσος των  $n$  πιο πρόσφατων παρατηρήσεων. Στις παραπάνω εξισώσεις, ορίζονται ίσοι συντελεστές στάθμισης σε κάθε μία από τις  $n$  πιο πρόσφατες παρατηρήσεις. Σε όλες τις άλλες ιστορικές παρατηρήσεις οι συντελεστές στάθμισης θεωρείται ότι είναι μηδενικοί. Σε κάθε παρατήρηση  $Z_t$  θα θεωρηθούν ως συντελεστές στάθμισης οι  $W_t$ . Οι τύποι ξαναγράφονται ως εξής:

$$M_t = \frac{W_t Z_t + W_{t-1} Z_{t-1} + \dots + W_{t-n+1} Z_{t-n+1}}{n} \quad (3.37)$$

και

$$M_t = M_{t-1} + \frac{W_t Z_t - W_{t-n} Z_{t-n}}{n} \quad (3.38)$$

όπου  $W_t$  είναι ο συντελεστής στάθμισης που αντιστοιχεί στην παρατήρηση  $Z_t$  σε χρόνο  $t$ . Καθώς θα έχουμε στη διάθεσή μας μια νέα τιμή των δεδομένων, αυτή θα περιλαμβάνεται στον μέσο όρο και η τιμή που αντιστοιχεί στη η-οστή προηγούμενη περίοδο θα αγνοείται. Επομένως,  $M_t$  είναι μια επικαιροποιημένη έκδοση της αμέσως προηγούμενης εκτίμησης. Η δεύτερη μορφή του  $M_t$  τονίζει αυτό το σημείο.

Οι κινητοί μέσοι όροι θα ανταποκρίνονται στις μεταβολές στο πρότυπο των δεδομένων σε βαθμό που εξαρτάται τόσο από τον αριθμό των περιόδων που περιλαμβάνονται στον κινητό μέσο όρο όσο και από τους συντελεστές στάθμισης, που αντιστοιχούν στη φόρμουλα. Σημειωτέον, ότι εάν οι συντελεστές στάθμισης θεωρηθούν οι ίδιοι, τότε οι τύποι για τους σταθμισμένους και τους μη σταθμισμένους

κινητούς μέσους όρους είναι ισοδύναμοι. Εάν δίνονται μεγάλοι συντελεστές στάθμισης στις τρέχουσες περιόδους και πολύ μικροί στις παλαιότερες, τότε η πρόσφατη παρατήρηση θα επιδρά περισσότερο στην τιμή του κινητού μέσου όρου. Επίσης, όταν συμπεριλαμβάνονται περισσότεροι περίοδοι στον κινητό μέσο (μεγαλύτερες τιμές του  $n$ ), τόσο λιγότερο ευαίσθητος μπορεί να γίνει ο κινητός μέσος όρος στις μεταβολές του προτύπου της κυκλικής κίνησης. Αντιθέτως, μια μικρή τιμή του  $n$ , οδηγεί σε έναν κινητό μέσο που ανταποκρίνεται πολύ γρήγορα στις μεταβολές. Τέλος, εάν κάποιος ήθελε να προβλέψει περιόδους πριν την τρέχουσα περίοδο  $t$ , η εξίσωση πρόβλεψης γίνεται:

$$\hat{Z}_{t+1} = M_t = \frac{W_t Z_t + W_{t-1} Z_{t-1} + \dots + W_{t-n+1} Z_{t-n+1}}{W_t + W_{t-1} + \dots + W_{t-n+1}} \quad (3.39)$$

όπου  $Z$  είναι η προβλεφθείσα τιμή του  $Z$  στο σημείο  $t+1$ .

Η γνώση της συνεισφοράς κάθε μιας από τις συστηματικές συνιστώσες, εποχική, τάση και κυκλικοί παράγοντες, που επιτυγχάνεται με την κλασική ανάλυση (classical decomposition) επιτρέπει στον προβλέποντα την επίτευξη του κυριότερου σκοπού της ανάλυσης χρονοσειρών. Ο σκοπός είναι η εξήγηση της διακύμανσης στα δεδομένα των παρατηρήσεων διαχρονικά.

Από τα παραπάνω γίνεται φανερό ότι η διαμόρφωση των προβλέψεων εξαρτάται σε μεγάλο βαθμό από τη διαδικασία αναγνώρισης των συνθετικών στοιχείων της χρονοσειράς. Όσο καλύτερη είναι η αναγνώριση των στοιχείων αυτών, τόσο καλύτερη αναμένεται να είναι και η πρόβλεψη των τιμών της χρονοσειράς. (Σούκος & Κάντζιος, 2006)

### 3.5 Ανάλυση ARIMA

#### 3.5.1 Μοντέλα χρονοσειρών ARIMA

Τα μοντέλα πολλαπλής γραμμικής παλινδρόμησης περιγράφονται από την εξίσωση (Κουγιουμτζής, 2005):

$$Y = b_0 + b_1 X_1 + b_2 X_2 + \dots + b_p X_p + e \quad (3.40)$$



όπου  $Y$  η εξαρτημένη μεταβλητή (υπό πρόβλεψη μεταβλητή) και  $X_1, X_2, \dots, X_p$  οι ανεξάρτητες μεταβλητές (επεξηγηματικές μεταβλητές).

Η εξίσωση  $Y_t = c + \phi_1 Y_{t-1} + e_t$  είναι μια εξίσωση πολλαπλής γραμμικής παλινδρόμησης, με τη διαφορά ότι οι ανεξάρτητες μεταβλητές είναι προηγούμενες τιμές της υπό πρόβλεψη μεταβλητής ή οι τιμές της μεταβλητής για χρονικές καθυστερήσεις  $1, 2, \dots, p$ . Το μοντέλο καλείται αυτοπαλινδρομικό (AutoRegression ή AR).

Τα αυτοπαλινδρομικά μοντέλα πρέπει να τύχουν διαφορετικού χειρισμού από τα κλασσικά παλινδρομικά μοντέλα για δύο λόγους :

- Η βασική υπόθεση της ανεξαρτησίας των σφαλμάτων (υπολοίπων) μπορεί εύκολα να παραβιαστεί στα αυτοπαλινδρομικά μοντέλα, αφού οι επεξηγηματικές μεταβλητές έχουν συνήθως μια εξαρτημένη σχέση καθώς περιγράφουν την εξέλιξη του ίδιου μεγέθους.
- Ο τρόπος προσδιορισμού του πλήθους των προηγούμενων τιμών της υπό πρόβλεψη μεταβλητής δεν είναι πάντοτε "ευθύς".

Όμοια είναι δυνατόν να θεωρηθεί ένα παλινδρομικό μοντέλο της μορφής :

$$Y = b_0 + b_1 e_{t-1} + b_2 e_{t-2} + \dots + b_p e_{t-p} + e_t \quad (3.41)$$

όπου οι επεξηγηματικές μεταβλητές, είναι οι τιμές των προηγούμενων σφαλμάτων.

Εδώ, σαφώς υπάρχει μια σχέση εξάρτησης μεταξύ διαδοχικών σφαλμάτων, και το μοντέλο καλείται κινητού μέσου όρου (Moving Average ή MA), επειδή ορίζεται σαν κινητός μέσος όρος της σειράς σφαλμάτων, et.

Τα αυτοπαλινδρομικά μοντέλα (AR) μπορούν να ενωθούν αποτελεσματικά με τα μοντέλα κινητού μέσου όρου (MA) και να σχηματίσουν μια χρήσιμη ομάδα μοντέλων χρονοσειρών, τα οποία ονομάζονται αυτοπαλινδρομικά μοντέλα κινητού μέσου όρου (autoregressive moving average ή ARMA models). Τα μοντέλα αυτά μπορούν να χρησιμοποιηθούν μόνο για στάσιμες χρονοσειρές αλλά μπορούν να επεκταθούν και σε μη στάσιμες χρονοσειρές, με τη χρήση της μεθόδου της διαφορίσης. Τότε ονομάζονται ολοκληρωμένα (integrated ή I) αυτοπαλινδρομικά μοντέλα κινητού μέσου όρου (autoregressive integrated moving average ή ARIMA models).

Τα μη εποχιακά ARIMA μοντέλα είναι γνωστά σαν ARIMA(p, d, q) όπου:

AR : p = η τάξη του AR όρου

I : d = η τάξη της διαφορίσης

MA : q = η τάξη του MA όρου

Το μοντέλο λευκού θορύβου ταξινομείται σαν ARIMA(0,0,0), ενώ το μοντέλο τυχαίου περιπάτου σαν ARIMA(0,1,0) (Κουγιουμτζής, 2005).

### Αυτοπαλινδρομικά μοντέλα πρώτης τάξης

Τα ARIMA(1,0,0) ή AR(1) περιγράφονται από την εξίσωση :

$$Y_t = c + \phi_1 Y_{t-1} + e_t \quad (3.42)$$

Η παρατήρηση  $Y_t$  εξαρτάται από την  $Y_{t-1}$ , ενώ ο συντελεστής  $\phi_1$  παίρνει τιμές ανάμεσα στο -1 και 1. Για  $\phi_1=0$  το μοντέλο είναι ισοδύναμο με ένα μοντέλο λευκού θορύβου, ενώ για  $\phi_1=1$  το μοντέλο είναι ισοδύναμο με ένα μοντέλο τυχαίου περιπάτου. Γενικά, μπορεί να χρησιμοποιηθεί ένα AR(1) μοντέλο εάν :

- οι συντελεστές αυτοσυσχέτισης φθίνουν εκθετικά
- υπάρχει μόνο ένας στατιστικά σημαντικός συντελεστής μερικής αυτοσυσχέτισης.

### Μοντέλα Κινητού Μέσου Όρου πρώτης τάξης

Τα ARIMA(0,0,1) ή MA(1) περιγράφονται από την εξίσωση :

$$Y_t = c + e_t - \theta_1 e_{t-1} \quad (3.43)$$

Η παρατήρηση  $Y_t$  εξαρτάται από τον όρο σφάλματος  $e_t$  και το προηγούμενο σφάλμα  $e_{t-1}$ , ενώ ο συντελεστής  $\theta_1$  παίρνει τιμές ανάμεσα στο -1 και 1. Τα σφάλματα  $e_t$  ακολουθούν την κανονική κατανομή με μέση τιμή 0 και διακύμανση 1. Οι συντελεστές μερικής αυτοσυσχέτισης φθίνουν εκθετικά στο μηδέν, ενώ υπάρχει μόνο μία σημαντική αυτοσυσχέτιση.

### Αυτοπαλινδρομικά μοντέλα μεγαλύτερης τάξης

Γενικότερα, ένα αυτοπαλινδρομικό μοντέλο p-τάξης ορίζεται ως εξής :

$$Y_t = c + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \dots + \phi_p Y_{t-p} + e_t \quad (3.44)$$

όπου

c = ο σταθερός όρος

$\phi_j$  = η παράμετρος j

$e_t$  = ο όρος σφάλματος τη χρονική στιγμή t.

Οι τιμές των παραμέτρων  $\phi_j$  ακολουθούν συγκεκριμένους περιορισμούς.

Για p = 1,  $-1 < \phi_1 < 1$

Για p = 2,  $-1 < \phi_2 < 1$ ,  $\phi_2 + \phi_1 < 1$ ,  $\phi_2 - \phi_1 < 1$ .

Για  $p \geq 3$ , επικρατούν πολυπλοκότερες συνθήκες.

Με τη χρήση του τελεστή ολίσθησης B, το μοντέλο μπορεί να γραφεί ως :

$$Y_t - \phi_1 Y_{t-1} - \phi_2 Y_{t-2} - \dots - \phi_p Y_{t-p} = c + e_t \quad (3.45)$$

ή

$$(1 - \phi_1 B - \dots - \phi_p B^p) Y_t = c + e_t \quad (3.46)$$

### Μοντέλα Κινητού Μέσου Όρου μεγαλύτερης τάξης

Ένα μοντέλο κινητού μέσου όρου q-τάξης ορίζεται ως εξής :

$$Y_t = c + e_t - \theta_1 e_{t-1} + \theta_2 e_{t-2} + \dots + \theta_p e_{t-p} \quad (3.47)$$

όπου

c = ο σταθερός όρος

$\theta_j$  = η παράμετρος j

$e_{t-k}$  = ο όρος σφάλματος τη χρονική στιγμή t-k.

Οι περιορισμοί που απαιτούνται για τα AR απαιτούνται και για τα MA μοντέλα.

Για q = 1,  $-1 < \theta_1 < 1$

Για  $q = 2$ ,  $-1 < \theta_2 < 1$ ,  $\theta_2 + \theta_1 < 1$ ,  $\theta_2 - \theta_1 < 1$ .

Για  $q \geq 3$ , επικρατούν πολυπλοκότερες συνθήκες.

Με τη χρήση του τελεστή ολίσθησης  $B$ , το μοντέλο μπορεί να γραφεί ως :

$$Y_t = c + (1 - \theta_1 B - \dots - \theta_q B^q) e_t \quad (3.48)$$

### Διαδικασίες ARMA

Τα βασικά στοιχεία των AR και MA μοντέλων μπορούν να συνδυαστούν για την ανάπτυξη πλήθους μοντέλων. Για παράδειγμα η παρακάτω εξίσωση συνδυάζει ένα AR μοντέλο πρώτης τάξης και ένα MA μοντέλο πρώτης τάξης. Το μοντέλο ονομάζεται ARMA(1,1) ή ARIMA(1,0,1), και χρησιμοποιείται για χρονοσειρές στάσιμες ως προς τη μέση τιμή και την διακύμανση.

$$Y_t = c + \phi_1 Y_{t-1} + e_t - \theta_1 e_{t-1} \quad (3.49)$$

Εδώ η  $Y_t$  εξαρτάται από την προηγούμενη τιμή  $Y_{t-1}$  και το προηγούμενο σφάλμα  $e_{t-1}$ . Με τη χρήση του τελεστή ολίσθησης  $B$ , το μοντέλο μπορεί να γραφεί ως :

$$(1 - \phi_1 B) Y_t = c + (1 - \theta_1 B) e_t \quad (3.50)$$

Η εξίσωση ενός ARMA μοντέλου με όρους μεγαλύτερης τάξης είναι :

$$(1 - \phi_1 B - \dots - \phi_p B^p) Y_t = c + (1 - \theta_1 B - \dots - \theta_p B^p) e_t \quad (3.51)$$

Για τις παραμέτρους ενός ARMA μοντέλου ισχύουν οι περιορισμοί που ισχύουν για τις παραμέτρους των AR και MA μοντέλων.

### Αναγνώριση της κατάλληλης AR ή MA διαδικασίας

Με τη χρήση των ACF και PACF, είναι δυνατή η αναγνώριση της κατάλληλης AR ή MA διαδικασίας. Στον ακόλουθο πίνακα περιγράφονται τα αναλυτικά χαρακτηριστικά των ACF και PACF των διαδικασιών αυτών.

**Πίνακας 3.2:** Εκτίμηση της τάξης των AR και MA με τη βοήθεια των ACF-PACF

Διαδικασία	ACF	PACF
<b>AR(1)</b>	Εκθετική μείωση : θετικές τιμές αν $\phi_1 > 0$ εναλλαγή πρόσημου ξεκινώντας από αρνητική τιμή, αν $\phi_1 < 0$	Απότομος μηδενισμός μετά την περίοδο 1. Η τιμή την περίοδο 1 είναι : θετική αν $\phi_1 > 0$ αρνητική αν $\phi_1 < 0$
<b>AR(p)</b>	Εκθετική μείωση ή πρότυπο φθίνουσας ημιτονοειδούς συνάρτησης. Το ακριβές πρότυπο εξαρτάται από το πρόσημο και το μέγεθος των $\phi_1 \phi_2, \dots, \phi_p$ .	Μη μηδενικές τιμές για τις πρώτες p περιόδους και στη συνέχεια απότομος μηδενισμός.
<b>MA(1)</b>	Απότομος μηδενισμός μετά την περίοδο 1. Η τιμή την περίοδο 1 είναι : θετική αν $\theta_1 < 0$ αρνητική αν $\theta_1 > 0$	Εκθετική μείωση : εναλλαγή πρόσημου ξεκινώντας από θετική τιμή αν $\theta_1 < 0$ αρνητικές τιμές αν $\theta_1 > 0$
<b>MA(q)</b>	Μη μηδενικές τιμές για τις πρώτες q περιόδους και στη συνέχεια απότομος μηδενισμός.	Εκθετική μείωση ή πρότυπο φθίνουσας ημιτονοειδούς συνάρτησης. Το ακριβές πρότυπο εξαρτάται από το πρόσημο και το μέγεθος των $\theta_1 \theta_2, \dots, \theta_q$ .

### Διαδικασίες ARIMA

Εάν η χρονοσειρά είναι μη στάσιμη τότε αντί για ένα ARMA μοντέλο, εφαρμόζεται ένα μοντέλο ARIMA(p,d,q). Η πιο απλή περίπτωση είναι το ARIMA(1,1,1) με εξίσωση :

$$(1 - \phi_1 B)(1 - B)Y_t = c + (1 - \theta_1 B)e_t \quad (3.52)$$

Οι συναρτήσεις αυτοσυσχέτισης (ACF) και μερικής αυτοσυσχέτισης (PACF) των μοντέλων ARIMA(p,d,q) ακολουθούν ένα πλήθος προτύπων που καθιστά αδύνατη τη θέσπιση κανόνων για την αναγνώριση του καταλληλότερου μοντέλου. Η επιλογή όμως ενός κατάλληλου μοντέλου δεν είναι τόσο δύσκολη καθώς διαφορετικά μοντέλα είναι δυνατόν να παράγουν "καλές" προβλέψεις. Στην πράξη σπάνια χρησιμοποιούνται μοντέλα με τιμές p, d, q διαφορετικές από 0, 1 και 2.

### Εποχικότητα και μοντέλα ARIMA.

Με τον ίδιο ακριβώς τρόπο που ακόλουθα δεδομένα παρουσιάζουν AR, MA, ARMA και ARIMA ιδιότητες, έτσι και δεδομένα που απέχουν s περιόδους (s η εποχικότητα) είναι δυνατόν να παρουσιάζουν τέτοιες ιδιότητες. Τα μοντέλα ARIMA που έχουν επεκταθεί ώστε να χειρίζονται την εποχικότητα συμβολίζονται ως :

**ARIMA (p, d, q) (P, D, Q) s**

Μη εποχιακοί όροι

εποχιακοί όροι

εποχικότητα

Στη συνέχεια περιγράφεται ένα μοντέλο ARIMA(1,1,1)(1,1,1)<sub>4</sub> με τη χρήση του τελεστή ολίσθησης B.

$$(1 - \phi_1 B)(1 - \Phi_1 B^4)(1 - B)(1 - B^4)Y_t = (1 - \theta_1 B)(1 - \theta_1 B^4)e_t \quad (3.52)$$

Οι εποχιακοί όροι μπορούν να αναγνωρισθούν από τις ACF και PACF.

Για παράδειγμα η ACF ενός εποχιακού MA μοντέλου, ARIMA(0,0,0)(0,0,1)<sub>12</sub>, θα παρουσιάζει μόνο μια στατιστικά σημαντική τιμή για καθυστέρηση 12 και η PACF θα φθίνει εκθετικά στις εποχιακές καθυστερήσεις 12, 24, 36, ... . Όμοια για ένα

εποχιακό AR μοντέλο,  $ARIMA(0,0,0)(1,0,0)_{12}$ , η ACF θα φθίνει εκθετικά στις εποχιακές καθυστερήσεις ενώ η PACF θα παρουσιάζει μόνο μια στατιστικά σημαντική τιμή για καθυστέρηση 12.

### Εκτίμηση των παραμέτρων

Η μέθοδος ελαχίστων τετραγώνων μπορεί να χρησιμοποιηθεί για την εκτίμηση των παραμέτρων ενός ARIMA μοντέλου, αλλά η εφαρμογή της είναι αρκετά δύσκολη όταν στο μοντέλο συμπεριλαμβάνονται MA όροι ( $q > 0$ ). Αντίθετα μπορεί να χρησιμοποιηθεί μια επαναληπτική μέθοδος. Αφού επιλεγθούν αρχικές τιμές για τις παραμέτρους, στη συνέχεια βελτιώνονται μέσω μιας επαναληπτικής διαδικασίας μέχρι να ελαχιστοποιηθεί το άθροισμα των τετραγώνων των σφαλμάτων.

Μια άλλη επαναληπτική μέθοδος η οποία χρησιμοποιείται συχνά είναι η μέθοδος μεγίστης πιθανοφάνειας. Η πιθανοφάνεια ενός συνόλου δεδομένων συμβολίζεται με  $L$  και είναι ανάλογη με την πιθανότητα να παραχθούν τα πραγματικά δεδομένα από το μοντέλο. Η μέθοδος μεγίστης πιθανοφάνειας χρησιμοποιείται για την εκτίμηση των τιμών των παραμέτρων που μεγιστοποιούν την πιθανότητα  $L$ .

Κάθε εκτιμούμενος συντελεστής έχει ένα τυπικό σφάλμα γιατί είναι μια στατιστική τιμή που βασίζεται σε πληροφορία από ένα μόνο δείγμα. Ένα διαφορετικό δείγμα πιθανόν να έδινε διαφορετικές εκτιμήσεις για τους συντελεστές του μοντέλου. Η σημαντικότητα των συντελεστών ελέγχεται μέσω του στατιστικού δείκτη  $t$ . Στην πράξη απορρίπτουμε κάθε εκτιμούμενη τιμή συντελεστή με απόλυτη  $t$ -τιμή μικρότερη του 2.0. Κάθε συντελεστής με απόλυτη τιμή  $t < 2$  δεν είναι σημαντικά διάφορος του μηδενός για επίπεδο σημαντικότητας 5% και οδηγεί στη δημιουργία μη φειδωλών μοντέλων και άρα σε λιγότερο ακριβείς προβλέψεις.

Τα περισσότερα στατιστικά υπολογιστικά πακέτα προσαρμόζουν αυτόματα ένα μοντέλο ARIMA σε μια χρονοσειρά, εκτελούν όλους τους αναγκαίους στατιστικούς ελέγχους και παράγουν εκθέσεις με τις τιμές των παραμέτρων, το τυπικό τους σφάλμα, την  $t$ -τιμή τους, το επίπεδο σημαντικότητας, την διακύμανση των υπολοίπων και γενικότερα τις τιμές όλων των στατιστικών δεικτών που χρησιμοποιούνται για τον έλεγχο της καταλληλότητας του μοντέλου. (Κουγιουμτζής,2005)

### 3.5.2 Επιλογή του καλύτερου μοντέλου ARIMA

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

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

Ένα ικανοποιητικό κριτήριο για την επιλογή του καλύτερου μοντέλου θα μπορούσε να είναι η ελαχιστοποίηση του αθροίσματος ελαχίστων τετραγώνων ή η μεγιστοποίηση της πιθανότητας. Η συγκεκριμένη προσέγγιση όμως δεν δίνει πάντα καλά αποτελέσματα, αφού το άθροισμα ελαχίστων τετραγώνων μπορεί να γίνει μικρότερο και η πιθανότητα μεγαλύτερη, αυξάνοντας απλά τον αριθμό των όρων του μοντέλου.

Ένα ικανοποιητικό κριτήριο είναι το Akaike's Information Criterion ή AIC, το οποίο "τιμωρεί" την εισαγωγή πρόσθετων όρων στο μοντέλο. Αν  $m = p + q + P + Q$  είναι το πλήθος των όρων, τότε επιλέγουμε εκείνες τις  $p, q, P, Q$  που ελαχιστοποιούν το AIC :

$AIC = -2\log L + 2m$ , όπου  $L$  η πιθανοφάνεια.

Διαφορά στις τιμές του AIC μικρότερη του 2 δεν θεωρείται σημαντική και μπορεί να χρησιμοποιηθεί το απλούστερο μοντέλο. (Κουγιουμτζής,2005)

Άλλο ένα ικανοποιητικό κριτήριο είναι το Bayesian Information Criterion ή BIC, το οποίο χρησιμοποιεί τους ίδιους όρους με το AIC, με τη διαφορά ότι ο όρος ποινικοποίησης (τιμωρίας) στο BIC εξαρτάται και από τον λογάριθμο του αριθμού των παρατηρήσεων

$$BIC = -2\ln(L) + k \ln(n), \quad (3.53)$$



Όπου  $n$  είναι ο αριθμός των παρατηρήσεων,  $k$  είναι ο αριθμός των ελεύθερων παραμέτρων που πρέπει να εκτιμηθούν (αν η εκτίμηση είναι ένα μοντέλο γραμμικής παλινδρόμησης τότε το  $k$  είναι ο αριθμός της παλινδρομούσας μεταβλητής συμπεριλαμβανομένης της σταθεράς) και  $L$  είναι η μέγιστη τιμή της συνάρτησης πιθανοφάνειας για το εκτιμώμενο μοντέλο.

Το μοντέλο με το μικρότερο BIC είναι αυτό που προτιμάται. (Acquah, 2010)

### 3.5.3 Διαγνωστικός έλεγχος

Αν και το επιλεγμένο μοντέλο θεωρείται το καλύτερο από αυτά που εξετάστηκαν, είναι επίσης αναγκαίο να επιβεβαιωθεί η επάρκεια του μοντέλου. Αυτό επιτυγχάνεται με την εξέταση των υπολοίπων (σφαλμάτων), προκειμένου να διαπιστωθεί εάν αυτά ακολουθούν κάποιο πρότυπο.

Τα υπόλοιπα ενός καλού μοντέλου πρόβλεψης πρέπει να είναι "λευκός θόρυβος" και συνεπώς οι ACF και PACF των υπολοίπων δεν πρέπει να παρουσιάζουν στατιστικά σημαντικές αυτοσυσχετίσεις και μερικές αυτοσυσχετίσεις αντίστοιχα. Οι συντελεστές αυτοσυσχέτισης των υπολοίπων μπορούν να εξετασθούν συνολικά με τη χρήση του στατιστικού δείκτη  $Q^*$  (Ljung-Box). Εάν η τιμή του  $Q^*$  δεν είναι στατιστικά σημαντική τα υπόλοιπα μπορούν να θεωρηθούν μια σειρά "λευκού θορύβου".

Εάν τα υπόλοιπα δεν είναι λευκός θόρυβος τότε το μοντέλο είναι ανεπαρκές και πρέπει να εξετασθούν άλλα μοντέλα ARIMA. Το πρότυπο που ακολουθούν οι στατιστικά σημαντικοί συντελεστές αυτοσυσχέτισης και μερικής αυτοσυσχέτισης των υπολοίπων, υποδεικνύουν τον τρόπο βελτίωσης του μοντέλου. Για παράδειγμα σημαντικές τιμές για εποχιακές καθυστερήσεις υποδεικνύουν την προσθήκη μιας εποχιακής συνιστώσας ή σημαντικές τιμές για μικρές καθυστερήσεις υποδεικνύουν την αύξηση των μη εποχιακών AR ή MA συνιστωσών του μοντέλου. Συνήθως τα μοντέλα με τις μικρότερες AIC ή BIC τιμές έχουν υπόλοιπα λευκού θορύβου. Περιστασιακά είναι δυνατόν να υιοθετηθεί όχι το μοντέλο με την μικρότερη AIC ή BIC τιμή αλλά αυτό με τα "καλύτερα" υπόλοιπα. (Κουγιουμτζής, 2005)

### 3.5.4 Πρόβλεψη με τη χρήση μοντέλων ARIMA

Αφού επιλεγεί το καλύτερο μοντέλο, αυτό χρησιμοποιείται για την παραγωγή προβλέψεων. Για παράδειγμα, ας υποθεθεί ότι το μοντέλο που επιλέχθηκε είναι το ARIMA(0,1,1)(0,1,1)<sub>12</sub> με εξίσωση :

$$(1-B)(1-B^{12})Y_t = (1-\theta_1 B)(1-\theta_1 B^{12})e_t \quad (3.54)$$

Αφού η εξίσωση αναλυθεί προκύπτει :

$$Y_t = Y_{t-1} + Y_{t-12} - Y_{t-13} + e_t - \theta_1 e_{t-1} + \theta_1 e_{t-12} + \theta_1 e_{t-13} \quad (3.55)$$

και προκειμένου να γίνει πρόβλεψη για χρονικό ορίζοντα 1, η εξίσωση γράφεται :

$$Y_{t+1} = Y_t + Y_{t-11} - Y_{t-12} + e_{t+1} - \theta_1 e_t - \theta_1 e_{t-11} - \theta_1 e_{t-12} \quad (3.56)$$

Ο όρος  $e_{t+1}$  δεν είναι γνωστός αφού αντιπροσωπεύει το μελλοντικό τυχαίο σφάλμα και θέτεται ίσος με μηδέν. Από το μοντέλο που έχει προσαρμοσθεί στην χρονοσειρά υπάρχουν τιμές για τους όρους  $e_t$ ,  $e_{t-11}$  και  $e_{t-12}$  που είναι τα υπόλοιπα που αντιστοιχούν στις περιόδους  $t$ ,  $t-11$  και  $t-12$ . Αν η πρόβλεψη πραγματοποιείται για μεγάλο χρονικό ορίζοντα και δεν υπάρχουν από το προσαρμοσμένο μοντέλο διαθέσιμες τιμές των όρων των υπολοίπων, αυτοί θέτονται ίσοι με μηδέν. Τα διαστήματα εμπιστοσύνης απαιτούν πολύπλοκους υπολογισμούς και συνήθως υπολογίζονται με τη βοήθεια κάποιου στατιστικού υπολογιστικού πακέτου.

#### Διαφόριση και Πρόβλεψη

Η μέθοδος της διαφόρισης έχει μεγάλη επίδραση στις προβλέψεις. Μοντέλα με διαφόριση διαφορετικής τάξης μπορεί να προσαρμόζονται το ίδιο καλά στα ιστορικά δεδομένα αλλά είναι δυνατόν να παράγουν προβλέψεις που διαφέρουν σημαντικά.

Ακολούθως περιγράφονται συνοπτικά τα χαρακτηριστικά των προβλέψεων που παράγουν τα διάφορα μοντέλα ARIMA (p, d, q) (P, D, Q)<sub>s</sub>, καθώς αυξάνει ο ορίζοντας πρόβλεψης.

- ✓ d=0. Οι προβλέψεις θα συγκλίνουν στη μέση τιμή των ιστορικών δεδομένων εάν συμπεριλαμβάνεται σταθερός όρος στο μοντέλο, διαφορετικά θα συγκλίνουν στο μηδέν. Η διακύμανση των προβλέψεων θα συγκλίνει στη

διακύμανση των ιστορικών δεδομένων και συνεπώς τα διαστήματα εμπιστοσύνης επίσης θα συγκλίνουν.

- ✓  $d=1$ . Οι προβλέψεις θα συγκλίνουν στη τιμή της τελευταίας παρατήρησης εάν δεν συμπεριλαμβάνεται σταθερός όρος στο μοντέλο. Εάν συμπεριλαμβάνεται σταθερός όρος οι μακροπρόθεσμες προβλέψεις θα ακολουθούν μια γραμμική τάση, με κλίση ίση με την τιμή του σταθερού όρου. Και στις δύο περιπτώσεις η διακύμανση των προβλέψεων θα αυξάνει με το χρονικό ορίζοντα και τα διαστήματα εμπιστοσύνης θα αποκλίνουν.
- ✓  $d=2$ . Εάν δεν συμπεριλαμβάνεται σταθερός όρος στο μοντέλο, οι προβλέψεις θα ακολουθούν μια γραμμική τάση, προεκτείνοντας την τάση που ακολουθούν οι τελευταίες παρατηρήσεις της χρονοσειράς. Εάν συμπεριλαμβάνεται σταθερός όρος οι μακροπρόθεσμες προβλέψεις θα ακολουθούν μια τετραγωνική τάση. Και στις δύο περιπτώσεις η διακύμανση των προβλέψεων θα αποκλίνει και τα διαστήματα εμπιστοσύνης επίσης θα αποκλίνουν.
- ✓ Εποχιακή διαφόριση. Οι επιδράσεις της εποχιακής διαφόρισης είναι παρόμοιες. Ειδικότερα, εάν έχει εφαρμοσθεί εποχιακή διαφόριση και στη συνέχεια διαφόριση πρώτης τάξης, οι προβλέψεις θα συμπεριφέρονται όμοια με αυτές του μοντέλου διαφόρισης δεύτερης τάξης ( $d=2$ ), με τη διαφορά ότι θα παρουσιάζουν εποχικότητα (Κουγιουμτζής, 2005).

## 3.6 Ανάλυση Νευρωνικού Δικτύου

### 3.6.1 Μονοστρωματικά Νευρωνικά Δίκτυα εμπρόσθιας τροφοδότησης

*Το πολυεπίπεδο perceptron – MLP*

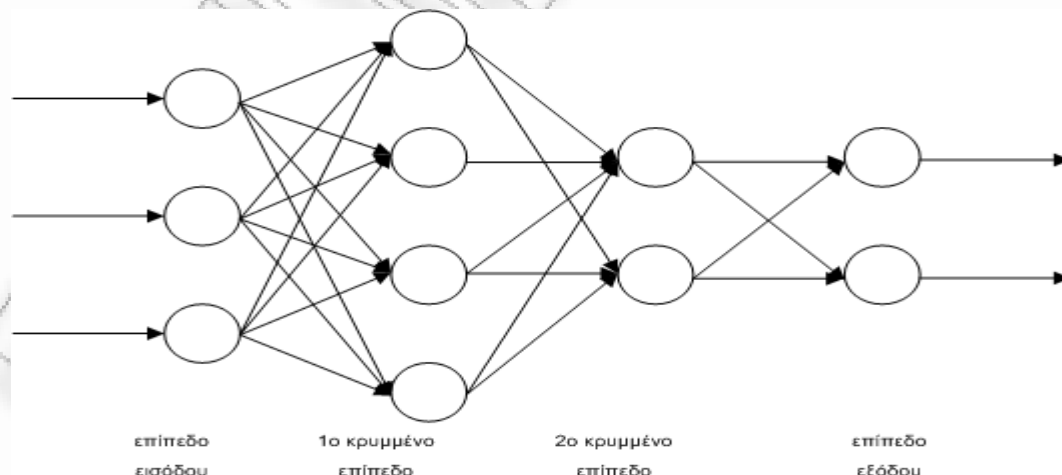
Από τα δίκτυα πρόσθιας τροφοδότησης, παρουσιάζει ενδιαφέρον η κατηγορία στην οποία οι νευρώνες είναι οργανωμένοι σε επίπεδα ή στρώματα (layers). Στην κατηγορία αυτή ανήκει και το πολυεπίπεδο perceptron (Multilayer Perceptron, MLP), μια αρχιτεκτονική του οποίου παρουσιάζεται στο Σχήμα 3.4.

Το χαρακτηριστικό των επιπέδων είναι ότι δεν υπάρχει διασύνδεση μεταξύ νευρώνων που ανήκουν στο ίδιο επίπεδο. Υπάρχει το επίπεδο εισόδου, το επίπεδο εξόδου και ένα ή και περισσότερα ενδιάμεσα επίπεδα, που ονομάζονται κρυμμένα (hidden). Συνήθως όλες οι συνδέσεις ξεκινάνε από κάποιο νευρώνα σε ένα επίπεδο και καταλήγουν σε νευρώνες του αμέσως επόμενου επιπέδου. Σημειώνουμε ότι οι νευρώνες στο επίπεδο εισόδου δεν πραγματοποιούν κάποιον υπολογισμό, αλλά απλά μεταδίδουν τις τιμές του διανύσματος εισόδου στους νευρώνες του επόμενου επιπέδου. (Λύκας,2008)

### Εκπαίδευση του MLP

Μια μέθοδος εκπαίδευσης του MLP ονομάζεται οπισθοδιάδοση σφάλματος (error backpropagation) και παρουσιάστηκε αρχικά το 1986 σε μια εργασία των Rumelhart, Hinton και Williams. (Rumelhart, Hinton and Williams, 1986)

Κατά τη μέθοδο αυτή παρουσιάζονται στο δίκτυο ένα σύνολο παραδειγμάτων εκπαίδευσης  $D = \{(x^n, t^n)\}, n = 1, 2, \dots, N$ , όπου  $x^n = (x_{n1}, \dots, x_{nd})^T \in \mathcal{R}^d$  η είσοδος του δικτύου και  $t^n = (t_{n1}, \dots, t_{nd})^T \in \mathcal{R}^p$  η επιθυμητή έξοδος. Το δίκτυο, με βάση τα διανύσματα εισόδου, υπολογίζει την έξοδό του  $o(x^n; w)$ , όπου  $x^n$  είναι το διάνυσμα εισόδου και  $w = (w_1, \dots, w_L)^T$  το διάνυσμα στο οποίο συγκεντρώνουμε όλα τα βάρη και τις πολώσεις του δικτύου. (Λύκας,2008)



Σχήμα 3.4: Πολυεπίπεδο perceptron με δύο κρυμμένα επίπεδα

Μπορούμε συνεπώς να ορίσουμε την τετραγωνική συνάρτηση σφάλματος ως (Λύκας,2008)

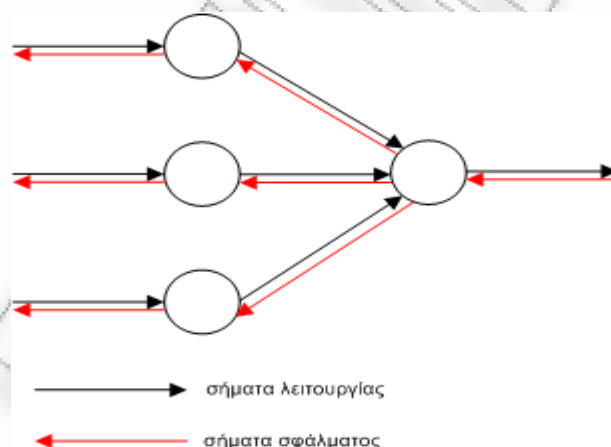
$$E(w) = \frac{1}{2} \sum_{n=1}^N \|t^n - o(x^n; w)\|^2 = \frac{1}{2} \sum_{n=1}^N \sum_{m=1}^p (t_{nm} - o_m(x^n; w))^2 \quad (3.57)$$

η οποία ως άθροισμα τετραγώνων των σφαλμάτων ανά παράδειγμα  $(\mathbf{x}_n, \mathbf{t}_n)$  έχει κάτω φράγμα την τιμή μηδέν, η οποία προκύπτει στην περίπτωση της τέλει εκπαίδευσης. Με  $E_n(w)$  συμβολίζουμε το τετραγωνικό σφάλμα ανά παράδειγμα  $(\mathbf{x}_n, \mathbf{t}_n)$ , οπότε ισχύει

$$E(w) = \sum_{n=1}^N E_n(w), E_n(w) = \frac{1}{2} \|t^n - o(x^n; w)\|^2 = \frac{1}{2} \sum_{m=1}^p (t_{nm} - o_m(x^n; w))^2 \quad (3.58)$$

Η εκπαίδευση, λοιπόν του MLP υλοποιείται μέσω της ενημέρωσης του διανύσματος των βαρών  $w$  με σκοπό την ελαχιστοποίηση του τετραγωνικού σφάλματος  $E(w)$  και αποτελεί εφαρμογή της μεθόδου καθόδου με βάση την κλίση (gradient descent), κατά την οποία ξεκινώντας από μια αρχική τιμή των βαρών  $w(0)$ , υπολογίζεται σε κάθε επανάληψη  $k$  η μεταβολή των βαρών  $\Delta w(k)$  κατά τέτοιο τρόπο, ώστε να μετακινούμαστε με μικρά βήματα στην κατεύθυνση κατά την οποία η συνάρτηση  $E(w)$  εμφανίζει το μέγιστο ρυθμό μείωσης. (Λύκας,2008)

Η εκπαίδευση πραγματοποιείται σε δύο περάσματα: ένα πέρασμα κατά την ευθεία φορά (forward propagation) και ένα κατά την αντίθετη φορά (back propagation), όπως φαίνεται στο Σχήμα 3.5. (Haykin,2009)



**Σχήμα 3.5:** Η κατεύθυνση ροής των δύο σημάτων κατά την εκπαίδευση του MLP

Αρχικά, κατά το ευθύ πέρασμα, υπολογίζονται και αποθηκεύονται οι τιμές των εξόδων των νευρώνων όλων των επιπέδων, οι οποίες ονομάζονται και σήματα λειτουργίας (function signals), με φορά από τις εισόδους προς τις εξόδους και στο τέλος παράγεται ένα σύνολο από εξόδους  $o_i, i = 1, 2, \dots, p$ , που αποτελούν την έξοδο του δικτύου. Στη συνέχεια η τιμή αυτή συγκρίνεται με την αντίστοιχη επιθυμητή  $t_i, i = 1, 2, \dots, p$  και παράγεται το σήμα σφάλματος (error signal) για τους νευρώνες εξόδου. Αυτό το σήμα στη συνέχεια μεταδίδεται κατά την αντίθετη φορά και υπολογίζονται τα επιμέρους σήματα σφάλματος που αντιστοιχούν στους κρυμμένους νευρώνες. (Λύκας, 2008)

Προκειμένου να περιγράψουμε τους υπολογισμούς που λαμβάνουν χώρα κατά την εκπαίδευση του MLP, θα χρειαστεί αρχικά να καθορίσουμε το συμβολισμό των διαφόρων μεγεθών. Ποιο συγκεκριμένα (Λύκας, 2008), θα χρησιμοποιήσουμε το συμβολισμό  $i^l$  για να αναφερόμαστε στο νευρώνα  $i$  του  $l$  επιπέδου. Κατά συνέπεια ορίζουμε:

- $u_i^{(l)}$  τη συνολική είσοδο του νευρώνα  $i^l$
- $y_i^{(l)}$  την έξοδο του νευρώνα  $i^l$
- $\delta_i^{(l)}$  το σφάλμα του νευρώνα  $i^l$
- $w_{i0}^{(l)}$  την πόλωση του νευρώνα  $i^l$
- $g_l$  τη συνάρτηση ενεργοποίησης των νευρώνων του επιπέδου  $l$
- $d_l$  τον αριθμό των νευρώνων του επιπέδου  $l$
- $w_{ij}^{(l)}$  το βάρος της σύνδεσης από το νευρώνα  $j^{l-1}$  στο νευρώνα  $i^l$

Τα επίπεδα αριθμούνται από την είσοδο προς την έξοδο και για  $l = 0$  έχουμε το επίπεδο εισόδου. Αν  $x^n = (x_1, \dots, x_d)^T$  είναι το διάνυσμα εισόδου, για τους νευρώνες στο επίπεδο εισόδου έχουμε  $y_i^{(0)} = x_i, x_0 = 1$ . Αν  $q$  είναι το επίπεδο εξόδου, τότε η έξοδος του MLP είναι  $o_i(x; w) = y_i^{(q)}$

Έστω λοιπόν ένα MLP με  $d$  εισόδους,  $p$  εξόδους και  $H$  κρυμμένα επίπεδα. Το επίπεδο εισόδου είναι το 0 και ισχύει ότι  $d_0 = d$ . Το επίπεδο εξόδου είναι το  $H + 1$

και ισχύει ότι  $d_{h+1} = p$ .

Με βάση τα παραπάνω, ο υπολογισμός του διανύσματος εξόδου  $o(x^n; w) = (o_1, \dots, o_p)^T$  για είσοδο  $x^n = (x_1, \dots, x_d)^T$  γίνεται ως εξής: (Λύκας, 2008)

### Ευθύ πέρασμα

✓ Επίπεδο εισόδου:  $y_i^{(0)} = x_i, y_0^{(0)} = x_0 = 1$  (3.59)

✓ Κρυμμένα επίπεδα και επίπεδο εξόδου:

$$u_i^{(h)} = \sum_{j=0}^{d_{h-1}} w_{ij}^{(h)} y_j^{(h-1)}, h = 1, \dots, H + 1, i = 1, \dots, d_h \quad (3.60)$$

$$y_i^{(h)} = g_h(u_i^{(h)}), h = 1, \dots, H + 1, i = 1, \dots, d_h, y_0^{(h)} = 1 \quad (3.61)$$

✓ Έξοδος του δικτύου:  $o_i = y_i^{(H+1)}, i = 1, \dots, p$  (3.62)

Αφού έχουμε υπολογίσει την έξοδο του δικτύου, στη συνέχεια εφαρμόζεται η οπισθοδιάδοση του σφάλματος ως εξής: (Λύκας, 2008)

### Οπισθοδιάδοση σφαλμάτων

✓ Υπολογισμός των σφαλμάτων στο επίπεδο εξόδου:

$$\delta_i^{H+1} = g'_{H+1}(u_i^{H+1}) \frac{\partial E^n}{\partial o_i}, i = 1, \dots, p \quad (3.63)$$

✓ Διάδοση των σφαλμάτων προς τα πίσω στο δίκτυο μέχρι τους νευρώνες του πρώτου κρυμμένου επιπέδου:

$$\delta_i^{(h)} = g'_h(u_i^h) \sum_{j=1}^{d_{h+1}} w_{ij}^{(h+1)} \delta_j^{(h+1)}, h = H, \dots, 1, i = 1, \dots, d_h \quad (3.64)$$

✓ Υπολογισμός των μερικών παραγώγων του σφάλματος  $E^n$  ως προς τα βάρη:

$$\frac{\partial E^n}{\partial w_{ij}^{(h)}} = \delta_i^{(h)} y_j^{(h-1)} \quad (3.65)$$

Στην περίπτωση που το  $E^n$  αντιστοιχεί στο τετραγωνικό σφάλμα ανά παράδειγμα σύμφωνα με τη σχέση (3.58) η σχέση (3.63) γίνεται

$$\delta_i^{H+1} = -g'_{H+1}(u_i^{H+1})(t_{ni} - o_i), i = 1, \dots, p \quad (3.66)$$

Στην περίπτωση που οι νευρώνες στα κρυμμένα επίπεδα έχουν λογιστική συνάρτηση ενεργοποίησης η σχέση (3.64) γίνεται

$$\delta_i^{(h)} = y_i^{(h)}(1 - y_i^{(h)}) \sum_{j=1}^{d_{h+1}} w_{ij}^{(h+1)} \delta_j^{(h+1)} \quad (3.67)$$

και δεν απαιτείται η αποθήκευση των  $u_i$  κατά το ευθύ πέρασμα. Το ίδιο ισχύει και για τους νευρώνες του επιπέδου εξόδου. Έτσι, αν η συνάρτηση ενεργοποίησής τους είναι η λογιστική, τότε:

$$\delta_i^{H+1} = -o_i(1 - o_i)(t_{ni} - o_i) \quad (3.68)$$

ενώ αν είναι η γραμμική

$$\delta_i^{H+1} = -(t_{ni} - o_i) \quad (3.69)$$

Έτσι, μπορούμε να περιγράψουμε στη συνέχεια τη μέθοδο εκπαίδευσης του MLP με ελαχιστοποίηση σφάλματος εκπαίδευσης με σειριακή ενημέρωση των βαρών: (Λύκας, 2008)

### Αλγόριθμος εκπαίδευσης του MLP με ελαχιστοποίηση σφάλματος εκπαίδευσης

1. Αρχικοποίηση των βαρών  $\mathbf{w}(0)$  τυχαία στο διάστημα  $(-1,1)$  και καθορισμός του ρυθμού μάθησης  $\eta$ . Αρχικοποίηση του μετρητή επαναλήψεων ( $k=0$ ) και του μετρητή εποχών ( $k=0$ )
2. Στην αρχή κάθε εποχής, έστω  $\mathbf{w}(k)$  το διάνυσμα βαρών του MLP.
  - i. Έναρξη εποχής  $k$ . Αποθήκευση του τρέχοντος διανύσματος βαρών  $\mathbf{w}_{old} = \mathbf{w}(k)$ .
  - ii. Για  $n=1, \dots, N$



a) Επιλογή του παραδείγματος εκπαίδευσης ( $\mathbf{x}^n, \mathbf{t}^n$ ) και εφαρμογή του κανόνα οπισθοδιάδοσης σφάλματος για τον υπολογισμό των μερικών παραγώγων  $\frac{\partial E^n}{\partial w_i}$

b) Ενημέρωση των βαρών  $w_i(k+1) = w_i(k) - n \frac{\partial E^n}{\partial w_i}$  (3.70)

c)  $k = k + 1$

iii. Τέλος εποχής  $k$ . Έλεγχος τερματισμού. Εάν ναι, τερματίζουμε.

iv.  $k = k + 1$ , μετάβαση στο βήμα 2.

Στη σχέση (3.70) μπορούμε να προσθέσουμε μια επιπλέον ποσότητα όπως φαίνεται στην παρακάτω εξίσωση (Λύκας,2008)

$$w_i(k+1) = w_i(k) + a\Delta w_i(k-1) - n \frac{\partial E^n}{\partial w_i} \quad (3.71)$$

Η ποσότητα αυτή ονομάζεται όρος ορμής (momentum term) και η σταθερά  $a$  ονομάζεται σταθερά ορμής. Ο όρος ορμής εισάγει μια μορφή μνήμης στον αλγόριθμο εκπαίδευσης σχετικά με τις μεταβολές  $\Delta w_i$  που πραγματοποιήθηκαν στα προηγούμενα βήματα του αλγορίθμου και έτσι επιτυγχάνεται ένα σταθεροποιητικό αποτέλεσμα και αποφεύγεται το φαινόμενο των ταλαντώσεων γύρω από την τιμή του ελαχίστου σφάλματος. (Λύκας,2008)

## ΚΕΦΑΛΑΙΟ 4

### Αποτελέσματα

#### 4.1 Εισαγωγή

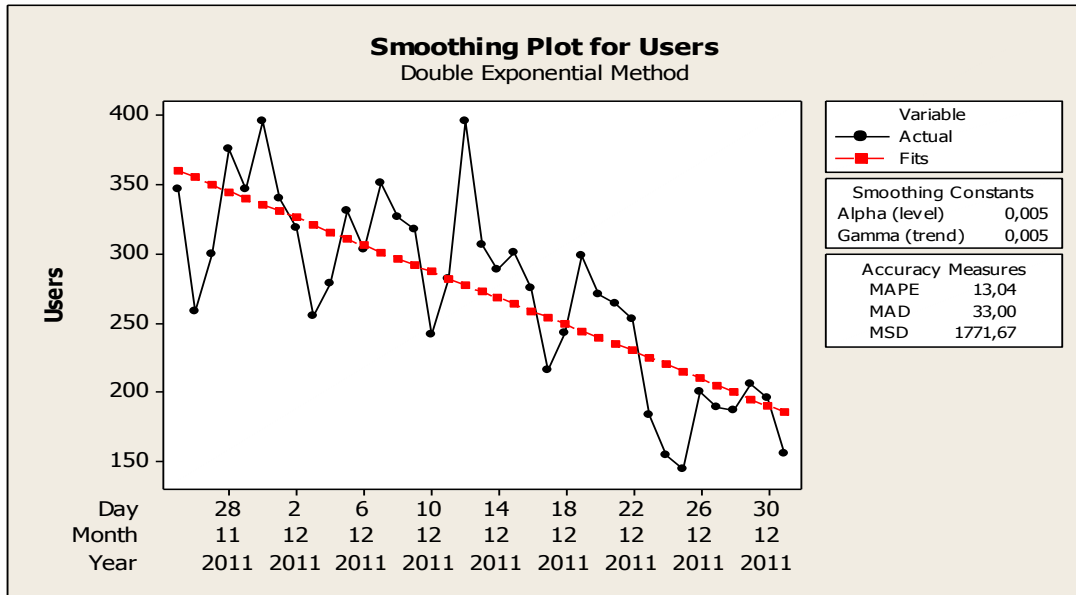
Στο Κεφάλαιο αυτό παρουσιάζονται αναλυτικά, χρησιμοποιώντας πίνακες και διαγράμματα, τα αποτελέσματα που προέκυψαν από την εφαρμογή τόσο των διαφόρων οικονομετρικών μοντέλων χρονοσειρών όσο και του νευρωνικού δικτύου, στην πρόβλεψη της επισκεψιμότητας ενός εκπαιδευτικού portal.

Για την εξαγωγή των πινάκων και των διαγραμμάτων χρησιμοποιήθηκαν τα λογισμικά πακέτα στατιστικής ανάλυσης Minitab και SPSS. Για την εκπαίδευση του νευρωνικού δικτύου χρησιμοποιήθηκε το λογισμικό WEKA.

#### 4.2 Αποτελέσματα μεθόδων εξομάλυνσης

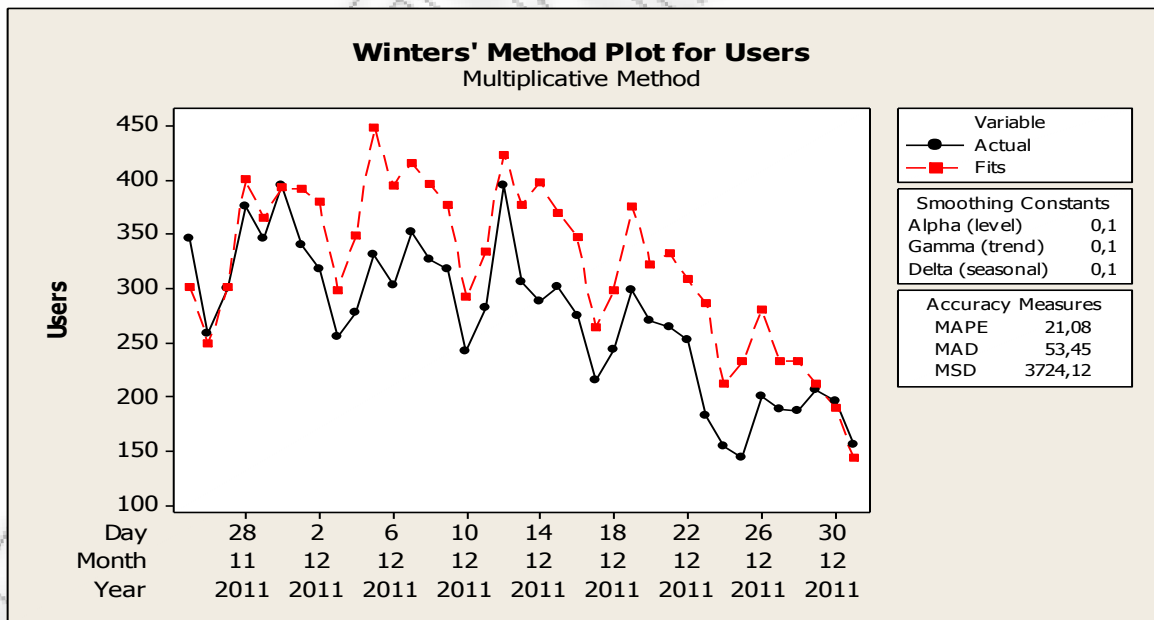
Τα μοντέλα που περιγράψαμε στην παράγραφο των μεθόδων εξομάλυνσης παραπάνω αφορούν μία βασική θεωρία ανάλυσης χρονοσειρών, με την οποία δεν θα ασχοληθούμε γιατί η χρονοσειρά μας είναι πιο περίπλοκη οπότε θα χρειαστούμε μια πολυπλοκότερη ανάλυση. Θα παραθέσουμε τα αποτελέσματα κάποιων μεθόδων απ' αυτές με τη βοήθεια του λογισμικού πακέτου minitab.

Αρχικά έχουμε τα αποτελέσματα της μεθόδου της *διπλής εκθετικής εξομάλυνσης*. Τα  $a$  και  $b$  δόθηκαν μετά από διάφορες δοκιμές.

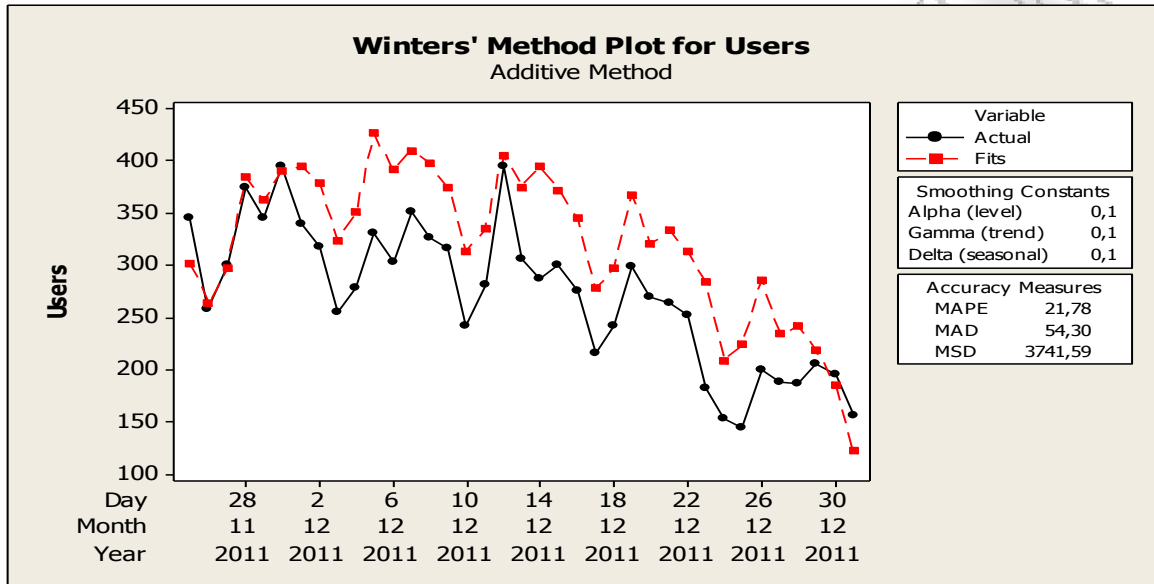


**Σχήμα 4.1:** Διπλή εκθετική εξομάλυνση χρονοσειράς με  $\alpha=0,005$  και  $\beta=0,005$  (TestSet)

Στη συνέχεια δίνονται τα αποτελέσματα των μεθόδων Winters. Στη μέθοδο αυτή ορίζουμε ως παράμετρο την εποχικότητα. Ορίζουμε την εποχικότητα σε 7 (εβδομαδιαία).



**Σχήμα 4.2:** Μέθοδος Winters (πολ/κο μοντέλο) στη χρονοσειρά με  $\alpha=0,1$ ,  $\beta=0,1$  και  $\gamma=0,1$  (TestSet)



**Σχήμα 4.3:** Μέθοδος Winters(προσθετικό μοντέλο) στη χρονοσειρά με  $\alpha=0,1$  ,  $\beta=0,1$  και  $\gamma=0,1$  (TestSet)

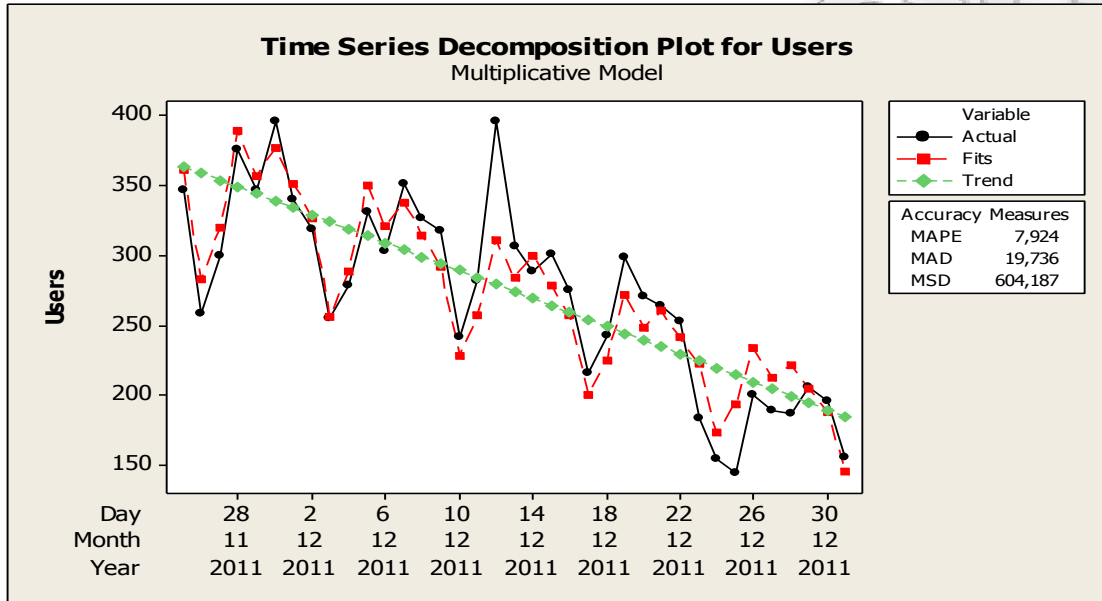
Ένας συνοπτικός πίνακας με τα αποτελέσματα των παραπάνω μεθόδων εξομάλυνσης δίνεται παρακάτω, έτσι ώστε να μπορούμε να τα συγκρίνουμε πιο εύκολα.

**Πίνακας 4.1:** Σύνοψη αποτελεσμάτων των μεθόδων εξομάλυνσης

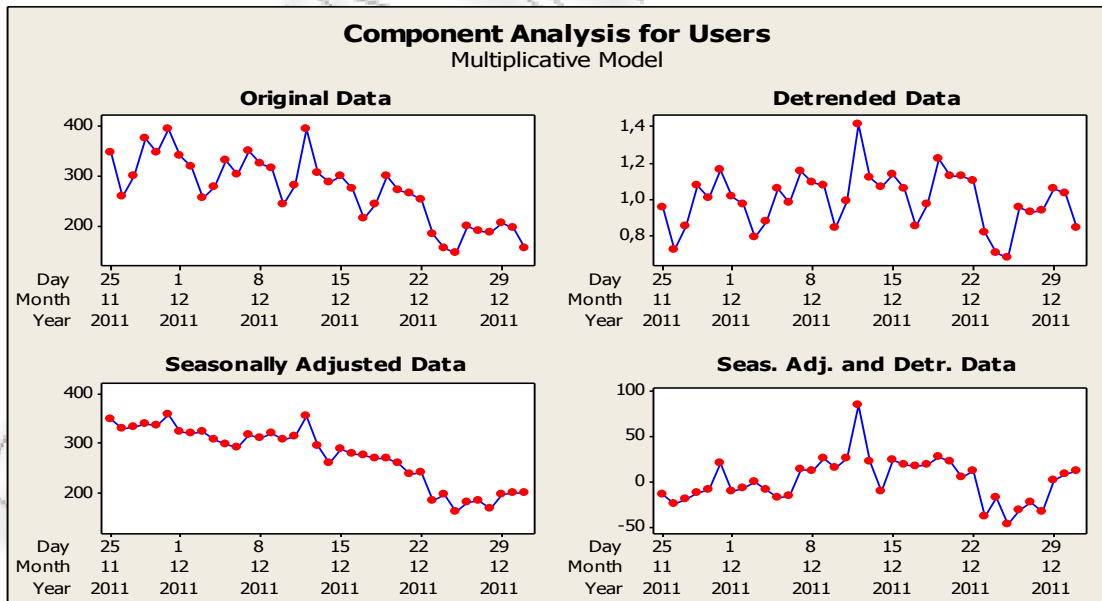
	Διπλή Εκθετική Εξομάλυνση		Winters (Multiplicative Method)		Winters (Additive Method)	
	Training Set	Test Set	Training Set	Test Set	Training Set	Test Set
<b>MAPE</b>	28.8	13.04	15.72	21.08	17.19	21.78
<b>MAD</b>	106.1	33.00	60.54	53.45	64.53	54.3
<b>RMSD</b>	153.67	42.09	90.81	61.03	94.82	61.17

### 4.3 Αποτελέσματα διάσπασης χρονοσειρών

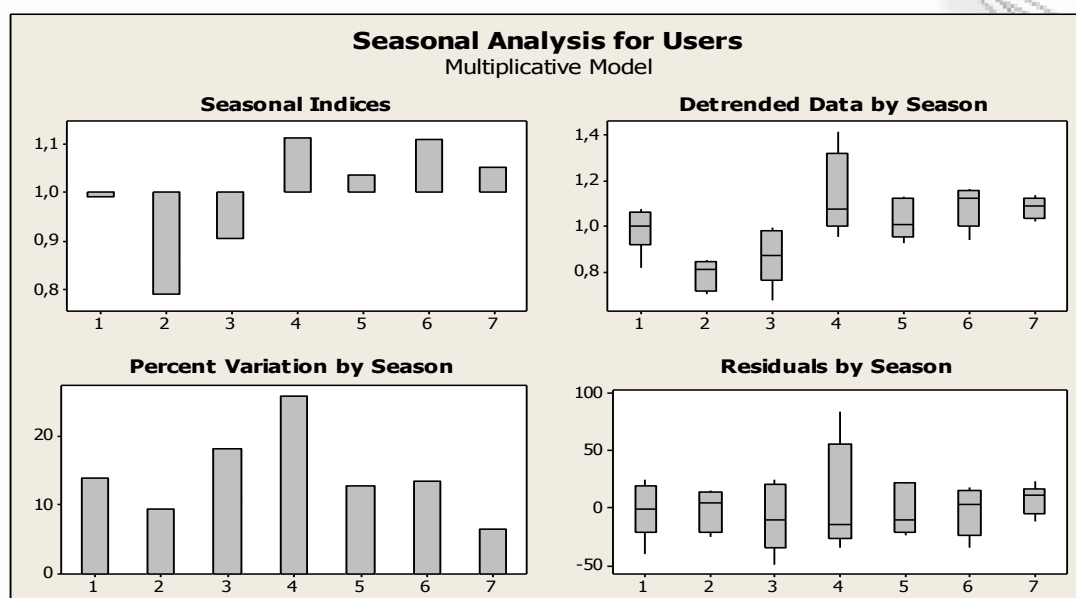
Χρησιμοποιώντας το στατιστικό λογισμικό Minitab θα παρουσιάσουμε συνοπτικά τη διάσπαση της χρονοσειράς μας στις συνιστώσες της με βάση το πολλαπλασιαστικό μοντέλο.



Σχήμα 4.4: Μέθοδος Διάσπασης Χρονοσειράς (πολλαπλασιαστικό μοντέλο)



Σχήμα 4.5: Συνιστώσες Χρονοσειράς (πολλαπλασιαστικό μοντέλο)



**Σχήμα 4.6:** Ανάλυση εποχικότητας Χρονοσειράς (πολλαπλασιαστικό μοντέλο)

Η σύνοψη των αποτελεσμάτων αυτών φαίνεται παρακάτω, τόσο για το training test όσο και για το test set. Τα διαγράμματα σε σχέση με τα αποτελέσματα των σφαλμάτων θα μας δείξουν την καταλληλότητα ή μη της μεθόδου.

**Πίνακας 4.2:** Σύνοψη αποτελεσμάτων μεθόδου διάσπασης

	Διάσπαση Χρονοσειράς	
	Training Set	Test Set
<b>MAPE</b>	27.2	7.924
<b>MAD</b>	97.7	19.736
<b>RMSD</b>	145.06	24.78

#### 4.4 Αποτελέσματα μοντέλου ARIMA

Εφαρμόζοντας τη μέθοδο Box-Jenkins έχουμε τα ακόλουθα αποτελέσματα με τη βοήθεια του λογισμικού SPSS. Το λογισμικό αυτό μας δίνει τη δυνατότητα να παραμετροποιήσουμε το μοντέλο ARIMA, το οποίο μπορεί να δώσει τις καλύτερες προβλέψεις. Η αξιολόγηση των διάφορων μοντέλων, από την αυτοματοποιημένη διαδικασία του λογισμικού, επιλέχθηκε να γίνεται με βάση το κριτήριο που έχει ήδη αναφερθεί στην υποενότητα 3.6.2 (Bayesian Information Criterion).

Το λογισμικό που χρησιμοποιήθηκε υπέδειξε το μοντέλο ARIMA(1,1,1)(1,1,1)7 ως το πλέον κατάλληλο για την ανάλυση και πρόβλεψη της συγκεκριμένης χρονοσειράς. Παρακάτω παρατίθενται πίνακες με τα αποτελέσματα του μοντέλου.

**Πίνακας 4.3:** Πίνακας τιμών σφαλμάτων μοντέλου ARIMA(1,1,1)(1,1,1)

Model Fit											
Fit Statistic	Mean	SE	Minimum	Maximum	Percentile						
					5	10	25	50	75	90	95
Stationary R-squared	,406	.	,406	,406	,406	,406	,406	,406	,406	,406	,406
R-squared	,812	.	,812	,812	,812	,812	,812	,812	,812	,812	,812
RMSE	71,115	.	71,115	71,115	71,115	71,115	71,115	71,115	71,115	71,115	71,115
MAPE	10,825	.	10,825	10,825	10,825	10,825	10,825	10,825	10,825	10,825	10,825
MaxAPE	76,822	.	76,822	76,822	76,822	76,822	76,822	76,822	76,822	76,822	76,822
MAE	43,354	.	43,354	43,354	43,354	43,354	43,354	43,354	43,354	43,354	43,354
MaxAE	480,838	.	480,838	480,838	480,838	480,838	480,838	480,838	480,838	480,838	480,838
Normalized BIC	8,565	.	8,565	8,565	8,565	8,565	8,565	8,565	8,565	8,565	8,565

**Πίνακας 4.4:** Πίνακας στατιστικών αποτελεσμάτων μοντέλου ARIMA(1,1,1)(1,1,1)

Model Statistics									
Model	Number of Predictors	Model Fit statistics				Ljung-Box Q(18)			Number of Outliers
		RMSE	MAPE	MAE	Normalized BIC	Statistics	DF	Sig.	
Users-Model_1	0	71,115	10,825	43,354	8,565	14,493	14	,414	0

**Πίνακας 4.5:** Πίνακας παραμέτρων μοντέλου ARIMA(1,1,1)(1,1,1)

ARIMA Model Parameters									
					Estimate	SE	t	Sig.	
Users-Model_1	Users	Natural Log	AR	Lag 1	,571	,052	11,065	,000	
			Difference			1			
			MA	Lag 1	,880	,031	28,250	,000	
			AR, Seasonal	Lag 1	,148	,050	2,984	,003	
			Seasonal Difference			1			
			MA, Seasonal	Lag 1	,836	,027	30,450	,000	

**Πίνακας 4.6:** Πίνακας υπολοίπων αυτοσυσχέτισης

Residual ACF									
Model		1	2	3	4	5	6	7	
Users-Model_1	ACF	-,010	,006	,024	-,002	,025	-,003	,008	
	SE	,037	,037	,037	,037	,037	,037	,037	

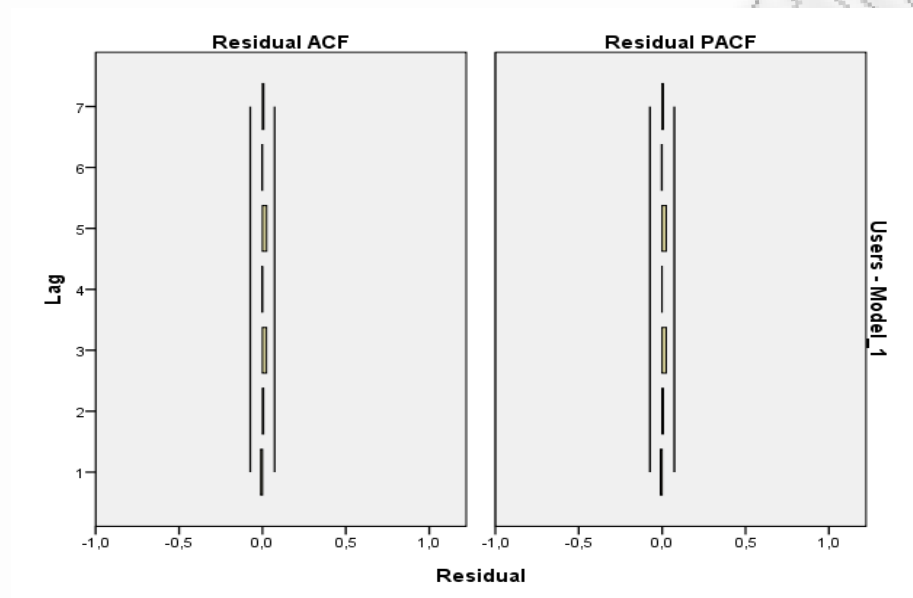
**Πίνακας 4.7:** Πίνακας υπολοίπων μερικής αυτοσυσχέτισης

Residual PACF									
Model		1	2	3	4	5	6	7	
Users-Model_1	PACF	-,010	,006	,024	-,001	,025	-,003	,007	
	SE	,037	,037	,037	,037	,037	,037	,037	

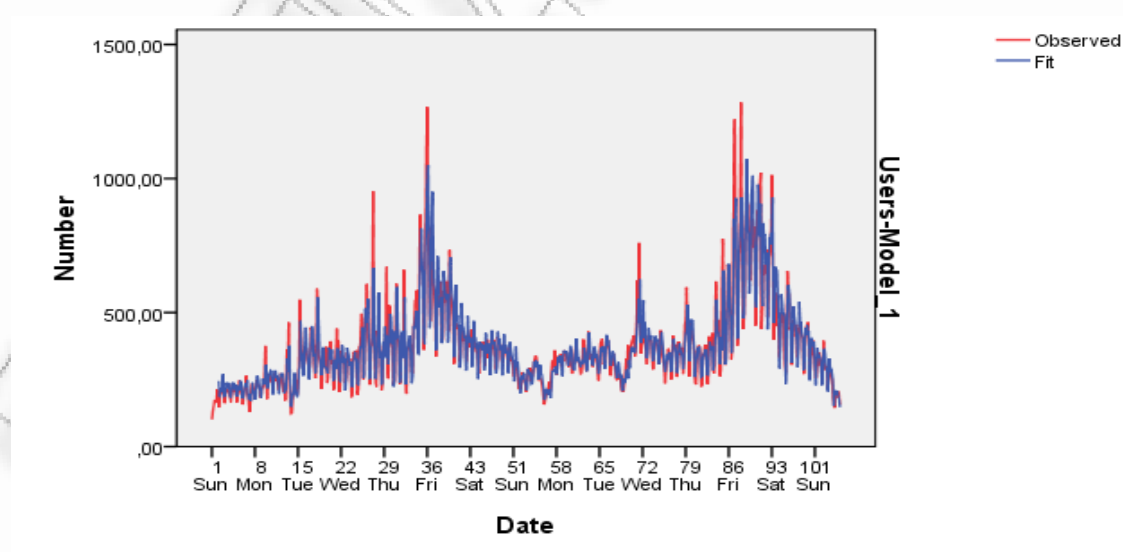


Όπως έχουμε ήδη αναφέρει η παρατήρηση των διαγραμμάτων παίζει σημαντικό ρόλο, καθώς υποδεικνύουν τους παράγοντες που προσδιορίζουν την καταλληλότητα ή μη ενός μοντέλου.

Παρακάτω δίνονται τα διαγράμματα Residual ACF και Residual PACF καθώς επίσης και του μοντέλου ARIMA που τελικά επιλέχθηκε με τις παρατηρούμενες τιμές (observed) και τις πραγματικές (fit) σε σχέση με το χρόνο.



**Σχήμα 4.7:** Διαγράμματα υπολοίπων αυτοσυσχέτισης (ACF) και μερικής αυτοσυσχέτισης (PACF)



**Σχήμα 4.8:** Μοντέλο ARIMA(1,1,1)(1,1,1)7

## 4.5 Αποτελέσματα Νευρωνικού δικτύου – MLP

Η αρχιτεκτονική του δικτύου MLP πραγματοποιήθηκε στο WEKA από την κλάση `weka.classifiers.function.MultilayerPerceptron` και από κλάσεις του πακέτου `weka.classifiers.function.neural` (Dieter Merkl). Το WEKA περιλαμβάνει δύο τύπους συναρτήσεων ενεργοποίησης, τη γραμμική και τη λογιστική (σιγμοειδή). Για τους νευρώνες εξόδου, που στην περίπτωση μας είναι αριθμητικοί (numeric) χρησιμοποιήθηκε η γραμμική συνάρτηση, ενώ για τους νευρώνες των κρυφών επιπέδων χρησιμοποιήθηκε η σιγμοειδής συνάρτηση. Η εκπαίδευση του δικτύου έγινε με την μέθοδο οπισθοδιάδοσης σφάλματος (backpropagation).

Οι παράμετροι του αλγορίθμου ρυθμίστηκαν σύμφωνα με τον παρακάτω πίνακα.

**Πίνακας 4.8:** Παράμετροι αλγορίθμου εκπαίδευσης νευρωνικού δικτύου

Παράμετρος	Τιμή	Σχόλια
HiddenLayers	a	Ο αριθμός των κρυφών νευρώνων ορίζεται αυτόματα και ίσο με το ημι-άθροισμα εισόδων και κλάσεων.
learningRate	0,3	Από προεπιλογή
Momentum	0,2	Από προεπιλογή
Seed	0	Από προεπιλογή
TrainingTime		Θα μπουν διαδοχικά οι τιμές 10,20,30,50 και 100.
validationSiteSize	0	Από προεπιλογή
validationThreshold	20	Από προεπιλογή

Ορίζουμε τα 36 τελευταία στοιχεία του δείγματός μας ως test data. Η αξιολόγηση θα γίνει βάσει των δεικτών που φαίνονται στον Πίνακα 4.8. Η πρόβλεψη έγινε για την επόμενη εβδομάδα.

Τρέξαμε τον αλγόριθμο 5 φορές για αριθμό κύκλων εκπαίδευσης ίσο με 10,20,30,50 και 100. Τα αποτελέσματα φαίνονται παρακάτω:

**Πίνακας 4.9:** Αποτελέσματα εκπαίδευσης νευρωνικού δικτύου

K	N	Mean absolute error	Root mean squared error	Mean absolute percentage error	Direction accuracy	Mean squared error
10	36	32.3697	40.6984	14.4374	77.1429	1656.3624
20	36	32.8787	40.8973	14.7375	80	1672.5909
30	36	32.3023	39.8611	14.3443	77.1429	1588.9064
50	36	32.5712	40.331	14.5313	74.2857	1626.5926
100	36	30.7345	38.8788	13.7742	77.1429	1511.5584

Επειδή το φαινόμενο της υπερεκπαίδευσης αρχίζει να εμφανίζεται για  $k=20$ , εξετάσαμε τιμές του  $k$  γύρω απ'αυτή την τιμή, για την εύρεση του καλύτερου αριθμού χρόνου εκπαίδευσης. Οπότε προκύπτει ο παρακάτω πίνακας:

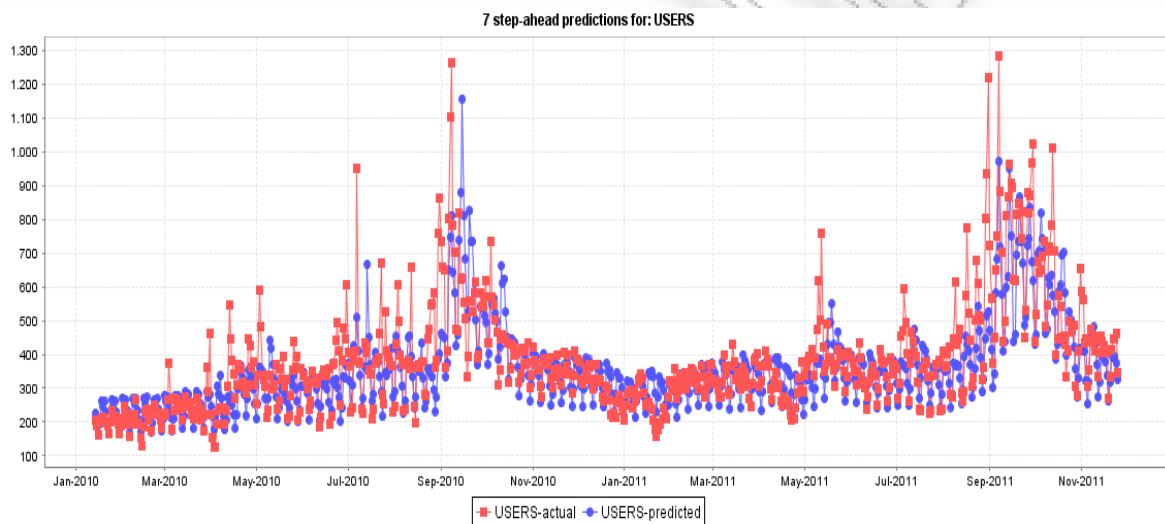
**Πίνακας 4.10:** Αποτελέσματα εκπαίδευσης νευρωνικού δικτύου

K	N	Mean absolute error	Root mean squared error	Mean absolute percentage error	Direction accuracy	Mean squared error
15	36	33.4315	41.5942	14.9655	77.1429	1730.075
16	36	33.3834	41.5425	14.9532	80	1725.7828
17	36	33.2599	41.4328	14.9064	80	1716.68
18	36	33.1253	41.2485	14.8516	80	1703.9108
19	36	32.9831	41.0948	14.7893	80	1688.7791
20	36	32.8787	40.8973	14.7375	80	1672.5909
21	36	32.7649	40.7001	14.6787	77.1429	1656.4981
22	36	32.6608	40.5142	14.62	77.1429	1641.4005

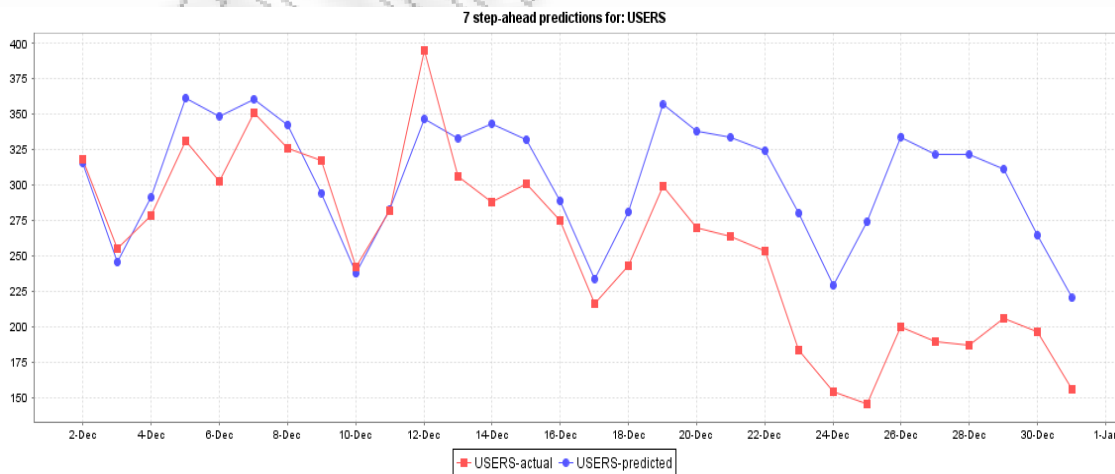
23	36	32.5808	40.3475	14.5669	77.1429	1627.92
----	----	---------	---------	---------	---------	---------

Θεωρούμε πως το κριτήριο επιλογής του βέλτιστου  $k$  αποτελεί το ποσοστό επιτυχίας (Direction Accuracy). Με βάση την παραδοχή αυτή, παρατηρούμε ότι ο βέλτιστος αριθμός των κύκλων εκπαίδευσης  $k$  είναι 16 αφού έχει το μεγαλύτερο ποσοστό επιτυχίας.

Επομένως επιλέγουμε και τρέχουμε το νευρωνικό για  $k=16$ . Στα σχήματα 4.9 και 4.10 φαίνονται οι τιμές πρόβλεψης σε σύγκριση με τις πραγματικές τιμές τόσο στο training set όσο και στο test set.

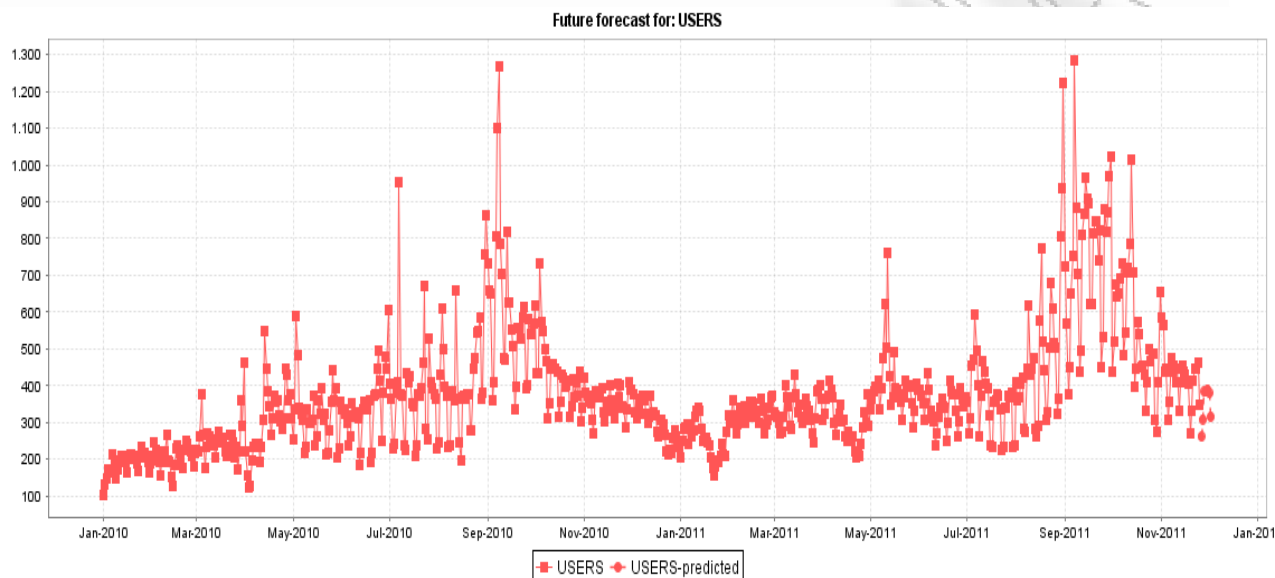


Σχήμα 4.9: Διάγραμμα πρόβλεψης επισκεπτών εβδομάδας (training set)

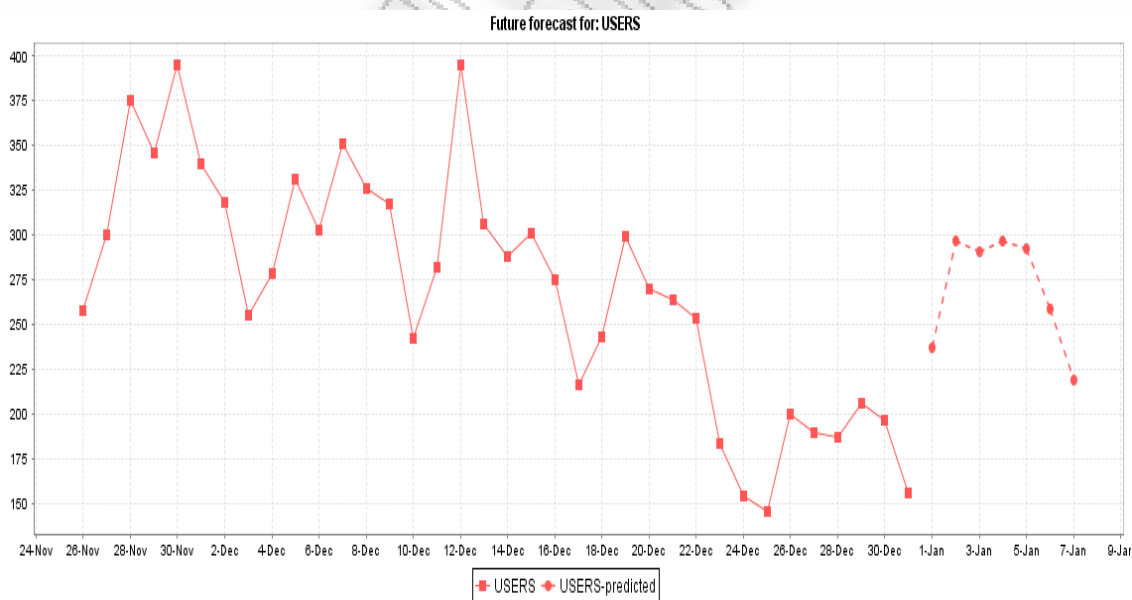


Σχήμα 4.10: Διάγραμμα πρόβλεψης επισκεπτών εβδομάδας (test set)

Ενώ στα σχήματα 4.11 και 4.12 εμφανίζονται μόνο οι τιμές πρόβλεψης για το training set και το test set αντίστοιχα.



Σχήμα 4.11: Διάγραμμα μελλοντικών προβλέψεων επισκεπτών (training set)



Σχήμα 4.12: Διάγραμμα μελλοντικών προβλέψεων επισκεπτών (test set)

## ΚΕΦΑΛΑΙΟ 5

### Συμπεράσματα

#### 5.1 Εισαγωγή

Όπως έχει ήδη αναφερθεί, η παρούσα εργασία ερευνά την εφαρμοσιμότητα των διαφόρων οικονομετρικών μοντέλων και μοντέλων νευρωνικών δικτύων στην βραχυπρόθεσμη πρόβλεψη επισκεψιμότητας χρηστών σε μία εκπαιδευτική πύλη (portal). Ο αριθμός των χρηστών που επισκέπτονται ένα εκπαιδευτικό portal, είναι ένας από τους σημαντικότερους παράγοντες που επηρεάζουν την ύπαρξή του.

Μετά την ανάλυση των δεδομένων και τα αποτελέσματα που προέκυψαν και παρουσιάστηκαν στα Κεφάλαια 3 και 4 αντίστοιχα, το Κεφάλαιο αυτό ασχολείται με την εξαγωγή συμπερασμάτων από την επιμέρους εφαρμογή των διαφόρων μοντέλων.

#### 5.2 Συμπεράσματα

Στα Κεφάλαια 3 και 4, μεθοδολογίας και αποτελεσμάτων αντίστοιχα, ασχοληθήκαμε με μεθόδους πρόβλεψης που συνιστούν τη γραμμική ανάλυση χρονοσειρών. Πιο συγκεκριμένα η πρόβλεψη των μελλοντικών τιμών της μεταβλητής μπορεί να προέλθει από τις εξής κατηγορίες: την εκθετική εξομάλυνση, τη διάσπαση χρονοσειρών και την ανάλυση ARIMA.

Αρχικά εφαρμόσαμε ορισμένες από τις πλέον γνωστές μεθόδους εξομάλυνσης. Με αυτές τις μεθόδους μπορούμε, χωρίς μεγάλο υπολογιστικό βαθμό δυσκολίας, να διαμορφώσουμε εύκολα και σχετικά γρήγορα προβλέψεις για μια μεταβλητή, που προέρχονται από την διαχρονική μελέτη του τρόπου δημιουργίας των τιμών της. Το πλεονέκτημα των μεθόδων αυτών είναι ότι μπορούν να χρησιμοποιηθούν ακόμα και για μικρά δείγματα παρατηρήσεων. Παρακάτω παρατίθεται πίνακας με τα συνοπτικά αποτελέσματα των παραπάνω μεθόδων.

**Πίνακας 5.1:** Σύνοψη αποτελεσμάτων μεθόδων πρόβλεψης

		<b>MAPE</b>	<b>MAD</b>	<b>RMSD</b>
<b>Διπλή Εκθετική Εξομάλυνση</b>	Training Set	28.8	106.1	153.67
	Test Set	13.04	33	42.09
<b>Winters (Multiplicative Method)</b>	Training Set	15.72	60.54	90.81
	Test Set	21.08	53.45	61.03
<b>Winters (Additive Method)</b>	Training Set	17.19	64.53	94.82
	Test Set	21.78	54.3	61.17
<b>Διάσπαση Χρονοσειράς</b>	Training Set	27.2	97.7	145.06
	Test Set	7.924	19.736	24.78
<b>ARIMA(1,1,1)(1,1,1)<sub>7</sub></b>	Training Set	13.82	48.67	77.45
	Test Set	6.34	38.6	53.67
<b>Νευρωνικό Δίκτυο MLP</b>	Training Set	11.52	45.91	73.48
	Test Set	14.95	33.38	41.54

Η αξιολόγηση των αποτελεσμάτων αυτών θα γίνει βάσει των κριτηρίων, MAPE, MAD και RMSD, τόσο για τα δεδομένα εκπαίδευσης όσο και για τα δεδομένα ελέγχου. Τα αποτελέσματα του test set μας δείχνουν πόσο τελικά απέδωσε η συγκεκριμένη μέθοδος σε πραγματικές προβλέψεις. Η πληροφορία, όμως, αυτή δεν είναι αρκετή από μόνη της για να μας δείξει κατά πόσο το μοντέλο δίνει σωστά αποτελέσματα. Έτσι θα χρησιμοποιήσουμε και τις τιμές σφαλμάτων του training set, το οποίο μας παρέχει μια ένδειξη της τυχαιότητας των αποτελεσμάτων. Για παράδειγμα, μία μέθοδος που μας δίνει πολύ μικρούς δείκτες σφάλματος για το test set και πολύ μεγάλους για το training set, δεν μπορούμε να την θεωρήσουμε αξιόπιστη. Αυτό συμβαίνει γιατί από το training set θα δούμε κατά πόσο καλά προσαρμόζεται η μέθοδος.

Από τον Πίνακα 5.1 μπορούμε να συμπεράνουμε ότι η μέθοδος Winters (πολλαπλασιαστική) δίνει τα πιο σταθερά και αξιόπιστα αποτελέσματα. Αυτό είναι κάτι το οποίο ήταν αναμενόμενο σύμφωνα με την θεωρία των μεθόδων αυτών. Συνοπτικά αναφέρουμε την επιπλέον ικανότητα της συγκεκριμένης μεθόδου να προσαρμόζεται στην εποχικότητα. Άλλωστε η εποχικότητα στα δεδομένα μας αποτελεί σημαντικό παράγοντα, σύμφωνα και με τα γραφήματα ACF

(αυτοσυσχετίσεων) και PACF (μερικών αυτοσυσχετίσεων). Για την αξιολόγηση ενός μοντέλου μεμονωμένα, δίνουμε περισσότερο βάρος στο κριτήριο RMSD, απ' ότι στο MAD κι αυτό γιατί στο RMSE το απόλυτο σφάλμα κάθε πρόβλεψης υψώνεται στο τετράγωνο, με άμεσο αποτέλεσμα να μην εμφανίζονται οι μεγάλες αποκλίσεις. Με το κριτήριο MAPE μπορούμε να αξιολογήσουμε κάθε μέθοδο ανεξάρτητα από τα δεδομένα που χρησιμοποιούμε.

Στη συνέχεια ασχοληθήκαμε με τη διάσπαση των χρονοσειρών. Με τη μέθοδο αυτή μπορούμε να πετύχουμε βελτίωση του σφάλματος πρόβλεψης, μέσω της διάσπασης του προτύπου της χρονοσειράς σε υπο-πρότυπα (συνιστώσες). Οι συνιστώσες αυτές που είναι η τάση, η εποχικότητα, οι κυκλικές κινήσεις και η μη-ικανότητα, αλληλεπιδρούν έτσι ώστε να εξηγήσουν τη διακύμανση στις τιμές των δεδομένων διαχρονικά. Λόγω των επιδράσεων αυτών παρατηρούμε συχνά ασταθή συμπεριφορά χρονοσειρών.

Από τα αντίστοιχα διαγράμματα στα σχήματα 4.4, 4.5 και 4.6, που έχουμε παραθέσει στα αποτελέσματα της συγκεκριμένης μεθόδου, παρατηρούμε ότι τα δεδομένα εμφανίζουν σταθερή καθοδική τάση και η ανάλυση της εποχικότητας, που φαίνεται καλύτερα στο διάγραμμα Seasonal Analysis, μας οδηγεί στο συμπέρασμα ότι η εποχικότητα υπάρχει όχι τόσο έντονα σε κάποιες περιόδους του χρόνου ενώ για κάποιες άλλες όχι (τα ποσοστά απόκλισης ανά περιόδους κυμαίνονται ανάλογα). Επίσης αν παρατηρήσουμε τον συνοπτικό πίνακα αποτελεσμάτων της μεθόδου αυτής, βλέπουμε ότι έχουμε το φαινόμενο πολύ μικρών τιμών σφάλματος στο test set σε σχέση με τα αντίστοιχα του training set κάτι το οποίο μας καθιστά το μοντέλο μη αξιόπιστο.

Το ARIMA με την σειρά του αποδεικνύεται ότι υπερτερεί των προηγούμενων μεθόδων σε όλες τις στατιστικές δίνοντας ιδιαίτερα μικρές τιμές σφάλματος. Η επιλογή του ARIMA(1,1,1)(1,1,1)<sub>7</sub> από την πλειάδα γενικών ARIMA μοντέλων δείχνει την εποχικότητα που διατρέχει τα δεδομένα μας καθώς και την αυτοσυσχέτιση τους. Το ARIMA όπως μπορεί να φανεί στην βιβλιογραφία έχει την ικανότητα να παράγει πολύ καλά και αξιόπιστα αποτελέσματα έναντι άλλων μεθόδων ανάλυσης.

Από την ανάλυση των παραμέτρων του μοντέλου, δίνοντας σημασία στα t-statistic παρατηρούμε ότι τα πιο σημαντικά είναι οι συντελεστές του mean average τόσο για αυτό της μη εποχικότητας όσο και για αυτό της εποχικότητας. Οι συντελεστές της



αυτοσυσχέτισης παρόλο που είναι στατιστικά σημαντικοί προσφέρουν λιγότερο στην επεξηγηματικότητα του μοντέλου μας απ'ότι ο κινούμενος μέσος.

Τέλος, τα αποτελέσματα που προέκυψαν από τη χρήση του νευρωνικού δικτύου, αποδεικνύονται υπέρτερα όλων των άλλων μοντέλων και παρότι είναι πολύπλοκο προβλέπει τις μελλοντικές τιμές με το μικρότερο σφάλμα.

Επιγραμματικά τα συμπεράσματα είναι, ότι τα επικρατούντα μοντέλα είναι αυτό του ARIMA και του Νευρωνικού δικτύου. Το πρώτο επιλέγεται για την απλότητά του, στην στοιχειοθέτηση και ερμηνεία των μεταβλητών καθώς επίσης και για την δυνατότητα να παράγει πολύ ακριβή αποτελέσματα. Το δεύτερο παρόλη την πολυπλοκότητά του και την δυσκολία ερμηνείας των παραμέτρων προσφέρει τις πιο ακριβείς προβλέψεις.

Οι χρήστες που επισκέπτονται το διαδίκτυο αυξάνονται διαχρονικά και συγκεκριμένα το τρίτο τρίμηνο κάθε έτους ο αριθμός τους πολλαπλασιάζεται. Τα μοντέλα που αναπτύχθηκαν σε αυτήν την πτυχιακή μπορούν να προβλέψουν με εξαιρετική ακρίβεια και σχετικά μικρό υπολογιστικό και πληροφοριακό κόστος την επισκεψιμότητα των portals. Η πληροφορία αυτή μπορεί να χρησιμοποιηθεί από τους διαχειριστές των portals αυτών αλλά και από όλους όσους ανήκουν στον χώρο της εκπαίδευσης για την κατασκευή καινοτόμων λύσεων αλλά και για την προσέλκυση, ανάπτυξη και ανάδειξή τους, κάνοντας πιο κατανοητή την απήχηση των ιστοτόπων αυτών.

### **5.3 Προτάσεις για περαιτέρω μελέτη**

Ως προτάσεις για περαιτέρω μελέτη θα μπορούσαμε να επιστήσουμε την προσοχή σε τρία βασικά θέματα.

Το πρώτο αποτελεί την χρησιμοποίηση περισσότερων ενδογενών χρονοσειρών (δηλαδή χρονοσειρών όπου υπάρχει συσχετισμός μεταξύ της μεταβλητής και του σφάλματος) από διαφορετικά εκπαιδευτικά portals από διάφορες χώρες, καθώς επίσης και εξωγενών μεταβλητών (δηλαδή μεταβλητών που είναι ανεξάρτητες από όλες τις άλλες τιμές του μοντέλου) που μπορούν να επηρεάσουν σημαντικά την επισκεψιμότητα στο διαδίκτυο. Με βάση το ARIMA μπορούμε να προχωρήσουμε

ένα βήμα παρακάτω και να εξετάσουμε τις μεταβλητές αυτές με μοντέλα ARIMAX κάνοντας χρήση των ιδιοτήτων του cointegration (συνολοκλήρωσης) μεταξύ δύο χρονοσειρών. Για να κάνουμε χρήση του cointegration θα πρέπει να εξετάσουμε τις χρονοσειρές για το αν διολισθαίνουν με τέτοιο τρόπο ώστε να μην απομακρύνονται μεταξύ τους, να υπάρχει δηλαδή ένας κοινός γραμμικός συνδυασμός των δύο χρονοσειρών. Οι μέθοδοι που χρησιμοποιούνται για να εξετάσουμε την ύπαρξη cointegration είναι οι επόμενες (Sørensen, 2005):

- Η δύο – βημάτων μέθοδος Engle-Granger (The Engle-Granger two-step method)
- Η διαδικασία Johansen (The Johansen procedure)
- Η εξέταση cointegration Phillips-Ouliaris (Phillips-Ouliaris Cointegration Test with R).

Το δεύτερο είναι η χρησιμοποίηση αυτών των μοντέλων σε ένα πιο ευρύ φάσμα εκπαιδευτικών μεταβλητών που θα μπορούν να περιγράψουν την δυναμική που αναδύεται στον χώρο του διαδικτύου, τις ευκαιρίες για εκμετάλλευση αυτής της ανάπτυξης αλλά και των παραμέτρων που συντελούν σε αυτήν την κατεύθυνση.

Το τρίτο αφορά την εφαρμογή της χρονοσειράς με γενετικούς αλγορίθμους για την μοντελοποίηση της χρονοσειράς. Τα αποτελέσματα τέτοιων μεθόδων δείχνουν ικανοποιητικά αν και έχουν βασικό πρόβλημα χρόνου εκτέλεσης. Επομένως άλλη μία υπο-πρόταση είναι η εύρεση κατάλληλων τεχνικών γενετικού προγραμματισμού έτσι ώστε να πάρουμε καλύτερα και πιο γρήγορα αποτελέσματα.

## Βιβλιογραφία

- Αγιακόγλου, Χ. & Οικονόμου Γ., 2004, *Μέθοδοι προβλέψεων και ανάλυσης αποφάσεων*, Γ. Μπένου
- Βλαχάβας, Ι., Κεφάλας, Π., Βασιλειάδης, Ν., Κόκκορας, Φ., Σακελλαρίου, Η., 2006. *Τεχνητή Νοημοσύνη*, Γ' Έκδοση, Β.Γκιούρδας Εκδοτική
- Κινγκ, Ρ., 2004. *Ευφυής Έλεγχος*, Τζιόλα
- Κουγιουμτζής, Δ., 2005. *Γραμμική ανάλυση χρονοσειρών*, Πανεπιστημιακές σημειώσεις
- Λύκας, Α., 2008. *Τεχνητά Νευρωνικά Δίκτυα - Εφαρμογές, Τεχνητή Νοημοσύνη Εφαρμογές*, Τόμος Β, Ελληνικό Ανοικτό Πανεπιστήμιο
- Σούκος Β. & Κάντζιος Δ., 2006. *Καταγραφή και παρουσίαση παραδοσιακών και σύγχρονων μεθόδων πρόβλεψης, επισκόπηση και εκτίμηση λογισμικών προβλέψεων*. Διπλωματική Εργασία
- Acquah, H de-Graft, 2010. 'Comparison of Akaike information criterion (AIC) and Bayesian information criteria (BIC) in selection of an asymmetric price relationship', *Journal of Development and Agricultural Economics* Vol. 2(1) pp. 001-006
- Armstrong, J. Scott, 2001. *Principles of Forecasting: A Handbook for Researchers and Practitioners*
- Burby, J., Brown, A., & WAA Standards Committee, 2007. *Web Analytics Definitions- Version 4.0*, Web Analytics Association
- Cave, M., Hanney, S., & Kogan, M., 1997. *The use of performance indicators in higher education: A critical analysis of developing practice* (3rd ed.). Bristol, PA: Jessica Kingsley Publishers.
- DAC, 2002, 'Glossary of Key Terms in Evaluation'
- Dorffner, G., 1996, 'Neural Networks for Time Series Processing', *Neural Network World*, vol. 4, pp. 447-468.
- Greaney, V., & Kellaghan, T., 1996. *Monitoring the learning outcomes of education systems*. Washington, D.C: World Bank.

Haddad, R., 2004. *Portal Usage Analytics: A Key Tool in the Battle Against the Empty Portal*, Project Performance Corporation

Hair, J., Black, W., Babin, B., Anderson, R., Tatham, R., 2006. *Multivariate Data Analysis*, 6<sup>th</sup> Edition, Pearson Education

Haykin, S., 2009, *Neural Networks and Learning Machines*, 3rd Edition, Pearson Education

Johnstone, J. N., 1981. *Indicators of education systems*. Paris: UNESCO.

Klassen, R. and Flores Benito, E., 2001. 'Forecasting practices of Canadian firms: survey results and comparisons', *International Journal of Production Economics*, Vol. 70 No. 2, pp. 163-74.

Land, K. C., & Spilerman, S., 1975. *Social indicator models*. New York: Russel Sage Foundation, In K. C. Land, & S. Spolerman (Eds.)

Merkl, D, *Machine Learning with WEKA*,  
(available at : [www.ec.tuwien.ac.at/~dieter/](http://www.ec.tuwien.ac.at/~dieter/))

Newbold, P. & Bos, T., 1994. *Introductory Business and Economic Forecasting*, 2nd ed., South-Western, Cincinnati, OH.

Owen, E., Hodgkinson, G. D., & Tuijnman, A., 1995. *Towards a strategic approach for developing international indicators of student achievement in OECD: Measuring what students learn*. Paris: OECD.

Rumelhart, D., Hinton, G., Williams, R., 1986, 'Learning Internal Representations by Error Propagation', *Parallel Distributed Processing*, pp.918-362, MIT Press.

Shavelson, R. J., McDonnell, L., Oakes, J., Carey, N., & Picus, L., 1997. *Indicators systems for measuring mathematics and science education*. Santa Monica, CA: The RAND corporation.

Sitte, R., Sitte, J., 2000, 'Analysis of the Predictive Ability of Time Delay Neural Networks Applied to the S&P 500 Time Series', *IEEE Transactions on Systems Man and Cybernetics*, vol. 30, no. 4, pp. 568-572.

Sørensen, E., Bent, 2005, *Cointegration*, Economics 266, Spring, 1997

Wacker, J. & Sprague, L., 1998. 'Forecasting accuracy: comparing the relative effectiveness of practices between seven developed countries', *Journal of Operations Management*, Vol. 16 No. 2-3, pp. 271-90.

Waddell, D. & Sohal, A., 1994. 'Forecasting: the key to managerial decision making', *Management Decision*, Vol. 32 No. 1, pp. 41-9.

ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΕΡΠΗ

## ΠΑΡΑΡΤΗΜΑ Α

### Περιγραφή Νευρωνικού Δικτύου

Στο παράρτημα αυτό περιλαμβάνετε ο σχεδιασμός του Νευρωνικού Δικτύου που υλοποιήθηκε. Δίνονται τα βάρη του κάθε νευρώνα σύμφωνα με το λογισμικό WEKA.

<b>Scheme:</b>	MultilayerPerceptron -L 0.3 -M 0.2 -N 16 -V 0 -S 0 -E 20 -H a
<b>Lagged and derived variable options:</b>	-F [USERS] -L 1 -M 7 -G DATES -dayofweek -weekend
<b>Relation:</b>	TimeSeriesTest
<b>Instances:</b>	730
<b>Attributes:</b>	2 DATES USERS
<b>Transformed training data:</b>	DATES USERS DayOfWeek Weekend DATES-remapped Lag_USERS-1 Lag_USERS-2 Lag_USERS-3 Lag_USERS-4 Lag_USERS-5 Lag_USERS-6 Lag_USERS-7 DATES-remapped^2

	DATES-remapped^3 DATES-remapped*Lag_USERS-1 DATES-remapped*Lag_USERS-2 DATES-remapped*Lag_USERS-3 DATES-remapped*Lag_USERS-4 DATES-remapped*Lag_USERS-5 DATES-remapped*Lag_USERS-6 DATES-remapped*Lag_USERS-7	
<b>USERS:</b>	Linear	Inputs      Weights
	Node 0	Threshold   -0.2578034213974621
		Node 1    0.4785009463722185
		Node 2    0.10096294005963304
		Node 3   -0.18482116623813172
		Node 4   -0.10461631445403934
		Node 5   -0.3577987499268957
		Node 6    0.6899425441986067
		Node 7    0.2712588445602023
		Node 8   -0.19009343747122326
		Node 9   -0.0012531708677928001
		Node 10   0.4860809974608061
		Node 11   0.5886931337878255
		Node 12   -0.0281223585242806
		Node 13   -0.6220332868174464
	Sigmoid	Inputs      Weights
	Node 1	Threshold   -0.4483133418686856
		Attrib DayOfWeek=sun   0.3222881250472569
		Attrib DayOfWeek=mon   0.5747133151494719

		Attrib DayOfWeek=tue 0.4519006757202803 Attrib DayOfWeek=wed 0.6227743878241149 Attrib DayOfWeek=thu 0.16495854639729016 Attrib DayOfWeek=fri 0.1057373990217604 Attrib DayOfWeek=sat -0.10581144318212335 Attrib Weekend -0.15514395655051846 Attrib DATES-remapped 0.062009426083346565 Attrib Lag_USERS-1 0.8110024701956741 Attrib Lag_USERS-2 0.20960004151719308 Attrib Lag_USERS-3 0.12068604629376083 Attrib Lag_USERS-4 0.138629669834121 Attrib Lag_USERS-5 0.15803332706571158 Attrib Lag_USERS-6 0.01526926166699445 Attrib Lag_USERS-7 0.4111495332659536 Attrib DATES-remapped^2 -0.0340984030745658 Attrib DATES-remapped^3 -0.036966769910024136 Attrib DATES-remapped*Lag_USERS-1 0.2989268820191832 Attrib DATES-remapped*Lag_USERS-2 -0.03116015984712579 Attrib DATES-remapped*Lag_USERS-3 -0.021938406753011744 Attrib DATES-remapped*Lag_USERS-4 -0.058897677326681316 Attrib DATES-remapped*Lag_USERS-5 -0.026817084227679335 Attrib DATES-remapped*Lag_USERS-6 -0.14623778016378353 Attrib DATES-remapped*Lag_USERS-7 0.2183275372736057
	Sigmoid Node 2	Inputs Weights Threshold -0.3763966091502817 Attrib DayOfWeek=sun 0.32159869234459215 Attrib DayOfWeek=mon 0.3164461852999614



		Attrib DayOfWeek=tue 0.3450961837666594 Attrib DayOfWeek=wed 0.3012748997833865 Attrib DayOfWeek=thu 0.2593177991519117 Attrib DayOfWeek=fri 0.2555954416515224 Attrib DayOfWeek=sat 0.21939948951378502 Attrib Weekend 0.11800503245966569 Attrib DATES-remapped -0.0862541390123169 Attrib Lag_USERS-1 0.21722035927177308 Attrib Lag_USERS-2 0.12939621913492685 Attrib Lag_USERS-3 0.1569327489969983 Attrib Lag_USERS-4 0.07135393050386221 Attrib Lag_USERS-5 0.09530842282727947 Attrib Lag_USERS-6 0.0969858155260657 Attrib Lag_USERS-7 0.1837277405676127 Attrib DATES-remapped^2 0.007501248991256356 Attrib DATES-remapped^3 0.030066444327215346 Attrib DATES-remapped*Lag_USERS-1 0.187281504826634 Attrib DATES-remapped*Lag_USERS-2 0.101083082643803 Attrib DATES-remapped*Lag_USERS-3 0.10048731643354603 Attrib DATES-remapped*Lag_USERS-4 0.1330594711990752 Attrib DATES-remapped*Lag_USERS-5 0.06478322937160064 Attrib DATES-remapped*Lag_USERS-6 0.1336875113522029 Attrib DATES-remapped*Lag_USERS-7 0.14114176498013645
	Sigmoid Node 3	Inputs                      Weights Threshold -0.38263395695898933 Attrib DayOfWeek=sun 0.250715643882618 Attrib DayOfWeek=mon 0.24191255261522585

		Attrib DayOfWeek=tue 0.3178074830891919 Attrib DayOfWeek=wed 0.3132989472424892 Attrib DayOfWeek=thu 0.337043117386044 Attrib DayOfWeek=fri 0.3331267028326602 Attrib DayOfWeek=sat 0.32032721396597963 Attrib Weekend 0.245838580463998 Attrib DATES-remapped -0.34759579960522985 Attrib Lag_USERS-1 -0.11998057352127395 Attrib Lag_USERS-2 0.012676265178133637 Attrib Lag_USERS-3 0.06433761481905831 Attrib Lag_USERS-4 0.1381659790208249 Attrib Lag_USERS-5 0.0411382262114793 Attrib Lag_USERS-6 0.06762995378407295 Attrib Lag_USERS-7 -0.0703659454897735 Attrib DATES-remapped^2 -0.15273152058497638 Attrib DATES-remapped^3 0.021791790581665194 Attrib DATES-remapped*Lag_USERS-1 0.1456825673114708 Attrib DATES-remapped*Lag_USERS-2 0.14743747685425343 Attrib DATES-remapped*Lag_USERS-3 0.2415227468511001 Attrib DATES-remapped*Lag_USERS-4 0.21882447632725174 Attrib DATES-remapped*Lag_USERS-5 0.14661444787365976 Attrib DATES-remapped*Lag_USERS-6 0.1294543035124297 Attrib DATES-remapped*Lag_USERS-7 0.07068539865838165
	Sigmoid Node 4	Inputs Weights Threshold -0.4118414929548902 Attrib DayOfWeek=sun 0.2720072559029518 Attrib DayOfWeek=mon 0.23003301486477387

		Attrib DayOfWeek=tue 0.3207113976858802 Attrib DayOfWeek=wed 0.2692732133099373 Attrib DayOfWeek=thu 0.30884857005729294 Attrib DayOfWeek=fri 0.26846048915462983 Attrib DayOfWeek=sat 0.33969948881608747 Attrib Weekend 0.20775074871963156 Attrib DATES-remapped -0.2493722803452301 Attrib Lag_USERS-1 0.0434579805991136 Attrib Lag_USERS-2 0.14026455429471263 Attrib Lag_USERS-3 0.15198853933908008 Attrib Lag_USERS-4 0.12981658454943046 Attrib Lag_USERS-5 0.13464682955998347 Attrib Lag_USERS-6 0.0946160461433201 Attrib Lag_USERS-7 0.0100268494812549 Attrib DATES-remapped^2 -0.029972563179514234 Attrib DATES-remapped^3 -0.015790298455402935 Attrib DATES-remapped*Lag_USERS-1 0.10150247279489512 Attrib DATES-remapped*Lag_USERS-2 0.18162657378726607 Attrib DATES-remapped*Lag_USERS-3 0.1497995219960636 Attrib DATES-remapped*Lag_USERS-4 0.14491913227436878 Attrib DATES-remapped*Lag_USERS-5 0.10342125366430767 Attrib DATES-remapped*Lag_USERS-6 0.13015589244968254 Attrib DATES-remapped*Lag_USERS-7 0.09758012917451436
	Sigmoid Node 5	Inputs           Weights Threshold -0.4120918661427312 Attrib DayOfWeek=sun 0.3543808195415931 Attrib DayOfWeek=mon 0.2484592186069413

		Attrib DayOfWeek=tue 0.28837681664177456
		Attrib DayOfWeek=wed 0.28277200710741524
		Attrib DayOfWeek=thu 0.36273342508289236
		Attrib DayOfWeek=fri 0.3494492586495705
		Attrib DayOfWeek=sat 0.23044802098702952
		Attrib Weekend 0.17674093868054302
		Attrib DATES-remapped -0.6200312546503836
		Attrib Lag_USERS-1 -0.3175173949424771
		Attrib Lag_USERS-2 -0.09210656694350812
		Attrib Lag_USERS-3 -0.043882630826839415
		Attrib Lag_USERS-4 0.04891473306264132
		Attrib Lag_USERS-5 0.11472021806272722
		Attrib Lag_USERS-6 0.10645638065509394
		Attrib Lag_USERS-7 -0.07322003400808373
		Attrib DATES-remapped^2 -0.2438747550830376
		Attrib DATES-remapped^3 -0.10555586361845075
		Attrib DATES-remapped*Lag_USERS-1 0.0681991127787611
		Attrib DATES-remapped*Lag_USERS-2 0.11839632604093474
		Attrib DATES-remapped*Lag_USERS-3 0.18005485249084555
		Attrib DATES-remapped*Lag_USERS-4 0.2033190452164077
		Attrib DATES-remapped*Lag_USERS-5 0.2454069705389179
		Attrib DATES-remapped*Lag_USERS-6 0.23732179765007136
		Attrib DATES-remapped*Lag_USERS-7 0.12342180621624982
	Sigmoid Node 6	Inputs Weights Threshold -0.503296749066039 Attrib DayOfWeek=sun 0.24023057937923029 Attrib DayOfWeek=mon 0.8075823015750183

		Attrib DayOfWeek=tue 0.5676367469853618 Attrib DayOfWeek=wed 0.8628861423205063 Attrib DayOfWeek=thu -0.012667236778939807 Attrib DayOfWeek=fri 0.07138503698388818 Attrib DayOfWeek=sat -0.14627828749282135 Attrib Weekend -0.2999567758919786 Attrib DATES-remapped -0.08303760913863288 Attrib Lag_USERS-1 1.167244302050216 Attrib Lag_USERS-2 0.17153936401950656 Attrib Lag_USERS-3 0.1299235818974972 Attrib Lag_USERS-4 0.18988741598766928 Attrib Lag_USERS-5 0.30406936805137336 Attrib Lag_USERS-6 -0.04780524605858436 Attrib Lag_USERS-7 0.543937141876285 Attrib DATES-remapped^2 -0.12026425257765018 Attrib DATES-remapped^3 -0.08835432457049539 Attrib DATES-remapped*Lag_USERS-1 0.39366777726736935 Attrib DATES-remapped*Lag_USERS-2 -0.17143059152912382 Attrib DATES-remapped*Lag_USERS-3 -0.15197219461133812 Attrib DATES-remapped*Lag_USERS-4 -0.12787310144831715 Attrib DATES-remapped*Lag_USERS-5 -0.10963253430678141 Attrib DATES-remapped*Lag_USERS-6 -0.28435247861750107 Attrib DATES-remapped*Lag_USERS-7 0.11868760000069792
	Sigmoid Node 7	Inputs           Weights Threshold -0.4169233864437071 Attrib DayOfWeek=sun 0.3274023000044988 Attrib DayOfWeek=mon 0.4107680816388396

		Attrib DayOfWeek=tue 0.3359133099771461 Attrib DayOfWeek=wed 0.3466318219647132 Attrib DayOfWeek=thu 0.31096323008331533 Attrib DayOfWeek=fri 0.1565330512081005 Attrib DayOfWeek=sat 0.01285323546531451 Attrib Weekend 0.013099776944656182 Attrib DATES-remapped -0.0073693955291870596 Attrib Lag_USERS-1 0.4229492060219972 Attrib Lag_USERS-2 0.12950205845141305 Attrib Lag_USERS-3 0.15116158454550638 Attrib Lag_USERS-4 0.07913014202828912 Attrib Lag_USERS-5 0.10684082286244302 Attrib Lag_USERS-6 0.1009380921394969 Attrib Lag_USERS-7 0.28844479939045975 Attrib DATES-remapped^2 -0.006292515324954868 Attrib DATES-remapped^3 0.09812936802951273 Attrib DATES-remapped*Lag_USERS-1 0.1959092013720133 Attrib DATES-remapped*Lag_USERS-2 0.11864172944639718 Attrib DATES-remapped*Lag_USERS-3 0.09641269850823862 Attrib DATES-remapped*Lag_USERS-4 0.11535434677833314 Attrib DATES-remapped*Lag_USERS-5 0.0682090022692837 Attrib DATES-remapped*Lag_USERS-6 0.05976436869167278 Attrib DATES-remapped*Lag_USERS-7 0.18270502527913024
	Sigmoid Node 8	Inputs           Weights Threshold -0.41932726813648613 Attrib DayOfWeek=sun 0.2306600193877859 Attrib DayOfWeek=mon 0.20119686867207937

		Attrib DayOfWeek=tue 0.31017224463038257 Attrib DayOfWeek=wed 0.27788290398741944 Attrib DayOfWeek=thu 0.2974065078074847 Attrib DayOfWeek=fri 0.30968490610890886 Attrib DayOfWeek=sat 0.3408803535605964 Attrib Weekend 0.2408894232789425 Attrib DATES-remapped -0.3618969558167914 Attrib Lag_USERS-1 -0.08047345262845312 Attrib Lag_USERS-2 0.13277760068032163 Attrib Lag_USERS-3 0.07224333386574013 Attrib Lag_USERS-4 0.12964847467067284 Attrib Lag_USERS-5 0.07462320320316222 Attrib Lag_USERS-6 0.04573566648143093 Attrib Lag_USERS-7 -0.0495262207105959 Attrib DATES-remapped^2 -0.0959035219700575 Attrib DATES-remapped^3 -0.0073647826950650605 Attrib DATES-remapped*Lag_USERS-1 0.15295341208023333 Attrib DATES-remapped*Lag_USERS-2 0.1567789486384827 Attrib DATES-remapped*Lag_USERS-3 0.19831255310081602 Attrib DATES-remapped*Lag_USERS-4 0.1811771421966453 Attrib DATES-remapped*Lag_USERS-5 0.22041417135443073 Attrib DATES-remapped*Lag_USERS-6 0.1855645553866785 Attrib DATES-remapped*Lag_USERS-7 0.0925105166048503
	Sigmoid Node 9	Inputs           Weights Threshold -0.38940380566921573 Attrib DayOfWeek=sun 0.28973257033300925 Attrib DayOfWeek=mon 0.24757548617863384

		Attrib DayOfWeek=tue 0.3097074709601204 Attrib DayOfWeek=wed 0.2686311516044652 Attrib DayOfWeek=thu 0.3309935584757291 Attrib DayOfWeek=fri 0.2530027294258616 Attrib DayOfWeek=sat 0.3100840643669512 Attrib Weekend 0.13489995079873127 Attrib DATES-remapped -0.17424211419454752 Attrib Lag_USERS-1 0.14406351620557067 Attrib Lag_USERS-2 0.13899787470310423 Attrib Lag_USERS-3 0.12545512810638892 Attrib Lag_USERS-4 0.07405833092714761 Attrib Lag_USERS-5 0.09760312561446081 Attrib Lag_USERS-6 0.13485390296132768 Attrib Lag_USERS-7 0.11483369623313441 Attrib DATES-remapped^2 -0.02021529226649142 Attrib DATES-remapped^3 0.03541455755571876 Attrib DATES-remapped*Lag_USERS-1 0.15859654642507515 Attrib DATES-remapped*Lag_USERS-2 0.15744321399448727 Attrib DATES-remapped*Lag_USERS-3 0.1410174512735985 Attrib DATES-remapped*Lag_USERS-4 0.15716520457023153 Attrib DATES-remapped*Lag_USERS-5 0.10757773553814286 Attrib DATES-remapped*Lag_USERS-6 0.10631967700909975 Attrib DATES-remapped*Lag_USERS-7 0.12488779588881431
	Sigmoid Node 10	Inputs           Weights Threshold -0.4107167048749462 Attrib DayOfWeek=sun 0.3382459381780998 Attrib DayOfWeek=mon 0.6403806562645352



		Attrib DayOfWeek=tue 0.39382008284215 Attrib DayOfWeek=wed 0.5802189407079228 Attrib DayOfWeek=thu 0.1826547786644482 Attrib DayOfWeek=fri 0.08503186975746206 Attrib DayOfWeek=sat -0.10249999461490539 Attrib Weekend -0.21930176560594142 Attrib DATES-remapped 0.017750683808297935 Attrib Lag_USERS-1 0.7858106488323153 Attrib Lag_USERS-2 0.15093769021223866 Attrib Lag_USERS-3 0.15671561526094152 Attrib Lag_USERS-4 0.16397967398601906 Attrib Lag_USERS-5 0.22841706882989343 Attrib Lag_USERS-6 -0.03134518508569189 Attrib Lag_USERS-7 0.4567685176419251 Attrib DATES-remapped^2 -0.018910045159453503 Attrib DATES-remapped^3 0.028928390265039267 Attrib DATES-remapped*Lag_USERS-1 0.3292494737215956 Attrib DATES-remapped*Lag_USERS-2 -0.0019530411733344956 Attrib DATES-remapped*Lag_USERS-3 -0.010744580420574228 Attrib DATES-remapped*Lag_USERS-4 -0.05014162404490348 Attrib DATES-remapped*Lag_USERS-5 -0.06312972231772111 Attrib DATES-remapped*Lag_USERS-6 -0.072990983940999 Attrib DATES-remapped*Lag_USERS-7 0.21584857860649645
	Sigmoid Node 11	Inputs           Weights Threshold -0.45427772496691465 Attrib DayOfWeek=sun 0.3116755479341905 Attrib DayOfWeek=mon 0.7342252862785511

		Attrib DayOfWeek=tue 0.4827741616674386 Attrib DayOfWeek=wed 0.6926297919106953 Attrib DayOfWeek=thu 0.07149638242095117 Attrib DayOfWeek=fri 0.05166713634124685 Attrib DayOfWeek=sat -0.181003044211579 Attrib Weekend -0.2520072059323856 Attrib DATES-remapped 0.029429850056880712 Attrib Lag_USERS-1 0.9861902495247483 Attrib Lag_USERS-2 0.16116231352128463 Attrib Lag_USERS-3 0.1639633601434892 Attrib Lag_USERS-4 0.17875974000402853 Attrib Lag_USERS-5 0.2872229352233621 Attrib Lag_USERS-6 0.020440748282750307 Attrib Lag_USERS-7 0.5082868026793731 Attrib DATES-remapped^2 -0.026866683637198463 Attrib DATES-remapped^3 -0.00872752961791897 Attrib DATES-remapped*Lag_USERS-1 0.33259634445497915 Attrib DATES-remapped*Lag_USERS-2 -0.10403309908354892 Attrib DATES-remapped*Lag_USERS-3 -0.08874435082881733 Attrib DATES-remapped*Lag_USERS-4 -0.09053498823926995 Attrib DATES-remapped*Lag_USERS-5 -0.05023054835510464 Attrib DATES-remapped*Lag_USERS-6 -0.21403833740315162 Attrib DATES-remapped*Lag_USERS-7 0.13785833187847252
	Sigmoid Node 12	Inputs Weights Threshold -0.4242189000984904 Attrib DayOfWeek=sun 0.25744474473932183 Attrib DayOfWeek=mon 0.2823061030513312

		Attrib DayOfWeek=tue 0.33473361308747535 Attrib DayOfWeek=wed 0.2432454960379391 Attrib DayOfWeek=thu 0.29438370245957535 Attrib DayOfWeek=fri 0.27670722050659746 Attrib DayOfWeek=sat 0.31206576854456364 Attrib Weekend 0.178924981818523 Attrib DATES-remapped -0.20309350543829408 Attrib Lag_USERS-1 0.056457107101103014 Attrib Lag_USERS-2 0.12124330261679356 Attrib Lag_USERS-3 0.08582746445715621 Attrib Lag_USERS-4 0.08261371270308361 Attrib Lag_USERS-5 0.08068311571425937 Attrib Lag_USERS-6 0.09108087941660753 Attrib Lag_USERS-7 0.12928734674351772 Attrib DATES-remapped^2 -0.005416438834614431 Attrib DATES-remapped^3 0.012888127719756524 Attrib DATES-remapped*Lag_USERS-1 0.17800033604095356 Attrib DATES-remapped*Lag_USERS-2 0.12954644444261493 Attrib DATES-remapped*Lag_USERS-3 0.1384041304115219 Attrib DATES-remapped*Lag_USERS-4 0.1433275361360275 Attrib DATES-remapped*Lag_USERS-5 0.14985427310665675 Attrib DATES-remapped*Lag_USERS-6 0.09697654169099876 Attrib DATES-remapped*Lag_USERS-7 0.11630713723423353
	Sigmoid Node 13	Inputs Weights Threshold -0.32893671467616487 Attrib DayOfWeek=sun -0.03452061765434829 Attrib DayOfWeek=mon 0.20346397074907377

		Attrib DayOfWeek=tue 0.20091237325402153 Attrib DayOfWeek=wed 0.43731614458130846 Attrib DayOfWeek=thu 0.023512054966419643 Attrib DayOfWeek=fri 0.26042758245461656 Attrib DayOfWeek=sat 0.3012919699210978 Attrib Weekend 0.016754767294003905 Attrib DATES-remapped -0.70639092415331 Attrib Lag_USERS-1 -2.081168633039932 Attrib Lag_USERS-2 -0.6073414604119864 Attrib Lag_USERS-3 -0.6282852217133581 Attrib Lag_USERS-4 -0.34182823862894 Attrib Lag_USERS-5 -0.473510798874747 Attrib Lag_USERS-6 -0.37280228715894204 Attrib Lag_USERS-7 -0.9235849696337937 Attrib DATES-remapped^2 0.07326729813596225 Attrib DATES-remapped^3 0.1878770230048716 Attrib DATES-remapped*Lag_USERS-1 -0.4013285718208915 Attrib DATES-remapped*Lag_USERS-2 0.27233492490989064 Attrib DATES-remapped*Lag_USERS-3 0.310104790607976 Attrib DATES-remapped*Lag_USERS-4 0.24733640281581093 Attrib DATES-remapped*Lag_USERS-5 0.29183015323958333 Attrib DATES-remapped*Lag_USERS-6 0.33328595100104447 Attrib DATES-remapped*Lag_USERS-7 0.32506045663452726
Class	Input Node 0	

## ΠΑΡΑΡΤΗΜΑ Β

### Προβλέψεις Νευρωνικού Δικτύου

Στο παράρτημα Β βλέπουμε τις προβλέψεις που έβγαλε το Νευρωνικό Δίκτυο. Με \* συμβολίζονται οι προβλέψεις για την πρώτη εβδομάδα του 2012 που δεν υπάρχει στο data set μας.

=== Future predictions from end of training data (Time / USERS) ===

2010-01-01	101	2010-01-20	209	2010-02-08	220
2010-01-02	132	2010-01-21	199	2010-02-09	222
2010-01-03	149	2010-01-22	189	2010-02-10	265
2010-01-04	173	2010-01-23	166	2010-02-11	198
2010-01-05	168	2010-01-24	203	2010-02-12	192
2010-01-06	164	2010-01-25	232	2010-02-13	152
2010-01-07	213	2010-01-26	207	2010-02-14	129
2010-01-08	184	2010-01-27	218	2010-02-15	184
2010-01-09	146	2010-01-28	207	2010-02-16	229
2010-01-10	173	2010-01-29	194	2010-02-17	236
2010-01-11	197	2010-01-30	165	2010-02-18	215
2010-01-12	210	2010-01-31	190	2010-02-19	206
2010-01-13	208	2010-02-01	210	2010-02-20	175
2010-01-14	206	2010-02-02	247	2010-02-21	225
2010-01-15	191	2010-02-03	224	2010-02-22	249
2010-01-16	162	2010-02-04	196	2010-02-23	240
2010-01-17	200	2010-02-05	192	2010-02-24	227
2010-01-18	214	2010-02-06	157	2010-02-25	223
2010-01-19	213	2010-02-07	192	2010-02-26	207

2010-02-27	182	2010-03-24	246	2010-04-18	311
2010-02-28	218	2010-03-25	204	2010-04-19	372
2010-03-01	224	2010-03-26	233	2010-04-20	357
2010-03-02	223	2010-03-27	172	2010-04-21	360
2010-03-03	264	2010-03-28	220	2010-04-22	319
2010-03-04	376	2010-03-29	362	2010-04-23	301
2010-03-05	265	2010-03-30	289	2010-04-24	284
2010-03-06	178	2010-03-31	463	2010-04-25	310
2010-03-07	234	2010-04-01	223	2010-04-26	445
2010-03-08	269	2010-04-02	155	2010-04-27	428
2010-03-09	264	2010-04-03	124	2010-04-28	362
2010-03-10	248	2010-04-04	127	2010-04-29	378
2010-03-11	252	2010-04-05	195	2010-04-30	311
2010-03-12	236	2010-04-06	232	2010-05-01	255
2010-03-13	206	2010-04-07	241	2010-05-02	337
2010-03-14	245	2010-04-08	240	2010-05-03	589
2010-03-15	274	2010-04-09	242	2010-05-04	483
2010-03-16	246	2010-04-10	193	2010-05-05	339
2010-03-17	250	2010-04-11	233	2010-05-06	324
2010-03-18	258	2010-04-12	305	2010-05-07	307
2010-03-19	223	2010-04-13	548	2010-05-08	215
2010-03-20	210	2010-04-14	447	2010-05-09	234
2010-03-21	235	2010-04-15	384	2010-05-10	336
2010-03-22	268	2010-04-16	342	2010-05-11	298
2010-03-23	265	2010-04-17	268	2010-05-12	299

2010-05-13	308	2010-06-07	351	2010-07-02	366
2010-05-14	372	2010-06-08	319	2010-07-03	231
2010-05-15	237	2010-06-09	317	2010-07-04	240
2010-05-16	262	2010-06-10	329	2010-07-05	381
2010-05-17	360	2010-06-11	312	2010-07-06	410
2010-05-18	325	2010-06-12	184	2010-07-07	953
2010-05-19	393	2010-06-13	215	2010-07-08	377
2010-05-20	307	2010-06-14	337	2010-07-09	373
2010-05-21	324	2010-06-15	354	2010-07-10	238
2010-05-22	211	2010-06-16	345	2010-07-11	224
2010-05-23	219	2010-06-17	337	2010-07-12	433
2010-05-24	278	2010-06-18	359	2010-07-13	408
2010-05-25	354	2010-06-19	194	2010-07-14	424
2010-05-26	440	2010-06-20	218	2010-07-15	352
2010-05-27	361	2010-06-21	372	2010-07-16	342
2010-05-28	394	2010-06-22	375	2010-07-17	210
2010-05-29	204	2010-06-23	444	2010-07-18	238
2010-05-30	230	2010-06-24	496	2010-07-19	377
2010-05-31	358	2010-06-25	412	2010-07-20	374
2010-06-01	336	2010-06-26	251	2010-07-21	392
2010-06-02	337	2010-06-27	381	2010-07-22	464
2010-06-03	323	2010-06-28	480	2010-07-23	671
2010-06-04	298	2010-06-29	445	2010-07-24	281
2010-06-05	238	2010-06-30	606	2010-07-25	254
2010-06-06	262	2010-07-01	405	2010-07-26	527

2010-07-27	410	2010-08-21	280	2010-09-15	625
2010-07-28	393	2010-08-22	278	2010-09-16	553
2010-07-29	377	2010-08-23	446	2010-09-17	506
2010-07-30	365	2010-08-24	473	2010-09-18	335
2010-07-31	230	2010-08-25	545	2010-09-19	395
2010-08-01	246	2010-08-26	549	2010-09-20	558
2010-08-02	431	2010-08-27	583	2010-09-21	558
2010-08-03	608	2010-08-28	364	2010-09-22	527
2010-08-04	500	2010-08-29	381	2010-09-23	584
2010-08-05	398	2010-08-30	757	2010-09-24	615
2010-08-06	372	2010-08-31	865	2010-09-25	392
2010-08-07	235	2010-09-01	734	2010-09-26	399
2010-08-08	239	2010-09-02	658	2010-09-27	581
2010-08-09	376	2010-09-03	649	2010-09-28	542
2010-08-10	383	2010-09-04	361	2010-09-29	542
2010-08-11	359	2010-09-05	410	2010-09-30	569
2010-08-12	659	2010-09-06	805	2010-10-01	617
2010-08-13	364	2010-09-07	1102	2010-10-02	435
2010-08-14	244	2010-09-08	1266	2010-10-03	434
2010-08-15	197	2010-09-09	784	2010-10-04	733
2010-08-16	368	2010-09-10	704	2010-10-05	571
2010-08-17	372	2010-09-11	473	2010-10-06	548
2010-08-18	363	2010-09-12	472	2010-10-07	501
2010-08-19	378	2010-09-13	818	2010-10-08	467
2010-08-20	378	2010-09-14	624	2010-10-09	309



2010-10-10	352	2010-11-04	355	2010-11-29	410
2010-10-11	448	2010-11-05	353	2010-11-30	387
2010-10-12	458	2010-11-06	306	2010-12-01	388
2010-10-13	446	2010-11-07	272	2010-12-02	378
2010-10-14	448	2010-11-08	385	2010-12-03	327
2010-10-15	436	2010-11-09	380	2010-12-04	312
2010-10-16	316	2010-11-10	370	2010-12-05	325
2010-10-17	358	2010-11-11	377	2010-12-06	351
2010-10-18	430	2010-11-12	391	2010-12-07	363
2010-10-19	422	2010-11-13	328	2010-12-08	372
2010-10-20	398	2010-11-14	302	2010-12-09	362
2010-10-21	401	2010-11-15	354	2010-12-10	325
2010-10-22	412	2010-11-16	332	2010-12-11	297
2010-10-23	316	2010-11-17	399	2010-12-12	317
2010-10-24	342	2010-11-18	359	2010-12-13	371
2010-10-25	419	2010-11-19	347	2010-12-14	319
2010-10-26	369	2010-11-20	311	2010-12-15	326
2010-10-27	410	2010-11-21	341	2010-12-16	321
2010-10-28	374	2010-11-22	407	2010-12-17	275
2010-10-29	436	2010-11-23	368	2010-12-18	260
2010-10-30	303	2010-11-24	399	2010-12-19	306
2010-10-31	339	2010-11-25	342	2010-12-20	307
2010-11-01	422	2010-11-26	341	2010-12-21	296
2010-11-02	382	2010-11-27	287	2010-12-22	267
2010-11-03	371	2010-11-28	336	2010-12-23	223

2010-12-24	214	2011-01-18	246	2011-02-12	332
2010-12-25	225	2011-01-19	236	2011-02-13	317
2010-12-26	215	2011-01-20	205	2011-02-14	358
2010-12-27	256	2011-01-21	178	2011-02-15	354
2010-12-28	278	2011-01-22	157	2011-02-16	351
2010-12-29	261	2011-01-23	179	2011-02-17	330
2010-12-30	245	2011-01-24	198	2011-02-18	317
2010-12-31	233	2011-01-25	193	2011-02-19	298
2011-01-01	204	2011-01-26	208	2011-02-20	332
2011-01-02	250	2011-01-27	240	2011-02-21	365
2011-01-03	283	2011-01-28	212	2011-02-22	335
2011-01-04	285	2011-01-29	210	2011-02-23	272
2011-01-05	294	2011-01-30	273	2011-02-24	295
2011-01-06	243	2011-01-31	320	2011-02-25	308
2011-01-07	288	2011-02-01	298	2011-02-26	339
2011-01-08	260	2011-02-02	319	2011-02-27	370
2011-01-09	278	2011-02-03	359	2011-02-28	371
2011-01-10	316	2011-02-04	321	2011-03-01	326
2011-01-11	330	2011-02-05	269	2011-03-02	320
2011-01-12	340	2011-02-06	295	2011-03-03	314
2011-01-13	331	2011-02-07	334	2011-03-04	297
2011-01-14	284	2011-02-08	339	2011-03-05	272
2011-01-15	251	2011-02-09	318	2011-03-06	274
2011-01-16	252	2011-02-10	345	2011-03-07	316
2011-01-17	257	2011-02-11	299	2011-03-08	400

2011-03-09	368	2011-04-03	365	2011-04-28	298
2011-03-10	345	2011-04-04	371	2011-04-29	378
2011-03-11	290	2011-04-05	412	2011-04-30	290
2011-03-12	282	2011-04-06	387	2011-05-01	337
2011-03-13	367	2011-04-07	367	2011-05-02	365
2011-03-14	429	2011-04-08	298	2011-05-03	383
2011-03-15	378	2011-04-09	265	2011-05-04	395
2011-03-16	327	2011-04-10	306	2011-05-05	391
2011-03-17	346	2011-04-11	322	2011-05-06	415
2011-03-18	314	2011-04-12	347	2011-05-07	336
2011-03-19	298	2011-04-13	303	2011-05-08	392
2011-03-20	316	2011-04-14	308	2011-05-09	474
2011-03-21	363	2011-04-15	260	2011-05-10	620
2011-03-22	345	2011-04-16	248	2011-05-11	503
2011-03-23	333	2011-04-17	274	2011-05-12	760
2011-03-24	305	2011-04-18	262	2011-05-13	424
2011-03-25	271	2011-04-19	251	2011-05-14	348
2011-03-26	247	2011-04-20	260	2011-05-15	370
2011-03-27	313	2011-04-21	222	2011-05-16	489
2011-03-28	383	2011-04-22	206	2011-05-17	392
2011-03-29	390	2011-04-23	223	2011-05-18	390
2011-03-30	402	2011-04-24	208	2011-05-19	378
2011-03-31	366	2011-04-25	242	2011-05-20	356
2011-04-01	306	2011-04-26	286	2011-05-21	307
2011-04-02	323	2011-04-27	329	2011-05-22	369

2011-05-23	415	2011-06-17	338	2011-07-12	439
2011-05-24	395	2011-06-18	250	2011-07-13	405
2011-05-25	403	2011-06-19	298	2011-07-14	393
2011-05-26	397	2011-06-20	415	2011-07-15	318
2011-05-27	351	2011-06-21	386	2011-07-16	238
2011-05-28	288	2011-06-22	376	2011-07-17	232
2011-05-29	333	2011-06-23	377	2011-07-18	358
2011-05-30	406	2011-06-24	332	2011-07-19	360
2011-05-31	390	2011-06-25	260	2011-07-20	377
2011-06-01	389	2011-06-26	301	2011-07-21	373
2011-06-02	374	2011-06-27	392	2011-07-22	335
2011-06-03	350	2011-06-28	363	2011-07-23	224
2011-06-04	308	2011-06-29	342	2011-07-24	233
2011-06-05	351	2011-06-30	370	2011-07-25	339
2011-06-06	434	2011-07-01	358	2011-07-26	338
2011-06-07	389	2011-07-02	271	2011-07-27	372
2011-06-08	332	2011-07-03	313	2011-07-28	369
2011-06-09	313	2011-07-04	453	2011-07-29	381
2011-06-10	303	2011-07-05	472	2011-07-30	234
2011-06-11	236	2011-07-06	594	2011-07-31	237
2011-06-12	272	2011-07-07	495	2011-08-01	411
2011-06-13	316	2011-07-08	402	2011-08-02	362
2011-06-14	340	2011-07-09	261	2011-08-03	368
2011-06-15	365	2011-07-10	373	2011-08-04	404
2011-06-16	343	2011-07-11	468	2011-08-05	423

2011-08-06	283	2011-08-31	1221	2011-09-25	530
2011-08-07	274	2011-09-01	722	2011-09-26	881
2011-08-08	448	2011-09-02	568	2011-09-27	818
2011-08-09	616	2011-09-03	376	2011-09-28	873
2011-08-10	429	2011-09-04	451	2011-09-29	968
2011-08-11	447	2011-09-05	649	2011-09-30	1022
2011-08-12	473	2011-09-06	751	2011-10-01	439
2011-08-13	284	2011-09-07	1284	2011-10-02	520
2011-08-14	262	2011-09-08	883	2011-10-03	675
2011-08-15	290	2011-09-09	704	2011-10-04	643
2011-08-16	576	2011-09-10	438	2011-10-05	651
2011-08-17	775	2011-09-11	497	2011-10-06	691
2011-08-18	521	2011-09-12	810	2011-10-07	734
2011-08-19	442	2011-09-13	869	2011-10-08	483
2011-08-20	305	2011-09-14	964	2011-10-09	546
2011-08-21	329	2011-09-15	907	2011-10-10	706
2011-08-22	504	2011-09-16	897	2011-10-11	718
2011-08-23	679	2011-09-17	623	2011-10-12	784
2011-08-24	610	2011-09-18	620	2011-10-13	1013
2011-08-25	515	2011-09-19	815	2011-10-14	706
2011-08-26	504	2011-09-20	847	2011-10-15	398
2011-08-27	325	2011-09-21	847	2011-10-16	446
2011-08-28	364	2011-09-22	741	2011-10-17	574
2011-08-29	805	2011-09-23	822	2011-10-18	542
2011-08-30	937	2011-09-24	451	2011-10-19	450

2011-10-20	456	2011-11-14	456
2011-10-21	430	2011-11-15	437
2011-10-22	332	2011-11-16	428
2011-10-23	410	2011-11-17	411
2011-10-24	500	2011-11-18	406
2011-10-25	465	2011-11-19	271
2011-10-26	486	2011-11-20	333
2011-10-27	488	2011-11-21	415
2011-10-28	306	2011-11-22	448
2011-10-29	276	2011-11-23	448
2011-10-30	410	2011-11-24	464
2011-10-31	655	2011-11-25	346
2011-11-01	586	2011-11-26*	261.5426
2011-11-02	563	2011-11-27*	308.2377
2011-11-03	436	2011-11-28*	383.5361
2011-11-04	446	2011-11-29*	383.89
2011-11-05	305	2011-11-30*	389.9479
2011-11-06	355	2011-12-01*	378.887
2011-11-07	473	2011-12-02*	315.4089
2011-11-08	437		
2011-11-09	455		
2011-11-10	410		
2011-11-11	444		
2011-11-12	332		
2011-11-13	408		

=== Future predictions from end of test data (Time / USERS) ===

2011-11-26	258	2011-12-20	270
2011-11-27	300	2011-12-21	264
2011-11-28	375	2011-12-22	253
2011-11-29	346	2011-12-23	183
2011-11-30	395	2011-12-24	154
2011-12-01	340	2011-12-25	145
2011-12-02	318	2011-12-26	200
2011-12-03	255	2011-12-27	189
2011-12-04	278	2011-12-28	187
2011-12-05	331	2011-12-29	206
2011-12-06	303	2011-12-30	196
2011-12-07	351	2011-12-31	156
2011-12-08	326	2012-01-01*	236.9237
2011-12-09	317	2012-01-02*	296.832
2011-12-10	242	2012-01-03*	290.7459
2011-12-11	282	2012-01-04*	296.3794
2011-12-12	395	2012-01-05*	292.4382
2011-12-13	306	2012-01-06*	258.8117
2011-12-14	288	2012-01-07*	218.7823
2011-12-15	301		
2011-12-16	275		
2011-12-17	216		
2011-12-18	243		
2011-12-19	299		