ΑΝΩΤΑΤΟ ΤΕΧΝΟΛΟΓΙΚΟ ΕΚΠΑΙΔΕΥΤΙΚΟ ΙΔΡΥΜΑ ΠΑΤΡΩΝ
ΤΜΗΜΑ ΕΠΙΧΕΙΡΗΜΑΤΙΚΟΥ ΣΧΕΔΙΑΣΜΟΥ ΚΑΙ ΠΛΗΡΟΦΟΡΙΑΚΩΝ ΣΥΣΤΗΜΑΤΩΝ

ΠΤΥΧΙΑΚΗ ΕΡΓΑΣΙΑ

ΒΑΣΙΚΕΣ ΜΕΘΟΔΟΙ ΑΝΑΛΥΣΗΣ ΧΡΟΝΟΣΕΙΡΩΝ

ΣΠΟΥΔΑΣΤΕΣ: ΓΑΣΤΟΥΝΙΩΤΗΣ ΝΙΚΟΛΑΟΣ
ΖΕΡΔΕΒΑ ΠΑΡΑΣΚΕΥΗ-ΜΑΡΙΑ

ΕΙΣΗΓΗΤΡΙΑ: ΚΑΡΥΩΤΗ ΒΑΣΙΛΙΚΗ

ΠΑΤΡΑ, ΝΟΕΜΒΡΙΟΣ 2005

ΤΕΧΝΟΛΟΓΙΚΟ ΕΚΠΑΙΔΕΥΤΙΚΟ ΙΔΡΥΜΑ ΠΑΤΡΑΣ

ΠΤΥΧΙΑΚΗ ΕΡΓΑΣΙΑ

ΒΑΣΙΚΕΣ ΜΕΘΟΔΟΙ ΑΝΑΛΥΣΗΣ ΧΡΟΝΟΣΕΙΡΩΝ

ΣΠΟΥΔΑΣΤΕΣ: ΓΑΣΤΟΥΝΙΩΤΗΣ ΝΙΚΟΛΑΟΣ
ΖΕΡΔΕΒΑ ΠΑΡΑΣΚΕΥΗ-ΜΑΡΙΑ

ΕΙΣΗΓΗΤΡΙΑ: ΚΑΡΥΩΤΗ-ΒΑΣΙΛΙΚΗ

ΠΑΤΡΑ, ΝΟΕΜΒΡΙΟΣ 2005

Πίνακας Περιεχομένων

Αφιέρωση

Στους γονείς μας, Θεόδωρος και Δήμητρα,
Αναστάσιος και Βασιλική.

Ευχαριστίες

Αρχικά, θα θέλαμε να ευχαριστήσουμε όλους αυτούς που μας συμπαραστάθηκαν κατά την διάρκεια της εργασίας, ειδικά στις δύσκολες στιγμές που υπήρξαν.

Όπως επίσης και το Κεντρικό Ταχυδρομείο Πατρών για την πολύτιμη βοήθεια του στην παροχή υλικού σχετικά με τις εφαρμογές.

Ειδικά, θέλουμε να ευχαριστήσουμε την καθηγήτρια μας κα. Καρυώτη Βασιλική για την καθοδήγηση και την βοήθεια της όλο αυτό το χρονικό διάστημα.

Τέλος, ένα "μεγάλο" ευχαριστώ στις οικογένειες μας για την υπομονή και την ανοχή που έδειξαν μέχρι το πέρας της εργασίας.

ΚΕΦΑΛΑΙΟ 1: ΕΙΣΑΓΩΓΗ

Η ανάλυση χρονοσειρών θεωρείται μια στατιστική μέθοδος ανάλυσης στην περίπτωση που τα δεδομένα προκύπτουν από παρατηρήσεις που συλλέγονται κατά τη διάρκεια μιας μακριάς χρονικής περιόδου.

Έτσι, χρονοσειρά ονομάζεται το σύνολο από χρονολογικά ταξινομημένες παρατηρήσεις σε σχέση με μια μεταβλητή κατά την διάρκεια διαδοχικών και ισόχρονων περιόδων. Βασικό χαρακτηριστικό κάθε χρονοσειράς είναι η εξάρτηση μεταξύ των διαδοχικών τιμών της. Η φύση της αλληλεξάρτησης μεταξύ των παρατηρήσεων μια χρονοσειράς είναι το αντικείμενο μελέτης και ανάλυσης του κλάδου χρονοσειρών. Οι τεχνικές που αναλύουν την μεταβλητότητα στη χρονοσειρά αναφέρονται ως μέθοδοι χρονοσειρών. Το πεδίο εφαρμογής αυτών των μεθόδων ανάλυσης χρονοσειρών δεν είναι περιορισμένο αλλά ευρύνεται σε πολλούς κλάδους της επιστήμης όπως: οικονομία, κλιματολογία, βιολογία, φυσική, οικολογία και ωκεανογραφία κ.τ.λ.

Οι κύριοι στόχοι ενός ερευνητή που ασχολείται με την ανάλυση χρονοσειρών είναι αρχικά να γίνει ο προσδιορισμός της φύσης του φαινόμενου που αντιπροσωπεύεται από την ακολουθία των παρατηρήσεων, και στο τέλος να γίνει πρόβλεψη των μελλοντικών τιμών της χρονοσειράς.

Η πρόβλεψη των μελλοντικών τιμών μιας μεταβλητής μπορεί να γίνει με διάφορες μεθόδους που διαφέρουν ως προς την πολυπλοκότητα, την ταχύτητα και το κόστος υπολογισμού τους, τη διαθεσιμότητα των απαραίτητων δεδομένων και των άλλων παραγόντων κατά περίπτωση.

Για να γίνει πρόβλεψη, θα πρέπει να εξεταστούν τα δεδομένα του παρελθόντος. Άρα, πρόβλεψη (Forecasting) είναι η διαδικασία εκείνη, που στηρίζεται στην ανάλυση - μέσω κατάλληλων στατιστικών μεθόδων - της πορείας ενός συγκεκριμένου μεγέθους κατά το παρελθόν και το παρόν έχοντας ως σκοπό να μας δώσει τις τιμές του μεγέθους αυτού στο μέλλον. Η διαδικασία που ακολουθείται είναι η εξής:

ΙΣΤΟΡΙΚΑ ΔΕΔΟΜΕΝΑ + ΜΕΘΟΔΟΛΟΓΙΑ → ΚΑΤΑΣΚΕΥΗ ΜΟΝΤΕΛΟΥ→ ΑΞΙΟΠΟΙΗΣΗ ΜΟΝΤΕΛΟΥ→ ΠΡΟΒΛΕΨΕΙΣ

Οι προβλέψεις χωρίζονται σε 4 κατηγορίες:

1. Βραχυπρόθεσμες προβλέψεις : τους επόμενους μήνες μέχρι και 3 χρόνια.
2. Μεσοπρόθεσμες προβλέψεις: διάστημα μεταξύ 3 και 5 χρόνων (σχετίζονται με τον προγραμματισμό μιας εταιρίας).
3. Μακροπρόθεσμες προβλέψεις: 5 με 15 χρόνια (σχετίζονται με τον τεχνολογικό εξοπλισμό μιας εταιρίας).
4. Και προβλέψεις που σχετίζονται με τεχνολογικές προβλέψεις 15 με 30 χρόνια

Οι στατιστικές μέθοδοι που χρησιμοποιούνται για διενέργεια προβλέψεων διακρίνονται στις εξής κατηγορίες:

1. Περιγραφικές
2. Ποιοτικές
3. Μέθοδοι Εκθετικής Εξομάλυνσης
4. Μέθοδοι Box-Jenkins
5. Φασματικές μέθοδοι
6. Οικονομετρικές

Σε οποιαδήποτε άσκηση πρόβλεψης μιας δεδομένης χρονοσειράς, θα πρέπει να σταθμίζεται το κόστος της συγκεκριμένης μεθόδου πρόβλεψης με τα αναμενόμενα οφέλη. Έτσι αν τα οφέλη από την χρησιμοποίηση μιας σύνθετης μεθόδου είναι υψηλά λόγω αυξημένης ακρίβειας των προβλέψεων, τότε αντισταθμίζεται το υψηλότερο κόστος της μεθόδου. Αν όμως τα οφέλη είναι οριακά, τότε είναι πολλές φορές προτιμότερο να επιλέγεται η απλούστερη αλλά και οικονομικότερη μέθοδος.

Το αντικείμενο μελέτης της παρούσας εργασίας είναι η έννοια της χρονοσειράς, βάση της οποίας θα γίνει η ανάλυση τής με διάφορες στατιστικές μεθόδους που θα αναφερθούν στα κεφάλαια που θα ακολουθήσουν.

Στο δεύτερο κεφάλαιο θα αναφερθούμε στις περιγραφικές μεθόδους (Descriptive Methods) οι οποίες αποτελούν αντικείμενο μελέτης της περιγραφικής στατιστικής. Κύρια εργαλεία των μεθόδων αυτών είναι η εκθετική εξομάλυνση και οι κινητοί μέσοι, βάση των οποίων επιτυγχάνεται ο διαχωρισμός μιας χρονοσειράς στις κύριες συνιστώσες της και η μελέτη της συμπεριφοράς αυτών.

Η ανάλυση στο πεδίο του χρόνου (Time Domain Analysis) είναι η μέθοδος με την οποία θα ασχοληθούμε στο τρίτο κεφάλαιο. Η μεθοδολογία αυτή αναπτύχθηκε και μορφοποιήθηκε από τους Box και Jenkins οι οποίοι μελέτησαν διάφορα μοντέλα χρονοσειρών (γνωστά ως ARIMA) και συστηματοποίησαν την όλη διαδικασία προσαρμογής μοντέλων σε παρατηρούμενες χρονοσειρές.

Στο τέταρτο κεφάλαιο θα εξετάσουμε τις οικονομετρικές μεθόδους (Econometric methods) που αποτελούν αντικείμενο μελέτης της οικονομετρίας και αφορούν την εκτίμηση εξισώσεων και συστημάτων αλληλοεξαρτώμενων εξισώσεων όπου οι μεταβλητές τους είναι οι χρονοσειρές.

Στο πέμπτο κεφάλαιο θα γίνει ανάλυση της μεθόδου στο πεδίο των συχνοτήτων (Frequency Domain Analysis). Η μέθοδος αυτή έχει σαν κύριο εργαλείο τη μελέτη της φασματικής πυκνότητας (Spectral Density) που συνδέεται με την έννοια της αυτοσυσχέτισης μέσα από έναν μετασχηματισμό Fourier. Το αποτέλεσμα που προκύπτει είναι η επονομαζόμενη ως Φασματική Ανάλυση (Spectral Analysis).

Τέλος, στο έκτο κεφάλαιο θα γίνει παρουσίαση εφαρμογών που αναφέρονται σε μοντέλα χρονοσειρών που έχουν εξεταστεί στα προηγούμενα κεφάλαια.

ΚΕΦΑΛΑΙΟ 2: ΠΕΡΙΓΡΑΦΙΚΕΣ ΜΕΘΟΔΟΙ

2.1 Εισαγωγή

Σ' αυτό το κεφάλαιο θα αναφερθούμε σε βασικές μεθόδους περιγραφικής ανάλυσης χρονοσειρών που αναφέρονται σε μη στοχαστικά μοντέλα. Αρχικά θα γίνει αναφορά στις τέσσερις συνιστώσες (τάση, κύκλος, εποχικότητα και τυχαία συνιστώσα) που επιδρούν στη χρονοσειρά. Θα εξετάσουμε τον τρόπο μέτρησής τους, όπως και την έννοια του σφάλματος ο οποίος είναι ένας πολύ σημαντικός εκτιμητής ακρίβειας της πρόβλεψης. Εν συνεχεία, στο τέλος του κεφαλαίου θα εξεταστούν βασικές μέθοδοι εξομάλυνσης της χρονοσειράς όπως: ο κινητός μέσος και η εκθετική εξομάλυνση.

2.2 Χρονοσειρά ή Χρονολογική σειρά

Χρονοσειρά ή Χρονολογική σειρά είναι η σειρά των τιμών της μεταβλητής που παίρνει σε διαδοχικές χρονικές στιγμές ή περιόδους. Μας ενδιαφέρει η μονάδα μέτρησης (σε χρόνια, μήνες, εβδομάδες) και συμβολίζεται με: Y_t ή X_t ή W_t ή Z_t (Καρυώτη Βασιλική, 2004).

2.3 Συνιστώσες Χρονοσειρών

Η κλασική μέθοδος ανάλυσης μίας χρονοσειράς αποβλέπει στο διαχωρισμό αυτής της χρονοσειράς στις επιμέρους συνιστώσες της, που είναι οι εξής:

2.3.1 Τάση (Trend)

Τάση είναι η μακροχρόνια συνεχής κίνηση που ακολουθεί η χρονοσειρά. Είναι η κατά μέσο όρο απαλλαγμένη από βραχυχρόνιες αυξομειώσεις και δείχνει την εξέλιξη της χρονοσειράς για μεγάλα διαστήματα, συνήθως πάνω από δέκα έτη (Μπόρα-Σέντα Ε., 1997). Γι΄ αυτό το λόγο καλείται και μακροχρόνια τάση (long–run trend). Μπορεί να έχει γραμμική ή μη-γραμμική μορφή με ανοδική ή καθοδική τάση. Η τάση ενσωματώνει τις μακροχρόνιες εξελίξεις του μεγέθους που εκφράζει η μεταβλητή και είναι αποτέλεσμα της εξέλιξης της οικονομίας, των τεχνολογικών μεταβολών και των μακροχρόνιων αλλαγών των διαφόρων βιομηχανικών κλάδων. Η τάση θεωρείται ανύπαρκτη όταν η κεντρική κίνηση της χρονοσειράς είναι παράλληλη προς τον άξονα του χρόνου t χωρίς να παρουσιάζει τάση προς αύξηση ή μείωση.

2.3.2 Κυκλική Συνιστώσα (Cyclical Component)

Η κυκλική συνιστώσα αντιπροσωπεύει εκείνες τις επαναλαμβανόμενες κυμάνσεις γύρω από τη τάση που η διάρκειά τους είναι μεγαλύτερη του έτους και προέρχονται από μεταβολές της γενικής οικονομικής δραστηριότητας. Αποτελεί τη διαφορά μεταξύ της αναμενόμενης τιμής της χρονοσειράς που προέρχεται από τη τάση και της πραγματικής τιμής της χρονοσειράς. Επομένως η κυκλική συνιστώσα, μπορεί να θεωρηθεί ως απόκλιση που κυμαίνεται γύρω από τη τάση σε σχέση με τη μεταβαλλόμενη γενική οικονομική δραστηριότητα.

2.3.3 Εποχική Συνιστώσα (Seasonal Component)

Η εποχική συνιστώσα είναι μία κυκλική κύμανση με περίοδο όμως το έτος, διότι μέσα σε αυτό εξαντλεί όλες τις ανοδικές και καθοδικές κινήσεις. Επίσης, είναι περιοδική διότι επαναλαμβάνεται ρυθμικά κάθε έτος (Μπόρα-Σέντα Ε., 1997). Είναι προφανές ότι η εποχική κύμανση, όταν βέβαια υπάρχει, εμφανίζεται μόνο σε χρονοσειρές με εποχικές παρατηρήσεις (μηνιαία, τριμηνιαία, κλπ. δεδομένα.). Η εποχική συνιστώσα θεωρείται η κύρια πηγή των εμπορικών και οικονομικών χρονοσειρών. Διαφέρει εκείνης της κυκλικής συνιστώσας από τη διάρκειά της και μπορεί να εμφανισθεί με διάρκεια εβδομαδιαία, μηνιαία, τριμηνιαία.

2.3.4 Τυχαία Συνιστώσα (Random Component)

Η τυχαία συνιστώσα μίας χρονοσειράς δεν επηρεάζεται από τις άλλες συστηματικές συνιστώσες και προκαλείται από μη περιοδικά ή απρόβλεπτα γεγονότα. Όπως οι ξαφνικές εξαγγελίες κυβερνητικών μέτρων, απρόβλεπτες αλλαγές τιμών στη διεθνή αγορά, ασυνήθιστες κλιματολογικές συνθήκες, πολιτικές κρίσεις και φυσικές καταστροφές. Τα γεγονότα αυτά δεν επαναλαμβάνονται όπως οι κύκλοι επιχειρηματικής δραστηριότητας που είναι επαναλαμβανόμενοι.

Παρατήρηση: Οι τρεις πρώτες συνιστώσες είναι συστηματικές, δηλαδή ακολουθούν συγκεκριμένο τρόπο επίδρασης και μπορούμε να τις μετρήσουμε ενώ η τέταρτη δεν είναι συστηματική και δεν μπορούμε να την προσδιορίσουμε.

2.4 Μοντέλα Χρονοσειρών

Πριν προχωρήσουμε στην περιγραφή των μεθόδων εκτίμησης των διαφόρων συνιστωσών των χρονοσειρών, πρέπει πρώτα να διευκρινίσουμε τους τρόπους με τους οποίους συνδέονται μεταξύ τους οι συνιστώσες για να δώσουν την τελική τιμή της σειράς. Οι τρόποι σύνδεσης των συνιστωσών των χρονοσειρών, που καλούνται μοντέλα χρονοσειρών, είναι δύο. Το προσθετικό και το πολλαπλασιαστικό μοντέλο.

2.4.1 Προσθετικό Μοντέλο

Στο προσθετικό μοντέλο η τιμή της μεταβλητής Y που παρατηρούμε στη χρονική περίοδο t, δηλαδή η Yt, προκύπτει από το άθροισμα των τιμών των τεσσάρων συνιστωσών που έχουν τη συγκεκριμένη αυτή περίοδο σύμφωνα με τον τύπο:

$$Y_t = T_t + C_t + S_t + I_t, \text{ όπου:}$$

Y_t = η τιμή της σειράς Y στη χρονική περίοδο t

T_t = η τιμή της τάσης

C_t = η επίδραση του κύκλου

S_t = η επίδραση της εποχικής συνιστώσας

I_t = η επίδραση της τυχαίας συνιστώσας

Το κύριο χαρακτηριστικό του προσθετικού μοντέλου είναι ότι όλες οι συνιστώσες εκφράζονται στην ίδια μονάδα μέτρησης, δηλαδή στη μονάδα μέτρησης της χρονοσειράς.

2.4.2 Πολλαπλασιαστικό Μοντέλο

Ένας άλλος τρόπος σύνδεσης των τεσσάρων συνιστωσών της χρονοσειράς είναι ο πολλαπλασιασμός τους για κάθε χρονική περίοδο. Δηλαδή στη περίοδο t η παρατήρηση Y_t προκύπτει σύμφωνα με το μοντέλο:

$$Y_t = T_t \cdot C_t \cdot S_t \cdot I_t$$

που ονομάζεται πολλαπλασιαστικό. Το βασικό χαρακτηριστικό του πολλαπλασιαστικού μοντέλου είναι ότι, σε αντίθεση με το προσθετικό, μόνο η συνιστώσα της τάσης (T) εκφράζεται στην ίδια μονάδα μέτρησης με τη μεταβλητή Y. Οι υπόλοιπες συνιστώσες (C, S, I) είναι δείκτες, δηλαδή καθαροί αριθμοί χωρίς μονάδα μέτρησης. Έτσι, η επίδραση των συνιστωσών αυτών εκφράζεται ως ποσοστό και όχι σε απόλυτους αριθμούς, όπως συμβαίνει στο προσθετικό μοντέλο.

Από τα δύο μοντέλα διαχωρισμού, το προσθετικό είναι πιο δύσκολο για παραπέρα ανάλυση. Επίσης, υποθέτει ότι οι συντελεστές μεταξύ τους είναι ανεξάρτητοι. Για παράδειγμα, αν χρησιμοποιήσουμε το προσθετικό μοντέλο, δεχόμαστε ότι η τάση δεν επηρεάζει τον εποχικό συντελεστή όποια τιμή και αν έχει αυτός. 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 αναλυτές της διαχωριστικής (decomposition) ανάλυσης το βρίσκουν πιο σωστό.

Κύριο αντικείμενο της ανάλυσης των χρονοσειρών είναι η εκτίμηση των τιμών των συνιστωσών για όλες τις χρονικές περιόδους που καλύπτει η χρονοσειρά. Πιο συγκεκριμένα, οι συστηματικές συνιστώσες (τάση, κύκλος και εποχικότητα) ακολουθούν κάποιο μοντέλο, ο στατιστικός προσδιορισμός του οποίου θα μας επιτρέψει να περιγράψουμε τον τρόπο προσδιορισμού των τιμών της χρονοσειράς. Εάν αυτό είναι εφικτό, τότε θα είμαστε σε θέση να προβλέψουμε την εξέλιξη της σειράς στο μέλλον με την υπόθεση ότι οι συστηματικές συνιστώσες θα συνεχίσουν και στο μέλλον να έχουν την ίδια συμπεριφορά.

Θα ξεκινήσουμε το στατιστικό προσδιορισμό των τιμών των συνιστωσών των χρονοσειρών με την εκτίμηση των μοντέλων που περιγράφουν τις μεταβολές της τάσης.

2.5 Εκτίμηση της Τάσης

Όπως αναφέραμε παραπάνω η τάση μίας χρονοσειράς, η οποία ανήκει στις συστηματικές συνιστώσες, μπορεί να είναι είτε γραμμική είτε μη-γραμμική. Η εκτίμηση της τάσης μπορεί να γίνει είτε με το γραμμικό ή εκθετικό μοντέλο που περιγράφει τη μεταβολή της τάσης, είτε με τη μέθοδο των κινητών μέσων όρων η οποία εξομαλύνει τη χρονοσειρά και την απαλλάσσει από τις βραχυχρόνιες αυξομειώσεις.

Το γραμμικό και το εκθετικό μοντέλο μπορούν να εκτιμηθούν με τη μέθοδο των ελαχίστων τετράγωνων (παλινδρόμηση) με εξαρτημένη μεταβλητή τις τιμές της χρονοσειράς και ανεξάρτητη μεταβλητή τη χρονική περίοδο t. Η μόνη διαφορά των δύο μοντέλων είναι ότι στο δεύτερο πρέπει πρώτα να λογαριθμίσουμε τη σχέση και ύστερα να χρησιμοποιήσουμε τη μέθοδο των ελαχίστων τετραγώνων. Αυτό συμβαίνει επειδή το συγκεκριμένο μοντέλο είναι μη-γραμμικό ως προς τις παραμέτρους.

Το γραμμικό μοντέλο έχει τη μορφή της απλής παλινδρόμησης

$$Y_t = a + bt$$

και σημαίνει ότι η μεταβλητή Y μεταβάλλεται κάθε περίοδο κατά το σταθερό ποσό b. Ενώ το εκθετικό μοντέλο περιγράφεται από τη σχέση:

$$Y_t = (\alpha b^{X_t})$$

όπου η μεταβλητή Y μεταβάλλεται με σταθερό ποσοστό.

2.5.1 Γραμμικό Μοντέλο

Μερικές φορές ο αναλυτής προκειμένου να εκτιμήσει τη τάση της χρονοσειράς χαράζει μία γραμμή δια μέσου των παρατηρήσεών της έτσι ώστε να αποκαλύψει το σχήμα και την κατεύθυνση της τάσης. Η μέθοδος αυτή περιέχει μεγάλο ρίσκο ως προς την ορθότητα των αποτελεσμάτων. Πρέπει λοιπόν ο αναλυτής να έχει πλήρη γνώση του θέματος και την ικανότητα να αναγνωρίζει την επαναλαμβανόμενη εποχική διακύμανση και τους κύκλους μέσω των οποίων διέρχεται η τάση. Εξαιτίας της αναξιοπιστίας της παραπάνω μεθόδου οι ερευνητές χρησιμοποιούν συνήθως την αντικειμενική μέθοδο των ελαχίστων τετραγώνων για την εκτίμηση της τάσης. Σύμφωνα με αυτή τη μέθοδο η γραμμική εξίσωση ορίζεται ως εξής:

$$\hat{Y}_t = \hat{a} + \hat{b}X_t \qquad \text{όπου}$$

$\hat{Y}$ είναι η προβλεφθείσα τιμή της τάσης για μία χρονική περίοδο t

$\hat{a}$ είναι ο εκτιμητής της τιμής της τάσης όταν X= 0

$\hat{b}$ είναι ο εκτιμητής της κλίσης της εξίσωσης και

X_t είναι η παρατήρηση για κάθε περίοδο t.

2.5.2 Εκθετικό Μοντέλο

Εάν η τάση είναι γραμμική σε ένα ημι-λογαριθμικό διάγραμμα, τότε αυτή προσαρμόζεται καλύτερα με μία εκθετική τάση. Το εκθετικό μοντέλο ορίζεται ως $Y_t = (\alpha b^{X_t})$. Εφαρμόζοντας τώρα τη μέθοδο των ελαχίστων τετραγώνων έχουμε την αντίστοιχη εξίσωση λογαριθμισμένη ως εξής: $\log\hat{Y}_t = \log a + X_t \log b$ και υπολογίζουμε:

$$\log a = \frac{\Sigma \log Y}{n}$$

$$\log b = \frac{\Sigma(X \log Y)}{\Sigma X^2}$$

2.5.3 Μέθοδος των Κινητών Μέσων Όρων

Η επικρατέστερη μέθοδος απεικόνισης της τάσης κυρίως της αξίας των χρηματιστηριακών τίτλων είναι η μέθοδος των κινητών μέσων όρων. Η μέθοδος αυτή εμπεριέχει τον υπολογισμό του μέσου όρου των παρατηρήσεων του δείγματος της χρονοσειράς καθώς και τη χρησιμοποίηση αυτού του μέσου σαν πρόβλεψη για την επόμενη χρονική περίοδο. Χρησιμοποιούμε τον όρο κινητός μέσος γιατί καθώς κάθε νέα παρατήρηση μπαίνει στο δείγμα, υπολογίζεται ένας νέος μέσος απορρίπτοντας τη παλαιότερη παρατήρηση του δείγματος συμπεριλαμβάνοντας τη νεότερη. Έτσι γίνεται πρόβλεψη για την επόμενη χρονική περίοδο πάντα με τον ίδιο αριθμό παρατηρήσεων του δείγματος της χρονοσειράς k διαδοχικών όρων. Δηλαδή εάν το μήκος του κινητού μέσου όρου είναι 3 (k= 3), τότε οι όροι του κινητού μέσου n τιμών της σειράς Y είναι $(Y_1+Y_2+Y_3)/3$, $(Y_2+Y_3+Y_4)/3$, …., $(Y_{n-2}+Y_{n-1}+Y_n)/3$. Εάν το k είναι μικρό η σειρά των κινητών μέσων ακολουθεί περισσότερο τις κινήσεις της αρχικής χρονοσειράς. Ενώ για μεγάλες τιμές του k έχουμε μεγαλύτερο εξομαλυντικό αποτέλεσμα και η σειρά των κινήσεων είναι πιο ομαλή. Η μέθοδος του κινητού μέσου όρου είναι μια πολύ εύκολη διαδικασία πρόβλεψης, αφού οι απαιτήσεις σε στατιστικά στοιχεία είναι μικρές.

Παρόλα αυτά όμως η μέθοδος των κινητών μέσων όρων έχει κάποια μειονεκτήματα.

i. Το κύριο μειονέκτημα είναι ότι οι κινητοί μέσοι όροι δεν μπορούν να υπολογιστούν για όλες τις τιμές της χρονοσειράς. Μπορεί να εκτιμηθεί για $n-k$ περιόδους, δηλαδή χάνονται $k/2$ τιμές στην αρχή και στο τέλος της χρονοσειράς.

ii. Ένα άλλο μειονέκτημα του κινητού μέσου όρου είναι ότι δεν μπορεί να χρησιμοποιηθεί για προβλέψεις λόγω του ότι δεν ακολουθεί κάποιο συγκεκριμένο μοντέλο. Γι΄ αυτό χρησιμοποιείται μόνο για την εξομάλυνση της χρονοσειράς και την απεικόνιση της κεντρικής της κίνησης κατά την περίοδο του δείγματος. Επίσης η επιλογή του μήκους του κινητού μέσου όρου είναι υποκειμενική και βασίζεται στην εμπειρία του αναλυτή.

2.5.4 Απομόνωση της Τάσης

Για να εντοπίσουμε τις υπόλοιπες συνιστώσες μίας χρονοσειράς απομακρίνουμε την επίδραση της τάσης στις τιμές της χρονοσειράς από τα στοιχεία του δείγματος. Στο πολλαπλασιαστικό μοντέλο απομακρίνουμε την τάση διαιρώντας τις τιμές της χρονοσειράς με τις αντίστοιχες τιμές της τάσης. Δηλαδή:

$$\frac{Y_t}{\mathrm{T}_t} = \frac{\mathrm{T}_t \cdot C_t \cdot S_t \cdot I_t}{\mathrm{T}_t} = C_t \cdot S_t \cdot I_t = Y^*,$$

όπου Y^* δίνει τις τιμές της χρονοσειράς απαλλαγμένη από την τάση.

2.6 Εκτίμηση του Κύκλου

Η εκτίμηση της κυκλικής συνιστώσας μίας χρονοσειράς μετράει τις αποκλίσεις των τιμών της χρονοσειράς γύρω από τη μακροχρόνια τάση της και έχει διάρκεια μεγαλύτερη του ενός έτους. Ο τρόπος απομόνωσης της, είναι παρόμοιος με αυτόν της απομόνωσης της τάσης. Δηλαδή σε ένα πολλαπλασιαστικό μοντέλο διαιρούμε τις τιμές της χρονοσειράς Y_t με τις τιμές της τάσης. Εάν όμως στη χρονοσειρά περιλαμβάνεται και η επίδραση της εποχικής συνιστώσας τότε απομονώνουμε πρώτα την εποχικότητα και εν συνεχεία την τάση. Δηλαδή:

$$\frac{Y_t}{S_t} = \frac{T_t \cdot C_t \cdot S_t \cdot I_t}{S_t} = T_t \cdot C_t \cdot I_t \qquad \text{(απομόνωση εποχικότητας),}$$

$$\frac{T_t \cdot C_t \cdot I_t}{T_t} = C_t \cdot I_t \qquad \text{(απομόνωση τάσης).}$$

Οι χρονοσειρές που προκύπτουν περιλαμβάνουν μόνο την επίδραση που σχετίζεται με την κυκλική και την τυχαία συνιστώσα.

2.7 Εκτίμηση της Εποχικότητας

Η εκτίμηση της εποχικής συνιστώσας βασίζεται σε εποχικούς δείκτες, η μέτρηση των οποίων γίνεται με τρόπο παρόμοιο όπως η εκτίμηση της κυκλικής συνιστώσας. Ας θεωρήσουμε ότι εξετάζουμε ένα πολλαπλασιαστικό μοντέλο:

$$Y_t = T_t \cdot C_t \cdot S_t \cdot I_t \Rightarrow \frac{Y_t}{T_t \cdot C_t} = S_t \cdot I_t$$

Επομένως σύμφωνα με την παραπάνω σχέση ο τρόπος εκτίμησης των εποχικών δεικτών γίνεται με διαίρεση των τιμών της τάσης και του κύκλου από τη μεταβλητή Y_t της χρονοσειράς. Αντίστοιχα σε ένα προσθετικό μοντέλο θα αφαιρούσαμε την επίδραση της τάσης και του κύκλου από τη χρονοσειρά Y_t. Από τον όρο $S_t \cdot I_t$ εκτιμούμε τη κατά μέσο όρο επίδραση κάθε εποχής. Οι επιδράσεις αυτές είναι οι δείκτες εποχικότητας. Να υπενθυμίσουμε ότι η εποχικότητα όσον αφορά το προσθετικό μοντέλο εκφράζεται στις ίδιες μονάδες μέτρησης με την αρχική μεταβλητή Y_t, ενώ για το πολλαπλασιαστικό μοντέλο οι εποχικές επιδράσεις εκφράζονται ως δείκτες. Παρακάτω θα αναφέρουμε αναλυτικά τον τρόπο εκτίμησης των εποχικών δεικτών.

2.7.1 Εποχικοί Δείκτες

Αναλυτικότερα οι εποχικοί δείκτες προσδιορίζονται με τη μέθοδο του κινητού μέσου (Moving Average). Στόχος είναι η εξάλειψη των άλλων συνιστωσών ώστε να απομείνουν μόνο οι εποχικές διακυμάνσεις. Η βασική παραδοχή που βρίσκεται πίσω από τη μέθοδο του κινητού μέσου όρου, είναι ότι το εποχικό πρότυπο δεν είναι

απόλυτα ακριβές κατά τα διαδοχικά έτη, ωστόσο υπάρχει μέσα σε λογικά όρια και ο κινητός μέσος όρος που ελαχιστοποιεί την επίδραση σε έναν τουλάχιστον βαθμό.
Για ευκολότερη κατανόηση του υπολογισμού των εποχικών δεικτών θα αναφερθεί αναλυτικά η διαδικασία σύμφωνα με ένα παράδειγμα.

Παράδειγμα

Τα δεδομένα αφορούν αφίξεις τουριστών στη Κύπρο ανά τρίμηνο κατά τη χρονική περίοδο 1996 έως 2000. Η διαδικασία υπολογισμού των εποχικών δεικτών γίνεται σε πέντε βήματα.

1ο Βήμα: Υπολογίζουμε τους κινητούς μέσους τεσσάρων τριμήνων για τα 5 έτη. Τα αποτελέσματα φαίνονται στη στήλη (ΚΜ), του πίνακα. Η πρώτη τιμή 48,5 είναι το άθροισμα των τεσσάρων πρώτων αριθμών των αφίξεων που αντιστοιχούν στα τέσσερα τρίμηνα του 1996 διαιρεμένο με το 4. Η δεύτερη τιμή 48,2 είναι το άθροισμα των τεσσάρων τριμήνων από το δεύτερο τρίμηνο του 1996 έως το πρώτο τρίμηνο του 1997 διαιρεμένο με το 4 κτλ.

2ο Βήμα: Υπολογίζουμε τον κεντρικό μέσο τεσσάρων τριμήνων, υπολογίζουμε δηλαδή τις επιδράσεις της τάσης και της κυκλικής συνιστώσας ώστε να τις απομονώσουμε από την αρχική χρονοσειρά. Το πρώτο αποτέλεσμα είναι 48,4 και το βρίσκουμε από το άθροισμα των πρώτων δύο παρατηρήσεων του κινητού μέσου της προηγούμενης στήλης και το διαιρούμε με το 2. Η δεύτερη τιμή είναι 48,4 και υπολογίζεται από το άθροισμα της δεύτερης και της τρίτης παρατήρησης της στήλης του κινητού μέσου διαιρεμένο με το 2 κτλ.

3ο Βήμα: Στη τελευταία στήλη υπολογίζουμε τους δείκτες ανά τρίμηνο. Δηλαδή διαιρούμε τη στήλη Y_t (αριθμός αφίξεων) με την αντίστοιχη τιμή της στήλης $T_t \cdot C_t$ (δείκτες).
Τα αποτελέσματα φαίνονται στον παρακάτω πίνακα και υπολογίστηκαν μέσω του προγράμματος Excel.

ΠΙΝΑΚΑΣ

Έτος	Τρίμηνο	Αριθμός αφίξεων σε χιλιάδες Y_t	Κινητός Μέσος $(KM)_t$	Κεντρικός Μέσος 4-τριμήνων $T_t \cdot C_t$	Δείκτης $S_t \cdot I_t$
1996	1	28,5	....	....	....
	2	57,5	48,5	....	....
	3	71,5	48,2	48,4	1,5
	4	36,5	48,5	48,4	0,8
1997	1	27,1	50,7	49,5	0,6
	2	58,5	52	51,4	1,2
	3	80,5	50,8	51,4	1,6
	4	41,5	53	51,9	0,5
1998	1	22,5	55,5	54,3	0,5
	2	67	55,5	55,5	1,2
	3	91	56,4	56,4	1,6
	4	41,5	58	57,9	0,7
1999	1	26	59	59,3	0,5
	2	73	61	60,7	1,2
	3	97	61	61,4	1,6
	4	47	63,5	63,6	0,7
2000	1	28,5	65,5	65,5	0,5
	2	82	67	66,8	1,2
	3	104,5	....	....	....
	4	52,5	....	....	....

4^{o} Βήμα: Φτιάχνουμε έναν δεύτερο πίνακα με τρίμηνα και έτη και γράφουμε τις τιμές της τελευταίας στήλης του προηγούμενου πίνακα αντίστοιχα. Αθροίζουμε τους δείκτες αυτούς κάθε στήλης και το αποτέλεσμα το διαιρούμε δια των αριθμό των ετών. Έτσι βρίσκουμε τον μέσο. Δηλαδή στο πρώτο τρίμηνο έχουμε το σύνολο 2,1 δια το 4 και το αποτέλεσμα είναι 0,525. Παρόλα αυτά, τόσο αυτή η τιμή όσο και οι υπόλοιπες τρεις δεν μπορούν να χρησιμοποιηθούν από μόνες τους σαν εποχικοί δείκτες γιατί πρέπει να έχουν άθροισμα 4, ενώ αυτές έχουν άθροισμα 3,975. Γι' αυτό πρέπει να γίνει μία διόρθωση και οι τελικές αυτές τιμές θα είναι οι εποχικοί δείκτες.

5^{o} Βήμα: Αν διαιρέσουμε το 4 με το 3,975 θα βρούμε αποτέλεσμα 0,99375. Αυτός ο αριθμός είναι ένας συντελεστής προσαρμογής. Πολλαπλασιάζουμε λοιπόν κάθε μέσο με τον συντελεστή προσαρμογής. Δηλαδή για το πρώτο τρίμηνο έχουμε $0,525 \cdot 0,99375 = 0,52171875$ κτλ. Αυτοί είναι οι εποχικοί δείκτες και αν τους αθροίσουμε το αποτέλεσμα θα είναι 4.

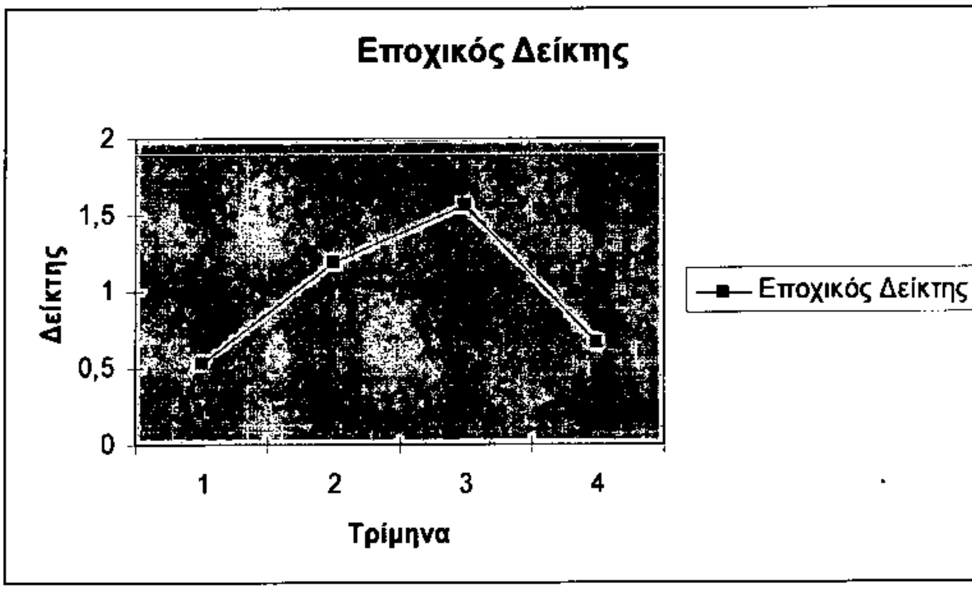
Άρα η απαλλαγμένη από εποχικότητα τιμή για το πρώτο τρίμηνο είναι: $\frac{28,5}{0,52171875} = 54,627 \times 100 = 5462,7$ (αριθμός αφίξεων). Έτσι διαιρούμε κάθε παρατήρηση με τον αντίστοιχο εποχικό δείκτη και πολλαπλασιάζουμε επί 100 ώστε να απαλλάξουμε όλα τα δεδομένα της χρονοσειράς από την εποχική συνιστώσα. Η χρονοσειρά που προκύπτει περιλαμβάνει μόνο την τάση, την κυκλική και την τυχαία συνιστώσα.

ΠΙΝΑΚΑΣ

Έτος		**Τρίμηνα**			
	1	**2**	**3**	**4**	**Σύνολο**
1996	....	....	1,5	0,8	
1997	0,6	1,2	1,6	0,5	
1998	0,5	1,2	1,6	0,7	
1999	0,5	1,2	1,6	0,7	
2000	0,5	1,2	....	....	
Σύνολο	2,1	4,8	6,3	2,7	
Μέσος	0,525	1,2	1,575	0,675	3,975
Εποχικός Δείκτης	0,521	1,192	1,565	0,670	4
		3,975/4=0,993			

Οι εποχικοί δείκτες που υπολογίσαμε πιο πάνω αναπαρίστανται και στο παρακάτω διάγραμμα και παρατηρούμε ότι στο πρώτο και στο τέταρτο τρίμηνο ο αριθμός αφίξεων των τουριστών στην Κύπρο ήταν χαμηλότερος σε σχέση με τα άλλα δύο τρίμηνα.

Αφίξεις τουριστών στη Κύπρο από το 1996 – 2000



ΠΗΓΗ: http//library.ucy.ac.cy

2.8 Έννοια του Σφάλματος

Πριν προχωρήσουμε στην ανάλυση των μεθόδων εξομάλυνσης, είναι σημαντικό να αναφέρουμε την έννοια του σφάλματος. Το σφάλμα (error) είναι η διαφορά της προβλεφθείσας τιμής F_t από την πραγματική τιμή X_t. Ο τύπος είναι: $e_t = X_t - F_t$. Το σφάλμα θεωρείται ένας πολύ σημαντικός εκτιμητής της ακρίβειας της πρόβλεψης και χρησιμοποιείται από τις περισσότερες τεχνικές προβλέψεων. Μια τεχνική για την αξιολόγηση της ακρίβειας των μεθόδων εξομάλυνσης είναι το μέσο σφάλμα τετραγώνου (MSE), που προκύπτει από τη διαίρεση του αθροίσματος τετραγώνων των σφαλμάτων πρόβλεψης (SSE) με τον αριθμό των παρατηρήσεων. Ο τύπος είναι:

$$MSE = \frac{\Sigma(e_t)^2}{n} = \frac{\Sigma(X_t - F_t)^2}{n}$$

Η μικρότερη τιμή του MSE είναι αυτή που λαμβάνεται υπόψη και θεωρείται το πιο αξιόπιστο κριτήριο για την επιλογή του καλυτέρου μοντέλου (Jeffrey Jarrett, 2002). Θα αναφερθούμε εκτενέστερα στην έννοια του σφάλματος και στον σημαντικό του ρόλο στην πρόβλεψη όταν θα μιλήσουμε παρακάτω για τις μεθόδους εξομάλυνσης.

2.9 Μέθοδοι Εξομάλυνσης

Σε πολλές περιπτώσεις η ανάπτυξη και η εφαρμογή κάποιας πολύπλοκης μεθόδου πρόβλεψης δεν είναι καθόλου πρακτική. Αυτό που χρειάζεται είναι μια τεχνική η οποία να μπορεί να εφαρμοστεί εύκολα για όλα τα είδη και να παρέχει ικανοποιητικές προβλέψεις στον βραχυπρόθεσμο ορίζοντα. Σε τέτοιες περιπτώσεις συχνά χρησιμοποιείται ένα είδος μεθόδων πρόβλεψης γνωστό ως μέθοδοι εξομάλυνσης (smoothing methods). Οι μέθοδοι εξομάλυνσης μιας χρονοσειράς αποσκοπούν στην απομάκρυνση της τυχαίας συνιστώσας και υποβαθμίζουν τόσο την τάση, όσο και την εποχική και την κυκλική συνιστώσα (Ανδρικόπουλος Ανδρέας, 1998).

Σε αυτές τις μεθόδους γίνεται χρήση των παλαιών δεδομένων για τον προσδιορισμό μιας "εξομαλυμένης" (smoothed) τιμής για τις χρονοσειρές. Στη συνέχεια αυτή η "εξομαλυμένη" τιμή προεκτείνεται ώστε να αποτελέσει πρόβλεψη για τη μελλοντική τιμή της χρονοσειράς. Στην πράξη, οι επικρατέστερες μέθοδοι εξομάλυνσης χρονοσειρών είναι:

- Η μέθοδος των κινητών μέσων (Moving Average Method)
- Η μέθοδος της εκθετικής εξομάλυνσης (Exponential Smoothing Method)

Πριν την ανάλυση της μεθόδου των κινητών μέσων, χρήσιμο είναι να αναφερθούμε στην μέθοδο του απλού μέσου. Όταν ο απλός μέσος πρόκειται να χρησιμοποιηθεί στη διενέργεια πρόβλεψης, τότε είναι κρίσιμο να καταλάβουμε τις συνθήκες που καθορίζουν την χρησιμότητα του για τον σκοπό αυτό. Οι συνθήκες αυτές προϋποθέτουν ότι τα δεδομένα κατανέμονται τυχαία και ακολουθούν σταθερή εξέλιξη ως προς το χρόνο. Λέγοντας στάσιμα δεδομένα, αναφερόμαστε στο πεδίο των χρονοσειρών που εμφανίζεται όμοιο με μια οριζόντια ευθεία γραμμή με την μεταβλητή του χρόνου να βρίσκεται στον άξονα των X. Εάν τα δεδομένα δεν είναι τυχαία, τότε υπάρχει συστηματική συμπεριφορά ή τάση σε αυτά. Σε αυτήν την περίπτωση άλλη μέθοδος πρόβλεψης μπορεί να προβλέψει την συμπεριφορά αυτή με μεγαλύτερη ακρίβεια από τον μέσο. Επομένως, ο μέσος είναι αξιόπιστος εκτιμητής

για στάσιμες τυχαίες σειρές, δηλαδή για σειρές που δεν παρουσιάζουν πρότυπο ή κάποιο στοιχείο περιοδικότητας.
Στην συνεχεία ακολουθούν παραδείγματα για την αξιολόγηση της χρησιμότητας της μεθόδου αυτής.

Παράδειγμα 1

Ένας διευθυντής μιας αποθήκης προϊόντων θέλει να μάθει πόσα δολάρια πληρώνει σε κάθε προμηθευτή για την παραλαβή σταθερής ποσότητας προϊόντος. Έτσι εξετάζει ένα τυχαίο δείγμα 12 προμηθευτών και παίρνει ως εκτιμητή το 10000. Αυτό συμβαίνει γιατί το ελάχιστο ποσό δολαρίων που δαπανεί είναι 7000$ και αντιστοιχεί στον όγδοο προμηθευτή, ενώ το μεγαλύτερο ποσό δολαρίων που δαπανεί είναι 13000$ και αντιστοιχεί στον ένατο προμηθευτή για αγορά ίδιων ποσοτήτων. Αν βρούμε το μέσο όρο αυτών των δύο ακραίων τιμών, θα έχουμε αποτέλεσμα 10000$. Ο μέσος αυτός είναι ο εκτιμητής που χρησιμοποιεί για κάθε καταναλωτή.
Είναι όμως το 10000 καλός εκτιμητής; Αυτό θα το εξακριβώσουμε αν βρούμε το μέσο τετραγωνικό σφάλμα του 10000 και το συγκρίνουμε με τα μέσα τετραγωνικά σφάλματα των τριών τυχαίων εκτιμητών (7000, 9000, 12000). Όποιος εκτιμητής έχει το μικρότερο μέσο τετραγωνικό σφάλμα αυτός είναι και ο καλύτερος.
Στο παρακάτω πίνακα έχουμε υπολογίσει το μέσο τετραγωνικό σφάλμα με εκτιμητή το 10000. Για ευκολία οι πράξεις έγιναν με απαλοιφή των χιλιάδων και υπολογίστηκαν μέσω του προγράμματος Excel.

Εκτιμητής 10

Προμηθευτής	Ποσό $(χιλ)	Σφάλμα	Τετραγωνικό Σφάλμα(SSE)
1	9	-1	1
2	8	-2	4
3	9	-1	1
4	12	2	4
5	9	-1	1
6	12	2	4
7	11	1	1
8	7	-3	9
9	13	3	9
10	9	-1	1
11	11	1	1
12	10	0	0
Σύνολο			36

Το τετραγωνικό σφάλμα (SSE) ισούται με 36.

Το μέσο τετραγωνικό σφάλμα (MSE) είναι το πηλίκο του τετραγωνικού σφάλματος δια τον αριθμό των παρατηρήσεων. Δηλαδή $\frac{36}{12}=3$.

Εάν υπολογίσουμε τα μέσα τετραγωνικά σφάλματα ορισμένων τυχαίων εκτιμητών (7, 9, 12) θα έχουμε τα αποτελέσματα που βλέπουμε στον παρακάτω πίνακα.

Εκτιμητής	7	9	10	12
Τετραγωνικό Σφάλμα (SSE)	144	48	36	84
Μέσο τετραγωνικό Σφάλμα (MSE)	12	4	3	7

Συμπεραίνουμε ότι ο εκτιμητής με το μικρότερο μέσο τετραγωνικό σφάλμα (MSE) είναι το 10. Επομένως ο διευθυντής προέβλεψε σωστά.

Παράδειγμα 2

Ο επόμενος πίνακας δίνει το καθαρό εισόδημα μιας κατασκευαστικής εταιρίας υπολογιστών από την χρονιά 1985 έως την χρονιά 1994.

ΧΡΟΝΙΑ	ΔΟΛΑΡΙΑ (ΕΚΑΤΟΜΜΥΡΙΑ)	ΜΕΣΟΣ	ΣΦΑΛΜΑ	ΣΦΑΛΜΑ ΤΕΤΡΑΓΩΝΟΥ
1985	46,163	48,776	-2,613	6,828
1986	46,998	48,776	-1,778	3,161
1987	47,816	48,776	-0,960	0,922
1988	48,311	48,776	-0,465	0,216
1989	48,758	48,776	-0,018	0,000
1990	49,164	48,776	0,388	0,151
1991	49,548	48,776	0,772	0,596
1992	48,915	48,776	0,139	0,019
1993	50,315	48,776	1,539	2,369
1994	50,768	48,776	1,992	3,968

Ο τύπος για να υπολογίσουμε τον αριθμητικό μέσο είναι:

$$\overline{X} = \frac{1}{n} \sum_{i=1}^{n} X_i$$

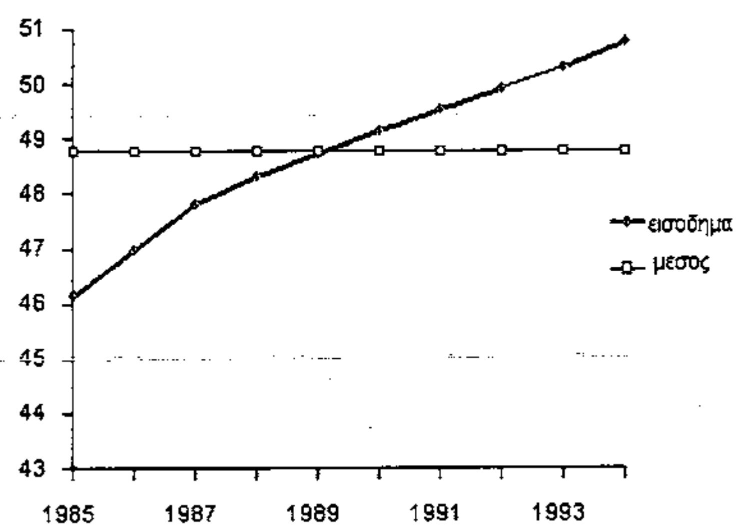
Όπου X_i οι παρατηρήσεις μας, δηλαδή τα δολάρια.

Όπου n ο αριθμός των παρατηρήσεων.

Ο μέσος μετά από υπολογισμό ισούται με 48.776.

Το μέσο σφάλμα του τετραγώνου (MSE) ισούται με 1,823.

<u>*Το ερώτημα που προκύπτει είναι:*</u> μπορούμε να χρησιμοποιήσουμε τον μέσο για να προβλέψουμε το εισόδημα εάν υποπτευόμαστε ότι υπάρχει τάση;
Το παρακάτω γράφημα δείχνει καθαρά ότι δεν μπορούμε να το προβλέψουμε.



ΠΗΓΗ: http://www.itl.nist.gov

Όπως βλέπουμε στο διάγραμμα, υπάρχει μεγάλη απόκλιση μεταξύ των γραμμών του μέσου και του εισοδήματος, άρα η προβλεπτική ικανότητα του μέσου είναι αμφιλεγόμενη.

Γενικότερα, η χρήση του αριθμητικού μέσου ως εργαλείο πρόβλεψης περιορίζεται εξαιτίας του μεγάλου αριθμού παρατηρήσεων ενός δείγματος. Για τα δεδομένα των χρονοσειρών, το μέγεθος του δείγματος αυξάνεται συνεχώς με τον χρόνο καθώς καταγράφονται νέες παρατηρήσεις. Οι παρατηρήσεις του νέου δείγματος συμπεριλαμβάνονται στον υπολογισμό του αριθμητικού μέσου, δημιουργώντας προβλήματα καταχώρησης, καθώς και υπολογιστικά όταν χρειάζονται προβλέψεις που αφορούν μεγάλο αριθμό μεγεθών.

Τέλος, ο αριθμητικός μέσος μιας χρονοσειράς δεν περιλαμβάνει τις μεταβολές της μέσης τιμής διαχρονικά. Για παράδειγμα, αν υπάρξει μια μεταβολή στα δεδομένα ο αριθμητικός μέσος δεν θα την λάβει υπόψη και η μέθοδος πρόβλεψης θα είναι σχετικά ανακριβής (όπως στο προηγούμενο παράδειγμα) για αρκετά μεγάλη χρονική περίοδο.

2.9.1 Μέθοδος των Κινητών Μέσων

Ένας τρόπος για την ελαχιστοποίηση της ανακριβείας είναι να κρατήσουμε τον ίδιο αριθμό παρατηρήσεων στο δείγμα κατά τον υπολογισμό του μέσου. Το αποτέλεσμα θα είναι μια διαδικασία υπολογισμού του μέσου των πιο πρόσφατων σταθερών παρατηρήσεων του δείγματος. Η διαδικασία αυτή όπως έχει αναφερθεί, ονομάζεται μέθοδος του κινητού μέσου όρου, λόγω του υπολογισμού ενός νέου μέσου όρου από την εισαγωγή κάθε νέας παρατήρησης στο δείγμα και απόρριψης της παλαιότερης. Συνεπώς κάθε πρόβλεψη χρησιμοποιεί τον ίδιο αριθμό παρατηρήσεων δείγματος από την χρονοσειρά και περιλαμβάνει μόνο την πλέον πρόσφατη παρατήρηση.
Η διαδικασία εξομάλυνσης χρονοσειρών με την μέθοδο των κινητών μέσων περιλαμβάνει τα παρακάτω στάδια:

ΣΤΑΔΙΟ 1: Από τις παρατηρήσεις της χρονοσειράς και βασιζόμενοι στην περίοδο, για την οποία επιθυμούμε τον υπολογισμό των κινητών μέσων (4-χρονικές περίοδοι, 3-χρονικές περίοδοι κ.τ.λ.), υπολογίζουμε τα κινητά σύνολα (moving totals).

ΣΤΑΔΙΟ 2:Τα κινητά σύνολα διαιρούνται με τον αριθμό των παρατηρήσεων σε κάθε υποσύνολο, το πηλίκο των οποίων μας δίνει τους κινητούς μέσους. Για χρονοσειρές με περιττό αριθμό παρατηρήσεων δεν υπολογίζονται κινητοί μέσοι για της πρώτες και τις τελευταίες (k-1)/2 χρονικές περιόδους. Όπου k ισούται ο αριθμός παρατηρήσεων για κάθε κινητό σύνολο.

ΣΤΑΔΙΟ 3: Παίρνουμε τη διαφορά ανάμεσα στις τιμές της χρονοσειράς και των κινητών μέσων. Αυτή η διαφορά μετράει τις αποκλίσεις της χρονοσειράς γύρω από τους κινητούς μέσους
Η μέθοδος του κινητού μέσου όρου, σε αλγεβρική μορφή, διατυπώνεται ως εξής:

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

όπου t είναι η πιο πρόσφατη παρατήρηση και t +1 είναι η επόμενη περίοδος. Στην μορφή αυτή απαιτείται ο αναλυτής να διαθέτει τις τιμές των παρατηρήσεων του παρελθόντος.

Με την προσθήκη μιας νέας παρατήρησης και την εξάλειψη της παλαιότερης, μπορούμε να αναδιατυπώσουμε τον τύπο ως εξής:

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

__Παράδειγμα__

Σύμφωνα με τα δεδομένα του «Παραδείγματος 1» θα υπολογίσουμε πάλι το μέσο τετραγωνικό σφάλμα (MSE) του κινητού μέσου και θα τα συγκρίνουμε. Υπολογίζουμε τον κινητό μέσο 3 όρων. Δηλαδή $(9+8+9)/3=8,667$ κτλ.

ΠΙΝΑΚΑΣ

Προμηθευτής	Ποσό $ (χιλ)	Κινητός Μέσος (MA)	Σφάλμα	Μέσο Τετραγωνικό Σφάλμα (MSE)
1	9	,,,,	,,,,	,,,,
2	8	,,,,	,,,,	,,,,
3	9	8,667	0,333	0,111
4	12	9,667	2,333	5,444
5	9	10,000	-1,000	1,000
6	12	11,000	1,000	1,000
7	11	10,667	0,333	0,111
8	7	10,000	-3,000	9,000
9	13	10,333	2,667	7,111
10	9	9,667	-0,667	0,444
11	11	11,000	0	0
12	10	10,000	0	0
Σύνολο				24,221

Αν διαιρέσουμε το μέσο τετραγωνικό σφάλμα (MSE) με το 12 το σύνολο των παρατηρήσεων θα έχουμε 2,018 αποτέλεσμα. Επομένως το μέσο τετραγωνικό σφάλμα του κινητού μέσου (2,018) είναι μικρότερο από το μέσο τετραγωνικό σφάλμα του απλού μέσου (3) που βρήκαμε στο προηγούμενο παράδειγμα. Άρα η μέθοδος του κινητού μέσου είναι καλύτερη.

Ο κινητός μέσος όρος συχνά χρησιμοποιείται για την πρόβλεψη περιορισμένων εμπορικών συναλλαγών που γίνονται κυρίως στα πλαίσια του προγράμματος ελέγχου των αποθεμάτων. Επίσης για προβλέψεις μικρού εύρους, για δραστηριότητες που χρειάζονται μεγάλο αριθμό μεγεθών όπως του σχεδιασμού της τιμολόγησης και της χρονικής κατανομής της παραγωγής.

Παρόλα αυτά η μέθοδος των κινητών μέσων παρουσιάζει τρία βασικά μειονεκτήματα που περιορίζουν την εφαρμογή της:

1. Δεν περιλαμβάνει κινητούς μέσους για μερικές από τις πρώτες και τις τελευταίες παρατηρήσεις της χρονοσειράς. Αυτό το γεγονός, ειδικά για περιπτώσεις που η χρονοσειρά έχει περιορισμένο αριθμό παρατηρήσεων, συντελεί στην απώλεια σχετικών με τη χρονοσειρά πληροφοριών.

2. Δε λαμβάνει υπόψη τιμές παρελθόντων ετών της χρονοσειράς, δίνει την ίδια βαρύτητα σε πρόσφατες τιμές και σε τιμές παλαιοτέρων χρονικών περιόδων και η αποτελεσματικότητα της μεθόδου εξαρτάται από το k, δηλαδή την χρονική διάρκεια υπολογισμού των κινητών μέσων.

3. 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.

Γενικά, η μέθοδος του κινητού μέσου όρου είναι πιο αποτελεσματική από την χρησιμοποίηση του αριθμητικού μέσου μιας ολόκληρης χρονοσειράς. Είναι πιο εύκαμπτη γιατί το n μπορεί να μεταβληθεί ώστε να ανταποκρίνεται στα παρατηρούμενα πρότυπα των δεδομένων.

Δυστυχώς, ούτε ο μέσος όρος όλων των στοιχείων ούτε ο κινητός μέσος όρος των πιο πρόσφατων τιμών, όταν χρησιμοποιούνται για προβλέψεις επόμενων περιόδων, δεν είναι σε θέση να αντιμετωπίσουν μια σημαντική τάση. Μια τέλεια γραμμική τάση στη χρονοσειρά έχει σαν αποτέλεσμα ένα συστηματικό σφάλμα που δεν μπορούμε να αποφύγουμε. Γι' αυτό δημιουργήθηκε μια παραλλαγή στη διαδικασία του κινητού μέσου που κάνει συχνά ένα καλύτερο υπολογισμό της τάσης, που ονομάζεται γραμμικός κινητός μέσος (Linear Moving Average).

Η μέθοδος αυτή απαιτεί τον υπολογισμό ενός δεύτερου κινητού μέσου, δηλαδή έναν κινητό μέσο του κινητού μέσου των τιμών που παρατηρηθήκαν. Αυτοί οι διπλοί κινητοί μεσοί (Double Moving Averages) υστερούν σε σύγκριση με τον απλό κινητό μέσο (SMA) κατά το ίδιο ποσό που ο απλός κινητός μέσος υστερεί σε σχέση με τις παρατηρηθείσες τιμές. Έτσι, οι διαφορές ανάμεσα στις παρατηρηθείσες τιμές και στον απλό κινητό μέσο θα είναι οι ίδιες με τις διαφορές ανάμεσα στον απλό και στον διπλό κινητό μέσο.

2.9.2 Μέθοδος της Εκθετικής Εξομάλυνσης

Η Εκθετική Εξομάλυνση είναι μία μέθοδος πρόβλεψης η οποία προεκτείνει στοιχεία του προτύπου των παρελθόντων δεδομένων, όπως τάσεις και εποχικούς κύκλους στο μέλλον. Οι προβλέψεις υπολογίζονται μετά από εξομάλυνση των δεδομένων, προκειμένου να απομονωθούν τα πραγματικά πρότυπα από τις καθαρά τυχαίες διακυμάνσεις. Η βασική αρχή των μεθόδων εκθετικής εξομάλυνσης είναι ότι όσο πιο πρόσφατα είναι τα δεδομένα, τόσο περισσότερη πληροφορία εμπεριέχουν. Η μέθοδος είναι παρόμοια με την μέθοδο του κινητού μέσου όρου, όμως επιτρέπει στους ερευνητές να διορθώσουν προγενέστερες ανακρίβειες στις προβλέψεις. Με άλλα λόγια, η μέθοδος αυτή έχει το πλεονέκτημα της διενέργειας μιας απλής, έγκαιρης πρόβλεψης, η οποία είναι όμοια με τη παλαιότερη πρόβλεψη μαζί με κάποια δηλωμένη αναλογία προβλεπτικού σφάλματος για την προηγούμενη περίοδο. Οι μέθοδοι της εκθετικής εξομάλυνσης χρησιμοποιούνται συνήθως για την πρόβλεψη μεγάλου αριθμού μεγεθών, όπως στην περίπτωση του σχεδιασμού των αναγκών σε υλικά, τον έλεγχο πωλήσεων, τις προβλέψεις κέρδους και άλλα οικονομικά δεδομένα.

Παρακάτω αναφέρονται διάφορες μέθοδοι εκθετικής εξομάλυνσης.

2.9.2.1 Εκθετική Εξομάλυνση μίας Παραμέτρου

Η μέθοδος αυτή απλοποιεί τους υπολογισμούς πρόβλεψης και έχει μικρές απαιτήσεις σε δεδομένα. Επίσης η μέθοδος αυτή δίνει αυτό-διορθούμενες προβλέψεις αφού διαθέτει μηχανισμό ενσωματωμένων προσαρμογών, ο οποίος ρυθμίζει τις τιμές αλλάζοντάς τες προς την αντίθετη κατεύθυνση από εκείνη που κινήθηκαν τα λάθη προηγούμενων περιόδων. Για να αναπτύξουμε αυτή τη μέθοδο αρχίζουμε με την εξίσωση:

$$F_{t+1} = \frac{X_t}{n} - \frac{F_t}{n} + F_t$$

Εάν κάνουμε την αντικατάσταση $F_t = X_{t-n}$, παραπέρα, η εξίσωση αυτή μπορεί να γραφεί ως:

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

Η προηγούμενη εξίσωση δίνει μια πρόβλεψη βασισμένη στη στάθμιση της πιο πρόσφατης παρατήρησης με συντελεστή $1/n$ και της πιο πρόσφατης πρόβλεψης με συντελεστή στάθμισης $\lfloor 1-(1/n) \rfloor$. 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 α, το βασικό μοντέλο γράφεται ως εξής:

$$F_{t+1} = \alpha X_t + (1-\alpha) F_t$$

Όπου:

t: είναι η τρέχουσα περίοδος,

F_{t+1}, F_t : είναι οι τιμές πρόβλεψης για την επόμενη κα την τρέχουσα περίοδο,

X_t : είναι η τιμή που παρατηρήθηκε την τρέχουσα περίοδο,

α : είναι η σταθερή εξομάλυνση και παίρνει τιμές $[0,1]$.

Η παραπάνω μέθοδος δεν είναι ακριβής όταν υπάρχει η επίδραση της τάσης στη χρονοσειρά. Στο ακόλουθο παράδειγμα παρατηρείται αυτή η ανακρίβεια.

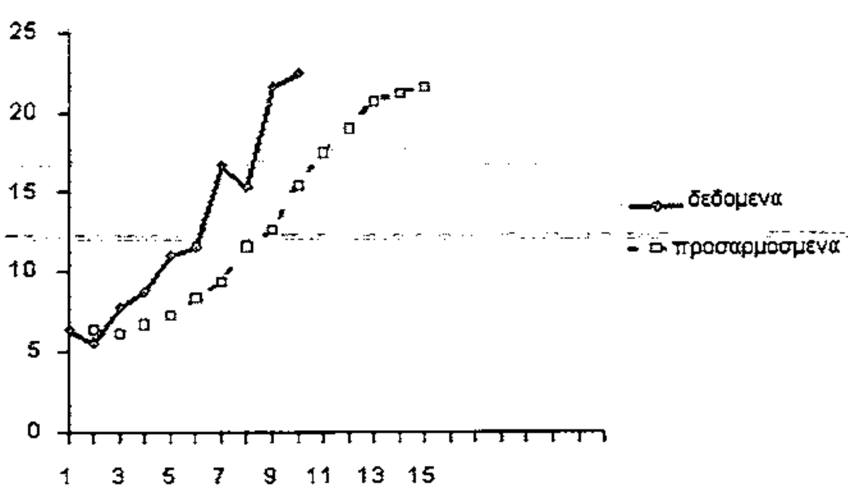
Παράδειγμα

Περίοδος	Παρατηρήσεις Xt	Πρόβλεψη F
1	6,4	...
2	5,6	6,4
3	7,8	6,2
4	8,8	6,7
5	11	7,3
6	11,6	8,4
7	16,7	9,4
8	15,3	11,6
9	21,6	12,7
10	22,4	15,4

Προσαρμόζοντας τα δεδομένα του πίνακα στον τύπο της εκθετικής εξομάλυνσης και με a=0,3 έχουμε π.χ για την πρόβλεψη της τρίτης περιόδου:

$$F_3 = aX_2 + (1-a)F_2 \Rightarrow$$
$$F_3 = 0,3 \cdot 5,6 + (1-0,3) \cdot 6,4 \Rightarrow$$
$$F_3 \simeq 6,20$$

Με το σκεπτικό αυτό παίρνουμε και τις προβλέψεις για τις επόμενες περιόδους. Ακολουθεί το διάγραμμα.



ΠΗΓΗ: http://www.itl.nist.gov

Το διάγραμμα μας δείχνει ότι υπάρχει ακαταλληλότητα της μεθόδου της εκθετικής εξομάλυνσης μιας παραμέτρου στην πρόβλεψη όταν υπάρχει τάση.

Μια καλύτερη μέθοδος πρόβλεψης όταν υπάρχει τάση είναι η μέθοδος της διπλής εκθετικής εξομάλυνσης. Προσαρμόζοντας τον τύπο της απλής εκθετικής εξομάλυνσης στη διπλή έχουμε:

$$F'_{t+1} = \alpha \cdot F_{t+1} + (1-\alpha) \cdot F'_t$$

Οι τιμές τόσο της απλής όσο και της διπλής εκθετικής εξομάλυνσης υστερούν σε σχέση με τις πραγματικές τιμές της χρονοσειράς.

Το κύριο προτέρημα της μεθόδου εκθετικής εξομάλυνσης μιας παραμέτρου είναι οι περιορισμένες απαιτήσεις σε δεδομένα και η σχετική της απλότητα. Παρά την ελκυστικότητα που εμφανίζει η μέθοδος αυτή παρατηρούνται ορισμένοι σημαντικοί περιορισμοί:

1. Οι προβλέψεις που προέρχονται από την εκθετική εξομάλυνση είναι ευαίσθητες στην εξειδίκευση της σταθεράς της εξομάλυνσης. Η επιλογή του α βασίζεται σε μια διαδικασία δοκιμασίας και εκτίμησης (trial-and-error) του σφάλματος.

2. Οι τεχνικές της εκθετικής εξομάλυνσης οδηγούν σε προβλέψεις που καθυστερούν ως προς τα σημεία καμπής των πραγματικών δεδομένων της χρονοσειράς.

Για να αξιολογήσουμε την χρησιμότητα της τεχνικής της εκθετικής εξομάλυνσης πρέπει να εκτιμήσουμε το σφάλμα πρόβλεψης. Μια εξέταση του σφάλματος στην πρόβλεψη επιτρέπει στον αναλυτή να αξιολογήσει αν η τεχνική προσεγγίζει ικανοποιητικά τα πρότυπα που ακολουθούν οι παρατηρήσεις του δείγματος. Αξιολόγηση της αξιοπιστίας οποιασδήποτε τεχνικής απαιτεί τον προσδιορισμό των κριτηρίων. Όμως, δεν υπάρχει γενικά παραδεχτό άριστο μέτρο. Υπάρχει ένας αριθμός δεικτών που χρησιμοποιούνται για την αξιολόγηση της ακρίβειας της πρόβλεψης. Μια τεχνική που είναι γενικά παραδεκτή για την αξιολόγηση των μεθόδων της εκθετικής εξομάλυνσης (και πολλών άλλων) είναι το μέσο σφάλμα τετραγώνου (Mean Square Error: MSE):

$$MSE = \frac{\Sigma(e_t)^2}{n} = \frac{\Sigma(X_t - F_t)^2}{n}$$

Το μέτρο αυτό ορίζει το σφάλμα ως το άθροισμα των τετραγώνων των σφαλμάτων της πρόβλεψης, e_t, διαιρεμένο με το μέγεθος του δείγματος, δηλαδή τον αριθμό των περιόδων της πρόβλεψης. Η χρήση αυτής της τεχνικής βασίζεται στον ορισμό ποια τιμή της σταθεράς a, είναι η καλύτερη, δηλαδή ποια έχει το μικρότερο μετρήσιμο σφάλμα. Άρα η επιλογή του συνόλου προβλέψεων βασίζεται σε ποιο σύνολο έχει το μικρότερο MSE.

Μια άλλη τεχνική, μέτρησης της αξιολόγησης του σφάλματος της πρόβλεψης είναι η απόλυτη μέση απόκλιση (Mean Absolute Deviation) και βασίζεται στην υπόθεση ότι η σοβαρότητα του σφάλματος σχετίζεται γραμμικά με το μέγεθος του. Αυτή ορίζεται από την σχέση:

$$MAD = \frac{\Sigma|e|}{n}$$

Είναι το άθροισμα των απολύτων τιμών του σφάλματος διαιρούμενο με τον αριθμό των προβλέψεων. Θεωρητικά σαν μέτρο του σφάλματος πρόβλεψης είναι λιγότερο αξιόπιστο από το MSE.

Όπως προαναφέραμε η μέθοδος του μέσου τετραγωνικού σφάλματος χρησιμοποιείται για τον καθορισμό της τιμής της σταθεράς a. Η σταθερά αυτή καθορίζει την έκταση που οι προβλέψεις του παρελθόντος επηρεάζουν την πρόβλεψη. Η επιλογή μιας μικρής ή μιας μεγάλης τιμής για την σταθερά εξομάλυνσης, ως εκτιμητή ακρίβειας της πρόβλεψης είναι αδιευκρίνιστη μέχρι στιγμής, χωρίς αξιοπιστία στα αποτελέσματα της. Για να λύσουμε το πρόβλημα της επιλογής εξετάζουμε το MSE στις ιστορικές χρονοσειρές, χρησιμοποιώντας διαφορετικές τιμές για την σταθερά της εξομάλυνσης. Για κάθε τιμή πάνω από ένα εύρος τιμών, προετοιμάζεται ένα πρότυπό πρόβλεψης με την μέθοδο της εκθετικής εξομάλυνσης και υπολογίζεται το κατάλληλο μέτρο της ακρίβειας της πρόβλεψης. Στην πράξη, μελέτες που έγιναν, δείχνουν ότι τιμές της σταθεράς a από 0,05 έως 0,30 ταιριάζουν πολύ καλά στα πρότυπα της εκθετικής εξομάλυνσης. Συνοψίζοντας, η σταθερά εξομάλυνσης που απορρέει από το τελευταίο MSE είναι η τιμή που θα έπρεπε να είχε επιλεγεί για την πρόβλεψη. Καθώς παίρνονται όλο και νέα δεδομένα, η διαδικασία επαναλαμβάνεται για να δημιουργηθούν καινούργιες «καλύτερες» σταθερές εξομάλυνσης. Εάν η «καλύτερη» σταθερά εξομάλυνσης είναι μεγαλύτερη από 0,3 τότε προτείνεται να ακολουθηθεί άλλο μοντέλο πρόβλεψης.

2.9.2.2 Γραμμική Εκθετική Εξομάλυνση Μίας Παραμέτρου του Brown

Όπως αναφέραμε η απλή και η διπλή εκθετική εξομάλυνση υστερούν σε σχέση με τις πραγματικές τιμές της χρονοσειράς. Στο σημείο αυτό η μέθοδος του Brown πλεονεκτεί έναντι της απλής και της διπλής εκθετικής εξομάλυνσης. Με τη μέθοδό του λοιπόν, η διαφορά αυτή ανάμεσα στις απλές και στις εξομαλυνθείσες τιμές, προστίθεται στην απλή εξομαλυνθείσα τιμή και έτσι γίνεται προσαρμογή των δεδομένων σε κάποιο πρότυπο (Jeffrey Jarrett, 2002).

Οι εξισώσεις για αυτές τις προσαρμογές είναι:

$$\alpha_t = S_t + (S_t - S'_t) = 2S_t - S'_t$$
$$b_t = \frac{\alpha}{1-\alpha}(S_t - S'_t)$$
$$F_{t+m} = \alpha_t + b_t m$$

Όπου

S_t είναι η τιμή μετά από την απλή εκθετική εξομάλυνση.

$S_t^{'}$ είναι η τιμή μετά από την διπλή εκθετική εξομάλυνση.

α_t είναι η διαφορά για κάθε περίοδο

b_t είναι ο συντελεστής ανάπτυξης και

m είναι ο αριθμός των μελλοντικών περιόδων που πρόκειται να προβλεφθούν.

Γενικά η διαφορά ανάμεσα στις τιμές της απλής και της διπλής εκθετικής εξομάλυνσης είναι ίδια με την διαφορά ανάμεσα στις πραγματικές και στις απλές εκθετικές τιμές. Αυτή τη διαφορά δηλώνει το α_t. Στη συνέχεια υπολογίζοντας το

b_t από τον κατάλληλο τύπο προσθέτουμε το αποτέλεσμα στο α_t για να προβλέψουμε μία περίοδο μπροστά. Αν θέλουμε να προβλέψουμε για περισσότερες από μία μελλοντικές περιόδους αντικαθιστούμε το m που θέλουμε στον τελευταίο τύπο ο οποίος μας δίνει την πρόβλεψη. Γνωρίζουμε βέβαια ότι όσο μεγαλύτερη πρόβλεψη θέλουμε να κάνουμε τόσο μεγαλώνει και η ανακρίβεια στο αποτέλεσμα που θα πάρουμε. Επομένως η πρόβλεψη και η ανακρίβεια είναι ποσά ανάλογα.

2.9.2.3 Γραμμική Εκθετική Εξομάλυνση Διπλής Παραμέτρου του Holt

Η γραμμική εκθετική εξομάλυνση διπλής παραμέτρου του Holt είναι μια επέκταση της μεθόδου του Brown όπου προστίθεται ένας συντελεστής ανάπτυξης στην εξίσωση εξομάλυνσης. Έτσι η μέθοδος αυτή εξομαλύνει τις τιμές της τάσης απευθείας. Άρα τώρα έχουμε δύο σταθερές εξομάλυνσης και τρεις εξισώσεις οι οποίες είναι:

$$S_t = \alpha X_t + (1-\alpha)(S_{t-1} + b_{t-1})$$

$$b_t = \beta(S_t - S_{t-1}) + (1-\beta)b_{t-1}$$

$$F_{t+m} = S_t + b_t m$$

Όπου:

b_t είναι ο συντελεστής ανάπτυξης και

β είναι η νέα σταθερά εξομάλυνσης για την τάση.

Ο πρώτος τύπος προσαρμόζει το S_t κατευθείαν για την αύξηση της προηγούμενης περιόδου b_{t-1}, προσθέτοντάς το στην εξομαλυνθείσα τιμή S_{t-1}, η οποία αφορά την προηγούμενη περίοδο. Επομένως η προηγούμενη εξομαλυνθείσα τιμή S_{t-1}, προσαρμόζεται άμεσα με τον συντελεστή της τάσης με αποτέλεσμα να ελαχιστοποιείται η καθυστέρηση στην εξομάλυνση. Η τάση γίνεται επίκαιρη όπως βλέπουμε στη δεύτερη εξίσωση και εκφράζεται ως η διαφορά ανάμεσα στις δύο τελευταίες εξομαλυνθείσες τιμές. Η τελευταία εξίσωση μας δίνει την πρόβλεψη των μελλοντικών περιόδων που επιθυμούμε προσθέτοντας στο S_t τον συντελεστή ανάπτυξης b_t, m φορές.

Παράδειγμα

Έχοντας τα ίδια δεδομένα με το προηγούμενο παράδειγμα, θα γίνει σύγκριση της εκθετικής εξομάλυνσης μιας παραμέτρου με την γραμμική εκθετική εξομάλυνση διπλής παραμέτρου του Holt. Για να γίνει σύγκριση όμως θα πρέπει να προσδιορίσουμε την τιμή του α με το μικρότερο MSE. Άρα στην περίπτωση της εκθετικής εξομάλυνσης μιας παραμέτρου το α =0,977 με MSE 8,8867. Ενώ για την γραμμική εξομάλυνση διπλής παραμέτρου η σταθερά εξομάλυνσης α, είναι ίση με 0,3623 και η σταθερά εξομάλυνσης για την τάση β, είναι ίση με 1 με MSE=3,7024. Η αρχική τιμή για το b_1 ισούται με 0,8, απορρέοντας από τον τύπο:

$$b_1 = \frac{\lfloor (y_2 - y_1) + (y_3 - y_2) + (y_4 - y_3) \rfloor}{3}$$

ΠΕΡΙΟΔΟΣ	ΠΑΡΑΤΗΡΗΣΕΙΣ	(St) α=0,977	(St) α=0,3623	(bt) β=1	ΠΡΟΒΛΕΨΗ m=1 Ft+m
1	6,4		6,4	0,8	
2	5,6	6,4	6,6	0,2	7,2
3	7,8	5,6	7,2	0,6	6,8
4	8,8	7,8	8,1	1	7,8
5	11	8,8	9,8	1,7	9,1
6	11,6	10,9	11,5	1,7	11,5
7	16,7	11,6	14,5	3	13,2
8	15,3	16,6	16,7	2,2	17,5
9	21,6	15,3	19,9	3,2	18,9
10	22,4	21,5	22,8	2,9	23,1

Παρατήρηση: Οι τιμές για τις σταθερές α και β μπορούν να προσδιορισθούν μέσω τεχνικών μη-γραμμικής βελτιστοποίησης, όπως του αλγορίθμου Marquardt.

Σύμφωνα με τους τύπους της γραμμικής εκθετικής εξομάλυνσης διπλής παραμέτρου:
Για την πρόβλεψη της δεύτερης περιόδου έχουμε:

$$F_2 = 6,4 + 0,8 = 7,2$$

Για την πρόβλεψη της τρίτης περιόδου έχουμε:

$$S_2 = 0,3623 \cdot 5,6 + (1-0,3623) \cdot (6,4+0,8) = 6,6$$
$$b_2 = 1 \cdot \left[6,6 - 6,4 + (1-1) \cdot 0,8\right] = 0,2$$
$$F_3 = 6,6 + 0,2 = 6,8$$

Για να δούμε την ακρίβεια των μεθόδων στην πρόβλεψη επόμενων περιόδων υπολογίζουμε τις πρώτες πέντε τιμές από την τελευταία παρατήρηση. Γιατί όπως ξέρουμε η εκθετική εξομάλυνση είναι μια μέθοδος, που η πρόβλεψη της τελευταίας περιόδου, είναι αυτή που πρέπει να διασωθεί.

ΠΕΡΙΟΔΟΣ	ΠΡΟΒΛΕΨΗ ΜΕ ΑΠΛΗ ΕΞΟΜΑΛΥΝΣΗ	ΠΡΟΒΛΕΨΗ ΜΕ ΔΙΠΛΗ ΕΞΟΜΑΛΥΝΣΗ
11	22,4	25,7
12	22,4	28,6
13	22,4	31,5
14	22,4	34,4
15	22,4	37,3

Για την απλή εκθετική εξομάλυνση της $11^{ης}$ περιόδου έχουμε:

$$S_{t+1} = \alpha X_t + (1-\alpha) S_t$$

$$S_{11} = 0,977 \cdot 22,4 + (1-0,977) \cdot 21,5 \Rightarrow$$
$$S_{11} = 21,88 + 0,49 = 22,4$$

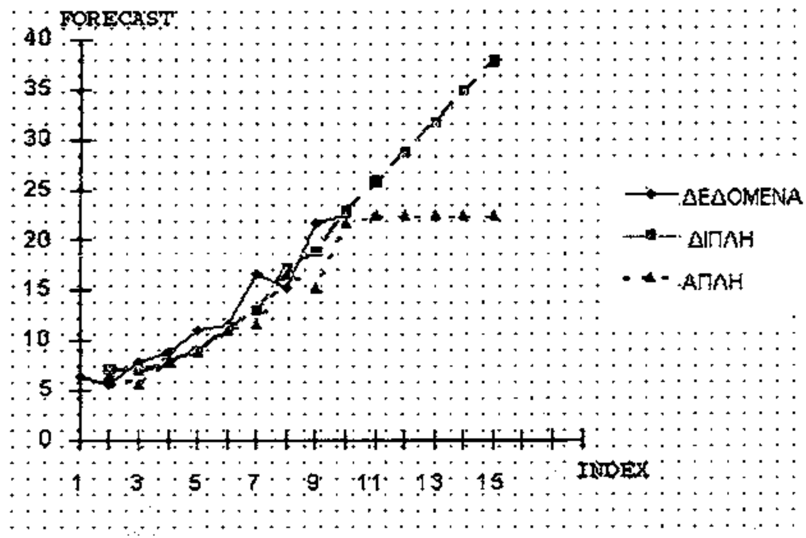
Το ίδιο αποτέλεσμα βρίσκουμε και για τις επόμενες 4 περιόδους.

Για την γραμμική εκθετική εξομάλυνση διπλής παραμέτρου έχουμε:

$F_{t+m} = S_t + b_t m$, δηλαδή

$$F_{11} = 22,8 + (2,9 \cdot 1) = 25,7$$
$$F_{12} = 22,8 + (2,9 \cdot 2) = 28,6$$

Με τον ίδιο τρόπο προβλέπουμε και τις επόμενες περιόδους αυξάνοντας κάθε φορά το m κατά 1.



ΠΗΓΗ: http://www.itl.nist.gov

Το γράφημα δείχνει ότι η μέθοδος της γραμμικής εκθετικής εξομάλυνσης διπλής παραμέτρου του Holt, ακολουθεί τα δεδομένα πιο πιστά από ότι η μέθοδος της απλής εξομάλυνσης, δηλαδή διαχειρίζεται τα δεδομένα με τάση με μεγαλύτερη ακρίβεια. Επιπλέον, η μέθοδος της απλής εξομάλυνσης δεν μπορεί να κάνει κάτι καλύτερο από ότι να προβλέπει μια ευθεία οριζόντια γραμμή, η οποία είναι πιθανόν να μην συμβαίνει στην πραγματικότητα. Άρα σε αυτήν την περίπτωση η μέθοδος της διπλής εξομάλυνσης προτιμάτε.

2.9.2.4 Εκθετική Εξομάλυνση σε Εποχικά Δεδομένα του Winters

Σύμφωνα με τον Jeffrey Jarrett η μέθοδος του Winters μπορεί να εφαρμοστεί για προβλέψεις με βάση μία χρονοσειρά που εμφανίζει και τάση και εποχικότητα. Το μοντέλο αυτό αποτελεί επέκταση του μοντέλου γραμμικής εκθετικής εξομάλυνσης διπλής παραμέτρου του Holt. Προσθέτει δηλαδή άλλη μία εξίσωση για τον υπολογισμό της εποχικότητας στην χρονοσειρά. Οι εξισώσεις είναι οι ακόλουθες:

$$S_t = \alpha \frac{X_t}{I_{t-L}} + (1-\alpha)(S_{t-1} + b_{t-1})$$

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

$$b_t = v(S_t - S_{t-1}) + (1-v)b_{t-1}$$

$$F_{t+m} = (S_t + b_t m)\mathrm{I}_{t-L+m}$$

Όπου:

L είναι το εύρος της εποχικότητας και

I είναι ο παράγοντας προσαρμογής της εποχικότητας.

v είναι η σταθερά εξομάλυνσης για την τάση.

Στην πρώτη εξίσωση, με μια μικρή διαφοροποίηση από την αντίστοιχη του Holt, γίνονται επίκαιρες οι εξομαλυνθείσες τιμές της χρονοσειράς.

Η δεύτερη εξίσωση δίνει την εκτίμηση της εποχικότητας, X_t / S_t, πολλαπλασιασμένη επί τη σταθερά β συν την παλιά εποχική εκτίμηση, I_{t-L}, πολλαπλασιασμένη επί $(1-\beta)$.

Η τρίτη εξίσωση εκφράζει την τωρινή εξίσωση της συνιστώσας της τάσης που υπολογίζεται με τη γνωστή διαδικασία της εκθετικής εξομάλυνσης.

Η τελευταία εξίσωση χρησιμοποιείται για την πρόβλεψη των μελλοντικών περιόδων. Είναι παρόμοια με αυτή του Holt, μόνο που η εκτίμηση για τη μελλοντική περίοδο $t+m$ πολλαπλασιάζεται επί I_{t-L+m}. Αυτός είναι ο τελικός διαθέσιμός εποχικός δείκτης και αποτελεί την προσαρμογή της πρόβλεψης για εποχικότητα.

2.10 Συμπέρασμα

Όπως αναφερθήκαμε στο παρόν κεφάλαιο, οι χρονοσειρές μελετώνται κάτω από την υπόθεση ότι με βάση την 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 στις συνιστώσες της έχει αξία στην διαδικασία της πρόβλεψης γιατί έτσι μπορεί να μειωθεί το σφάλμα πρόβλεψης. Βέβαια, κάθε αναλυτής που θέλει να κάνει μια πρόβλεψη αξιολογεί με τα δικά του κριτήρια αυτές τις διαδικασίες ανάλυσης και διάσπασης χρονοσειρών. Επομένως, απαιτείται καλή αξιολογική κρίση προκειμένου να χρησιμοποιηθούν αυτές οι μέθοδοι σε προβλήματα πρόβλεψης.

Βασικό ρόλο στην δημιουργία ενός αξιόπιστου προτύπου όσον αφορά την ακρίβεια της πρόβλεψης, έχουν οι μέθοδοι εξομάλυνσης (Smoothed Methods).

Αυτές οι μέθοδοι προσαρμόζουν τις όποιες συνιστώσες της χρονοσειράς στο βασικό πρότυπο, είτε απομακρύνοντας, είτε υποβαθμίζοντας τες. Οι βασικές μέθοδοι εξομάλυνσης είναι του κινητού μέσου και της εκθετικής εξομάλυνσης.

Αποτελέσματα εμπειρικών μελετών αποδεικνύουν ότι οι μέθοδοι εκθετικής εξομάλυνσης παρουσιάζουν ικανοποιητικά ποσοστά ακριβείας σε σχέση με πιο πολύπλοκες μεθόδους πρόβλεψης. Το γεγονός αυτό οφείλεται στο ότι οι μέθοδοι εκθετικής εξομάλυνσης δεν επηρεάζονται από τις ιδιομορφίες των προτύπων των δεδομένων ή από περιστασιακά εμφανιζόμενες ακραίες τιμές, οι οποίες παρατηρούνται σε επιχειρησιακά δεδομένα.

Όπως αναφέραμε και παραπάνω η επιλογή της μεθόδου, υπόκειται στην υποκειμενική κρίση του αναλυτή. Συμπερασματικά, ακολουθούν κάποια βήματα για την διενέργεια μιας σωστής πρόβλεψης.

i. Επιλέγουμε μια μέθοδο πρόβλεψης βασισμένη σε ήδη αποκτημένη γνώση γύρω από το παρατηρηθέν μοντέλο της χρονοσειράς.
ii. Χρησιμοποιούμε τη μέθοδο πρόβλεψης για την απόκτηση εκτιμημένης τιμής των δεδομένων.
iii. Υπολογίζουμε το σφάλμα πρόβλεψης.
iv. Παίρνουμε απόφαση για την καταλληλότητα του μοντέλου που βασίζεται στη μέτρηση του σφάλματος πρόβλεψης.

ΚΕΦΑΛΑΙΟ 3: ΑΝΑΛΥΣΗ ΣΤΟ ΠΕΔΙΟ ΤΟΥ ΧΡΟΝΟΥ

3.1 Εισαγωγή

Στο κεφάλαιο αυτό θα παρουσιαστούν στοχαστικά μαθηματικά μοντέλα, με τα οποία θα προσπαθήσουμε να περιγράψουμε τη διαχρονική εξέλιξη κάποιου φυσικού μεγέθους. Τέτοια μοντέλα είναι τα Ολοκληρωμένα Αυτοπαλινδρομικά Μοντέλα Κινητού Μέσου όρου (Autoregressive Integrated – Moving Average, ARIMA) τα οποία έχουν μελετηθεί εκτεταμένα κατά την τελευταία εικοσαετία από τους Box και Jenkins. Αυτοί πρότειναν μια οικογένεια αλγεβρικών μοντέλων πρόβλεψης για την επιλογή του καταλληλότερου μοντέλου μιας δεδομένης γραμμικής χρονοσειράς. Τα μοντέλα αυτά (γνωστά ως ARIMA) αποτελούν μια σύνθετη τεχνική που απαιτούν εμπειρία από τον αναλυτή και δίνουν ικανοποιητικά αποτελέσματα.

3.2 Στοχαστική Διαδικασία και Χρονοσειρά

Στοχαστική διαδικασία είναι μια οικογένεια τυχαίων μεταβλητών Y_t, $t = 0, \pm 1, \pm 2, \ldots$, διατεταγμένων στον χρόνο οι οποίες εξελίσσονται διαχρονικά σύμφωνα με ορισμένο νόμο πιθανότητας.

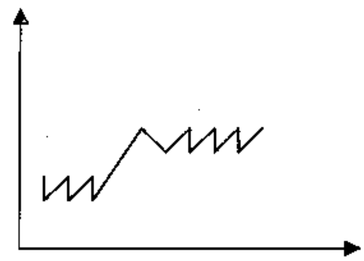
Η χρονοσειρά Ν διαδοχικών παρατηρήσεων $y_1, y_2, \ldots, y_N$ αποτελεί πραγματοποίηση μιας στοχαστικής διαδικασίας γιατί θεωρείται σαν ένα σύνολο από κοινού κατανεμημένων τυχαίων μεταβλητών (Δημελή Σοφία, 2002).

Στη χρονοσειρά η παράμετρος t εκφράζει τον χρόνο, ενώ στη στοχαστική διαδικασία η παράμετρος t δεν εκφράζει απαραίτητα τον χρόνο. Είναι δηλαδή μια απλή παρατήρηση από μια πολυμετάβλητη κατανομή πιθανότητας.

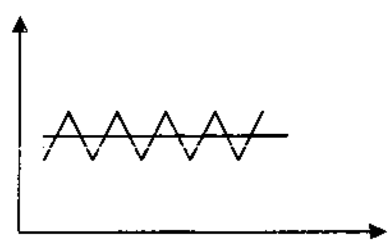
3.3 Στάσιμη – Μη Στάσιμη

Ένα σημαντικό ερώτημα που τίθεται στην εξέταση των χρονοσειρών είναι «αν είναι στάσιμες ή μη – στάσιμες».

Μη στάσιμη (non stationary) καλείται μια χρονοσειρά που τα χαρακτηριστικά της μεταβάλλονται διαχρονικά και δεν μπορεί να περιγραφεί από κάποιο αλγεβρικό μοντέλο.



<u>Στάσιμη (stationary)</u> καλείται μια χρονοσειρά που τα χαρακτηριστικά της παραμένουν διαδοχικά σε ισορροπία γύρω από ένα σταθερό μέσο επίπεδο. Δηλαδή τα δεδομένα κυμαίνονται γύρω από ένα σταθερό μέσο ανεξάρτητα του χρόνου και η διακύμανση παραμένει σταθερή.



Υπάρχουν δύο είδη στάσιμων χρονοσειρών, η ισχυρή και η ασθενής.

Μία χρονοσειρά y_t όπου $t \in \mathrm{T}$ θα λέγεται <u>*αυστηρά στάσιμη*</u> (strict stationary) εάν η από κοινού κατανομή F των $y_{t_1}, y_{t_2}, \ldots, y_{t_n}$ είναι ανεξάρτητη της αρχής των συντεταγμένων.

<u>*Στάσιμη ασθενώς*</u> (weakly stationary) ή <u>*Στάσιμη δεύτερης τάξης*</u> χρονοσειρά, καλείται η χρονοσειρά της οποίας η μέση τιμή, η διασπορά και η συνδιακύμανση είναι σταθερές και ανεξάρτητες του χρόνου.

<u>Σταθερή Μέση Τιμή</u>

$$\mathrm{E}(Y_t) = \mu$$

<u>Σταθερή Διασπορά</u>

$$Var(Y_t) = \sigma^2_x = \gamma_0, \quad t = 1, \ldots, \mathrm{T}$$

<u>Σταθερή Συνδιακύμανση</u>

$$Cov(Y_t, Y_{t+k}) = \mathrm{E}\left[(Y_t - \mu)(Y_{t+k} - \mu)\right] = \gamma_k, \quad k = 1, 2, \ldots$$

Σημείωση: Όταν θα κάνουμε λόγο για στάσιμη χρονοσειρά θα αναφερόμαστε στην ασθενώς χρονοσειρά.

Για τον έλεγχο της στασιμότητας ή μη της χρονοσειράς συνήθως χρησιμοποιείται η γραφική παράσταση.

- 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.
- Εάν παρατηρείται κάποια αλλαγή κατά μήκος του χρόνου, τότε η χρονοσειρά δεν είναι στάσιμη.

Ένας άλλος τρόπος για τον έλεγχο της στασιμότητας μιας χρονοσειράς ως προς τη μέση τιμή είναι το διάγραμμα των αυτοσυσχετίσεων. Όταν η χρονοσειρά είναι στάσιμη οι αυτοσυσχετίσεις φθίνουν στο μηδέν με γρήγορο ρυθμό, ενώ όταν η χρονοσειρά δεν είναι στάσιμη οι αυτοσυσχετίσεις φθίνουν με αργό ρυθμό καθώς αυξάνει ο αριθμός των υστερήσεων.

3.4 <u>Συνάρτηση Αυτοσυσχέτισης (Autocorrelation Function)</u>

Λόγω της δυσκολίας στην πλήρη περιγραφή μιας στοχαστικής διαδικασίας από στατιστικούς δείκτες όπως ο μέσος, η διακύμανση και η τυπική απόκλιση, ο συντελεστής αυτοσυσχέτισης είναι ο πλέον χρήσιμος δείκτης στην ανάλυση χρονοσειρών. Γενικά, η συνάρτηση αυτοσυσχέτισης μετράει τον βαθμό συσχέτισης ανάμεσα στις γειτονικές παρατηρήσεις της χρονοσειράς.

Ορίζοντας την αυτοσυσχέτιση με υστέρηση k έχουμε:

$$\rho_k = \frac{\mathrm{cov}(y_t, y_{t-k})}{\sigma_{y_t} \cdot \sigma_{y_{t-k}}}$$

όπου

cov η συνδιακύμανση μεταξύ y_t και y_{t-k}

και σ η τυπική απόκλιση της y_t και y_{t-k}

επίσης ισχύει:

1. $\rho(0)=1$
2. $\rho(k)=\rho(-k)$ επειδή είναι άρτια συνάρτηση, οι αρνητικές τιμές ταυτίζονται με τις θετικές τιμές
3. $|\rho(k)|\leq 1$

Όλοι οι παραπάνω τύποι αναφέρονται σε πληθυσμό. Ο συμβολισμός της αυτοσυσχέτισης για τα δείγματα είναι ο εξής:

$$r_k=\frac{c_k}{c_0} \qquad k=0,1,2\ldots$$

Όπου
$$c_k=\frac{1}{N}\sum_{t=1}^{n-k}(Y_t-\overline{Y})\cdot(Y_{t-k}-\overline{Y})$$

$$c_0=\frac{1}{N}\sum_{t=1}^{n}(Y_t-\overline{Y})^2$$

$$\overline{Y}=\frac{1}{N}\sum_{t=1}^{n}y_t$$

Ο r_1 υποδηλώνει πώς οι διαδοχικές παρατηρήσεις (υστέρησης 1) της χρονοσειράς σχετίζονται μεταξύ τους, ο r_2 πώς σχετίζονται οι παρατηρήσεις που απέχουν δύο χρονικές περιόδους (υστέρησης 2) κλπ. Οι συντελεστές αυτοσυσχέτισης για υστερήσεις 1,2,..., δημιουργούν τη συνάρτηση αυτοσυσχέτισης (Autocorrelation Function) ή ACF. 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. 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.
Η στατιστική θεωρία διδάσκει ότι η κατανομή δειγματοληψίας συντελεστών αυτοσυσχέτισης τυχαίων δεδομένων θεωρείται κανονική όταν:

- ο μέσος των αυτοσυσχετίσεων ισούται με το 0
- το τυπικό σφάλμα των αυτοσυσχετίσεων ισούται με $\frac{1}{\sqrt{n}}$

Η πληροφορία αυτή μπορεί να χρησιμοποιηθεί για την ανάπτυξη ελέγχων υποθέσεων. Άρα, για να θεωρούνται τα δεδομένα τυχαία, δηλαδή να μην είναι στατιστικά διάφορα του μηδενός, θα πρέπει το 95% των συντελεστών αυτοσυσχέτισης του δείγματος να ανήκουν στο διάστημα $\simeq \pm 2/\sqrt{n}$.

3.5 Συντελεστής Μερικής Αυτοσυσχέτισης (Particial Autocorrelation Coefficient)

Οι συντελεστές μερικής αυτοσυσχέτισης μετρούν το βαθμό της σχέσης μεταξύ των Y_t και Y_{t-k}, όταν το αποτέλεσμα των άλλων χρονικών υστερήσεων στο Y_t, παραμένει σταθερό. Η συσχέτιση αυτών των μεταβλητών δεν επηρεάζεται από τις ενδιάμεσες μεταβλητές και ο συντελεστής τους συμβολίζεται με $\phi_{\kappa\kappa}$ (k τάξεως).
Οι μερικές αυτοσυσχετίσεις ορίζονται από την παρακάτω εξίσωση παλινδρόμησης:

$$y_t = \phi_{1k} y_{t-1} + \phi_{2k} y_{t-2} + \ldots + \phi_{kk} y_{t-k} + e_t$$

όπου e_t το σφάλμα παλινδρόμησης.

Σε μια εξίσωση $y_t = \phi_{12} y_{t-1} + \phi_{22} y_{t-2} + e_t$ όπου ϕ_{12} το 1 δηλώνει τις χρονικές υστερήσεις της μεταβλητής (εδώ 1^{η} υστέρηση) και το 2 δηλώνει τη μέγιστη τάξη παλινδρόμησης (η οποία παραμένει σταθερή στην εξίσωση). Πρέπει να σημειωθεί ότι ο πρώτος συντελεστής μερικής αυτοσυσχέτισης ϕ_{11} είναι πάντα ίσος με τον πρώτο συντελεστή αυτοσυσχέτισης r_1. Το διάγραμμα για τον συντελεστή μερικής αυτοσυσχέτισης λέγεται PACF όπως και για τον αντίστοιχο συντελεστή αυτοσυσχέτισης ονομάζεται ACF. Ο σκοπός και ο ρόλος των διαγραμμάτων στην διαδικασία μοντελοποίησης μιας χρονοσειράς θα αναφερθεί στην παραπέρα ανάλυση του κεφαλαίου.

Μια τελευταία υπενθύμιση αφορά τη μέτρηση αυτοσυσχετίσεων και γραφημάτων των συντελεστών. Για μια τέτοια δουλειά υπάρχει πληθώρα από software. Για τέτοιους υπολογισμούς είναι διαθέσιμα προγράμματα όπως SAS/Econometric and Time series Library, SPSS-X, BMDP όπως και το Minitab για τους μικρο-υπολογιστές. Η δυνατότητα επιλογής είναι μεγάλη και τα οφέλη, από την χρήση τέτοιων software πολλά.

3.6 Μετατροπή Μη-στάσιμης Χρονοσειράς σε Στάσιμη

Όπως αναφέραμε παραπάνω μια μη-στάσιμη χρονοσειρά δεν μπορεί να περιγραφεί από κάποιο αλγεβρικό μοντέλο. Επομένως είναι πολύ σημαντικό πρώτα από όλα να ελέγχεται η στασιμότητα της χρονοσειράς. Σε μια χρονοσειρά τάσεις ή άλλα μη-στάσιμα πρότυπα στο επίπεδο της, έχουν σαν αποτέλεσμα θετικές αυτοσυσχετίσεις οι οποίες επικρατούν στο διάγραμμα των αυτοσυσχετίσεων. Γι' αυτό λοιπόν πρέπει να εξαλειφθούν τα μη-στάσιμα πρότυπα για να αναπτυχθεί το κατάλληλο μοντέλο. Ένας τρόπος για να επιτευχθεί αυτό είναι με τη χρήση της μεθόδου της διαφόρισης.

Οι χρονοσειρές «διαφορών πρώτης τάξης» προκύπτουν από τις διαφορές των διαδοχικών παρατηρήσεων της αρχικής χρονοσειράς:

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

όπου το ∇ δείχνει τη διαφορά μεταξύ του Y στην περίοδο t και στην προηγούμενη περίοδο $(t-1)$. Οι χρονοσειρές διαφορών πρώτης τάξης έχουν $n-1$ δεδομένα.

Οι χρονοσειρές «διαφορών δεύτερης τάξης» προκύπτουν από τις διαφορές των πρώτων διαφορών και έχουν $n-2$ δεδομένα :

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

όπου το ∇^2 δείχνει τη διαφορά μεταξύ του Y στην περίοδο $(t-1)$ και στην προηγούμενη περίοδο $(t-2)$.

Με τον τρόπο αυτό μετατρέπεται μια χρονοσειρά από μη-στάσιμη σε στάσιμη διαφορίζοντάς την όσες φορές χρειάζεται.

Για να κατανοήσουμε καλύτερα την έννοια της διαφόρισης θα εισάγουμε την έννοια του τελεστή ολίσθησης B προς τα πίσω (Back Shift Operator) ο οποίος ορίζεται ως:

$$\mathrm{B}Y_t = Y_{t-1}$$

Δηλαδή ο χρονικός δείκτης κάθε μεταβλητής η οποία πολλαπλασιάζεται με τον τελεστή, μετατοπίζεται τόσες περιόδους προς τα πίσω όσες είναι ο εκθέτης k του τελεστή: $\mathrm{B}^k Y_t = Y_{t-k}$

Για παράδειγμα: $\mathrm{B}^2 Y_t = Y_{t-2}$,

$\mathrm{B}^3 Y_t = Y_{t-3}$ κτλ.

Με τη χρήση του τελεστή ολίσθησης προς τα πίσω είναι εύκολο να περιγραφεί η διαδικασία της διαφόρισης οποιασδήποτε τάξης θέλουμε.

Διαφόριση 1ης τάξης:

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

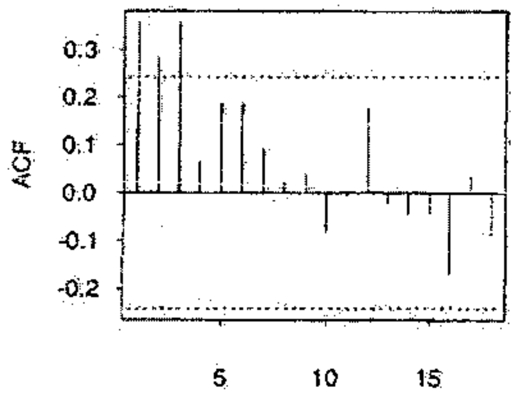
Διαφόριση 2ης τάξης:

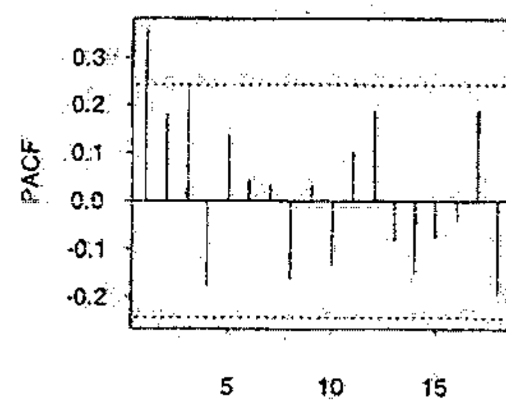
$$Y''_t = Y'_t - Y'_{t-1} = (Y_t - Y_{t-1}) - (Y_{t-1} - Y_{t-2}) = Y_t - 2Y_{t-1} + Y_{t-2} = (1 - 2\mathrm{B} + \mathrm{B}^2)Y_t = (1-\mathrm{B})^2 Y_t$$

Γενικότερα η διαφόριση d τάξης μπορεί να περιγραφεί ως $(1-\mathrm{B})^d Y_t$.

3.7 Αναγνώριση Εποχικότητας

Μια χρονοσειρά θεωρείται ότι έχει εποχικότητα όταν επαναλαμβάνεται κατά την διάρκεια μιας συγκεκριμένης χρονικής περιόδου, συνήθως ετήσια. Με την εξέταση των συντελεστών αυτοσυσχέτισης, μπορούμε να αναγνωρίσουμε ένα εποχικό μοντέλο. Μετράμε περισσότερες από δυο ή τρεις χρονικές υστερήσεις ώστε να ορίσουμε εάν είναι στατιστικά διάφορες του μηδενός. Οι αυτοσυσχετίσεις που είναι σημαντικά διάφορες του μηδενός, δηλώνουν την ύπαρξη ενός προτύπου. Για την εύρεση της εποχικότητας πρέπει να αναζητήσουμε σημαντικές αυτοσυσχετίσεις υψηλότερων τάξεων. Έτσι, για μηνιαία δεδομένα που παρουσιάζουν εποχικότητα, υψηλές αυτοσυσχετίσεις μπορούν να παρατηρηθούν για υστέρηση 12 αλλά και για καθυστερήσεις 24 και 36. Όταν τα δεδομένα είναι τριμηνιαία και παρουσιάζουν εποχικότητα, υψηλές αυτοσυσχετίσεις θα παρουσιαστούν για υστέρηση 4, 8, 12.





ΠΗΓΗ:http://fsu.ece.ntua.gr/old/gr/texnikes_2004.html

Όπως βλέπουμε στις παραπάνω γραφικές παραστάσεις, στην ACF παρατηρούνται υψηλές αυτοσυσχετίσεις για υστερήσεις 1,2 και 3 και στην PACF σημαντικές μερικές αυτοσυσχετίσεις για υστερήσεις 1 και 3. 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.

Η εποχικότητα είναι συνήθως εύκολο να αναγνωριστεί όταν τα δεδομένα δεν ακολουθούν ένα πρότυπο π.χ τάση, που να δημιουργεί μη στάσιμα δεδομένα. Εφόσον υπάρχει πρότυπο, πρέπει πρώτα να μετασχηματισθούν τα μη-στάσιμα δεδομένα σε στάσιμα και στην συνέχεια να εξετασθούν για την παρουσία εποχικότητας.

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 εποχική διαφορά ορίζεται :

$$Y_t = Y_t - Y_{t-12} = Y_t - B^{12}Y_t = (1 - B^{12})Y_t$$

Γενικά, οι σειρές εποχικών 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 διαφόριση μπορεί να επαναληφθεί ώστε να προκύψουν δεύτερης τάξης εποχιακές διαφορές, αν και αυτό σπάνια κρίνεται αναγκαίο.

3.8 Μοντέλα Χρονοσειρών ARIMA

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. 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 πολλαπλής γραμμικής παλινδρόμησης περιγράφονται από την εξίσωση :

$$Y_t = b_0 + b_1 X_1 + b_2 X_2 + \ldots + b_p X_p + e_t$$

όπου Y_t η εξαρτημένη μεταβλητή (υπό πρόβλεψη μεταβλητή) και X_1 , X_2 , ... , X_p οι ανεξάρτητες μεταβλητές (επεξηγηματικές μεταβλητές).

Η εξίσωση $Y_t = b_0 + b_1 Y_{t-1} + b_2 Y_{t-2} + \ldots + b_p Y_{t-p} + e_t$ 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 υστερήσεις 1,2, ... , p.

Το μοντέλο αυτό καλείται *Αυτοπαλινδρομικό* (*Autoregression* ή AR) τάξεως p.

Τα Αυτοπαλινδρομικά Μοντέλα πρέπει να έχουν διαφορετικό χειρισμό από τα κλασσικά παλινδρομικά μοντέλα για δύο λόγους :

1. 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.

2. 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 :

$$Y_t = b_0 + b_1 e_{t-1} + b_2 e_{t-2} + \ldots + b_q e_{t-q} + e_t$$

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, e_t.

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 διαφόρισης. Τότε ονομάζονται *Ολοκληρωμένα* (*Integrated* ή I) *Αυτοπαλινδρομικά Μοντέλα Κινητού Μέσου Όρου* (*Autoregressive Integrated Moving Average* ή ARIMA).

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 διαφόρισης
MA : q = η τάξη του MA όρου

3.9 Μοντέλο Λευκού Θορύβου (White Noise)

Μια χρονοσειρά που απαρτίζεται από τυχαία δεδομένα ονομάζεται "λευκός θόρυβος", όταν δηλαδή η χρονοσειρά δεν ακολουθεί κάποιο πρότυπο. Οι προϋποθέσεις που πρέπει να ισχύουν για να υπάρχει "λευκός θόρυβος" είναι ότι η μέση τιμή πρέπει να είναι σταθερή (συνήθως μηδέν), όπως και η διακύμανση. Συμβολίζεται με e_t και οι τιμές των μεταβλητών δεν συσχετίζονται.

$$E(e_t) = \mu$$

$$Var(e_t) = \sigma^2_e$$

$$\mathrm{cov}(e_t, e_{t+k}) = 0 \text{, αν } k \neq 0$$

Μια τέτοια τυχαία χρονοσειρά είναι πάντα στάσιμη και αποτελεί θεμελιώδες μοντέλο σε πολλές τεχνικές ανάλυσης χρονοσειρών.

3.10 Τυχαίος Περίπατος (Random Walk)

Μια διαδικασία Y_t, $t = 1, 2, ..., \mathrm{T}$ καλείται τυχαίος περίπατος αν:

$$Y_t = Y_{t-1} + e_t \; ,$$

όπου e_t είναι λευκός θόρυβος.

Η παρατήρηση Y_t εξαρτάται μόνο από την προηγούμενη παρατήρηση Y_{t-1}. Δηλαδή κατά την χρονική περίοδο t η παρατήρηση Y_t θα παριστά την θέση ενός περιπατητή και η απόφαση για το που θα πάει την επόμενη περίοδο (μπροστά ή πίσω) δεν εξαρτάται από το που βρίσκεται αυτή τη στιγμή.

Οι τύποι που ισχύουν είναι:

Μέση τιμή $\quad \mathrm{E}(y_t) = \mathrm{E}(y_0)$

Διακύμανση $\quad Var(y_t) = t\sigma_\varepsilon^2$

Συνδιακύμανση $\quad Cov(y_t, y_{t-k}) = (t-k)\sigma^2_\varepsilon$

Η συνδιακύμανση επειδή εξαρτάται από τον χρόνο δεν είναι στάσιμη. Για να μετατραπεί σε στάσιμη παίρνουμε τις πρώτες διαφορές και τότε γίνεται ίση με τον λευκό θόρυβο.

Το μοντέλο της τυχαίας διαδρομής μπορεί να έχει και μια σταθερά β.

Άρα:

$$Y_t = \beta + e_t \text{ (Random Withdrift)},$$

όπου το β μας δείχνει μετατόπιση της χρονοσειράς προς τα πάνω (Makridakis/Wheelright/McGee, 1983).

Όπως αναφέραμε το μοντέλο αυτό χρησιμοποιείται για μη στάσιμα δεδομένα και παρουσιάζει μεγάλες περιόδους εμφανών τάσεων, οι οποίες μπορούν να αλλάξουν κατεύθυνση με απρόβλεπτο τρόπο. Επομένως ένα καλό παράδειγμα τέτοιων χρονοσειρών που έχουν τη μορφή τυχαίου περιπάτου αποτελούν οι μετοχές.

Αναφέρθηκαν ήδη στοχαστικές διαδικασίες, η μια λευκού θορύβου ARIMA(0,0,0) (στάσιμη) και η άλλη ο τυχαίος περίπατος ARIMA(0,1,0) (μη στάσιμη). Εν συνεχεία θα αναφερθούμε στις συνήθεις στάσιμες στοχαστικές διαδικασίες που χρησιμοποιεί η ανάλυση χρονοσειρών. Αυτές είναι η *Αυτοπαλινδρομούμενη Διαδικασία (AR)*, η *Κινητού Μέσου Διαδικασία (MA)* και η *Μικτή (ARMA)*.

3.11 Αυτοπαλινδρομικά Μοντέλα (AR)

Όπως ήδη έχει αναφερθεί, ένα *Αυτοπαλινδρομικό Μοντέλο* χρονοσειρών διαφέρει από τα γενικά γραμμικά παλινδρομικά μοντέλα στο ότι οι ανεξάρτητες μεταβλητές είναι παλαιότερες τιμές της εξαρτημένης μεταβλητής. Το γενικό μοντέλο **AR(p)** έχει τη μορφή:

$$Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \ldots + \phi_p Y_{t-p} + e_t \quad \text{ή}$$

με την χρήση του τελεστή ολίσθησης B το μοντέλο μπορεί να γραφεί ως:

$$Y_t - \phi_1 Y_{t-1} - \ldots - \phi_p Y_{t-p} = e_t \text{ ή } (1 - \phi_1 B - \ldots - \phi_p B^p) Y_t = e_t$$

Η παρατήρηση Y_t εξαρτάται από τις Y_{t-p} και όχι από μια ανεξάρτητη μεταβλητή, ενώ τα σφάλματα e_t ακολουθούν την κανονική κατανομή με μέση τιμή 0 και διακύμανση 1. Για να θεωρηθεί η αυτοπαλινδρούμενη διαδικασία στάσιμη θα πρέπει οι ρίζες της εξίσωσης Φ(Β)=0 να βρίσκονται έξω από το μοναδιαίο κύκλο (Παπαδήμας 'Οθωνας, 1994). Το γράμμα p στο AR(p) αναφέρεται στην τάξη του μοντέλου, και καθορίζεται από εξέταση των μερικών αυτοσυσχετίσεων. Σημειώνεται, ότι η τάξη του μοντέλου είναι ίση με τον αριθμό των στατιστικά σημαντικών μερικών αυτοσυσχετίσεων.

ΕΙΔΙΚΕΣ ΠΕΡΙΠΤΩΣΕΙΣ

Όταν για p=1, το αυτοπαλινδρομικό μοντέλο AR(1) έχει την μορφή: $Y_t = \phi_1 Y_{t-1} + e_t$

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. Για να είναι ένα μοντέλο AR(1) στάσιμο πρέπει να ικανοποιεί την παρακάτω συνθήκη.

$$|\phi_1| < 1$$

Για AR(1) διαδικασίες, η δειγματική συνάρτηση αυτοσυσχέτισης φθίνει εκθετικά, ενώ υπάρχει μόνο ένας στατιστικά σημαντικός συντελεστής μερικής αυτοσυσχέτισης, δηλαδή η δειγματική συνάρτηση μερικής αυτοσυσχέτισης σταματά να είναι σημαντική μετά από μια υστέρηση.

Όταν p=2, το αυτοπαλινδρομικό μοντέλο AR(2) έχει την μορφή:

$$Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + e_t$$

Η συνθήκη στασιμότητας για το AR(2) μοντέλο που πρέπει να ισχύει είναι:

$$\phi_2 + \phi_1 < 1$$
$$\phi_2 - \phi_1 < 1$$
$$|\phi_2| < 1$$

Οι αρχικές τιμές των παραμέτρων ενός AR(p) μοντέλου μπορούν να υπολογισθούν με την χρήση των εξισώσεων Yule-Walker:

$$\begin{aligned}
\rho_1 &= \varphi_1 + \varphi_2 \rho_1 + \ldots + \varphi_p \rho_{p-1} \\
\rho_2 &= \varphi_1 \rho_1 + \varphi_2 + \ldots + \varphi_p \rho_{p-2} \\
&\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\
\rho_p &= \varphi_1 \rho_{p-1} + \varphi_2 \rho_{p-2} + \ldots + \varphi_p
\end{aligned}$$

όπου $\rho_1, \rho_2, \ldots, \rho_p$ οι θεωρητικές αυτοσυσχετίσεις για υστερήσεις 1, 2, ... , p αντίστοιχα και $\varphi_1, \varphi_2, \varphi_p$ οι συντελεστές (παράμετροι) του μοντέλου.

Για **AR(1)** μοντέλο έχουμε:

Για $p = 1$

$$\rho_1 = \varphi_1$$

Εάν η θεωρητική αυτοσυσχέτιση ρ_1 αντικατασταθεί με την r_1 τότε προκύπτει

$$\hat{\varphi}_1 = r_1$$

Για **AR(2)** μοντέλο έχουμε:

Για $p = 2$

$$\rho_1 = \varphi_1 + \varphi_2\rho_1$$

$$\rho_2 = \varphi_1\rho_1 + \varphi_2$$

Αντικαθιστώντας τα ρ_1 , ρ_2 με τα r_1 , r_2 τότε:

$$\hat{\varphi}_1 = \frac{r_1(1-r_2)}{1-r_1^2}$$

και

$$\hat{\varphi}_2 = \frac{r_2 - r_1^2}{1-r_1^2}$$

3.12 Μοντέλα Κινητών Μέσων Όρων (ΜΑ)

Τα μοντέλα κινητών μέσων όρων δίνουν προβλέψεις Y_t βασισμένες σε ένα γραμμικό συνδυασμό τρεχόντων και παρελθόντων σφαλμάτων πρόβλεψης. Το γενικό μοντέλο **MA(q)** έχει τη μορφή:

$$Y_t = e_t - \theta_1 e_{t-1} - \theta_2 e_{t-2} - \ldots - \theta_q e_{t-q} \quad \text{ή}$$

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

$$Y_t = \left(1 - \theta_1 \mathrm{B} - \ldots - \theta_q \mathrm{B}^q\right) e_t .$$

Η παρατήρηση Y_t εξαρτάται από τον όρο σφάλματος e_t και το προηγούμενο σφάλμα e_{t-q}. Τα σφάλματα e_t ακολουθούν την κανονική κατανομή με μέση τιμή 0 και διακύμανση 1. Μια διαδικασία κινητού μέσου ΜΑ είναι πάντα στάσιμη, ενώ είναι και αντιστρέψιμη μόνο αν οι ρίζες του πολυωνύμου $\Theta(B)=0$ βρίσκονται έξω από τον μοναδιαίο κύκλο (Παπαδήμας 'Οθωνας, 1994). Αντιστρέψιμη λέγεται μια ΜΑ διαδικασία όταν μπορεί να αντιστραφεί η έκφρασή της, δηλαδή να εκφραστούν τα e_t σαν συνάρτηση των $Y_t, Y_{t-1}...$ με μοναδικό τρόπο. Άρα ως αντιστρεψιμότητα αναφέρεται η δυνατότητα ενός ΜΑ τύπου μοντέλου πεπερασμένης τάξης έστω MA(q), να μετατρέπεται σε ένα AR μοντέλο απείρου τάξης, AR(∞). Αντίστοιχα θα λέμε ότι ένα AR(p) μοντέλο είναι αντιστρέψιμο αν μπορεί να λάβει τη μορφή ενός μοντέλου MA(∞). Το γράμμα q στο MA(q) αναφέρεται στην τάξη του μοντέλου και καθορίζεται από την εξέταση των αυτοσυσχετίσεων. Δηλαδή αν το $q=1$, έχουμε μια στατιστικά σημαντική υστέρηση, άρα έχουμε πρώτης τάξης μοντέλο κτλ.

ΕΙΔΙΚΕΣ ΠΕΡΙΠΤΩΣΕΙΣ

Για $q=1$, το μοντέλο των κινητών μέσων MA(1) έχει την μορφή:

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

Για οποιαδήποτε τιμή της παραμέτρου θ_1, το μοντέλο MA(1) περιγράφει τη συμπεριφορά μιας στάσιμης χρονοσειράς. Όμως απαιτούνται ορισμένοι περιορισμοί για τις παραμέτρους των μοντέλων ΜΑ προκειμένου να διασφαλιστεί η αντιστρεψιμότητα. Για ένα MA(1) μοντέλο η συνθήκη αντιστρεψιμότητας που πρέπει να ικανοποιείται είναι:

$$|\theta_1| < 1$$

Η δειγματική συνάρτηση μερικής αυτοσυσχέτισης (PACF) φθίνει προς το μηδέν εκθετικά, ενώ υπάρχει μόνο ένας στατιστικά σημαντικός συντελεστής αυτοσυσχέτισης. Δηλαδή η δειγματική συνάρτηση αυτοσυσχέτισης μετά από μια υστέρηση δεν είναι στατιστικά σημαντική.

Για $q=2$ το μοντέλο κινητών μέσων MA(2) έχει τη μορφή:

$$Y_t = e_t - \theta_1 e_{t-1} - \theta_2 e_{t-2}$$

Η συνθήκη αντιστρεψιμότητας για το MA(2) μοντέλο που πρέπει να ισχύει είναι:

$$\theta_1 + \theta_2 < 1$$
$$\theta_2 - \theta_1 < 1$$
$$|\theta_2| < 1$$

Οι αρχικές τιμές των παραμέτρων ενός MA(q) μοντέλου μπορούν να υπολογισθούν με τις παρακάτω εξισώσεις:

$$\rho_\kappa = \begin{cases} \dfrac{-\theta_\kappa + \theta_1\theta_{\kappa+1} + \ldots + \theta_{q-\kappa}\theta_q}{1+\theta_1^2 + \ldots + \theta_q^2}, & \kappa = 1,2,\ldots,q \\ 0 & , \kappa = q \end{cases}$$

Για **MA(1)** μοντέλο έχουμε:
Για $q = 1$

$$\rho_1 = \begin{cases} \dfrac{-\theta_1}{1+\theta_1^2}, & \kappa = 1 \\ 0 & , \kappa \geq 2 \end{cases}$$

Για **MA(2)** μοντέλο έχουμε:
Για $q = 2$

$$\rho_1 = \frac{-\theta_1(1-\theta_2)}{1+\theta_1^2+\theta_2^2}$$

$$\rho_2 = \frac{-\theta_2}{1+\theta_1^2+\theta_2^2}$$

Ενώ για $q > 2$:

$$\rho_\kappa = 0 \quad , \kappa \geq 3$$

3.13 Μικτά μοντέλα ARMA(p,q)

Τα μικτά μοντέλα ARMA(p,q) είναι συνδυασμός μιας AR(p) και μιας MA(q) διαδικασίας. Το μοντέλο έχει τη μορφή:

$$Y_t - \phi_1 Y_{t-1} - \ldots - \phi_p Y_{t-p} = e_t - \theta_1 e_{t-1} - \ldots - \theta_q e_{t-q}$$

ή με τη χρήση του τελεστή B το μοντέλο μπορεί να γραφεί ως:

$$(1-\phi_1 B - \ldots - \phi_p B^p)Y_t = (1-\theta_1\mathrm{B} - \ldots - \theta_q \mathrm{B}^q)e_t \quad ή \quad \Phi(\mathrm{B})Y_t = \Theta(\mathrm{B})e_t.$$

Η παρατήρηση Y_t εξαρτάται από την προηγούμενη τιμή Y_{t-p} και από το προηγούμενο σφάλμα e_{t-q}. Το μοντέλο αυτό είναι στάσιμο όταν οι ρίζες του πολυωνύμου είναι έξω από τον μοναδιαίο κύκλο και αντιστρέψιμο όταν οι ρίζες του πολυωνύμου είναι έξω από τον μοναδιαίο κύκλο (Παπαδημας Οθωνας, 1994). Για τον προσδιορισμό του μοντέλου, χρησιμοποιούνται οι δειγματικές αυτοσυσχετίσεις και οι μερικές δειγματικές αυτοσυσχετίσεις. Δηλαδή η τάξη του μοντέλου είναι ίση με τον αριθμό των στατιστικά σημαντικών μερικών αυτοσυσχετίσεων όπως και με τον αριθμό των στατιστικά σημαντικών αυτοσυσχετίσεων.

Ένα μοντέλο ARMA(1,1) έχει τη μορφή:

$$Y_t = \phi_1 Y_{t-1} + e_t - \theta_1 e_{t-1}$$

Για να είναι στάσιμο θα πρέπει να ισχύει η συνθήκη $|\phi_1|<1$ και αντιστρέψιμο η συνθήκη $|\theta_1|<1$.

Οι εκτιμήσεις των παραμέτρων ενός ARMA μοντέλου μπορούν να υπολογισθούν από τους τύπους:

Για $k=1$

$$\rho_1 = \frac{(1-\phi_1\theta_1)(\phi_1-\theta_1)}{1-\theta_1^2-2\theta_1\phi_1}$$

Για $k=2$

$$\rho_2 = \phi_1\rho_1$$

Για $k \geq 3$

$$\rho_k = \phi_1\rho_{k-1}$$

3.14 Μοντέλα ARIMA(p,d,q)

Ένα ARIMA(p,d,q) είναι ένα μοντέλο, το οποίο διαφοριζόμενο d φορές παράγει ένα μοντέλο ARMA(p,d,q). Εάν η υπό εξέταση χρονοσειρά δεν είναι στάσιμη τότε αντί για ένα ARMA μοντέλο εφαρμόζεται ένα ARIMA(p,d,q) μοντέλο. Η γενική μορφή του είναι: $\phi(\mathrm{B})\nabla^d Y_t = \theta(\mathrm{B})e_t$. Η πιο απλή περίπτωση είναι το ARIMA(1,1,1) με εξίσωση:

$$(1-\phi_1 B)(1-\mathrm{B})Y_t = (1-\theta_1\mathrm{B})e_t$$

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 του κατάλληλου μοντέλου γίνεται μέσω διαφορετικών μοντέλων που έχουν τη δυνατότητα να παράγουν "καλές" προβλέψεις.

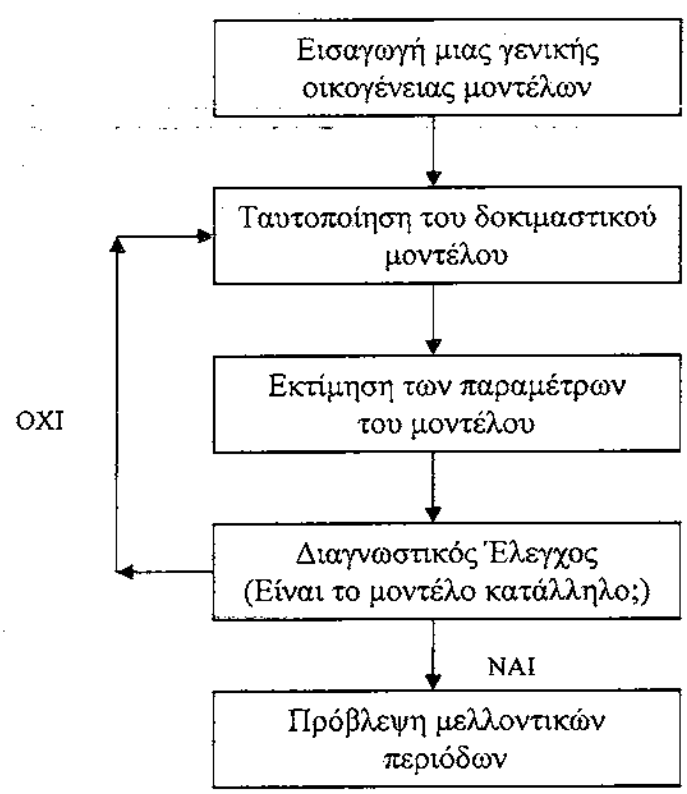
ΠΙΝΑΚΑΣ

Μοντέλο	Συνθήκες Σταθερότητας	Συνθήκες Αντιστρεψιμότητας	Συντελεστές Αυτοσυσχέτισης	Συντελεστές Μερικής Αυτοσυσχέτισης
AR(1)	$\lvert\theta_1\rvert<1$	Καμία	Φθίνει εκθετικά προς τα κάτω	Διακόπτεται μετά από μία χρονική υστέρηση
AR(2)	$\lvert\phi_1\rvert<1$ $\phi_1+\phi_2<1$ $\phi_2-\phi_1<1$	Καμία	Φθίνει εκθετικά προς τα κάτω	Διακόπτεται μετά από δύο χρονικές υστερήσεις
MA(1)	Καμία	$\lvert\phi_1\rvert<1$	Διακόπτεται μετά από μία χρονική υστέρηση	Φθίνει εκθετικά προς τα κάτω
MA(2)	Καμία	$\lvert\theta_1\rvert<1$ $\theta_1+\theta_2<1$ $\theta_2-\theta_1<1$	Διακόπτεται μετά από δύο χρονικές υστερήσεις	Φθίνει εκθετικά προς τα κάτω
ARMA(p,q)	$\lvert\phi_1\rvert<1$	$\lvert\theta_1\rvert<1$	Φθίνει εκθετικά προς τα κάτω	Φθίνει εκθετικά προς τα κάτω

3.15 Μεθοδολογία Box-Jenkins

Οι Box-Jenkins (1976) ενοποίησαν με επιτυχία τη μεθοδολογία που εκφράζει τις ανάγκες των γνώσεων για την κατανόηση και χρησιμοποίηση απλών μοντέλων ARMA με μια μεταβλητή (univariate). Η βάση αυτής της προσέγγισης περιλαμβάνει τέσσερις διαφορετικές φάσεις.

1. **Ταυτοποίηση** του προκαταρκτικού του μοντέλου.
2. **Εκτίμηση** των παραμέτρων του μοντέλου.
3. **Διαγνωστικός έλεγχος** του μοντέλου για την καταλληλότητα του.
4. **Πρόβλεψη** μελλοντικών περιόδων.



Ταυτοποίηση

Σκοπός της φάσης της ταυτοποίησης είναι η επιλογή ενός ιδιαίτερου μοντέλου ARMA από την γενική ομάδα των μοντέλων ARIMA(p,d,q). Η επιλογή των κατάλληλων τιμών των p και q προϋποθέτει την εξέταση των συντελεστών μερικής αυτοσυσχέτισης και αυτοσυσχέτισης αντίστοιχα, που έχουν υπολογιστεί για τα δεδομένα.

Στο πρώτο στάδιο ταυτοποίησης του μοντέλου θα πρέπει να γίνει έλεγχος στασιμότητας της χρονοσειράς. Όπως έχουμε αναφέρει τα διαγράμματα των αυτοσυσχετίσεων (ACF) μας υποδεικνύουν αν η χρονοσειρά είναι στάσιμη ή μη. Αν η χρονοσειρά δεν είναι στάσιμη, μπορεί εύκολα να μετατραπεί με την μέθοδο της διαφόρισης. Συνήθως οι διαφορές πρώτου ή δεύτερου βαθμού αρκούν για να επιτευχθεί η σταθερότητα των δεδομένων. Για την ανάλυση των μεθόδων Box-Jenkins τα δεδομένα πρέπει να εξετάζονται για την στασιμότητα ή την αντιστρεψιμότητα τους, ανάλογα πάντα με την μορφή του προκαταρκτικού μοντέλου που έχει αναγνωριστεί, αυτό πρέπει να γίνει για να ακολουθήσουν στην συνέχεια οι απαραίτητοι υπολογισμοί στα μετασχηματισμένα δεδομένα.

Μια ακόμα χρησιμότητα του διαγράμματος αυτοσυσχετίσεων (ACF) είναι ο έλεγχος για την τυχαιότητα των παρατηρούμενων δεδομένων της χρονοσειράς. Ο Bartlett (1946) έδειξε ότι για μια διαδικασία "λευκού θορύβου" για ένα μεγάλο δείγμα δεδομένων, η κατά προσέγγιση δειγματική κατανομή των αυτοσυσχετίσεων (r_k) είναι κανονική όταν $E(r_k)=0$ και $Var(r_k)=1/n$. Με την χρησιμοποίηση των ορίων $\pm 2/\sqrt{n}$ γίνεται εκτίμηση των στατιστικά σημαντικών αυτοσυσχετίσεων που είναι διάφορες του μηδενός. Αυτά τα όρια μπορούν να χρησιμοποιηθούν ώστε να μας δώσουν χρήσιμες πληροφορίες από το ACF, για την σημασία των στατιστικών αποτελεσμάτων.

Σε αυτό το σημείο, αξίζει να σημειώσουμε μερικές χρήσιμες παρατηρήσεις για την ταυτοποίηση ενός μοντέλου. Ένα μοντέλο AR ενδείκνυται όταν οι αυτοσυσχετίσεις φθίνουν εκθετικά προς το μηδέν. Με την ίδια λογική, ένα μοντέλο MA ενδείκνυται όταν οι μερικές αυτοσυσχετίσεις φθίνουν προς το μηδέν. Εάν και η αυτοσυσχέτιση και η μερική αυτοσυσχέτιση φθίνουν προς το μηδέν, τότε το κατάλληλο μοντέλο είναι ARMA. Τέλος, η τάξη του μοντέλου AR υποδηλώνεται από τον αριθμό των μερικών αυτοσυσχετίσεων, και η τάξη του μοντέλου MA από τον αριθμό των αυτοσυσχετίσεων, που είναι στατιστικά διάφορες του μηδενός.

Εκτίμηση των Παραμέτρων

Αφού επιλεγεί το προκαταρκτικό μοντέλο ARMA(p,q), στη συνέχεια πρέπει να εκτιμήσουμε τις παραμέτρους του μοντέλου.

Δηλαδή σε ένα μοντέλο:

$$Y_t = \phi_1 Y_{t-1} + \ldots + \phi_p Y_{t-p} + e_t - \theta_1 e_{t-1} - \ldots - \theta_q e_{t-q},$$

πρέπει να εκτιμήσουμε τα $\phi_1, \ldots, \phi_p, \theta_1, \ldots, \theta_q$.

Υπάρχουν δύο βασικές μέθοδοι εκτίμησης των παραμέτρων.

Η πρώτη είναι η <u>μη γραμμική διαδικασία ελαχίστων τετραγώνων</u> ενός μοντέλου ARMA(p,q). Είναι μη γραμμική γιατί οφείλεται στους κινητούς μέσους MA. Ο σκοπός της είναι η εύρεση των εκτιμήσεων της παραμέτρου που ελαχιστοποιούν το άθροισμα των τετραγώνων των σφαλμάτων. Ο τύπος είναι:

$$\min \sum_{t=1}^{n} e_t^2$$

όπου $$e_t = Y_t - \phi_1 Y_{t-1} - \ldots - \phi_p Y_{t-p} + \theta_1 e_{t-1} + \ldots + \theta_q e_{t-q},$$

$\phi_1, \ldots, \phi_p, \theta_1, \ldots, \theta_q$ οι εκτιμητές των παραμέτρων και

n ο αριθμός των παρατηρήσεων.

Η διαδικασία αυτή επαναλαμβάνεται σαν τη λειτουργία ενός αλγόριθμου, έως ότου οι τελικές τιμές των εκτιμήσεων των παραμέτρων να είναι σημαντικά διαφορετικές από τις αρχικές τιμές της διαδικασίας της εκτίμησης. Εάν μετά από κάποιο αριθμό επαναλήψεων του αλγορίθμου, δεν επιτευχθεί αυτό, τότε θα πρέπει να γίνουν αλλαγές με αποτέλεσμα να γίνουν νέες αρχικές εκτιμήσεις.

Η δεύτερη μέθοδος είναι αυτή της <u>πιθανοφάνειας</u> (loglikelihood), η οποία εκτιμά την τιμή που μεγιστοποιεί τη συνάρτηση. Οι τιμές των παραμέτρων που μεγιστοποιούν το μοντέλο αυτό ονομάζονται μέγιστοι εκτιμητές πιθανοφάνειας (MLE). Ο τύπος είναι:

$$\ln L\left(\phi, \theta, \sigma_e^2\right) = -\frac{n}{2}\left(\ln 2\pi\sigma_e^2 - \frac{S(\phi, \mu, \theta)}{2\sigma_e^2}\right)$$

όπου $S(\phi, \mu, \theta) = \sum_{t=p+1}^{n} e_t^2$ είναι το άθροισμα των τετραγώνων των σφαλμάτων.

Δύο άλλα κριτήρια που μπορούμε να χρησιμοποιήσουμε για την εκλογή της τάξης του (κατάλληλου) μοντέλου είναι:

1. Το κριτήριο του Akaike (AIC)

   $AIC = l + 2$ (ο αριθμός των εκτιμηθέντων παραμέτρων)

   όπου $l = -2\ln L$ και L η πιθανοφάνεια.

Το κριτήριο αυτό δεν επιτρέπει την εισαγωγή πρόσθετων όρων στο μοντέλο.

2. Το κριτήριο Schwartz Bayesian (SBC)

   $SBC = l + (\ln n)$ (ο αριθμός των εκτιμηθέντων παραμέτρων)

   όπου n είναι ο αριθμός των συνολικών παρατηρήσεων (Karioti Vassiliki, 1997).

Το κριτήριο αυτό αποτελεί παραλλαγή του AIC κριτηρίου και έχει μια μικρή ανοδική διαφορά στην τιμή του σε σχέση με το πρώτο. Επιλέγουμε πάντα αυτό με τον μικρότερο αριθμό παραμέτρων που συνήθως είναι το AIC.

Διαγνωστικός Έλεγχος

Αν και το μοντέλο που έχει επιλεγεί θεωρείται το καλύτερο από αυτά που εξετάστηκαν, είναι αναγκαίο να ελεγχθεί η επάρκεια του πριν χρησιμοποιηθεί για πρόβλεψη. Αυτός ο διαγνωστικός έλεγχος γίνεται με την εξέταση των υπολοίπων (σφαλμάτων), τα οποία όπως ξέρουμε είναι η διαφορά μεταξύ των πραγματικών και των μετασχηματισμένων δεδομένων. Ο κύριος σκοπός του διαγνωστικού ελέγχου είναι να ανακαλύψουμε αν τα στατιστικά χαρακτηριστικά αυτών των υπολοίπων είναι "λευκός θόρυβος". Αυτό έχει ως συνέπεια ότι τα υπόλοιπα του μοντέλου θα πρέπει να ικανοποιούν συνθήκες όπως:1) **της στασιμότητας**, 2) **της τυχαιότητας**, 3) **της κανονικότητας.**

Η *στασιμότητα* (stasionarity) των σφαλμάτων του μοντέλου ελέγχεται από τα διαγράμματα ACF και PACF, στα οποία δεν πρέπει να παρουσιάζονται στατιστικά σημαντικές αυτοσυσχετίσεις όπως και μερικές αυτοσυσχετίσεις αντίστοιχα. Γι' αυτό τα υπόλοιπα των αυτοσυσχετίσεων πρέπει να βρίσκονται μέσα στο εύρος που ορίζει το $\pm 2/\sqrt{n}$ (μέσα στο 95% του διαστήματος εμπιστοσύνης). Εάν τα υπόλοιπα είναι έξω από το διάστημα εμπιστοσύνης, θεωρούνται στατιστικά διάφορα του μηδενός και το μοντέλο θεωρείται ανεπαρκές.

Η *τυχαιότητα* (ανεξαρτησία) των σφαλμάτων βασίζεται πάνω σε στατιστικούς δείκτες όπως του Box-Pierce (1970) και του Ljung-Box (1978).

Ο δείκτης του Box-Pierce περιγράφεται από τον τύπο:

$$Q = n\sum_{k=1}^{K} r_k^2$$

όπου n το πλήθος των παρατηρήσεων και

όπου K ο αριθμός των αυτοσυσχετίσεων που έχουν ελεγχθεί.

Ο δείκτης Ljung-Box περιγράφεται από τον παρακάτω τύπο:

$$Q^* = n(n+2)\sum_{k=1}^{K}\left(\frac{1}{n-k}\right)r_k^2$$

όπου n το πλήθος των παρατηρήσεων και

όπου K ο αριθμός των αρχικών αυτοσυσχετίσεων που έχουν ελεγχθεί.

Για τον έλεγχο υποθέσεων, πιο χρήσιμος φαίνεται ο δείκτης Ljung-Box.

Το Q^* ακολουθεί την κατανομή X^2 με k-m βαθμούς ελευθερίας (degrees of freedom) όπου m ο αριθμός των παραμέτρων του μοντέλου. Εάν η υπολογισθείσα τιμή του Q^* είναι μεγαλύτερη από το X^2 για k-m βαθμούς ελευθερίας, τότε το μοντέλο θα πρέπει να θεωρηθεί ανεπαρκές. Ο αναλυτής θα πρέπει να αναζητήσει ένα εναλλακτικό μοντέλο και να συνεχίσει την ανάλυση Box-Jenkins μέχρι να βρεθεί ένα ικανοποιητικό μοντέλο.

Η *κανονικότητα* (normality) των σφαλμάτων εξετάζεται μέσω της μελέτης της κατανομής των υπολοίπων στο Q - Q Plot του στατιστικού πακέτου SPSS. Εάν τα σφάλματα (τελίτσες) συμπίπτουν με την ευθεία του διαγράμματος, τότε θεωρούμε ότι τα σφάλματα ακολουθούν την κανονική κατανομή.

Μετά το πέρας της διαδικασίας του διαγνωστικού ελέγχου, αν δυο ή περισσότερα μοντέλα θεωρούνται ίδια, παρόλο που κανένα μοντέλο δεν δείχνει ακριβή προσαρμογή, κατά την επιλογή του πρέπει να επικρατήσει η ιδιότητα της φειδωλότητας (parsimony).

Παρατήρηση

Γενικά τα μοντέλα που αναφέρθηκαν στη μεθοδολογία Box-Jenkins περιέχουν λίγες μόνο παραμέτρους, το πολύ μέχρι τρεις. Ένας από τους στόχους που θέτει η διαδικασία ανάπτυξης μοντέλων πρόβλεψης είναι η ανάπτυξη φειδωλών (parsimonious) μοντέλων (Jeffrey Jarett, 2002). Δηλαδή εάν δύο διαφορετικά μοντέλα περιγράφουν εξίσου καλά μία χρονοσειρά πρέπει να προτιμήσουμε το μοντέλο με τις λιγότερες παραμέτρους. Έτσι έχοντας να εκτιμήσουμε λιγότερες παραμέτρους, η διαδικασία της εκτίμησης γίνεται απλή όπως και η ερμηνεία του μοντέλου, αλλά και τα σφάλματα πρόβλεψης λιγότερα. Τέλος, με λιγότερες παραμέτρους, είναι πιο εύκολο να βρούμε εκτιμήσεις των παραμέτρων που είναι στατιστικά σημαντικές.

Πρόβλεψη

Αφού επιλεγεί το κατάλληλο μοντέλο στα προηγούμενα στάδια της διαδικασίας Box-Jenkins, αυτό χρησιμοποιείται για να γίνουν οι προβλέψεις. Υποθέτουμε ότι έχουμε ένα απλό γραμμικό αυτοπαλινδρομούμενο μοντέλο AR(1)

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

και προβλέπουμε την επόμενη χρονική περίοδο 1. Η εξίσωση γίνεται:

$$Y_{t+1} = \phi Y_t + e_{t+1}$$

$$\widehat{Y}_{t+1} = \phi Y$$

Ο όρος e_{t+1} δεν είναι γνωστός γιατί αναφέρεται στο μελλοντικό τυχαίο σφάλμα γι' αυτό θέτεται ίσος με το μηδέν και αντιστοιχεί στην περίοδο Y_{t+1}.

Εάν θέλουμε να προβλέψουμε τις δύο επόμενες περιόδους, έχουμε:

$$Y_{t+2} = \phi Y_{t+1} + e_{t+2}$$

$$\widehat{Y}_{t+2} = \phi Y_{t+1} = \phi^2 Y_t$$

Το αντίστοιχο συμβαίνει και εδώ με τον όρο e_{t+2} όπως στην προηγούμενη χρονική περίοδο πρόβλεψης που είχαμε e_{t+1} όρους.

Η γενική μορφή της πρόβλεψης για m περιόδους έχει τη μορφή:

$$Y_{t+m} = \phi Y_{t+m-1} + e_{t+m}$$

$$\widehat{Y}_{t+m} = \phi^m Y_t$$

Καθώς η περίοδος της πρόβλεψης γίνεται πιο μακρινή μέσα στο μέλλον, οι πιθανότητες σφάλματος πρόβλεψης μεγαλώνουν. Αν γι' αυτό το μεγάλο διάστημα πρόβλεψης δεν υπάρχουν από το μοντέλο διαθέσιμες τιμές των υπολοίπων, τότε θέτονται ίσα με το μηδέν.

3.16 <u>Εποχικά Μοντέλα (SARIMA)</u>

Όπως αναφέραμε και πιο πάνω στην αναγνώριση της εποχικότητας, υπάρχουν χρονοσειρές που παρουσιάζουν σε ίσα χρονικά διαστήματα την ίδια συμπεριφορά. Όπως τα μοντέλα ARIMA(p,d,q), έτσι και τα μοντέλα που εμφανίζουν εποχικότητα s παρουσιάζουν παρόμοιες ιδιότητες. Τα μοντέλα ARIMA που έχουν επεκταθεί για να χειρίζονται την εποχικότητα συμβολίζονται ως:

$$\mathrm{ARIMA}(p,d,q)\times(P,D,Q)_s$$,όπου s η εποχικότητα.

$\left\{\begin{matrix}\mu\eta\ \varepsilon\pi o\chi\iota\kappa o\iota \\ o\rho o\iota\end{matrix}\right\}$ $\left\{\begin{matrix}\varepsilon\pi o\chi\iota\kappa o\iota \\ o\rho o\iota\end{matrix}\right\}$

Ας υποθέσουμε ότι έχουμε ένα μοντέλο $\mathrm{SARIMA}(1,1,1)(1,1,1)_4$ το οποίο με τη βοήθεια του τελεστή ολίσθησης Β γράφεται ως:

$$(1-\phi_1 B)(1-\Phi_1 B^4)(1-B)(1-B^4)Y_t=(1-\theta_1 B)(1-\Theta_1 B^4)e_t$$

$\left\{\begin{matrix}\varepsilon\pi o\chi\iota\kappa o\iota \\ o\rho o\iota\end{matrix}\right\}$

Οι εποχικοί όροι αναγνωρίζονται από τις αυτοσυσχετίσεις ACF και τις μερικές αυτοσυσχετίσεις PACF. Για ένα εποχικό AR μοντέλο, $\mathrm{ARIMA}(0,0,0)(1,0,0,)_{12}$, η ACF θα φθίνει εκθετικά στις εποχικές υστερήσεις ενώ η PACF θα παρουσιάζει μόνο μια στατιστικά σημαντική τιμή για υστέρηση 12. Αντίστοιχα για ένα εποχιακό MA μοντέλο, $\mathrm{ARIMA}(0,0,0)(0,0,1)_{12}$, η ACF θα παρουσιάζει μόνο μια στατιστικά σημαντική τιμή για υστέρηση 12 και η PACF θα φθίνει εκθετικά στις εποχικές υστερήσεις 12, 24, 36,...

3.17 <u>Συμπέρασμα</u>

Η μεθοδολογία Box-Jenkins αποτελεί μια στατιστικά εξειδικευμένη προσέγγιση στην ανάλυση και την κατασκευή ενός μοντέλου πρόβλεψης με στόχο την όσο το δυνατόν αναπαράσταση μιας χρονοσειράς. Η προσέγγιση αυτή έχει έναν αριθμό πλεονεκτημάτων έναντι άλλων μεθόδων ανάλυσης χρονοσειρών.

1. Η μέθοδος είναι λογική και στατιστικά ακριβής.
2. Αποσπά ένα μεγάλο αριθμό πληροφοριών από τα παλαιότερα δεδομένα της χρονοσειράς.
3. Επιδρά στην αύξηση της ακρίβεια της πρόβλεψης, ενώ συγκρατεί τον αριθμό των παραμέτρων σε ένα ελάχιστο επίπεδο, σε σύγκριση με άλλες παρόμοιες διαδικασίες κατασκευής μοντέλων.

Τα πεδία εφαρμογής της μεθοδολογίας αναφέρονται είτε σε βραχυπρόθεσμο είτε σε μεσοπρόθεσμο ορίζοντα όσο αναφορά την πρόβλεψη. Παράδειγμα αποτελεί η πρόβλεψη των κερδών μιας επιχείρησης, η τιμολόγηση προϊόντων όπως και ο έλεγχος αποθεμάτων.

Η εφαρμογή της μεθοδολογίας Box-Jenkins χρειάζεται μεγάλη προσοχή και δεν θα πρέπει να εφαρμόζεται απερίσκεπτα, ούτε αυτόματα σε όλα τα είδη των χρονοσειρών. Τα αποτελέσματα ωστόσο, μπορούν να είναι πολύ καλά στα χέρια ενός έμπειρου αναλυτή που στοχεύει σε υψηλό επίπεδο ακρίβειας και στη κατανόηση των διαδικασιών που διαμορφώνουν τη χρονοσειρά.

ΚΕΦΑΛΑΙΟ 4: ΟΙΚΟΝΟΜΕΤΡΙΚΕΣ ΜΕΘΟΔΟΙ

4.1 Εισαγωγή

Σε αυτό το κεφάλαιο θα ασχοληθούμε με τα δυναμικά μοντέλα, δηλαδή με μοντέλα που περιλαμβάνουν μεταβλητές σε διαφορετικά χρονικά σημεία (variables at different points of time), και συγκεκριμένα μεταβλητές με χρονικές υστερήσεις. Θα γίνει αναφορά σε πολύ συνηθισμένες περιπτώσεις, όπου οι παρελθούσες τιμές των παρατηρήσεων της εξαρτημένης μεταβλητής αποτελούν μια από τις σημαντικότερες συνιστώσες στην διαμόρφωση της τρέχουσας τιμής της. Όπως επίσης και στις περιπτώσεις όπου οι επιδράσεις από μεταβολές στις ανεξάρτητες μεταβλητές δεν εξαντλούνται στην ίδια χρονική περίοδο, αλλά επηρεάζουν (για διάφορους λόγους) και μελλοντικές τιμές της εξαρτημένης μεταβλητής. Ακόμα κατά την διάρκεια του κεφαλαίου θα γίνει ανάλυση της έννοιας της συνολοκλήρωσης (cointegration) όπως και των σχέσεων και των ελέγχων που διέπουν την έννοια αυτή.

4.2 Λόγοι Ύπαρξης Χρονικών Υστερήσεων

Τα μοντέλα με χρονικές υστερήσεις περιέχουν ως ερμηνευτικές μεταβλητές τιμές με χρονική υστέρηση των ανεξάρτητων μεταβλητών ή και της εξαρτημένης μεταβλητής. Η οικονομική θεωρία είναι συνήθως στατική (ή πιο συγκεκριμένα, τα όρια των χρονικών περιόδων που ορίζει είναι ασαφή), και σπάνια αναφέρεται στις χρονικές υστερήσεις που εκφράζουν τη δυναμική του οικονομικού συστήματος. Τα στατιστικά όμως στοιχεία εκφράζουν στην ουσία την οικονομική δράση ατόμων και κοινωνικών ομάδων, καθώς και τις αντιδράσεις τους σε μεταβολές οικονομικών και μη-οικονομικών παραγόντων. Έτσι οι οικονομικές διεργασίες στην πραγματικότητα είναι δυναμικές, με βαθμιαίες προσαρμογές που υποβοηθούνται ή παρεμποδίζονται από διάφορους παράγοντες όπως:

- Ψυχολογικούς
- Τεχνολογικούς
- Θεσμικούς
- Τρόπους μέτρησης των στατιστικών στοιχείων

4.3 Δομή Χρονικών Υστερήσεων

Όπως είναι γνωστό, η επίδραση μιας ερμηνευτικής μεταβλητής σε μια οικονομική σχέση στην ερμηνευόμενη μεταβλητή δεν είναι άμεση (instantaneous). Ειδικότερα, σε μοντέλα με χρονοσειρές, ο χρόνος που απαιτείται μεταξύ μιας οικονομικής απόφασης και του τελικού αποτελέσματος είναι αρκετά μεγάλος. Στην πράξη, αυτό το γεγονός αντιμετωπίζεται με την εισαγωγή στο μοντέλο μεταβλητών με χρονικές υστερήσεις. Τα μοντέλα αυτής της μορφής εμπίπτουν σε δυο βασικές κατηγορίες (Κασκαρέλης Ιωάννης, 1999):

- Μοντέλα με απεριόριστο αριθμό χρονικών υστερήσεων που περιλαμβάνουν μοντέλα:

    1. Γεωμετρικής μορφής χρονικών υστερήσεων (Geometric Distributed Lag models).
    2. Αναπροσαρμοζόμενων προσδοκιών (Adaptive Expectation Models).
    3. Μερικής προσαρμογής (Partial Adjustment models).
    4. Συνδυασμού αναπροσαρμοζόμενων προσδοκιών και μερικής προσαρμογής (Combination of rehabilitated expectations and partial adaptation).

- Μοντέλα με περιορισμένο αριθμό χρονικών υστερήσεων. Σε αυτήν την κατηγορία ανήκουν μοντέλα:

    1. Πολυωνυμιακής μορφής (Polynomial Distributed Lag Models).
    2. Αριθμητικής μορφής (Arithmetic Distributed Lag Models).
    3. Αντίστροφης Μορφής (Inverted Distributed Lag Models).

4.4 Έννοιες και Ορισμοί

4.4.1 Μοντέλο Κατανεμημένων Χρονικών Υστερήσεων

Μοντέλο κατανεμημένων χρονικών υστερήσεων καλείται το μοντέλο το οποίο η εξαρτημένη μεταβλητή φαίνεται να επηρεάζεται όχι μόνο από την τιμή της ανεξάρτητης μεταβλητής στην ίδια χρονική περίοδο, αλλά και από τις τιμές της ανεξάρτητης μεταβλητής σε προηγούμενες p χρονικές περιόδους, δηλαδή για t = 1, 2, ... p

$$Y_t = a + \beta_0 X_t + \beta_1 X_{t-1} + \beta_2 X_{t-2} + \ldots + \beta_p X_{t-p} + \varepsilon_t$$

Βέβαια σε αυτό το μοντέλο μερικοί από τους συντελεστές $\beta_0, \beta_1, \ldots, \beta_p$ μπορούν να είναι και μηδέν, που σημαίνει ότι οι αντίστοιχες χρονικές υστερήσεις της X δεν επιδρούν στην διαμόρφωση της Y. Ο αριθμός των χρονικών υστερήσεων μπορεί να είναι πεπερασμένος ή να είναι άπειρος αλλά, για να αποφύγουμε την περίπτωση μη-ελεγχόμενης συμπεριφοράς της Y, υποθέτουμε γενικά ότι:

$$\lim_{p \to \infty} \sum_{i=0}^{p} \beta_i = |\mathrm{K}|$$

Η βραχυχρόνια επίδραση της X πάνω στην Y ορίζεται σαν $\beta_0 = \partial Y_t / \partial X_t$. Ενώ σαν μακροχρόνια επίδραση ορίζεται η συνολική επίδραση του X_t και όλων των χρονικών του υστερήσεων (πεπερασμένων ή απείρων) μέχρι εκείνη την περίοδο $t-p$, η οποία έχει οριστεί ότι σηματοδοτεί το ακρότατο χρονικό σημείο υστερήσεων της X_t που έχει επίδραση στην Y_t, δηλαδή η μακροχρόνια επίδραση θα είναι: $\sum_{i=0}^{k} \beta_i$

4.4.2 Μοντέλο Αυτοπαλινδρόμησης

Μοντέλο αυτοπαλινδρόμησης ορίζεται το μοντέλο όπου η εξαρτημένη μεταβλητή επηρεάζεται όχι μόνο από την τρέχουσα τιμή της ανεξάρτητης X_t, αλλά και από τις παρελθούσες τιμές της ίδιας της εξαρτημένης μεταβλητής, δηλαδή:

$$Y_t = a + \beta X_t + \gamma_1 Y_{t-1} + \gamma_2 Y_{t-2} + ... \gamma_p Y_{t-p} + \varepsilon_t$$

το οποίο για $\beta = 0$ αποτελεί το αυτοπαλίνδρομο σχήμα p-τάξης AR(p). Το μοντέλο αυτό γράφεται με την χρήση του τελεστή B ως:

$$(1 - \gamma_1 B - \gamma_2 B^2 - ... - \gamma_p B^p) Y_t = a + \beta X_t + \varepsilon_t \rightarrow C(B) Y_t = a + \beta X_t + \varepsilon_t$$

το οποίο μας δίνει σαν βραχυχρόνιο πολλαπλασιαστή το $\beta = \partial Y_t / \partial X_t$ και σαν μακροχρόνιο πολλαπλασιαστή το: $\dfrac{\beta}{C(1)}$

Για να υπάρχει σε αυτό το μοντέλο στασιμότητα και να έχουμε σύγκλιση της Y σε ένα μακροχρόνιο επίπεδο ισορροπίας μετά από μια μεταβολή της X, θα πρέπει να ισχύουν κάποιες συνθήκες στασιμότητας. Έτσι ορίζεται, στη γενική μορφή του πολυωνύμου $C(B) = 1 - \gamma_1 B - \gamma_2 B^2 - ... - \gamma_p B^p$, ότι για να είναι το μοντέλο στάσιμο θα πρέπει οι χαρακτηριστικές ρίζες της εξίσωσης του αυτοπαλίνδρομου μέρους του μοντέλου, έστω

$$C(z) = 1 - \gamma_1 z - \gamma_2 z^2 - ... - \gamma_p z^p = 0$$

να είναι όλες σε απόλυτη τιμή μεγαλύτερες της μονάδος. Έτσι αν $p = 1$, στο μοντέλο:

$$Y_t = a + \beta X_t + \gamma_1 Y_{t-1} + \varepsilon_t$$

η χαρακτηριστική εξίσωση είναι η $C(z) = 1 - \gamma z = 0$ και η μοναδική ρίζα είναι η $z = 1/\gamma$, η οποία είναι μεγαλύτερη της μονάδος όταν $|\gamma| < 1$.

Στην περίπτωση που $p=2$, δηλαδή $(1-\gamma_1 B-\gamma_2 B^2)Y_t = a+\beta X_t + \varepsilon_t$ για την οποία η χαρακτηριστική εξίσωση του πολυωνύμου είναι: $C(z)=1-\gamma_1 z-\gamma_2 z^2=0$ και αν Δ είναι η διακρίνουσα της χαρακτηριστικής εξίσωσης: $\Delta=\gamma_1^2+4\gamma_2$ τότε οι συνθήκες στασιμότητας του μοντέλου ορίζονται ως εξής:

για Δ>0 και $\gamma_1 > 0$ πρεπει $\gamma_1+\gamma_2<1$
για Δ>0 και $\gamma_1<0$ πρεπει $\gamma_2 < 1+\gamma_1$
για Δ<0 πρεπει $\gamma_2<-1$

4.4.3 Μοντέλα Var (Δυναμικά Μοντέλα)

Ο συνδυασμός των μοντέλων κατανεμημένων χρονικών υστερήσεων και των μοντέλων αυτοπαλινδρόμησης μας δίνει το γενικότερο τύπο των δυναμικών μοντέλων, σύμφωνα με τα οποία η εξαρτημένη μεταβλητή Y_t, δέχεται επιδράσεις όχι μόνο από την/τις ανεξάρτητη/τες στην τρέχουσα χρονική περίοδο αλλά και από τις τιμές ανεξάρτητων και της ίδιας της εξαρτημένης σε προηγούμενες χρονικές περιόδους.

Για την καλύτερη κατανόηση του μοντέλου ας θεωρήσουμε το παράδειγμα ενός διμετάβλητου διανύσματος αυτοπαλινδρομήσεων (VAR) με μεταβλητές το προϊόν, X_t και την προσφορά του χρήματος, M_t. Τότε η συμπεριφορά των δυο αυτών μεταβλητών θα διέπεται από το σύστημα:

$$X_t = a_{10}+a_{11}X_{t-1}+a_{12}X_{t-2}+\beta_{11}M_{t-1}+\beta_{12}M_{t-2}+e_{t1} \quad (4.1\alpha)$$

$$M_t = a_{20}+a_{21}X_{t-1}+a_{22}X_{t-2}+\beta_{21}M_{t-1}+\beta_{22}M_{t-2}+e_{t2} \quad (4.1\beta)$$

όπου έχουμε θεωρήσει δυο υστερήσεις για κάθε ενδογενή μεταβλητή. Αν γράψουμε το σύστημα σε διανυσματική μορφή τότε θα έχουμε:

$$\begin{bmatrix} X_t \\ M_t \end{bmatrix} = \begin{bmatrix} a_{10} \\ a_{20} \end{bmatrix} + \begin{bmatrix} a_{11} & \beta_{11} \\ a_{21} & \beta_{21} \end{bmatrix} \begin{bmatrix} X_{t-1} \\ M_{t-1} \end{bmatrix} + \begin{bmatrix} a_{12} & \beta_{12} \\ a_{22} & \beta_{22} \end{bmatrix} \begin{bmatrix} X_{t-2} \\ M_{t-2} \end{bmatrix} + \begin{bmatrix} e_{t1} \\ e_{t2} \end{bmatrix}$$

ή αλλιώς

$$Y_t = \delta + A_1 Y_{t-1} + A_2 Y_{t-2} + e_t \quad (4.2)$$

όπου

Y_t είναι το διάνυσμα των ενδογενών μεταβλητών,

δ είναι το διάνυσμα των σταθερών όρων,

A_i ($i = 1,2$) είναι οι μήτρες των συντελεστών των ενδογενών μεταβλητών χρονικής υστέρησης και

e_t είναι το διάνυσμα των υπολοίπων.

Η εξίσωση (4.2) είναι ένα AR(2) μοντέλο στο διμετάβλητο διάνυσμα Y_t γι' αυτό και ονομάζεται, όπως αναφερθήκαμε νωρίτερα, αυτοπαλινδρομικό διάνυσμα (Vector Autoregression).

Για την ορθή εκτίμηση ενός VAR μοντέλου, θα πρέπει να ικανοποιούνται ορισμένες βασικές υποθέσεις τόσο για τις ενδογενείς μεταβλητές όσο και για τα υπόλοιπα. Έτσι, υποθέτουμε ότι το διάνυσμα υπολοίπων e_t ενός VAR συστήματος έχει μέσο μηδέν και ότι το υπόλοιπο κάθε εξίσωσης χωριστά έχει σταθερή διακύμανση που οι τιμές του δεν αυτοσυσχετίζονται, αλλά το υπόλοιπο αυτό μπορεί να συσχετίζεται με το υπόλοιπο άλλης εξίσωσης. Εάν ισχύουν αυτές οι προϋποθέσεις, όπως γνωρίζουμε τα υπόλοιπα της κάθε εξίσωσης θα είναι λευκός θόρυβος, ενώ ταυτόχρονα τα υπόλοιπα των εξισώσεων μπορεί να συσχετίζονται μεταξύ τους κατά την τρέχουσα περίοδο.

Επιπλέον θα πρέπει να υποθέσουμε ότι το VAR είναι στάσιμο. Αυτό σημαίνει ότι το διάνυσμα των ενδογενών μεταβλητών Y_t πρέπει να έχει σταθερό μέσο, σταθερή διακύμανση και οι μήτρες συνδιακυμάνσεων μεταξύ Y_t και Y_{t+k} να εξαρτώνται μόνο από την απόσταση k μεταξύ των τιμών και όχι από τον χρόνο t. Αν πληρούνται οι υποθέσεις που προαναφέρθηκαν, τότε η εκτίμηση των ελαχίστων τετραγώνων (least squares) των εξισώσεων του VAR συστήματος είναι μια καλή μέθοδος, που δίνει συνεπείς και αποτελεσματικούς εκτιμητές των παραμέτρων του συστήματος. Εντούτοις είναι δύσκολο να ερμηνευτούν, από οικονομικής πλευράς, οι εκτιμώμενοι συντελεστές ενός VAR συστήματος. Γι'αυτό το λόγο, τα VAR συστήματα μετατρέπονται συνήθως σε μορφή κινητών μέσων έτσι ώστε να μπορεί κανείς να υπολογίσει τις επιδράσεις τις οποίες μπορεί να επιφέρει μια τυχαία διαταραχή στα λάθη στις τιμές των ενδογενών μεταβλητών.

Για να υπολογίσουμε αυτές τις μεταβολές χρησιμοποιούμε τις συναρτήσεις *αιφνίδιων αντιδράσεων* (impulse response functions). Μια τέτοια συνάρτηση προσδιορίζει τις αντιδράσεις των ενδογενών μεταβλητών σε διαταραχές προερχόμενες από συγκεκριμένες μεταβλητές. Έτσι μπορούμε να καταγράψουμε την επίδραση από μια απότομη διαταραχή (shock) μιας ή περισσότερων τυπικών αποκλίσεων πάνω στις πρόσφατες και μελλοντικές τιμές των ενδογενών μεταβλητών (Δημελή Σοφια, 2002).

Εάν θεωρήσουμε την απλούστερη περίπτωση που τα λάθη των εξισώσεων είναι ασυσχέτιστα μεταξύ τους, τότε μπορούμε εύκολα να ερμηνεύσουμε τα αποτελέσματα. Έτσι στο παράδειγμα του μοντέλου (4.1), αν τα e_{t1} και e_{t2} είναι ασυσχέτιστα, τότε το e_{t1} εκφράζει τις τυχαίες διαταραχές (shocks ή innovations) του προϊόντος X_t, ενώ το e_{t2} αυτές τις νομισματικής μεταβλητής M_t. Η εξίσωση λοιπόν αιφνιδίων αντιδράσεων του e_{t2} μετρά την επίδραση μια τυχαίας νομισματικής διαταραχής σήμερα πάνω στις πρόσφατες και μελλοντικές τιμές των X_t και M_t. Αντίστοιχη είναι η ερμηνεία στην περίπτωση μιας αιφνίδιας διαταραχής e_{t1} από την πλευρά της μεταβλητής του προϊόντος.

Παρατήρηση: Με τον όρο ενδογενείς εννοούμε τις τιμές των μεταβλητών που προσδιορίζονται μέσα στο σύστημα, ενώ με τον όρο εξωγενείς, τις τιμές των μεταβλητών που καθορίζονται εκτός συστήματος.

4.5 Επιλογή του Αριθμού των Χρονικών Υστερήσεων βάσει Στατιστικών Κριτηρίων

Αν ο αριθμός των χρονικών υστερήσεων (p, θ) είναι πεπερασμένος αλλά δεν καθορίζεται μέσω της οικονομικής θεωρίας, ή μέσω στατιστικών ή μαθηματικών μοντέλων, τότε η εκτίμηση του μοντέλου με χρονικές υστερήσεις παρουσιάζει τις εξής δυσκολίες: 1) Αν ο αριθμός p και θ των χρονικών υστερήσεων που τελικά θα επιλέξουμε είναι μεγάλος, η εκτίμηση των p + θ + 2 παραμέτρων της συνάρτησης από τις n-παρατηρήσεις μειώνει σημαντικά τους βαθμούς ελευθερίας, και 2) Επειδή συνήθως στις περισσότερες οικονομικές χρονοσειρές υπάρχει υψηλή συσχέτιση μεταξύ της τρέχουσας και των παρελθοντικών τιμών μιας μεταβλητής. Γι' αυτό η εκτίμηση ενός μοντέλου με χρονικές υστερήσεις πιθανόν να έχει σαν αποτέλεσμα την εμφάνιση σοβαρού προβλήματος πολυσυγγραμικότητας, με συνέπειες ως προς την σταθερότητα και, γενικότερα ως προς την αξιοπιστία των εκτιμητών.

Σε περίπτωση που η οικονομική θεωρία δεν μας παρέχει κάποιες ενδείξεις για τον αριθμό των χρονικών υστερήσεων ή για περιορισμούς μεταξύ των συντελεστών τους, τότε η επιλογή των χρονικών υστερήσεων που θα εισαχθούν στο μοντέλο πρέπει να καθοριστεί με συγκεκριμένα στατιστικά κριτήρια. Ένα τέτοιο κριτήριο που χρησιμοποιείται πολύ συχνά, είναι το κριτήριο πληροφορίας του Akaike το οποίο ορίζεται ως:

$$AIC = N \ln\left[\hat{\sigma}_{\varepsilon}^{2}\right] + 2k$$

Όπου: N = Αριθμός των παρατηρήσεων,

k = Αριθμός συντελεστών στο μοντέλο,

$\hat{\sigma}_{\varepsilon}^{2}$ = Εκτιμητής της διακύμανσης.

Το κριτήριο αυτό αντισταθμίζει την αύξηση της ερμηνευτικής ικανότητας από την είσοδο μιας ακόμα υστέρησης με την απώλεια του αντίστοιχου βαθμού ελευθερίας. Έτσι θα επιλέξουμε εκείνο το μοντέλο το οποίο μας δίνει την μικρότερη τιμή στο κριτήριο AIC.

Με το ίδιο κριτήριο μπορούμε να επιλέξουμε τον άριστο αριθμό των χρονικών υστερήσεων στα μοντέλα αυτοσυσχέτισης, καθώς και να αποφανθούμε στα δυναμικά μοντέλα (τα οποία δεν προέρχονται από συγκεκριμένη οικονομική θεωρία η οποία να υποδεικνύει συγκεκριμένες σχέσεις) περί **αιτιότητας κατά Granger** (Κασκαρέλης Ιωάννης, 1999). Δηλαδή μπορούμε να αποφανθούμε μέσω του κριτηρίου AIC, κατά πόσο μια μεταβλητή και οι χρονικές υστερήσεις της ερμηνεύουν στατιστικά σημαντικό μέρος μιας μεταβλητής που εμφανίζεται σαν εξαρτημένη στην συγκεκριμένη εξίσωση, ή κατά πόσο ισχύει το αντίστροφο, όταν δηλαδή εμφανισθεί η πρώτη σαν εξαρτημένη μεταβλητή.

4.6 Αιτιότητα κατά Granger

Κατά την διάρκεια της εξειδικεύσεως ενός μοντέλου ένα σημαντικό ερώτημα που εμφανίζεται είναι, ο προσδιορισμός του κατά πόσο μια μεταβλητή αιτιάζει μια άλλη ή αιτιάζεται από αυτήν ή ακόμα αν είναι ανεξάρτητη από τις άλλες. Στην οικονομική επιστήμη όμως μια τέτοια σχέση είναι σχεδόν αδύνατο να καθοριστεί. Γι΄αυτό συχνά θεωρούμε εκ των προτέρων δεδομένη μια συγκεκριμένη σχέση αιτίου και αποτελέσματος προκειμένου να εφαρμόσουμε τις κλασικές οικονομετρικές μεθόδους εκτίμησης ενός μοντέλου.

Για παράδειγμα εάν έχουμε δυο μεταβλητές Χ και Υ και σύμφωνα με την οικονομική θεωρία η μεταβλητή Χ προσδιορίζει την συμπεριφορά της Υ. Το ερώτημα που τίθεται είναι αν πράγματι μια τέτοια σχέση υπάρχει. Η πιο γνωστή διαδικασία για να απαντηθεί το ερώτημα αυτό είναι να παλινδρομήσουμε την μεταβλητή Υ πάνω στη Χ χρησιμοποιώντας τα υπάρχοντα στατιστικά δεδομένα και να ελέγξουμε τη στατιστική σημαντικότητα του συντελεστή της Χ.

Εντούτοις η ύπαρξη υψηλής συσχέτισης μεταξύ δυο μεταβλητών με βάση μη πειραματικά δεδομένα δεν αποτελεί σε καμιά περίπτωση και απόδειξη ότι υπάρχει σχέση αιτιότητας μεταξύ των υπό μελέτη μεταβλητών. Οι δυσκολίες αυτές ως προς τον καθορισμό μιας αληθινής σχέσης αιτιότητας μεταξύ οικονομικών μεταβλητών οδήγησαν τον Granger (1969) στην ανάπτυξη της οικονομικής έννοιας της αιτιότητας γνωστή ως **"αιτιότητα κατά Granger"** (Granger causality). Γενικά, θα λέμε ότι μια μεταβλητή X_t αιτιάζει κατά Granger μια άλλη Y_t, αν όλη η πρόσφατη και προηγούμενη πληροφόρηση γύρω από τις τιμές της μεταβλητής αυτής βοηθούν στην καλύτερη πρόβλεψη των τιμών της Y_t.

Για να χρησιμοποιηθεί όμως αυτός ο ορισμός της αιτιότητας είναι απαραίτητο να προσδιοριστούν τόσο ο τρόπος διεξαγωγής των προβλέψεων όσο και ο τρόπος μέτρησης της ακρίβειας των προβλέψεων που αφορούν τις τιμές της υπό εξέταση μεταβλητής. Στον ορισμό του ο Granger περιορίζεται στις αμερόληπτες προβλέψεις ελαχίστων τετραγώνων (unbiased least squares predictions) και χρησιμοποιεί για την μέτρηση της ακρίβειας των προβλέψεων αυτών, τη διακύμανση των λαθών πρόβλεψης μιας περιόδου στο μέλλον (one-step prediction error).

Έτσι σύμφωνα με τον ορισμό του Granger, η μεταβλητή X_t αιτιάζει την Y_t αν η πρόβλεψη της Y_t για μια περίοδο στο μέλλον, έστω $\overline{Y}_t$, που προέκυψε με βάση όλη την προηγούμενη πληροφόρηση έχει μικρότερο μέσο τετραγωνικό σφάλμα (MSE) από την πρόβλεψη του Y_t που γίνεται με βάση όλη την προηγούμενη πληροφόρηση πλην εκείνης που αφορά την μεταβλητή X_t (Δημελή Σοφία, 2002).

$$MSE(\overline{Y}/\overline{U}) < MSE(\overline{Y}/\overline{U}-\overline{X})$$

όπου $\overline{U}$ παριστάνει όλη την προηγούμενη πληροφόρηση για την πρόβλεψη $\overline{Y}$ της Y_t και $\overline{X}$ την προηγούμενη πληροφόρηση που αφορά την μεταβλητή X_t. Τα σύνολα αυτά μπορεί να περιλαμβάνουν όχι μόνο την προηγούμενη αλλά και την πρόσφατη πληροφόρηση οπότε τότε λέμε ότι το X αιτιάζει το Y και στιγμιαία (instantaneous causality).

4.7 Έλεγχος Αιτιότητας κατά Granger

Τα μοντέλα VAR τα οποία χρησιμοποιούνται και για προβλέψεις, όπως και κάθε μοντέλο ανάλυσης χρονοσειρών, μπορούν να μας δώσουν ενδείξεις γύρω από τις σχέσεις εξάρτησης μεταξύ των εξεταζόμενων μεταβλητών με την βοήθεια των ελέγχων αιτιότητας κατά Granger. Αυτοί οι έλεγχοι μπορούν να γίνουν κατά εξίσωση, εκτιμώντας τις εξισώσεις ξεχωριστά σαν ανεξάρτητες ή σαν μοντέλα εξισώσεων, που αλληλεξαρτώνται μέσω των διακυμάνσεων των όρων σφάλματος αυτών.

Έτσι, στο προηγούμενο παράδειγμα του VAR μοντέλου μεταξύ προϊόντος X_t και προσφοράς χρήματος M_t, θα λέμε ότι η μεταβλητή του χρήματος M_t αιτιάζει το προϊόν, αν και μόνο αν, ταυτόχρονα οι συντελεστές β_{11} και β_{12} είναι από κοινού στατιστικά σημαντικοί. Δηλαδή, για να αιτιάζει μια μεταβλητή X_t μια άλλη Y_t, θα πρέπει οι συντελεστές όλων των χρονικών υστερήσεων της X_t στην εξίσωση της Y_t να διαφέρουν στατιστικά σημαντικά από το μηδέν, ενώ οι συντελεστές των χρονικών υστερήσεων της Y_t στην εξίσωση της X_t να μην διαφέρουν σημαντικά από το μηδέν. Ο έλεγχος αυτός μπορεί να γίνει με το γνωστό στατιστικό κριτήριο της κατανομής F για την από κοινού σημαντικότητα των παραμέτρων των χρονικών υστερήσεων των αντίστοιχων μεταβλητών.

Έτσι για να ελέγξουμε αν προκύπτουν αιτιώδεις σχέσεις μεταξύ των μεταβλητών του παραδείγματος, θα προχωρήσουμε στον από κοινού στατιστικό έλεγχο των εξής υποθέσεων:

$$H_{0(1)}: \beta_{11} = \beta_{12} = 0 \quad \text{και} \quad H_{1(1)}: \beta_{11} \neq 0, \beta_{12} \neq 0$$
$$H_{0(2)}: a_{21} = a_{22} = 0 \quad \text{και} \quad H_{1(2)}: a_{21} \neq 0, a_{22} \neq 0$$

Αν ο στατιστικός έλεγχος δείξει ότι:

- Η $H_{1(1)}$ και η $H_{0(2)}$ γίνουν αμφότερες δεκτές, τότε συμπεραίνεται ότι η μεταβλητή M_t επηρεάζει την μεταβλητή X_t.
- Η $H_{0(1)}$ και η $H_{1(2)}$ γίνουν αμφότερες δεκτές, τότε συμπεραίνεται ότι η μεταβλητή X_t επηρεάζει την μεταβλητή M_t .
- Η $H_{1(1)}$ και η $H_{1(2)}$ γίνουν αμφότερες δεκτές, τότε υπάρχει δίπλευρη (bilateral) αιτιότητα μεταξύ των μεταβλητών M_t και X_t .
- Η $H_{0(1)}$ και η $H_{0(2)}$ γίνουν αμφότερες δεκτές, τότε συμπεραίνεται ότι δεν υπάρχει αλληλεξάρτηση ανάμεσα στις μεταβλητές M_t και X_t.

Ο παραπάνω έλεγχος πραγματοποιείται με την βοήθεια του παρακάτω κριτηρίου:

$$F = \frac{(SSE^R - SSE^U)/k}{SSE^U/f}$$

όπου

SSE^U =Άθροισμα τετραγώνων υπολοίπων που προκύπτουν από την εκτίμηση της εξίσωσης παλινδρόμησης (4.1α) (μη περιορισμένη εξίσωση).

SSE^R =Άθροισμα τετραγώνων υπολοίπων που προκύπτουν από την εκτίμηση της εξίσωσης (4.1α) υπό τον περιορισμό της $H_{0(1)}$ ότι $\beta_{11} = \beta_{12} = 0$ (περιορισμένη εξίσωση).

Η παράμετρος k δηλώνει τον αριθμό των περιορισμών, άρα $k = 2$ στο συγκεκριμένο παράδειγμα. Η παράμετρος f ισούται με τους βαθμούς ελευθερίας στην μη περιορισμένη εξίσωση, δηλαδή στην (4.1α). Άρα, εδώ f=T-1-4, όπου T είναι το πλήθος των παρατηρήσεων και το 1 αντιστοιχεί στην σταθερά. Πρέπει να σημειώσουμε, ότι η χρήση του στατιστικού κριτηρίου F γίνεται στην περίπτωση της εξίσωσης του προϊόντος. Με την ίδια λογική θα εργαζόμαστε και για την εξίσωση της προσφοράς χρήματος. Αν η τιμή του F που υπολογίζουμε με την παραπάνω εξίσωση είναι μεγαλύτερη από τη τιμή των πινάκων $F_{k,f}$ σε ένα επίπεδο σημαντικότητας a% τότε η υπόθεση $H_{0(1)}$ απορρίπτεται και συμπεραίνουμε ότι οι υστερήσεις της μεταβλητής M_t επηρεάζουν σημαντικά την συμπεριφορά της X_t.

Τέλος, όμως θα πρέπει να υπάρξει προσοχή ως προς τα συμπεράσματα μας γιατί η ισχύς αυτών των ελέγχων αιτιότητας κατά Granger είναι αμφισβητήσιμη, καθώς τα αποτελέσματα μπορεί να επηρεασθούν από τον αριθμό των υστερήσεων που επιλέγουμε καθώς και από την συχνότητα των στοιχείων που χρησιμοποιούμε.

4.8 Έλεγχος Στασιμότητας μέσω Κριτηρίου των Dickey- Fuller

Στη συνέχεια του κεφαλαίου πριν προχωρήσουμε στην έννοια της συνολοκλήρωσης θα αναφερθούμε στη βασική έννοια της στασιμότητας μιας χρονοσειράς με το κριτήριο των Dickey- Fuller όπως και με το επαυξημένο κριτήριο των Dickey- Fuller. Στον ορισμό της συνολοκλήρωσης θα δούμε ότι για να συνολοκληρώνονται δύο μεταβλητές θα πρέπει ο γραμμικός τους συνδυασμός να είναι στάσιμος. Ακολουθούν δύο από τα σημαντικότερα κριτήρια αυτού του ελέγχου.

4.8.1 Το Κριτήριο των Dickey-Fuller

Ο έλεγχος στασιμότητας μιας χρονοσειράς βασίζεται στις στατιστικές t και F και εξαρτάται από τη μορφή του μοντέλου. Συγκεκριμένα διακρίνουμε τις παρακάτω τρεις περιπτώσεις:

$$Y_t = \alpha + \beta T + \rho Y_{t-1} + \varepsilon_t$$

$$Y_t = \rho Y_{t-1} + \varepsilon_t$$

$$Y_t = \alpha + \rho Y_{t-1} + \varepsilon_t$$

όπου

α = η παράμετρος της χρονοσειράς,

T = η τάση,

ρ = ο συντελεστής αυτοσυσχέτισης και

ε_t = λευκός θόρυβος.

4.8.1.1 <u>t-στατιστική</u>

Ο έλεγχος της χρονοσειράς για στασιμότητα με την t-στατιστική αφορά τον έλεγχο του συντελεστή ρ στα δύο πρώτα μοντέλα. Η διαδικασία ελέγχου για την ύπαρξη ή μη μοναδιαίας ρίζας (unit root test) έχει ως εξής:

- Το αντίστοιχο μοντέλο εκτιμάται με την OLS (μέθοδος ελαχίστων τετραγώνων). Για παράδειγμα η εκτίμηση του μοντέλου δύο $Y_t = \rho Y_{t-1} + \varepsilon_t$ έχει ως εξής:

$$Y_t = \rho Y_{t-1}$$

- Ελέγχουμε με την t-στατιστική την υπόθεση:

$$H_0 : \rho = 1 \text{ και } H_1 : \rho < 1$$

Αν $t_\rho < t(\alpha, N-2)$, η υπόθεση H_0 απορρίπτεται, που σημαίνει ότι η χρονοσειρά είναι στάσιμη.

Αν $t_\rho > t(\alpha, N-2)$, η υπόθεση H_0 γίνεται δεκτή, που σημαίνει ότι η χρονοσειρά δεν είναι στάσιμη.

Παρόλα αυτά η διαδικασία ελέγχου της στασιμότητας με την t-στατιστική δεν ισχύει όταν η πραγματική τιμή του συντελεστή ρ ισούται με τη μονάδα. Για να αντιμετωπιστεί αυτό το πρόβλημα οι Dickey-Fuller ανέπτυξαν μια κατανομή για τον εκτιμητή του συντελεστή ρ που ισχύει ακόμα και στην περίπτωση που η πραγματική τιμή του συντελεστή ρ ισούται με την μονάδα. Στην περίπτωση που η H_0 γίνει δεκτή, δηλαδή η χρονοσειρά δεν είναι στάσιμη, το μοντέλο μετασχηματίζεται κατά τέτοιο τρόπο ώστε να αποφευχθούν οι συνέπειες της μη στασιμότητας. Ένας τέτοιος μετασχηματισμός επιτυγχάνεται αν το μοντέλο εκφρασθεί σε πρώτες διαφορές. Συγκεκριμένα το δεύτερο μοντέλο μπορεί να γραφεί ως εξής:

$$Y_t - Y_{t-1} = \rho Y_{t-1} - Y_{t-1} + \varepsilon_t$$

ή

$$\Delta Y_t = (\rho - 1) Y_{t-1} + \varepsilon_t = \gamma Y_{t-1} + \varepsilon_t$$

Αυτή η μετατροπή οδηγεί στον έλεγχο της παρακάτω υπόθεσης:

$$H_0 : \gamma = 0 \quad \text{ή} \quad H_0 : \rho - 1 = 0$$

Εκτιμώντας το μοντέλο $Y_t - Y_{t-1} = \rho Y_{t-1} - Y_{t-1} + \varepsilon_t$ με την OLS, ελέγχουμε τη σημαντικότητα του συντελεστή γ με την t-στατιστική.

Αν η H_0 γίνει δεκτή, τότε $\rho = 1$ άρα η χρονοσειρά δεν είναι στάσιμη.

Αν $\gamma < 0$, τότε $\rho < 1$ άρα η χρονοσειρά είναι στάσιμη.

Αν $\gamma > 0$, τότε $\rho > 1$ άρα η χρονοσειρά δεν είναι στάσιμη.

Η ίδια ακριβώς διαδικασία μπορεί να εφαρμοσθεί για τον έλεγχο μοναδιαίας ή μη ρίζας στα μοντέλα $Y_t = \alpha + \beta T + \rho Y_{t-1} + \varepsilon_t$ και $Y_t = \alpha + \rho Y_{t-1} + \varepsilon_t$.

4.8.1.2 F- Στατιστική

Η F--Στατιστική εφαρμόζεται στην περίπτωση που ο έλεγχος αφορά τον από κοινού έλεγχο όλων των συντελεστών του μοντέλου. Έστω το μοντέλο $Y_t = \alpha + \beta T + \rho Y_{t-1} + \varepsilon_t$.

Η υπόθεση μηδέν εκφράζεται ως εξής:

$$H_0 : \alpha = \beta = 0 \text{ και } \rho = 1$$

Η H_0 απορρίπτεται αν η υπολογιζόμενη F-Στατιστική είναι μεγαλύτερη της θεωρητικής της τιμής. Αποδοχή της H_0 σημαίνει ότι η χρονοσειρά δεν είναι στάσιμη. Στασιμότητα της χρονοσειράς θα μπορούσε να επιτευχθεί αν το μοντέλο εκφρασθεί σε μορφή πρώτων διαφορών. Δηλαδή:

$$\begin{aligned} Y_t - Y_{t-1} &= \alpha + \beta T + \rho Y_{t-1} - Y_{t-1} + \varepsilon_t \\ &= \alpha + \beta T + (\rho - 1) Y_{t-1} + \varepsilon_t \\ &= \alpha + \beta T + \gamma Y_{t-1} + \varepsilon_t \end{aligned}$$

Το επαυξημένο κριτήριο των Dickey-Fuller (ADF) εφαρμόζεται για την ύπαρξη ή μη μοναδιαίας ρίζας όταν το αρχικό μοντέλο περιλαμβάνει περισσότερες από μια χρονικές υστερήσεις.

4.9 <u>Συνολοκλήρωση Χρονοσειρών</u>

Την έννοια της συνολοκλήρωσης (cointegration) εισήγαγε αρχικά ο Granger (1981). Αν έχουμε δύο μη στάσιμες I(1) μεταβλητές και υπάρχει γραμμικός συνδυασμός των μεταβλητών αυτών που να είναι στάσιμος I(0) τότε θα λέμε ότι οι δύο αυτές μεταβλητές συνολοκληρώνονται.

Γενικά, στην περίπτωση που δύο μεταβλητές είναι διαφορετικού βαθμού ολοκληρωσιμότητας, ο γραμμικός τους συνδυασμός θα είναι ολοκληρωμένος στον υψηλότερο βαθμό από τους δύο.

Από την άλλη μεριά αν δύο ή περισσότερες μη στάσιμες μεταβλητές είναι του ίδιου βαθμού ολοκληρωσιμότητας d, είναι δηλαδή I(d) και υπάρχει γραμμικός συνδυασμός μεταξύ τους ή διάνυσμα γραμμικών τους συνδυασμών που να είναι βαθμού ολοκλήρωσης b μικρότερου του βαθμού ολοκλήρωσης $d(b<d)$, για $b>0$, των μεταβλητών αυτών τότε σύμφωνα με τους Engle και Granger (1987) οι χρονοσειρές αυτές συνολοκληρώνονται ή είναι συνολοκληρωμένες (cointegrated) τάξης (d-b) και συμβολίζονται ως CI(d-b). Στην περίπτωση που υπάρχει ένας τέτοιος γραμμικός συνδυασμός, μπορούμε να πούμε ότι υπάρχει και μια μακροχρόνια σχέση μεταξύ των μεταβλητών αυτών, αν και οι βραχυχρόνιες διακυμάνσεις τους μπορεί να μη συσχετίζονται μεταξύ τους. Δηλαδή στο μακροχρόνιο επίπεδο, αυτές οι μεταβλητές συνδιακυμαίνονται ή παρουσιάζουν κοινές μακροχρόνιες τάσεις. Άρα η εξέταση σε αυτήν την περίπτωση των αιτιωδών οικονομικών σχέσεων μεταξύ των μεταβλητών αυτών, πρέπει οπωσδήποτε να λαμβάνει υπόψη της και την πληροφορία περί κοινής τάσης στο μακροχρόνιο επίπεδο, αλλιώς διαπράττουμε σφάλμα εξειδίκευσης. Ο τρόπος να συμπεριλάβουμε αυτή την πληροφορία είναι μέσω των **Διανυσματικών Μοντέλων Διόρθωσης Λαθών** (Vector Error Correction, VEC Models).

Το πλεονέκτημα των VEC είναι ότι μελετούν τις βραχυχρόνιες μεταβολές των μεταβλητών ενώ ταυτόχρονα περιορίζουν τις μη στάσιμες συνολοκληρωμένες χρονοσειρές να συγκλίνουν στη μακροχρόνια σχέση ισορροπίας. Δηλαδή τα σφάλματα των VAR μοντέλων διορθώνονται σταδιακά μέσω των βραχυχρόνιων προσαρμογών των επιμέρους μεταβλητών του συστήματος.

Ο Granger (1983) και οι Engle και Granger (1987) έχουν δείξει ότι, αν οι μεταβλητές Υ και Χ είναι I(1) και συνολοκληρωμένες, υπάρχει ένα μοντέλο διόρθωσης λαθών με την ακόλουθη μορφή:

$$\Delta Y_t = -\rho_1 \varepsilon_{t-1} + \text{υστερησεις}(\Delta Y, \Delta X) + \varepsilon_{1t}$$
$$\Delta X_t = -\rho_2 \varepsilon_{t-1} + \text{υστερησεις}(\Delta Y, \Delta X) + \varepsilon_{2t}$$

για $|\rho_1| + |\rho_2| \neq 0$

Το ε_{t-1} είναι το λάθος υστέρησης μιας περιόδου από την παλινδρόμηση συνολοκλήρωσης $y_t = \beta x_t + e_t$ και $\varepsilon_{it}, i = 1,2$ είναι τα δύο υπόλοιπα των εξισώσεων διόρθωσης λαθών.

Η διαδικασία της εκτίμησης της παλινδρόμησης συνολοκλήρωσης, $y_t = \beta x_t + e_t$, έχει ως εξής:

- Με την OLS (μέθοδος ελαχίστων τετραγώνων) εκτιμούμε το παραπάνω μοντέλο διόρθωσης λαθών από το οποίο παίρνουμε τα υπόλοιπα e_t.
- Τα υπολογιζόμενα υπόλοιπα αντικαθίστανται στο μοντέλο της παλινδρόμησης συνολοκλήρωσης, τα οποία στη συνέχεια εκτιμώνται με την OLS. Οι υπολογιζόμενες εξισώσεις χρησιμοποιούντα για μελλοντικές προβλέψεις των χρονοσειρών.

Οι συντελεστές ρ_1 και ρ_2 μπορούν να ερμηνευθούν και ως συντελεστές της ταχύτητας προσαρμογής (speed of adjustment) των λαθών ισορροπίας. Επομένως όσο πιο μεγάλοι είναι οι συντελεστές αυτοί, τόσο πιο γρήγορη θα είναι η προσαρμογή της ενδογενούς μεταβλητής στη διόρθωση του σφάλματος της προηγούμενης περιόδου και το αντίθετο ενώ στην περίπτωση που οι συντελεστές αυτοί είναι μηδενικοί, τότε δεν υπάρχει σχέση συνολοκλήρωσης δηλαδή οι μεταβλητές αυτές δεν συνολοκληρώνονται. Η ανάλυση αυτή γενικεύεται εύκολα και για μοντέλα με περισσότερες από δύο μεταβλητές.

Το σημαντικό είναι ότι όχι μόνο η συνολοκλήρωση συνεπάγεται την ύπαρξη ενός μοντέλου διόρθωσης λαθών αλλά ισχύει και το αντίθετο. Δηλαδή η ύπαρξη ενός μοντέλου διόρθωσης λαθών συνεπάγεται την συνολοκλήρωση των μεταβλητών. Αυτό το αποτέλεσμα είναι το λεγόμενο **Αντιπροσωπευτικό Θεώρημα του Granger (Granger Representation Theorem)**. Σημαντικό είναι ότι με τα μοντέλα διόρθωσης λαθών VEC γίνεται απευθείας έλεγχος ύπαρξης συνολοκλήρωσης ώστε στη συνέχεια να γίνει η εκτίμηση της υπάρχουσας σχέσης αυτής.

4.9.1 Έλεγχοι Συνολοκλήρωσης

Για να ελέγξουμε αν δύο ή περισσότερες χρονοσειρές συνολοκληρώνονται υπάρχουν δύο βασικές κατηγορίες μεθόδων που μπορούμε να ακολουθήσουμε:

1. Η μέθοδος της μιας εξίσωσης και
2. Η μέθοδος των συστημάτων εξισώσεων

Η πρώτη κατηγορία περιλαμβάνει τους αρχικούς ελέγχους συνολοκλήρωσης που πρότειναν οι Engle και Granger (1987) ως εφαρμογή του Αντιπροσωπευτικού Θεωρήματος που διατύπωσαν.

Η δεύτερη κατηγορία προσδιορίζει το μέγιστο αριθμό σχέσεων συνολοκλήρωσης που διέπουν τις μεταβλητές. Η πιο διαδεδομένη μέθοδος από την κατηγορία αυτή είναι η μέθοδος του Johansen.

4.9.2 Μέθοδος Granger - Engle δύο Σταδίων για έλεγχο Συνολοκλήρωσης

Οι Granger και Engle πρότειναν μια διαδικασία δύο σταδίων (two-step method) για τον έλεγχο περί τη συνολοκλήρωση και την εύρεση ενός ασυμπτωτικά αποτελεσματικού εκτιμητή του όρου διόρθωσης σφάλματος, με τη χρήση της μακροχρόνιας σχέσης μεταξύ δύο μεταβλητών. Πρώτα εξετάζουμε κατά πόσον οι δύο μεταβλητές μας y_t και x_t έχουν τον ίδιο βαθμό ολοκλήρωσης έστω I(1) μέσω των κριτηρίων ADF. Στη συνέχεια εκτιμούμε τη μακροχρόνια σχέση αγνοώντας το πρόβλημα της μη-στασιμότητας:

$$y_t = \beta x_t + \varepsilon_t = \widehat{\beta} x_t + \widehat{\varepsilon}_t$$

Για να είναι οι δύο μεταβλητές συνολοκληρωμένες, θα πρέπει τα υπόλοιπα $\widehat{\varepsilon}_t$ από αυτήν την μακροχρόνια σχέση να προσφέρουν επιπλέον πληροφορίες. Δηλαδή αν y_t και x_t, I(1) σειρές δεν συνολοκληρώνονται, τότε τα $\widehat{\varepsilon}_t$ ακολουθούν ένα ARIMA(0,1,0), δηλαδή δεν υπάρχουν στην περίπτωση αυτή αυτοπαλίνδρομα ή σχήματα κινητού μέσου των υπολοίπων που να δίνουν επιπλέον πληροφορίες. Άρα αν:

$$\widehat{\varepsilon}_t = \rho \widehat{\varepsilon}_{t-1} + v_t - \theta v_{t-1}$$

η υπόθεση περί μη-συνολοκλήρωσης μεταξύ των y_t και x_t διατυπώνονται ως:

$$H_0 : \rho = 1 \text{ και } \theta = 0$$

Δηλαδή ισοδυναμεί με έλεγχο Dickey-Fuller για την ύπαρξη μοναδιαίας ρίζας στα υπόλοιπα, που προκύπτουν από την εκτίμηση της μακροχρόνιας σχέσης $y_t = \hat{\beta} x_t + \hat{\varepsilon}_t$. Αν τα $\hat{\varepsilon}_t$ είναι ολοκληρωμένα με I(1), τότε λέμε ότι οι y_t και x_t δεν είναι συνολοκληρωμένες, ενώ αν τα $\hat{\varepsilon}_t$ είναι ολοκληρωμένα με I(0), τότε οι y_t και x_t είναι συνολοκληρωμένες και υπάρχει μεταξύ τους γραμμικός συνδυασμός μικρότερου βαθμού ολοκλήρωσης. Θα πρέπει να παρατηρήσουμε ότι εδώ εκτιμάται μια συνάρτηση, η οποία υποτίθεται ότι αντιπροσωπεύει μια μακροχρόνια σχέση μεταξύ των προς εξέταση μεταβλητών. Αν ο αριθμός των παρατηρήσεων δεν είναι πολύ μεγάλος, υπάρχει πρόβλημα με τη δύναμη του ελέγχου. Οι Banerjee et al (1986) πρότειναν το R^2 της μακροχρόνιας αυτής σχέσης σαν προσεγγιστικό κριτήριο της δύναμης ελέγχου. Δηλαδή το R^2 πρέπει να προσεγγίζει τη μονάδα ώστε ο έλεγχος μιας τέτοιας μακροχρόνιας σχέσης να θεωρείται ότι είναι επαρκής.

4.10 Συμπέρασμα

Σε αυτό το κεφάλαιο έγινε αναφορά σε περιπτώσεις μοντέλων με χρονικές υστερήσεις. Τέτοια μοντέλα είναι και τα δυναμικά (VAR), που αναφερθήκαμε πρωτίστως κατά την διάρκεια του κεφαλαίου. Τα μοντέλα VAR αποτελούν επέκταση των μονομεταβλητών αυτοπαλινδρομικών μοντέλων (AR(p)).

Τα δυναμικά μοντέλα εκτός από την αποτελεσματικότητα και την επιτυχία τους σε προβλέψεις συστημάτων αλληλοεξαρτώμενων μεταβλητών, θεωρούνται εύκολα και γρήγορα στην χρήση τους. Για τους λόγους αυτούς, η μεθοδολογία των VAR μοντέλων εφαρμόζεται ευρέως σε εμπειρικές μελέτες όσον αφορά τον οικονομικό τομέα.

Λόγω των μεταβλητών σε διαφορετικά χρονικά σημεία που περιλαμβάνουν τα δυναμικά μοντέλα, βασικό ρόλο έχει η επιλογή του αριθμού των υστερήσεων. Εκτός από αυτό, οι σχέσεις εξάρτησης μεταξύ των μεταβλητών των μοντέλων καθορίζονται από κρίσιμες έννοιες όπως της συνολοκλήρωσης και ελέγχονται από κριτήρια στασιμότητας όπως του Dickey-Fuller κ.α.

ΚΕΦΑΛΑΙΟ 5: ΑΝΑΛΥΣΗ ΣΤΟ ΠΕΔΙΟ ΤΩΝ ΣΥΧΝΟΤΗΤΩΝ

5.1 Εισαγωγή

Στο κεφάλαιο αυτό θα αναφερθούμε στην ανάλυση χρονοσειρών στο πεδίο των συχνοτήτων, σε αντίθεση με τα προηγούμενα κεφάλαια, στα οποία εξετάσαμε τις χρονοσειρές στο πεδίο του χρόνου. Η χρονοσειρά αναλύεται στο πεδίο των συχνοτήτων για τον λόγο ότι έχει περιοδική κυματομορφή. Η ανάλυση αυτή είναι συμπληρωματική της ανάλυσης στο πεδίο του χρόνου. Μπορεί όμως να διερευνηθούν χαρακτηριστικά που δεν εντοπίζονται εύκολα με την ανάλυση στο πεδίο του χρόνου. Αυτά τα χαρακτηριστικά έχουν κυρίως σχέση με περιοδικότητες που συνυπάρχουν στην χρονοσειρά. Στην ανάλυση του πεδίου των συχνοτήτων η χρονοσειρά θεωρείται στάσιμη. Βασικό στοιχείο της γραμμικής ανάλυσης είναι η μελέτη της αυτοσυσχέτισης (ή αυτοσυνδιασποράς) που συνοψίζει τις συσχετίσεις σε διάφορες υστερήσεις, δηλαδή χρόνους. Ισοδύναμα μπορούμε να μελετήσουμε το φάσμα ισχύος, δηλαδή την κατανομή της ισχύος της χρονοσειράς σε όλες τις δυνατές συχνότητες.

5.2 Βασικές 'Εννοιες – Συμβολισμοί

Οι συμβολισμοί που θα χρησιμοποιηθούν σε αυτό το κεφάλαιο είναι:

- $\{X_n\}$: η χρονοσειρά ορισμένη θεωρητικά ως στοχαστική διαδικασία σε διακριτό χρόνο n (συνήθως υποθέτουμε $-\infty < n < \infty$).
- $\{X(t)\}$: η χρονοσειρά ορισμένη θεωρητικά ως στοχαστική διαδικασία σε συνεχή χρόνο t (συνήθως υποθέτουμε $-\infty < t < \infty$).
- $\{\chi_n\}$: η παρατηρούμενη χρονοσειρά, δηλαδή μια πραγματοποίηση της διακριτής στοχαστικής διαδικασίας $\{X_n\}$, ή της συνεχούς στοχαστικής διαδικασίας $\{X(t)\}$ που παρατηρείται σε διακριτές χρονικές στιγμές $n = t\tau_s$, όπου τ_s είναι ο χρόνος δειγματοληψίας.
- N : το μήκος της παρατηρούμενης χρονοσειράς $\{\chi_n\}$. Συνήθως θα θεωρούμε ότι η χρονοσειρά είναι $\{\chi_0, \chi_1, \dots, \chi_{N-1}\}$.

- ν : Η συχνότητα ορίζεται ως το αντίστροφο της περιόδου. Οι αντίστοιχες συχνότητες είναι $\frac{1}{2},\frac{1}{3},...,\frac{1}{T}$ ή σε γωνιακές συχνότητες (σε ακτίνες ανά μονάδα χρόνου) $\frac{2\pi}{2},\frac{2\pi}{3},...,\frac{2\pi}{T}$. Η θεμελιώδης συχνότητα ταλάντωσης, δηλαδή η συχνότητα της πρώτης αρμονικής ταλάντωσης, είναι $\nu = 1/T$ και αντίστοιχα η θεμελιώδης γωνιακή συχνότητα είναι $\omega = 2\pi/T = 2\pi\nu$.

5.3 Σειρές Fourier

Υποθέτουμε μια χρονοσειρά μήκους N ως μια σειρά από κύκλους περιόδου $2,3,...,T$.

Μπορούμε λοιπόν να ορίσουμε μια χρονοσειρά ως μια περιοδική κυματομορφή (periodic waveform) περιόδου το πολύ T που δίνεται από τη σειρά Fourier:

$$X_n = \alpha_0 + \sum_{k=1}^{M}\left(\alpha_k \cos(2\pi k\nu n) + b_k \sin(2\pi k\nu n)\right),$$

όπου α_0 είναι η μέση τιμή, α_k και b_k είναι τα πλάτη για την κάθε συνημιτονοειδή και ημιτονοειδή ταλάντωση στις αρμονικές συχνότητες $k\omega = 2\pi k\nu$ αντίστοιχα και το M μπορεί να τείνει στο άπειρο. Για μία μη περιοδική χρονοσειρά μήκους N η υψηλότερη δυνατή περίοδος είναι $T = N\tau_s$. Στην πράξη ο αριθμός των ταλαντώσεων M περιορίζεται από τη χαμηλότερη συχνότητα $V = 1/(N\tau_s)$ (που είναι η θεμελιώδης συχνότητα) και από την υψηλότερη συχνότητα $V_s = 1/(2\tau_s)$.

Για ευκολία στους μαθηματικούς υπολογισμούς χρησιμοποιούμε την ισοδύναμη εκθετική μορφή της σειράς Fourier, (της προηγούμενης εξίσωσης)

$$X_n = \sum_{k=-M}^{M} d_k e^{i2\pi k\nu n},$$

όπου

$$d_k = \begin{cases} (\alpha_k + ib_k)/2, & k<0 \\ \alpha_0, & k=0 \\ (\alpha_k - ib_k)/2, & k>0 \end{cases}$$

$$e^{-ix} = \cos(x) - i\sin(x)$$

Με αυτόν τον τρόπο ορίζουμε σε κάθε αρμονική συχνότητα $k\omega = 2\pi k\nu$ μια τριγωνομετρική μορφή και ισοδύναμα μια μιγαδική μορφή. Το πλάτος της μιγαδικής μορφής είναι $|d_k| = \sqrt{\alpha_k^2 + b_k^2}$ και η φασική γωνία είναι $\phi_k = -\tan^{-1}(\alpha_k/b_k)$.

Τα πλάτη $|d_k|$ (για όλο το φάσμα των συχνοτήτων) εκφράζουν τα γραμμικά χαρακτηριστικά της χρονοσειράς ενώ αν υπάρχουν επιπλέον μη-γραμμικές συσχετίσεις αυτές διατηρούνται στις φασικές γωνίες ϕ_k.

5.4 Μετασχηματισμός Fourier

Αν υποθέσουμε ότι ο χρόνος είναι συνεχής και η χρονοσειρά είναι μια συνεχής κυματομορφή με υψηλότερη περίοδο T, τότε το πλάτος κάθε μιας από τις M αρμονικές ταλαντώσεις είναι

$$d_k = \frac{1}{t}\int_{-T/2}^{T/2} X(t) e^{-i2\pi k\nu t} dt.$$

Καθώς η περίοδος T αυξάνει, το διάστημα $d\nu = 1/T$ μεταξύ των συχνοτήτων των ταλαντώσεων μικραίνει. Αφήνοντας την περίοδο να τείνει στο άπειρο, θεωρώντας δηλαδή ότι η κυματομορφή δεν είναι περιοδική και διαιρώντας με $d\nu$ στην παραπάνω σχέση, ορίζουμε το **μετασχηματισμό Fourier** για ένα συνεχές φάσμα συχνοτήτων ν

$$F(\nu) = \int_{-\infty}^{\infty} X(t) e^{-i2\pi\nu t} dt.$$

Αν ο χρόνος δεν είναι συνεχής και έχουμε μια χρονοσειρά N στοιχείων, τότε ορίζεται ο **διακριτός μετασχηματισμός Fourier** με επίσης N στοιχεία ως

$$F_D(\nu) = \sum_{n=1}^{N} X_n e^{-i2\pi\nu n}, \quad -1/2 < \nu < 1/2.$$

Συνήθως υποθέτουμε ότι η συχνότητα παίρνει τιμές στο διάστημα $(-1/2,1/2)$ αλλά όταν δίνεται ο χρόνος δειγματοληψίας τ_s η συχνότητα ορίζεται στο $\left(-1/(2\tau_s),1/(2\tau_s)\right)$ και το άθροισμα στην παραπάνω εξίσωση πολλαπλασιάζεται με τ_s.

Όταν το πλήθος των παρατηρήσεων N είναι δύναμη του 2, ο υπολογισμός του $F_D(\nu)$ μπορεί να γίνει με πολύ λιγότερες πράξεις (NlogN αντί για N^2) με τη χρήση του αλγορίθμου του **Γρήγορου Μετασχηματισμού Fourier (Fast Fourier Transform, FFT)**. Ακόμα και όταν το μήκος της χρονοσειράς δεν είναι δύναμη του 2, μπορούμε να προσθέσουμε κατάλληλο αριθμό μηδενικών στο τέλος της χρονοσειράς για να το πετύχουμε (αυτό δεν επηρεάζει τη συνάρτηση $F_D(\nu)$ παρά μόνο την ευκρίνειά της ως προς τη συχνότητα ν).

Τα στοιχεία του μετασχηματισμού Fourier $F(\nu)$ ή του διακριτού μετασχηματισμού Fourier $F_D(\nu)$ είναι μιγαδικοί αριθμοί. Κάθε μιγαδικός αριθμός $F(\nu)$ δίνεται ως $F(\nu)=R(\nu)+i\mathrm{I}(\nu)$, έχει μέτρο $|F(\nu)|=\sqrt{R(\nu)^2+i\mathrm{I}(\nu)^2}$ και φασική γωνία $\phi(\nu)=\tan^{-1}\left(\mathrm{I}(\nu)/R(\nu)\right)$.

Αντίστοιχα ορίζεται και ο αντίστροφος μετασχηματισμός Fourier που μεταφέρει την πληροφορία που περιέχεται στο φάσμα συχνοτήτων $\left(F(\nu) \text{ ή } F_D(\nu)\right)$ πίσω στο πεδίο του χρόνου (X_n).

5.5 Φάσμα Ισχύος Στοχαστικής Διαδικασίας

Θεωρούμε τη διακριτή εργοδική στοχαστική διαδικασία $\{X_n\}$ για $-\infty<n<\infty$ με συνάρτηση αυτοσυνδιασποράς $\gamma_x(k)$ που ορίζεται για θετικές και αρνητικές τιμές της υστέρησης k και ικανοποιεί την συνθήκη $\Sigma_{\kappa=-\infty}^{\infty}|\gamma_x(k)|<\infty$. Επίσης θα θεωρούμε πάντα πως η μέση τιμή της $\{X_n\}$ είναι 0.

Το **φάσμα ισχύος** ορίζεται από το θεώρημα **Wiener-Khintchine** ως ο (διακριτός) μετασχηματισμός Fourier της συνάρτησης αυτοσυνδιασποράς $\gamma_x(k)$

$$P_x(\nu) = \sum_{\kappa=-\infty}^{\infty} \gamma_x(k) e^{-i2\pi\nu k}, \qquad -1/2 < \nu < 1/2.$$

Επίσης το φάσμα ισχύος δίνεται από το τετράγωνο του μέτρου των μιγαδικών τιμών του διακριτού μετασχηματισμού Fourier της X_n, δηλαδή ως

$$P_x(\nu) = \lim_{M\to\infty} E\left[\frac{1}{2M+1} \left| \sum_{n=-M}^{M} X_n e^{-i2\pi\nu n} \right|^2 \right]$$

όπου $E[x]$ είναι η μέση τιμή του x. Αυτός ο ορισμός είναι η βάση για την εκτίμηση του φάσματος ισχύος με την μέθοδο του **περιοδογράμματος**. Οι δυο ορισμοί του φάσματος ισχύος των προηγούμενων εξισώσεων είναι ισοδύναμοι όταν η συνάρτηση της αυτοσυσχέτισης φθίνει ικανοποιητικά γρήγορα. Στην συνέχεια ακολουθούν κάποιες ιδιότητες και παρατηρήσεις για το φάσμα ισχύος:

- $P_x(\nu)$ είναι συνεχής μη-αρνητική συνάρτηση και παίρνει πραγματικές τιμές
- $P_x(\nu)$ είναι άρτια συνάρτηση, $P_x(\nu) = P_x(-\nu)$
- Η αυτοσυνδιασπορά δίνεται ως συνάρτηση του $P_x(\nu)$ από τον αντίστροφο μετασχηματισμό Fourier, δηλαδή ως

$$\gamma_x(k) = \int_{-1/2}^{1/2} P_x(\nu) e^{i2\pi\nu k} d\nu$$

Ο τύπος αυτός χρησιμοποιείται για τον υπολογισμό του $\gamma_x(k)$ επειδή ο υπολογισμός του $P_x(\nu)$ είναι πιο γρήγορος, ειδικά με την χρήση του FFT.

- Η διασπορά της διαδικασίας (ή ολική ισχύ όπως λέγεται με αναφορά στο πεδίο των συχνοτήτων) ορίζεται εναλλακτικά από το $P_x(v)$ και ισχύει

$$\gamma_x(0) = \int_{-1/2}^{1/2} P_x(v)dv = \Sigma_{n=-\infty}^{\infty} x_n^2 = \sigma_x^2,$$

δηλαδή η ολική ισχύ είναι ιδία είτε την υπολογίσουμε στο πεδίο του χρόνου είτε στο πεδίο των συχνοτήτων. Το παραπάνω αποτέλεσμα είναι ως γνωστό ως το **θεώρημα του Parseval** (Κουγιουμτζής Δημήτριος, Α.Π.Θ. 2004)

- Ο όρος $P_x(v)dv$ δηλώνει την συνεισφορά στη διαφορά από τα στοιχεία της διαδικασίας με συχνότητες στο διάστημα $(v, v+dv)$. Αν σε αυτό το διάστημα αντιστοιχεί κορυφή του $P_x(v)$ αυτό δηλώνει ότι τα στοιχεία σε αυτό το διάστημα συχνοτήτων συνεισφέρουν σημαντικά στην ολική ισχύ.

- Υπάρχει αντιστοιχία του $P_x(v)$ με την συνάρτηση πυκνότητας πιθανότητας $f_x(x)$. Δηλαδή δεν έχει νόημα να υπολογίσουμε τη συνάρτηση για κάποια συγκεκριμένη τιμή αλλά για ένα διάστημα τιμών .

- Μπορούμε να ορίσουμε το $P_x(v)$ από την αυτοσυσχέτιση $\rho_x(k)$ αντί της αυτοσυνδιασποράς $\gamma_x(k)$. Τότε $\int_{-1/2}^{1/2} Px(v)dv = 1$, δηλαδή η ολική ισχύς ισούται με 1. Συνήθως όμως θεωρούμε την $P_x(v)$ ως μετασχηματισμό Fourier της $\gamma_x(k)$.

Στην συνέχεια παρουσιάζονται τα φάσματα κάποιων απλών συστημάτων.

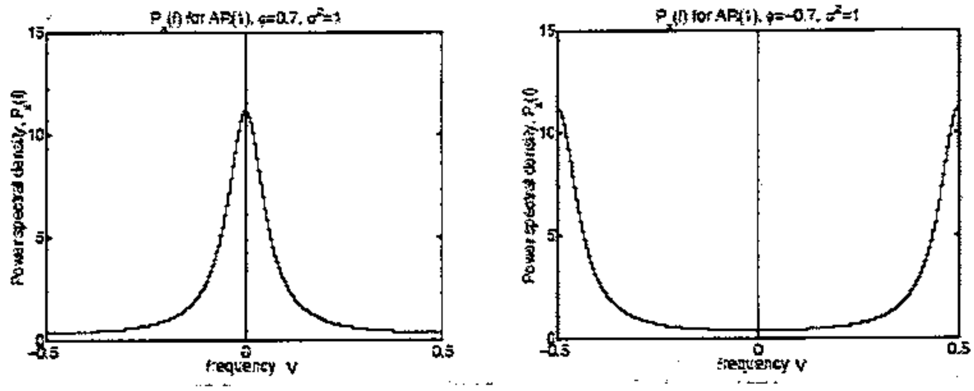
5.6 Φάσμα Ισχύος Λευκού Θορύβου

Το φάσμα ισχύος λευκού θορύβου, $\{Z_t\} \sim WN(0,\sigma_z^2)$ είναι $P_z(\nu)=\sigma_z^2$, $-1/2<\nu<1/2$, δηλαδή κάθε συχνότητα του φάσματος συχνοτήτων συνεισφέρει το ίδιο στη διασπορά της διαδικασίας.

5.7 Φάσμα Ισχύος Διαδικασίας AR(1)

$$P_Z(\nu)=\frac{\sigma_z^2}{1-2\phi\cos(2\pi\nu)+\phi^2}, \qquad -1/2<\nu<1/2,$$

Στα δυο παρακάτω σχήματα δίνεται το $P_X(\nu)$ του AR(1) για φ=0.7και φ=-0.7.



ΠΗΓΗ: http://users.auth.gr/~dkugiu/Teach/TimeSeries/

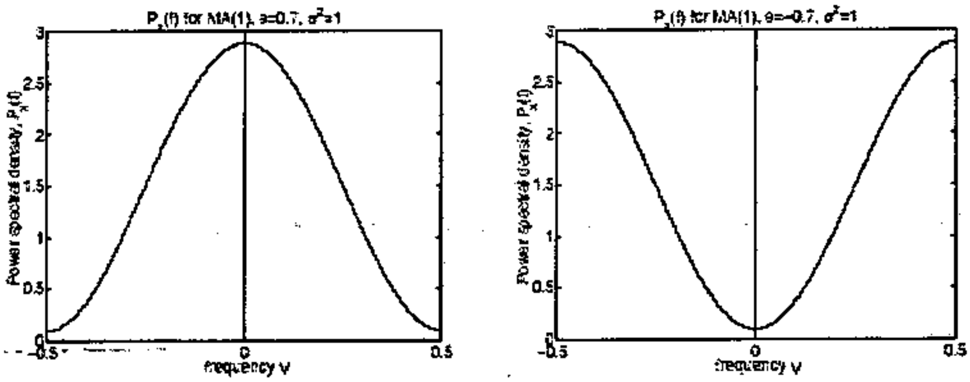
Για θετικό συντελεστή φ του AR(1) η ισχύς δίνεται από τις χαμηλές συχνότητες (η χρονοσειρά φαίνεται πιο ομαλοποιημένη), ενώ για αρνητικό φ η ισχύς βρίσκεται στις υψηλές συχνότητες (η χρονοσειρά δείχνει διαταραχές σε μικρή χρονική κλίμακα).

5.8 Φάσμα Ισχύος Διαδικασίας MA(1)

Το φάσμα ισχύος μιας διαδικασίας MA(1) $(x_t = z_t + \theta z_{t-1}, z_t,\ WN(0,\sigma_z^2))$ είναι

$$P_Z(\nu) = \sigma_z^2(1 + 2\theta\cos(2\pi\nu) + \theta^2), \qquad -1/2 < \nu < 1/2,$$

Στα δυο παρακάτω σχήματα δίνεται το $P_X(\nu)$ του MA(1) για θ=0.7 και θ=-0.7



ΠΗΓΗ: http://users.auth.gr/~dkugiu/Teach/TimeSeries/

Όμοια είναι τα συμπεράσματα για θετικό και αρνητικό συντελεστή θ του MA(1) αλλά η κορυφή στο ν=0 για θ=0.7 του MA(1) είναι πιο χαμηλή και πλατιά από αυτήν για κορυφή ϕ=0.7 του AR(1). Επίσης στην κορυφή του AR(1) για ν=0 και ϕ=0.7 αντιστοιχεί κοιλάδα του MA(1) για ν=0 και θ=-0.7

5.9 Εκτίμηση Φάσματος Ισχύος

Λόγω της σημασίας του φάσματος ισχύος στην μελέτη χρονοσειράς έχουν αναπτυχθεί πολλές μέθοδοι εκτίμησής του. Οι βασικότερες είναι:

1. **Κλασικές ή μη-παραμετρικές μέθοδοι εκτίμησης:** Με αυτές τις μεθόδους η εκτίμηση του φάσματος ισχύος γίνεται απευθείας από την χρονοσειρά (σήμα). Σε αυτήν την κατηγορία ανήκει για παράδειγμα το περιοδόγραμμα (periodogram) και η μέθοδος Welch.

2. **Μοντέρνες ή παραμετρικές μέθοδοι εκτίμησης:** Εδώ η εκτίμηση του φάσματος ισχύος γίνεται με τη βοήθεια ενός γραμμικού μοντέλου που προσαρμόζεται στην χρονοσειρά και εκτιμώνται οι παράμετροι αυτού του μοντέλου. Τέτοιες μέθοδοι είναι αυτές που βασίζονται στην εκτίμηση των παραμέτρων του AR μοντέλου, όπως η μέθοδος Yule-Walker και η μέθοδος Burg.

3. **Μέθοδοι υποχώρου ή μέθοδοι υψηλής ευκρίνειας:** Οι μέθοδοι αυτοί χρησιμοποιούνται για τον εντοπισμό σημαντικών συχνοτήτων για περιοδικές χρονοσειρές. Η εκτίμηση αφορά τις συχνότητες που έχουν υψηλή ισχύ παρά το φάσμα ισχύος και βασίζεται στην ανάλυση ιδιοτιμών του πίνακα συσχέτισης. Τέτοιες μέθοδοι είναι η *ταξινόμηση πολλαπλών σημάτων* (multiple signal classification (MUSIC) Method) και η μέθοδος *των ιδιοδιανυσμάτων* (eigenvector method (EV)).

5.9.1 Κλασική Εκτίμηση Φάσματος Ισχύος

Οι μέθοδοι κλασικής εκτίμησης βασίζονται στους δυο ορισμούς του φάσματος ισχύος, το περιοδόγραμμα και τον μετασχηματισμό Fourier της αυτοσυνδιασποράς.

5.9.1.1 Περιοδόγραμμα

Η εκτίμηση με το περιοδόγραμμα προκύπτει από την εξίσωση:

$$P_x(\nu) = \lim_{M \to \infty} E\left[\frac{1}{2M+1} \left| \sum_{n=-M}^{M} X_n e^{-i2\pi\nu n} \right|^2 \right]$$

παραλείποντας τη μέση τιμή και χρησιμοποιώντας μόνο τις διαθέσιμες παρατηρήσεις $x_0, x_1, \ldots, x_{N-1}$ (υποθέτουμε πως οι παρατηρήσεις για αρνητικούς χρόνους ή χρόνους μεγαλύτερους του N είναι μηδέν και επίσης αφαιρούμε από τις παρατηρήσεις το μέσο όρο).

Η εκτίμηση είναι: $P_{per}(\nu) = \frac{1}{N} \left| \sum_{n=0}^{N-1} x_n e^{-i2\pi\nu n} \right|^2 \qquad -1/2 < \nu < 1/2$.

Η αντίστοιχη ισοδύναμη εκτίμηση από τον ορισμό $P_x(\nu)=\sum_{\kappa=-\infty}^{\infty}\gamma_x(k)e^{-i2\pi\nu k}$ είναι:

$$P_{per}(\nu)=\sum_{k=-(N-1)}^{N-1}\gamma_x(k)e^{-i2\pi\nu k} \qquad -1/2<\nu<1/2$$

όπου $\gamma_x(k)=\frac{1}{N}\sum_{n=0}^{N-k-1}x_{n+k}\cdot x_n$ για $0<k<N-1$ και $\gamma_x(-k)=\gamma_x(k)$.

Το περιοδόγραμμα εκφράζει το τετράγωνο του πλάτους του διακριτού μετασχηματισμού Fourier μιας πραγματοποίησης της υπό μελέτης διαδικασίας.

5.9.1.2 Ιδιότητες του Περιοδογράμματος

Για να δούμε αν το περιοδόγραμμα αποτελεί ικανοποιητική εκτίμηση του φάσματος ισχύος της διαδικασίας $P_x(\nu)$, εξετάζουμε τις στατιστικές ιδιότητες του.

Μέση τιμή του $P_{per}(\nu)$:

$$E\left[P_{per}(\nu)\right]=\sum_{k=-(N-1)}^{N-1}E\left[\gamma_x(k)\right]e^{-i2\pi\nu k}=\sum_{k=-(N-1)}^{N-1}\left(1-\frac{|k|}{N}\right)\gamma_x(k)e^{-i2\pi\nu k}$$

$$=\sum_{k=-(N-1)}^{N-1}w_B(k)\gamma_x(k)e^{-i2\pi\nu k}=w_B(\nu)\cdot P_x(\nu)$$

όπου $w_B(k)$ ονομάζεται το **τριγωνικό ή Bartlett παράθυρο υστέρησης** (lag window) και ορίζεται ως: $w_B(k)=\begin{cases}1-\frac{|\kappa|}{N} & |k|<N\\ 0 & \alpha\lambda\lambda o\upsilon\end{cases}$

Το $w_B(\nu)$ είναι το **Bartlett παράθυρο φάσματος** και προκύπτει από τον μετασχηματισμό Fourier του $w_B(k)$

$$w_B(\nu)=F\left[w_B(k)\right]=\frac{1}{N}\left(\frac{\sin(\pi\nu N)}{\sin(\pi\nu)}\right)^2$$

5.9.1.3 Διασπορά και Συνδιασπορά του Περιοδογράμματος

Ο ακριβής υπολογισμός της συνδιασποράς (για δυο συχνότητες ν_1 και ν_2) και της διασποράς του $P_{per}(\nu)$ γενικά δεν είναι δυνατός για μια οποιαδήποτε διαδικασία. Η συνδιασπορά μπορεί όμως να υπολογιστεί προσεγγιστικά από τους τύπους για Gaussian λευκό θόρυβο κα ισχύει ότι:

$$Cov\left[P_{per}(\nu_1), P_{per}(\nu_2)\right] \approx P_x(\nu_1)P_x(\nu_2)\left\{\left(\frac{\sin(\pi(\nu_1+\nu_2)N)}{N\sin(\pi(\nu_1+\nu_2))}\right)^2 + \left(\frac{\sin(\pi(\nu_1-\nu_2)N)}{N\sin(\pi(\nu_1-\nu_2))}\right)^2\right\}$$

Η διασπορά του $P_{per}(\nu)$ προκύπτει από τον παραπάνω τύπο για $\nu_1=\nu_2$

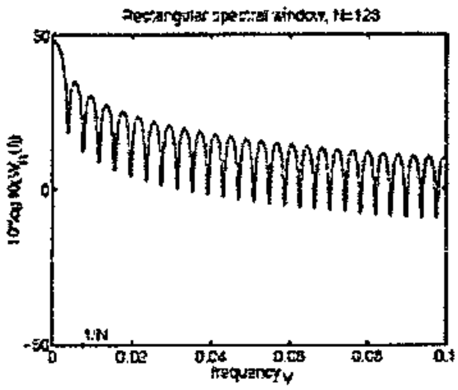
$$Var\left[P_{per}(\nu)\right] \approx \left(P_x(\nu)\right)^2\left(1+\left(\frac{\sin(2\pi\nu N)}{N\sin(2\pi\nu)}\right)^2\right)$$

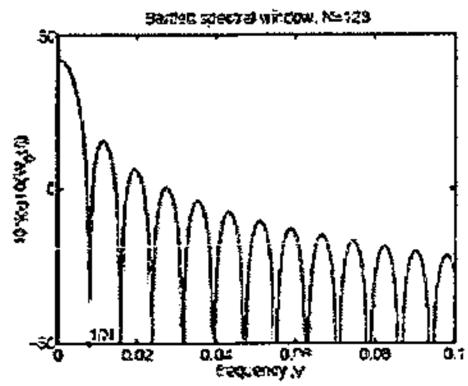
5.9.1.4 Φασματική Διαρροή

Για να εξηγήσουμε την εμφάνιση του παραθύρου Bartlett στην μέση τιμή του $P_{per}(\nu)$ θεωρούμε την χρονοσειρά $\{x_n\}_0^{N-1}$ ως το γινόμενο τη άπειρης χρονοσειράς $\{x_{\infty,n}\}_{-\infty}^{\infty}$ με το **ορθογώνιο παράθυρο δεδομένων** (data window) $w_R(n)$ που έχει μήκος N και κάθε στοιχείο ίσο με την μονάδα.

$$x_n = x_{\infty,n} \cdot w_R(n)$$

Αν πολλαπλασιάσουμε τα στοιχεία x_n μεταξύ τους στην εξίσωση $P_x(\nu) = \lim_{M\to\infty} E\left[\frac{1}{2M+1}\left|\sum_{n=-M}^{M} X_n e^{-i2\pi\nu n}\right|^2\right]$ (με την οποία εκτιμούμε το περιοδόγραμμα), προκύπτει το **τριγωνικό παράθυρο υστέρησης Bartlett**. Στα παρακάτω σχήματα παρουσιάζεται το φασματικά ορθογώνιο παράθυρο και το φασματικό Bartlett παράθυρο.





ΠΗΓΗ: http://users.auth.gr/~dkugiu/Teach/TimeSeries/

Για καλύτερη ερμηνεία ενός διαγράμματος φάσματος ισχύος, δίνεται συνήθως φάσμα ισχύος $P(\nu)$ σε dB, δηλαδή ως $10\log_{10} P(\nu)$. Τα δυο σχήματα δείχνουν ένα κύριο λοβό (main lobe) γύρω από την συχνότητα 0 και μια ακολουθία από πλευρικούς λοβούς (site lobes) με το ύψους τους να μικραίνει καθώς απομακρύνονται από το $\nu = 0$. Οι πλευρικοί λοβοί σχηματίζουν αυτό που ονομάζεται **φασματική διαρροή** (spectral leakage). Συγκρίνοντας τα δυο φασματικά παράθυρα παρατηρούμε ότι στο παράθυρο Bartlett (δεξιά σχήμα), το ύψος των πλευρικών λοβών μικραίνει και τείνει πιο γρήγορα στο μηδέν αλλά το πλάτος τους είναι διπλάσιο από αυτό του ορθογώνιου φασματικού παραθύρου. Ως πλάτος λοβού (lobe width) ορίζουμε τη διαφορά της συχνότητας της κορυφής του λοβού από την συχνότητα που αντιστοιχεί σε μείωση της κορυφής στο μισό, δηλαδή κατά 3dB. Το πλάτος του λοβού είναι της τάξης 1/N.

Η δημιουργία των πλευρικών λοβών και της φασματικής διαρροής δε σχετίζεται με τον αριθμό των συχνοτήτων που χρησιμοποιούνται για τον υπολογισμό του περιοδογράμματος αλλά μόνο με το μήκος της χρονοσειράς. Η ύπαρξη πλευρικών λοβών δηλώνει την μεροληψία της εκτίμησης του $P_x(\nu)$.

5.9.1.5 Ευκρίνεια

Για να διακρίνονται δυο ημιτονοειδής ταλαντώσεις που συνυπάρχουν στο σήμα και έχουν κοντινές συχνότητες θα πρέπει η διαφορά των δυο συχνοτήτων τους να είναι μικρότερη του πλάτους του κύριου λοβού του κάθε ημιτονοειδούς. Συγκεκριμένα για δυο συχνότητες ν_1 και ν_2 η συνθήκη ευκρίνειας είναι:

$$\Delta\nu = \nu_1 - \nu_2 > \frac{1}{N}$$

Γενικά η **ευκρίνεια** (resolution) αναφέρεται στην ικανότητα να μπορεί η εκτιμήτρια του φάσματος ισχύος να ξεχωρίσει τα φασματικά χαρακτηριστικά. Επίσης συνδέεται με την συνδιασπορά (ή συσχέτιση) του $P_{per}(\nu)$ και βελτιώνεται με την αύξηση του N.

5.10 Παραμετρική Εκτίμηση Φάσματος Ισχύος

Η χρονοσειρά προέρχεται από μια γραμμική ARMA(p,q) διαδικασία που μπορεί να έχει μια από τις παρακάτω τρεις μορφές :

$$\chi_n = \phi_1\chi_{n-1} + \dots + \phi_p\chi_{n-p} + z_n + \theta_1 z_{n-1} + \dots + \theta_q z_{n-q}$$

$$\phi_p(\mathrm{B})\chi_n = \theta_q(\mathrm{B}) z_n$$

$$\chi_n = \frac{\theta_q(\mathrm{B})}{\phi_p(\mathrm{B})} z_n = h(\mathrm{B}) z_n$$

όπου z_n είναι λευκός θόρυβος, $\phi_p(\mathrm{B}) = 1 - \phi_1\mathrm{B} - \dots - \phi_p\mathrm{B}^p$ και $\theta_q(\mathrm{B}) = 1 + \theta_1\mathrm{B} + \dots + \theta_q\mathrm{B}^q$ είναι τα χαρακτηριστικά πολυώνυμα του AR και MA μέρους αντίστοιχα. Τέλος $h(\mathrm{B})$ είναι το γραμμικό σύστημα με είσοδο το λευκό θόρυβο και έξοδο τη χρονοσειρά που παρατηρούμε η οποία ονομάζεται **συνάρτηση μεταφοράς** (transfer function) του συστήματος $H(\nu)$ και προκύπτει από τον μετασχηματισμό Fourier του $h(\mathrm{B})$.

Η συνάρτηση $H(\nu)$ δίνεται από τον τύπο:

$$H(\nu) = H\left(e^{-i2\pi\nu}\right) = \frac{1 + \sum_{k=1}^{q} \theta_k e^{-i2\pi\nu k}}{1 - \sum_{k=1}^{p} \phi_k e^{-i2\pi\nu k}}$$

Το φάσμα ισχύος δίνεται από τον τύπο:

$$P_x(\nu)\left|H(\nu)\right|^2 P_z(\nu) = \left|H(\nu)\right|^2 \sigma_z^2 = H\left(e^{-i2\pi\nu}\right) H\left(e^{i2\pi\nu}\right)\sigma_z^2$$

Το $P_x(\nu)$ μιας ARMA διαδικασίας υπολογίζεται αν γνωρίζουμε τις παραμέτρους της διαδικασίας αυτής δηλαδή τα $\phi_1,\ldots,\phi_p,\theta_1,\ldots,\theta_p,\sigma_z^2$. Στην πράξη η παραμετρική εκτίμηση του $P_x(\nu)$ γίνεται από το μοντέλο τύπου ARMA που επιλέγουμε από τους δύο παραπάνω τύπους εκτιμώντας τις παραμέτρους του από την χρονοσειρά. Αν $\hat{\phi}_1,\ldots,\hat{\phi}_p,\hat{\theta}_1,\ldots,\hat{\theta}_p,\sigma_z^2$ είναι οι εκτιμήσεις των παραμέτρων του ARMA μοντέλου η εκτίμηση του $P_x(\nu)$ είναι

$$P_{ARMA}(\nu)=\frac{\hat{\sigma}_z^2\left|1+\sum_{k=1}^{q}\hat{\theta}_k e^{-i2\pi\nu k}\right|^2}{\left|1-\sum_{k=1}^{p}\hat{\phi}_k e^{-i2\pi\nu k}\right|^2}=\frac{\hat{\sigma}_z^2\left(1+\sum_{k=1}^{q}\hat{\theta}_k e^{-i2\pi\nu k}\right)\left(1+\sum_{k=1}^{q}\hat{\theta}_k e^{i2\pi\nu k}\right)}{\left(1-\sum_{k=1}^{p}\hat{\phi}_k e^{-i2\pi\nu k}\right)\left(1-\sum_{k=1}^{p}\hat{\phi}_k e^{i2\pi\nu k}\right)}$$

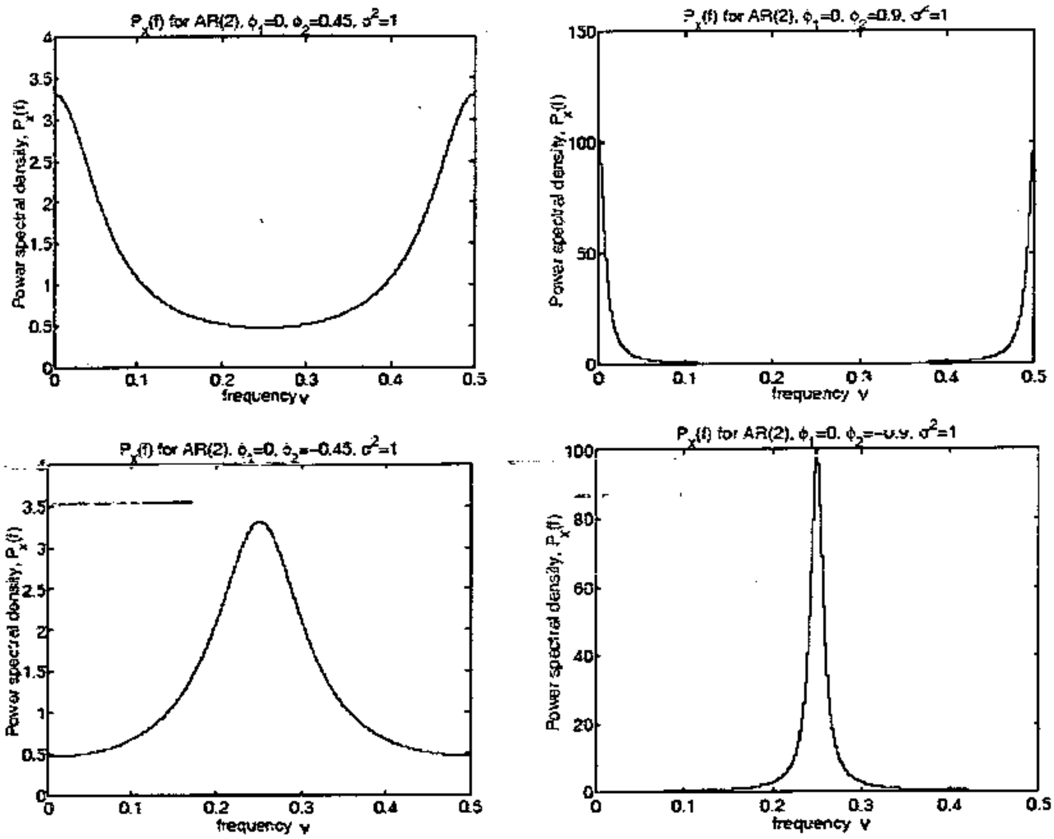
Αυτός ο τύπος μας εκφράζει τις εκτιμήτριες $P_{AR}(\nu)$ και $P_{MA}(\nu)$ του μοντέλου AR και του μοντέλου MA αντίστοιχα.

5.11 Φάσμα ισχύος AR(2)

Για το φάσμα ισχύος $P_x(\nu)$ μιας AR(2) διαδικασίας από την προηγούμενη εξίσωση έχουμε:

$$P_x(\nu)=\frac{\sigma_z^2}{\left(1-\phi_1 e^{i2\pi\nu}-\phi_1 e^{i4\pi\nu}\right)\left(1-\phi_1 e^{-i2\pi\nu}-\phi_1 e^{-i4\pi\nu}\right)}$$
$$=\frac{\sigma_z^2}{1+\phi_1^2+2\phi_2+\phi_2^2+2\left(\phi_1\phi_2-\phi_1\right)\cos\left(2\pi\nu\right)-4\phi_2\cos^2\left(2\pi\nu\right)}$$

Στα παρακάτω σχήματα δίνεται το $P_x(\nu)$ για τέσσερις διαδικασίες AR(2) με $\phi_1=0$, $\sigma_z^2=1$ και $\phi_2=0.45, 0.90, -0.45, -0.90$ αντίστοιχα. Για τις θετικές τιμές του ϕ_2 οι ρίζες του χαρακτηριστικού πολυωνύμου $\phi_2(\mathrm{B})=1-\phi_1\mathrm{B}-\phi_2\mathrm{B}^2$ είναι πραγματικοί αριθμοί ενώ για τις αρνητικές τιμές του ϕ_2 είναι μιγαδικοί αριθμοί. Όταν οι ρίζες του $\phi_2(\mathrm{B})$ είναι πραγματικές το φάσμα ισχύος $P_x(\nu)$ έχει κορυφή μόνο στη συχνότητα $\nu=0$ ή στη συχνότητα $\nu=1/2$ ή και στις δύο ανάλογα με τις τιμές των ϕ_1 και ϕ_2. Όταν οι ρίζες είναι συζυγείς μιγαδικές το $P_x(\nu)$ έχει κορυφή σε κάποια άλλη συχνότητα ν_0 που μπορεί να βρεθεί από τη συνάρτηση του φάσματος ισχύος στην παραπάνω εξίσωση. Στο αριθμητικό παράδειγμα, όταν η τιμή του ϕ_2 μεγαλώνει κατά απόλυτη τιμή η αυτοσυσχέτιση γίνεται πιο ισχυρή με αποτέλεσμα το φάσμα να έχει υψηλότερες κορυφές. Για $\phi_2 = -0.90$ η διαδικασία AR(2) γίνεται σχεδόν περιοδική.



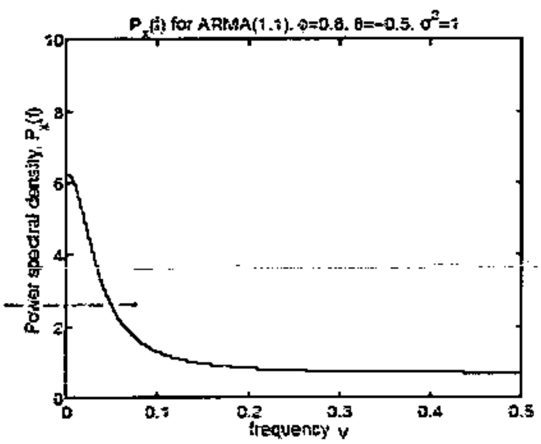
ΠΗΓΗ: http://users.auth.gr/~dkugiu/Teach/TimeSeries/

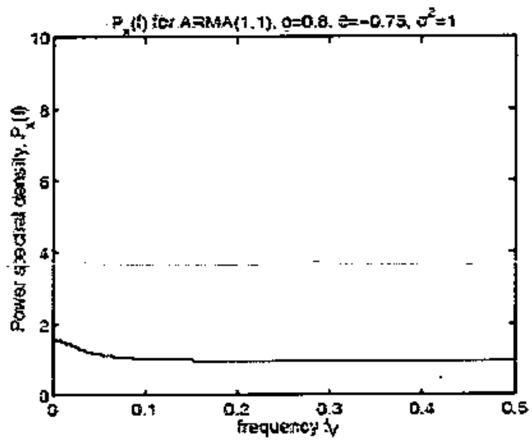
5.12 Φάσμα Ισχύος ARMA(1,1)

Για μια διαδικασία ARMA(1,1) έχουμε:

$$P_x(\nu) = \frac{\sigma_z^2\left(1+\theta e^{-i2\pi\nu}\right)\left(1+\theta e^{i2\pi\nu}\right)}{\left(1-\phi e^{i2\pi\nu}\right)\left(1-\phi e^{-i2\pi\nu}\right)} = \frac{\sigma_z^2\left(1+2\theta\cos(2\pi\nu)+\theta^2\right)}{\left(1-2\phi\cos(2\pi\nu)+\phi^2\right)}.$$

Η μορφή του $P_x(\nu)$ μιας διαδικασίας ARMA(1,1) μοιάζει με αυτό των AR(1) και MA(1). Όταν όμως οι συντελεστές $-\phi$ και θ τείνουν προς την ίδια τιμή το $P_x(\nu)$ τείνει να γίνει επίπεδο χωρίς κορυφή. Αυτό συμβαίνει γιατί το MA(1) μέρος ορίζει «κοιλάδα» στο φάσμα ισχύος στην ίδια συχνότητα (εδώ είναι $\nu = 0$) που το AR(1) μέρος ορίζει κορυφή. Αποτέλεσμα είναι το ένα μέρος (AR ή MA) να μηδενίζει το άλλο όταν οι τιμές $-\phi$ και θ είναι ίδιες, αλλιώς σχηματίζεται κορυφή ή «κοιλάδα» ανάλογα με το ποια από τις δύο παραμέτρους είναι μεγαλύτερη.





ΠΗΓΗ: http://users.auth.gr/~dkugiu/Teach/TimeSeries/

Παρατηρήσαμε ορισμένες μορφές του φάσματος ισχύος που μπορούν να δώσουν τα μοντέλα ARMA για μικρές τάξεις. Για μεγάλες τάξεις, το αντίστοιχο φάσμα ισχύος έχει πιο πολύπλοκη μορφή.

Γενικά από τα παραπάνω μπορούμε να συμπεράνουμε ότι η εκτίμηση του φάσματος ισχύος με μοντέλα τύπου AR προτιμάται όταν το φάσμα ισχύος παρουσιάζει κορυφές γι΄ αυτό και χρησιμοποιούνται συχνά στις εφαρμογές. Επίσης ότι τα μοντέλα τύπου MA χρησιμοποιούνται λιγότερο. Είναι πιο κατάλληλα όταν το φάσμα ισχύος που θέλουμε να εκτιμήσουμε παρουσιάζει «κοιλάδες». Η εκτίμηση των παραμέτρων του μοντέλου AR γίνεται με κάποιες από τις γνωστές μεθόδους Yule-Walker, αυτοδιασποράς (covariance) και τροποποιημένης αυτοδιασποράς (modified covariance). Τέλος οι παραμετρικές μέθοδοι δίνουν πιο ομαλά και ακριβή φάσματα ισχύος από τις κλασικές μεθόδους αλλά υποθέτουν κάποιο συγκεκριμένο μοντέλο για τη χρονοσειρά που μπορεί να μην είναι κατάλληλο.

5.13 <u>Συμπέρασμα</u>

Σε αυτό το κεφάλαιο αναπτύχθηκε ένα πεδίο μεθόδων για την ανάλυση των χρονοσειρών στο πεδίο των συχνοτήτων. Οι μέθοδοι αυτοί περιγράφουν τις μεταβολές των χρονοσειρών σε όρους της κυκλικής συνιστώσας των ημίτονων και των συνημίτονων διαφορετικών συχνοτήτων.

Αν και η φασματική ανάλυση δίνεται σε όρους των συχνοτήτων ή των περιόδων της κυκλικής συνιστώσας, υπάρχει μια ακριβής αλλά περίπλοκη σχέση μεταξύ της απεικόνισης των συχνοτήτων και της αυτοσυσχέτισης των χρονοσειρών. Η ίδια πληροφορία παρουσιάζεται σε διαφορετικές μορφές από το περιοδόγραμμα.

Η φασματική ανάλυση θεωρείται από τους πιο προχωρημένους κλάδους της στατιστικής με εφαρμογές κυρίως στη φυσική, στους τεχνικούς κλάδους αλλά και στην οικονομική επιστήμη.

ΚΕΦΑΛΑΙΟ 6: ΕΦΑΡΜΟΓΕΣ

6.1 Εισαγωγή

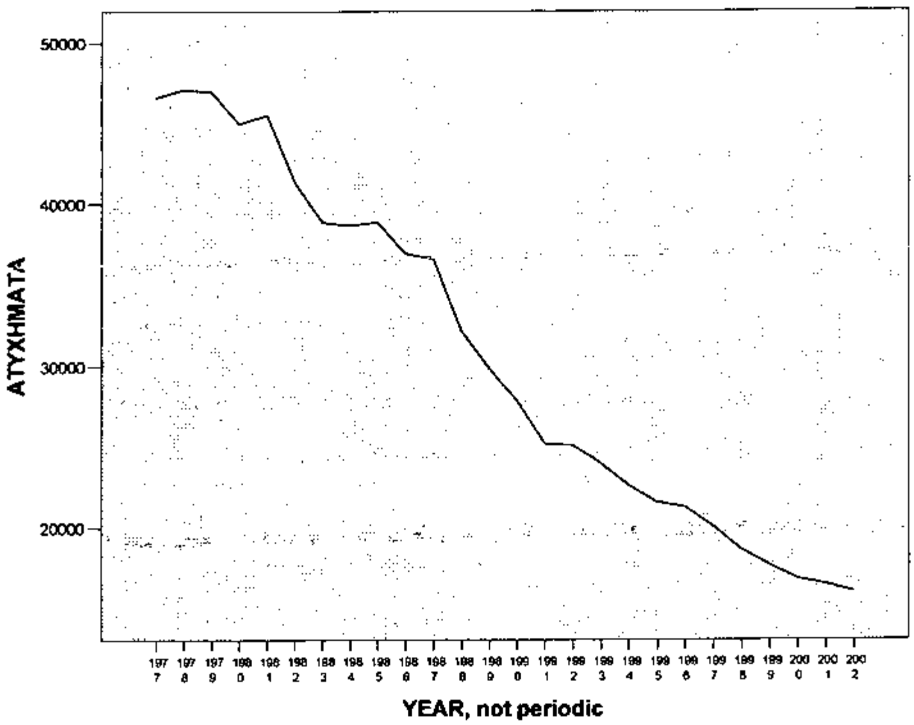
Στα προηγούμενα κεφάλαια παρουσιάστηκαν διάφορα μοντέλα ανάλυσης χρονοσειρών. Στο παρόν κεφάλαιο θα γίνει εφαρμογή ορισμένων από αυτά τα μοντέλα. Οι εφαρμογές βασίστηκαν σε πραγματικά δεδομένα. Η διαδικασία μοντελοποίησης έγινε με τη χρήση του στατιστικού προγράμματος SPSS 13.0. Χρησιμοποιήθηκαν μοντέλα παλινδρόμησης πολυωνυμικής μορφής όπως και μοντέλα με την μεθοδολογία Box-Jenkins σε μορφή ARIMA. Επίσης, μοντέλα ARIMA με την παρεμβολή ανεξάρτητης μεταβλητής. Τέλος έγινε εφαρμογή μοντέλου με τη χρήση του περιοδογράμματος και της φασματικής πυκνότητας.

6.2 Εφαρμογή - Ατυχήματα

Τα δεδομένα αφορούν τα ετήσια ατυχήματα των ασφαλισμένων στην Ελλάδα από το 1977 έως το 2002.

Από τη γραφική παράσταση (sequence), παρατηρούμε ότι η εξίσωση έχει καθοδική γραμμική μορφή. Επομένως αναφερόμαστε στην απλή γραμμική παλινδρόμηση με γενική μορφή της εξίσωσης:

$$X_t = \alpha + \beta t$$



Model Summary

R	R Square	Adjusted R Square	Std. Error of the Estimate
,988	,976	,975	1718,524

ANOVA

	Sum of Squares	df	Mean Square	F	Sig.
Regression	2916758586,675	1	2916758586,675	987,619	,000
Residual	70879779,825	24	2953324,159		
Total	2987638366,500	25			

Coefficients

	Unstandardized Coefficients		Standardized Coefficients	t	Sig.
	B	Std. Error	Beta		
Case Sequence	-1412,220	44,937	-,988	-31,426	,000
(Constant)	49735,468	693,988		71,666	,000

Linear

1ος Πίνακας "Model Summary"

- Το R-Square είναι ο δείκτης προσδιορισμού με τιμή 0,976. Δηλαδή κατά 97,6% η μεταβλητότητα της μεταβλητής " Ατυχήματα" ερμηνεύεται από τον χρόνο.
- Το Adjusted Rsquare είναι η διορθωμένη τιμή του δείκτη προσδιορισμού με τιμή 0,975.
- Το Std. Error είναι το τυπικό σφάλμα της εκτίμησης και είναι 1718,524.

2ος Πίνακας "ANOVA"

Εδώ βλέπουμε τα αποτελέσματα από την ανάλυση της διακύμανσης, Analysis of Variance (ANOVA).

- Sig.: Είναι η κρίσιμη τιμή με βάση την οποία απορρίπτουμε ή αποδεχόμαστε την μηδενική υπόθεση. Ο έλεγχος είναι:

Ho = όλοι οι συντελεστές του μοντέλου είναι ίσοι με μηδέν

vs

H_1 = υπάρχει ένας έστω συντελεστής διάφορος του μηδενός.

Η τιμή εδώ του Sig F είναι 0 δηλαδή είναι μικρότερη από την κρίσιμη τιμή α= 0,05, το οποίο είναι το επίπεδο σημαντικότητας στο οποίο εξετάζουμε τα δεδομένα μας. Επομένως απορρίπτουμε την μηδενική υπόθεση H_0 και αποδεχόμαστε την εναλλακτική H_1. Δηλαδή, οι συντελεστές του δείγματος είναι στατιστικά σημαντικοί και υπάρχει γραμμικά σχέση μεταξύ τους.

3ος Πίνακας "Coefficients"

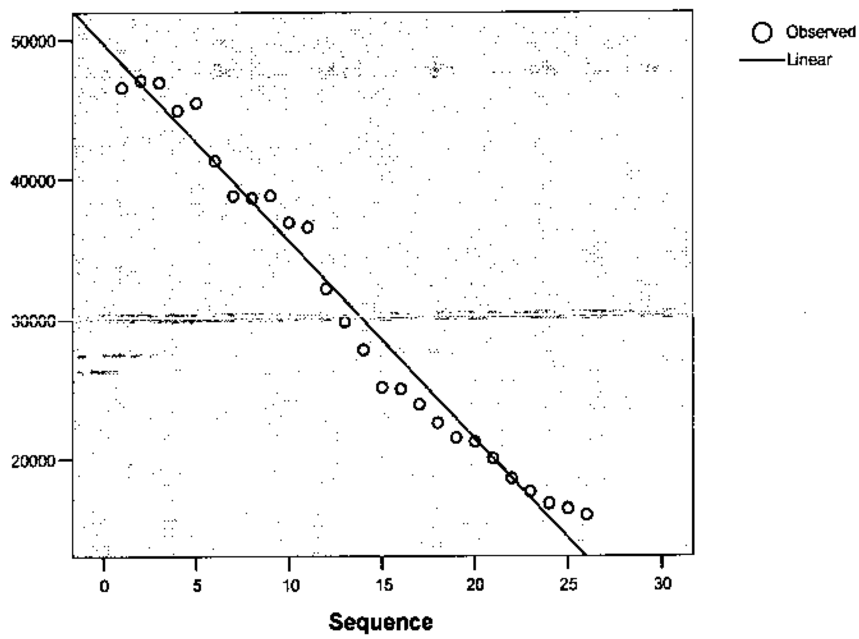
$\beta_1 = -1412,220$
$\alpha = 49735,468$
Άρα η εξίσωση είναι:

$$X_t = 49735,468 - 1412,220t$$

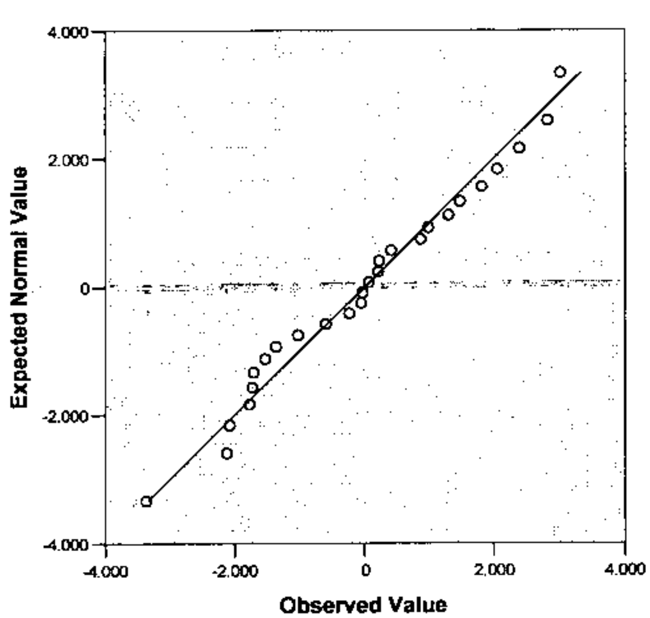
Το p-value του β_1 συντελεστή είναι 0. Με επίπεδο σημαντικότητας 0,05 είναι μικρότερο. Άρα είναι στατιστικά σημαντικός συντελεστής.

Το p-value του α συντελεστή είναι 0. Με επίπεδο σημαντικότητας 0,05 είναι μικρότερο. Άρα είναι στατιστικά σημαντικός συντελεστής.

ATYXHMATA



Normal Q-Q Plot of Error for ATYXHMATA from CURVEFIT, MOD_2 LINEAR



Από το πρώτο διάγραμμα είναι προφανές ότι η ευθεία παλινδρόμησης

$$X_t = 49735,468 - 1412,220t$$

έχει αρνητική κλίση από τα αριστερά προς τα δεξιά κάτω.

Οι κουκίδες δείχνουν τα δεδομένα, από τα οποία τα περισσότερα εφάπτονται στην ευθεία γραμμή.

Στο δεύτερο διάγραμμα Q-Q βλέπουμε ότι τα υπόλοιπα είναι κανονικά κατανεμημένα γιατί τα περισσότερα από αυτά πέφτουν επάνω στην διαγώνιο.

ΠΡΟΒΛΕΨΗ

Τα παρακάτω αποτελέσματα είναι τα αποτελέσματα της πρόβλεψης για τα επόμενα 3 χρόνια 2003, 2004, 2005.

PERIOD	FIT	LCL	UCL
2003	11605,53	7780,38533	15430,6793
2004	10193,31	6336,81132	14049,8136
2005	8781,093	4891,27816	12670,9071

Το έτος 2003 η τιμή της πρόβλεψης είναι 11605,53. Το κατώτερο όριο που ανήκει στο 95% διάστημα εμπιστοσύνης είναι 7780,38533 και το ανώτερο15430,6793.

Το έτος 2004 η τιμή της πρόβλεψης είναι 10193,31. Το κατώτερο όριο που ανήκει στο 95% διάστημα εμπιστοσύνης είναι 6336,81132 και το ανώτερο14049,8136.

Το έτος 2005 η τιμή της πρόβλεψης είναι 8781. Το κατώτερο όριο που ανήκει στο 95% διάστημα εμπιστοσύνης είναι 4891,27816 και το ανώτερο12670,9071.

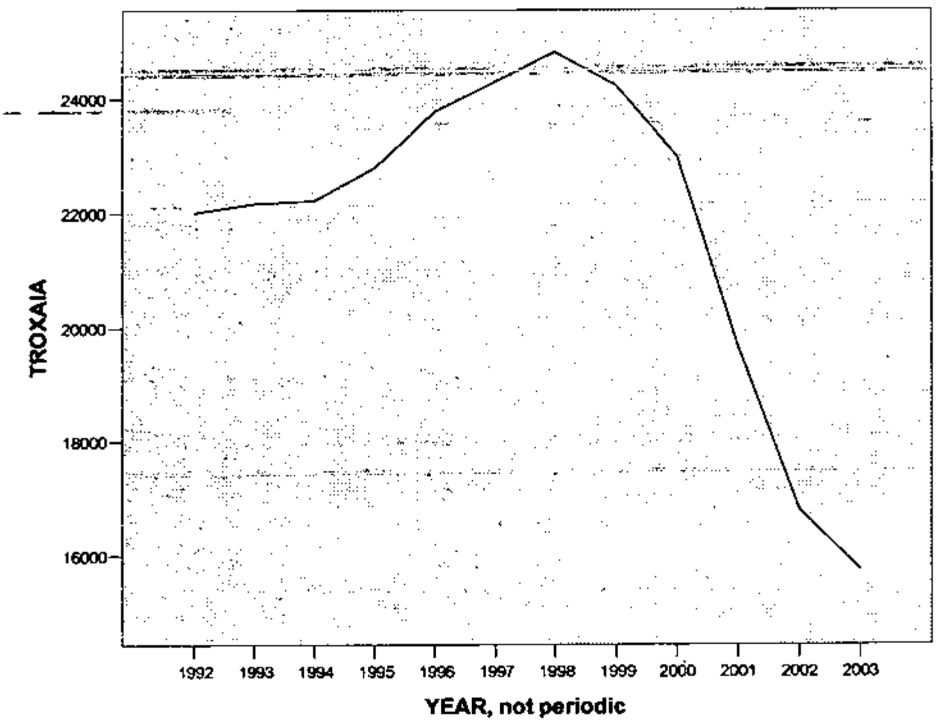
Παρατηρούμε ότι τα ατυχήματα των ασφαλισμένων θα μειωθούν σε μεγάλο βαθμό σε σχέση με τα προηγούμενα χρόνια.

6.3 Εφαρμογή - Τροχαία

Τα δεδομένα αφορούν τα ετήσια τροχαία ατυχήματα στην Ελλάδα από το 1992 έως το 2003.

Από τη γραφική παράσταση (sequence), παρατηρούμε ότι η καμπύλη παλινδρόμησης έχει πολυωνυμική μορφή και η γενική μορφή της εξίσωσης είναι:

$$X_t = \alpha + \beta_1 t + \beta_2 t^2 + \beta_3 t^3$$



Model Summary

R	R Square	Adjusted R Square	Std. Error of the Estimate
,967	,936	,911	869,665

ANOVA

	Sum of Squares	df	Mean Square	F	Sig.
Regression	87924230,930	3	29308076,977	38,751	,000
Residual	6050543,320	8	756317,915		
Total	93974774,250	11			

Coefficients

	Unstandardized Coefficients		Standardized Coefficients		
	B	Std. Error	Beta	t	Sig.
Case Sequence	19,353	909,842	,024	,021	,984
Case Sequence ** 2	190,333	159,359	3,135	1,194	,267
Case Sequence ** 3	-19,967	8,081	-3,936	-2,471	,039
(Constant)	21482,889	1422,845		15,099	,000

CUBIC

Πίνακας "Coefficients"

$\beta_1 = 19,353$

$\beta_2 = 190,333$

$\beta_3 = -19,967$

$\alpha = 21482,889$

Άρα η εξίσωση είναι:

$$X_t = 19,353t + 190,333t^2 - 19,967t^3 + 21482,889$$

Το p-value του β_1 συντελεστή είναι 0,984. Με επίπεδο σημαντικότητας 0,05 είναι μεγαλύτερο. Άρα δεν είναι στατιστικά σημαντικός συντελεστής.

Το p-value του β_2 συντελεστή είναι 0,267. Με επίπεδο σημαντικότητας 0,05 είναι μεγαλύτερο. Άρα δεν είναι στατιστικά σημαντικός συντελεστής.

Το p-value του β_3 συντελεστή είναι 0,039. Με επίπεδο σημαντικότητας 0,05 είναι μικρότερο. Άρα είναι στατιστικά σημαντικός συντελεστής.

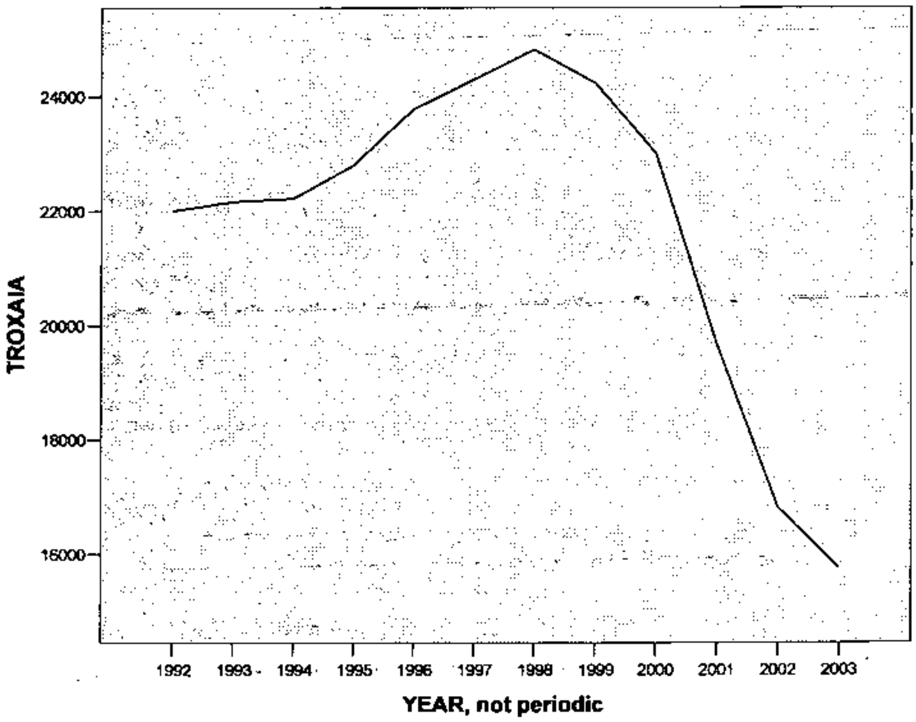
Το p-value του α συντελεστή είναι 0. Με επίπεδο σημαντικότητας 0,05 είναι μικρότερο. Άρα είναι στατιστικά σημαντικός συντελεστής.

Επειδή δεν είναι στατιστικά σημαντικοί όλοι οι συντελεστές πέφτουμε σε μοντέλο χαμηλότερης πολυωνυμικής μορφής.

Από τη γραφική παράσταση (sequence), παρατηρούμε ότι η καμπύλη παλινδρόμησης έχει πολυωνυμική μορφή και η γενική μορφή της εξίσωσης είναι:

$$X_t = \alpha + \beta_1 t + \beta_2 t^2$$

Άρα προσαρμόζουμε ένα μη-γραμμικό μοντέλο όπως και πριν, αλλά μιας δύναμης χαμηλότερο.



Model Summary

R	R Square	Adjusted R Square	Std. Error of the Estimate
,942	,886	,861	1088,747

ANOVA

	Sum of Squares	df	Mean Square	F	Sig.
Regression	83306438,100	2	41653219,050	35,139	,000
Residual	10668336,150	9	1185370,683		
Total	93974774,250	11			

Coefficients

	Unstandardized Coefficients		Standardized Coefficients		
	B	Std. Error	Beta	t	Sig.
Case Sequence	2125,842	397,976	2,622	5,342	,000
Case Sequence ** 2	-199,018	29,802	-3,278	-6,678	,000
(Constant)	18757,432	1125,252		16,670	,000

Quadratic

1ος Πίνακας "Model Summary"

- Το R-Square είναι ο δείκτης προσδιορισμού με τιμή 0,886. Δηλαδή κατά 88,6% η μεταβλητότητα της μεταβλητής "Τροχαία" ερμηνεύεται από τον χρόνο.
- Το Adjusted Rsquare είναι η διορθωμένη τιμή του δείκτη προσδιορισμού με τιμή 0,861.
- Το Std. Error είναι το τυπικό σφάλμα της εκτίμησης και είναι 1088,747.

2ος Πίνακας "ANOVA"

Εδώ βλέπουμε τα αποτελέσματα από την ανάλυση της διακύμανσης, Analysis of Variance (ANOVA).

- Sig.: Είναι η κρίσιμη τιμή με βάση την οποία απορρίπτουμε ή αποδεχόμαστε την μηδενική υπόθεση. Ο έλεγχος είναι:

Ho = όλοι οι συντελεστές του μοντέλου είναι ίσοι με μηδέν

vs

H_1 = υπάρχει ένας έστω συντελεστής διάφορος του μηδενός.

Η τιμή εδώ του Sig F είναι 0 δηλαδή είναι μικρότερη από την κρίσιμη τιμή α= 0,05, το οποίο είναι το επίπεδο σημαντικότητας στο οποίο εξετάζουμε τα δεδομένα μας. Επομένως απορρίπτουμε την μηδενική υπόθεση H_0 και αποδεχόμαστε την εναλλακτική H_1. Δηλαδή οι συντελεστές του δείγματος είναι στατιστικά σημαντικοί και υπάρχει μη-γραμμική σχέση μεταξύ τους.

3ος Πίνακας "Coefficients"

$\beta_1 = 2125,842$
$\beta_2 = -199,018$
$\alpha = 18757,432$

Άρα η εξίσωση είναι:

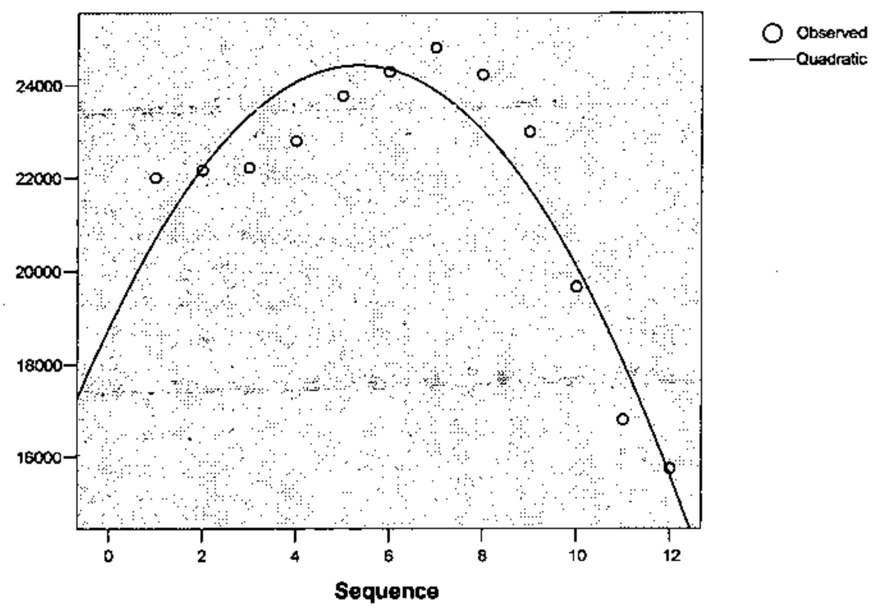
$$X_t = 2125,842t - 199,018t^2 + 18757,432$$

Το p-value του β_1 συντελεστή είναι 0. Με επίπεδο σημαντικότητας 0,05 είναι μικρότερο. Άρα είναι στατιστικά σημαντικός συντελεστής.

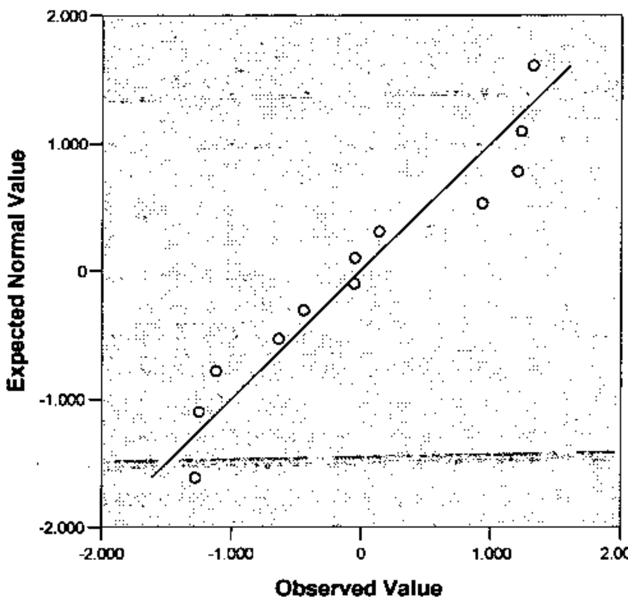
Το p-value του β_2 συντελεστή είναι 0. Με επίπεδο σημαντικότητας 0,05 είναι μικρότερο. Άρα είναι στατιστικά σημαντικός συντελεστής.

Το p-value του α συντελεστή είναι 0. Με επίπεδο σημαντικότητας 0,05 είναι μικρότερο. Άρα είναι στατιστικά σημαντικός συντελεστής.

TROXAIA



Normal Q-Q Plot of Error for TROXAIA from CURVEFIT, MOD_2 QUADRATIC



Από το πρώτο διάγραμμα είναι προφανές ότι η καμπύλη παλινδρόμησης

$$X_t = 2125,842t - 199,018t^2 + 18757,432$$

έχει θετική κλίση προς τα δεξιά και μετά μειώνεται.

Οι κουκίδες δείχνουν τα δεδομένα, από τα οποία τα περισσότερα εφάπτονται στην καμπύλη.

Στο δεύτερο διάγραμμα Q-Q βλέπουμε ότι τα υπόλοιπα. είναι κανονικά κατανεμημένα γιατί τα περισσότερα από αυτά πέφτουν επάνω στην διαγώνιο.

ΠΡΟΒΛΕΨΗ

Τα παρακάτω αποτελέσματα είναι τα αποτελέσματα της πρόβλεψης για τα επόμενα 2 χρόνια 2004, 2005.

PERIOD	FIT	LCL	UCL
2004	12759,3	9217,331	16301,26
2005	9511,645	5281,947	13741,34

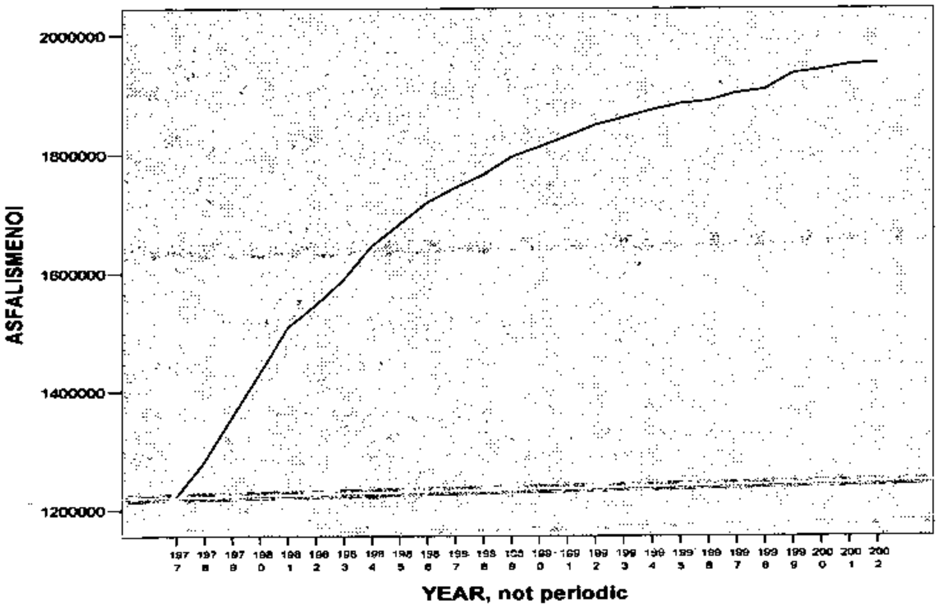
Το έτος 2004 η τιμή της πρόβλεψης είναι 12759,3. Το κατώτερο όριο που ανήκει στο 95% διάστημα εμπιστοσύνης είναι 9217,331 και το ανώτερο16301,26.

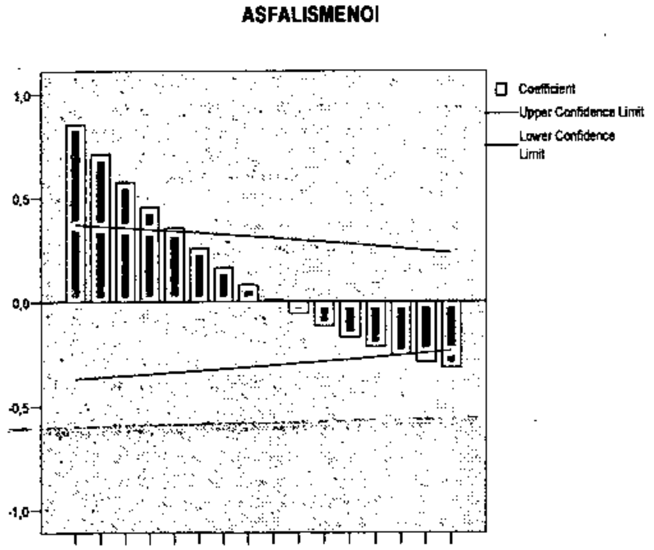
Το έτος 2005 η τιμή της πρόβλεψης είναι 9511,645. Το κατώτερο όριο που ανήκει στο 95% διάστημα εμπιστοσύνης είναι 5281,947 και το ανώτερο13741,34.

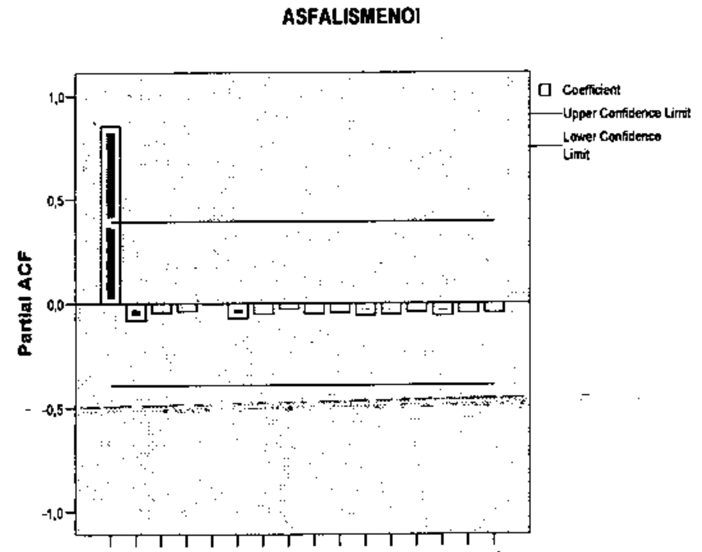
Το συμπέρασμα είναι ότι τα τροχαία στην Ελλάδα θα μειωθούν σε μεγάλο βαθμό σε σχέση με τα προηγούμενα χρόνια. Αυτό είναι φανερό αν συγκρίνει κανείς τις τιμές της πρόβλεψης με τις πραγματικές τιμές που φαίνονται στο παράρτημα, στο τέλος της εργασίας.

6.4 Εφαρμογή - Ασφαλισμένοι

Τα δεδομένα που χρησιμοποιούνται στο μοντέλο αναφέρονται στους ασφαλισμένους του ΙΚΑ στην Ελλάδα και οι τιμές είναι ετήσιες από το διάστημα 1977 έως 2002.





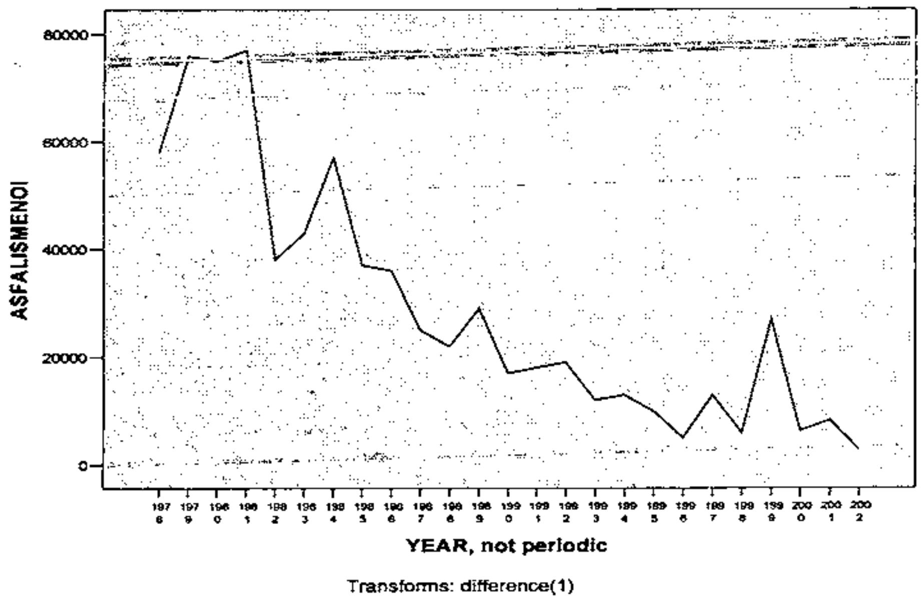


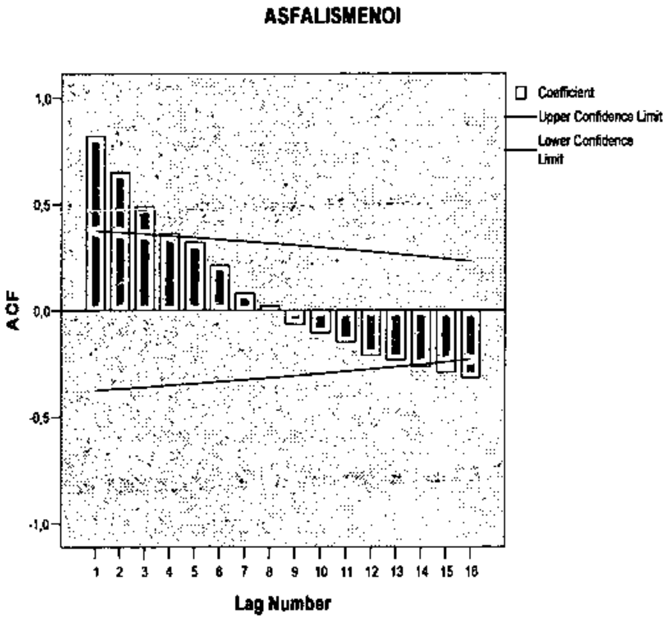
ΕΡΜΗΝΕΙΑ

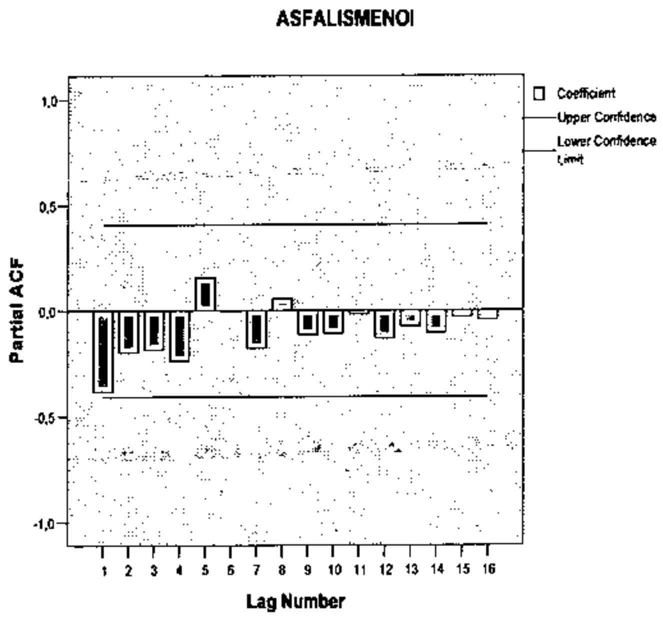
Παρατηρώντας κανείς την πρώτη γραφική παράσταση sequence βλέπει ότι η χρονοσειρά δεν είναι στάσιμη.

Στο διάγραμμα των αυτοσυσχετίσεων (ACF) και των μερικών αυτοσυσχετίσεων (PACF) αντίστοιχα βλέπουμε ότι οι παρατηρήσεις φθίνουν αργά επομένως η χρονοσειρά δεν είναι στάσιμη.

Για τον λόγο αυτό, επειδή δηλαδή η χρονοσειρά δεν είναι στάσιμη, θα την διαφορίσουμε. Τα αποτελέσματα φαίνονται στους παρακάτω πίνακες.

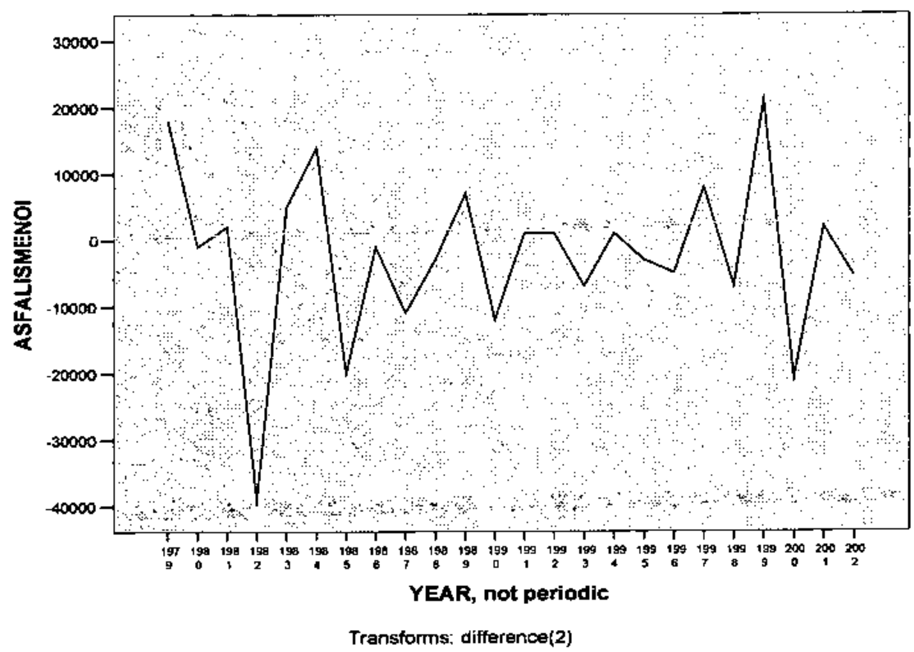


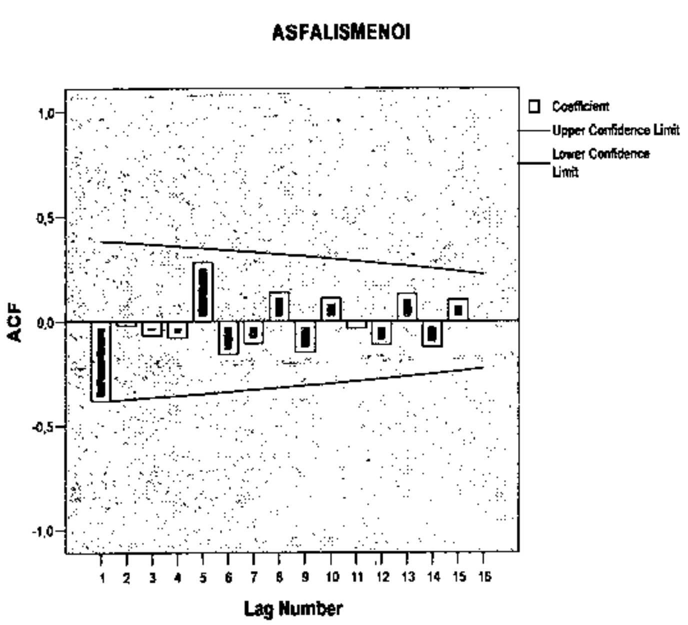


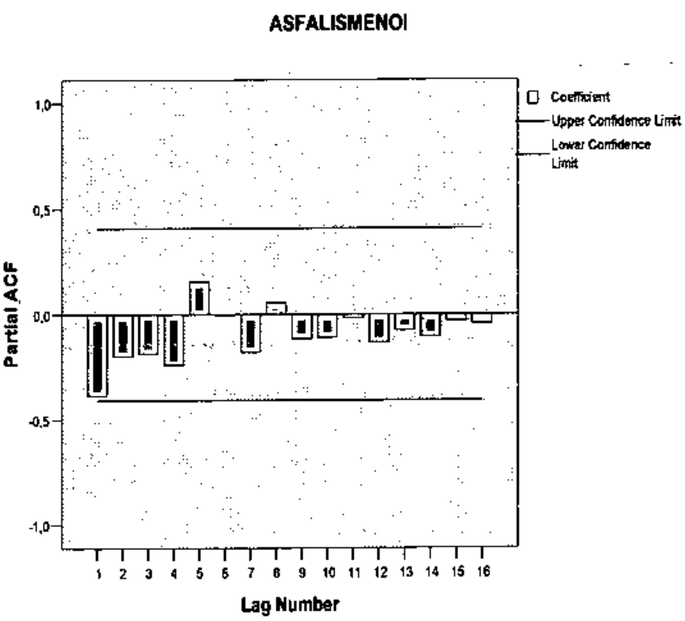


Από το παραπάνω διάγραμμα sequence παρατηρούμε ότι παρόλο που η χρονοσειρά έχει διαφοριστεί μια φορά δεν επιτεύχθηκε η στασιμότητα.

Το ίδιο παρατηρούμε και στο ACF διάγραμμα. Η χρονοσειρά φθίνει αργά άρα δεν είναι στάσιμη. Εξαιτίας λοιπόν της μη-στασιμότητας θα διαφορίσουμε την χρονοσειρά μας για δεύτερη φορά.







Μετά την δεύτερη διαφόριση από το διάγραμμα sequence είναι φανερό ότι η χρονοσειρά "μαζεύτηκε" αρκετά και προς τα επάνω σε σχέση με το προηγούμενο διάγραμμα. Τα όρια του διαγράμματος από τις 80000 κυμαίνονται τώρα στις 30000 και οι κορυφές του παρέμειναν σχεδόν στο ίδιο ύψος. Άρα το διάγραμμα μας δείχνει ότι η διαδικασία αυτή είναι στάσιμη γιατί έχουμε ένα διασκορπισμένο ορθογώνιο γύρω από το μηδενικό οριζόντιο επίπεδο χωρίς πιθανές τάσεις.

Εν συνεχεία το διάγραμμα των αυτοσυχετίσεων (ACF) φθίνει γρήγορα πράγμα που δείχνει την στασιμότητα της χρονοσειράς. Επίσης επιλέγουμε q=1 γιατί η πρώτη υστέρηση βρίσκεται ακριβώς στα όρια.

Από το PACF διάγραμμα επιλέγουμε p=1 γιατί η πρώτη υστέρηση βρίσκεται στα όρια.

Επομένως έχουμε ένα ARIMA (1,2,1) μοντέλο με τα εξής παρακάτω αποτελέσματα.

Residual Diagnostics

Number of Residuals	24
Number of Parameters	2
Residual df	21
Adjusted Residual Sum of Squares	2831608283,156
Residual Sum of Squares	2898914980,212
Residual Variance	132225822,920
Model Std. Error	11498,949
Log-Likelihood	-257,162
Akaike's Information Criterion (AIC)	520,325
Schwarz's Bayesian Criterion (BIC)	523,859

Parameter Estimates

		Estimates	Std Error	t	Approx Sig
Non-Seasonal Lags	AR1	,022	,361	,062	,951
	MA1	,626	,298	2,102	,048
Constant		-2702,126	966,960	-2,794	,011

Melard's algorithm was used for estimation.

Από τον πίνακα "Parameter Estimates" έχουμε τα εξής αποτελέσματα:

- $\hat{\phi} = 0,022$ δηλαδή ο συντελεστής του AR(1), με p-value=0,951. Συγκρίνοντας το τελευταίο με το 0,05 επίπεδο σημαντικότητας 0,951>0,05 βλέπουμε ότι είναι μεγαλύτερο. Άρα το $\hat{\phi}$ δεν είναι στατιστικά σημαντικός συντελεστής.
- $\hat{\theta} = 0,626$ δηλαδή ο συντελεστής του MA(1) με p-value=0,048. Συγκρίνοντας το τελευταίο με το 0,05 επίπεδο σημαντικότητας 0,048<0,05 είναι μικρότερο. Άρα το $\hat{\theta}$ είναι στατιστικά σημαντικός συντελεστής.

- $\hat{a} = -2702,126$ δηλαδή η σταθερά, με p-value=0,011. Εάν το συγκρίνουμε με το 0,05 επίπεδο σημαντικότητας είναι μικρότερο. Άρα το $\hat{a}$ είναι στατιστικά σημαντικός συντελεστής.

Άρα θα επαναλάβουμε την ίδια διαδικασία αφαιρώντας τον συντελεστή AR(1) από το μοντέλο επειδή δεν είναι στατιστικά σημαντικός.

Τα αποτελέσματα μετά την αφαίρεση του συντελεστή $\hat{\phi}$ τα βλέπουμε στους δύο πίνακες που ακολουθούν.

Residual Diagnostics

Number of Residuals	24
Number of Parameters	1
Residual df	22
Adjusted Residual Sum of Squares	2832101282,714
Residual Sum of Squares	2928179838,062
Residual Variance	126231052,716
Model Std. Error	11235,259
Log-Likelihood	-257,116
Akaike's Information Criterion (AIC)	518,232
Schwarz's Bayesian Criterion (BIC)	520,588

Parameter Estimates

		Estimates	Std Error	t	Approx Sig
Non-Seasonal Lags	MA1	,613	,181	3,379	,003
Constant		-2703,463	953,082	-2,837	,010

Melard's algorithm was used for estimation.

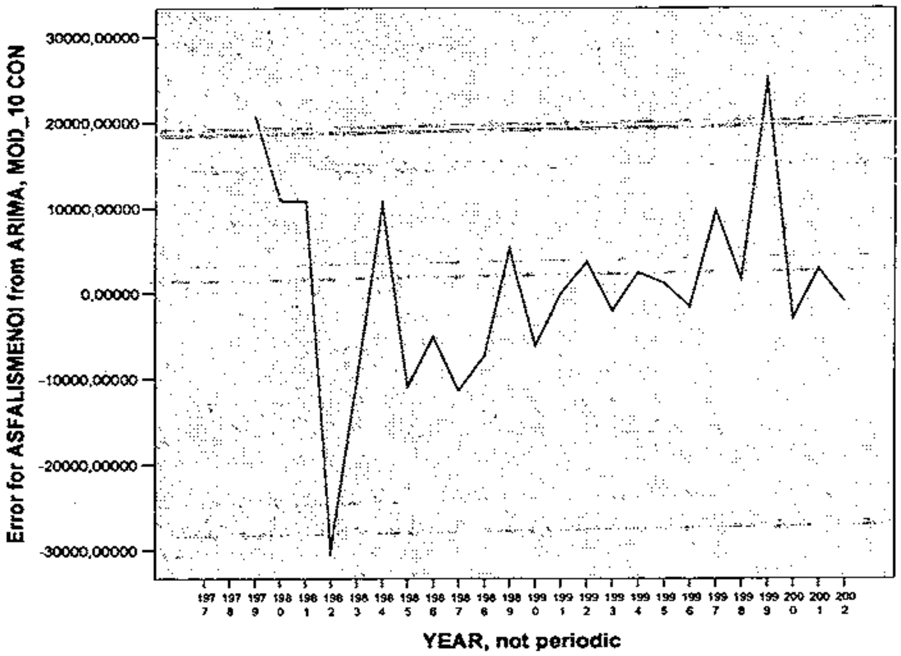
Από τον πίνακα "Parameter Estimates" έχουμε τα εξής αποτελέσματα:

- $\hat{\theta} = 0,613$ δηλαδή ο συντελεστής του MA(1) με p-value=0,003. Συγκρίνοντας το τελευταίο με το 0,05 επίπεδο σημαντικότητας 0,003<0,05 είναι μικρότερο. Άρα το $\hat{\theta}$ είναι στατιστικά σημαντικός συντελεστής.
- $\hat{a} = -2703,463$ δηλαδή η σταθερά, με p-value=0,010. Εάν το συγκρίνουμε με το 0,05 επίπεδο σημαντικότητας είναι μικρότερο. Άρα το $\hat{a}$ είναι στατιστικά σημαντικός συντελεστής.

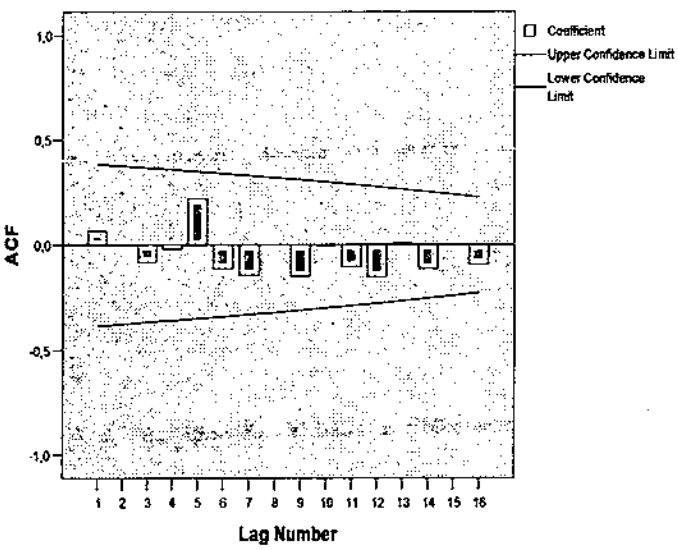
Το μοντέλο μας είναι ARIMA(0,2,1) με εξίσωση:

$$(1-B)^2 X_t = (1-0{,}613B)e_t - 2703{,}463$$

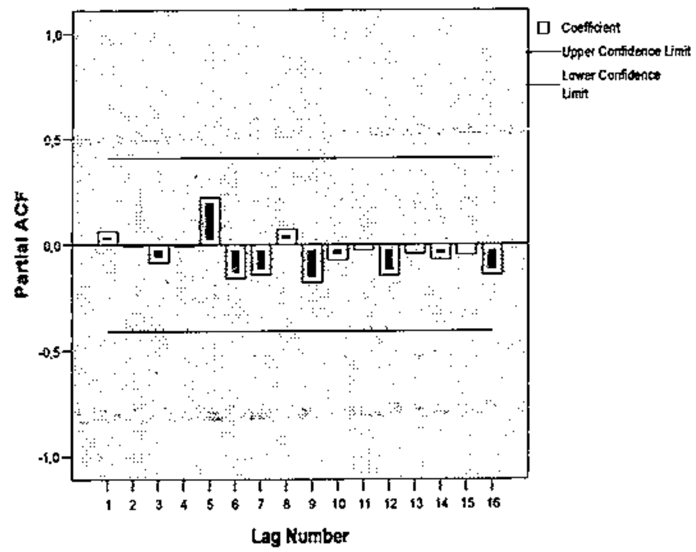
Εν συνεχεία εξετάζουμε τα υπόλοιπα.



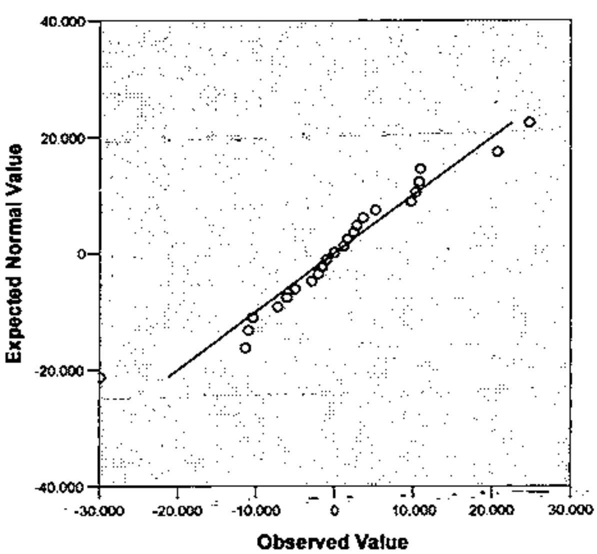
Error for ASFALISMENOI from ARIMA, MOD_10 CON



Error for ASFALISMENOI from ARIMA, MOD_10 CON



Normal Q-Q Plot of Error for ASFALISMENOI from ARIMA, MOD_10 CON



Από το πρώτο διάγραμμα παρατηρούμε ότι τα υπόλοιπα είναι κατανεμημένα κανονικά γιατί έχουμε ένα διασκορπισμένο ορθογώνιο γύρω από το μηδενικό οριζόντιο επίπεδο χωρίς πιθανές τάσεις.

Επίσης από τα διαγράμματα ACF και PACF παρατηρούμε ότι όλες οι τιμές είναι εντός των ορίων άρα "λευκός θόρυβος".

Εξετάζοντας τώρα το Q-Q διάγραμμα βλέπουμε ότι τα υπόλοιπα είναι κανονικά κατανεμημένα γιατί τα περισσότερα από αυτά πέφτουν επάνω στην διαγώνιο. Άρα τα δεδομένα μας ακολουθούν κανονική κατανομή.

ΠΡΟΒΛΕΨΗ

PERIOD	FIT	LCL	UCL
2003	1952884	1929031	1976737
2004	1950832	1909165	1992499
2005	1946077	1884853	2007300

Το έτος 2003 η τιμή της πρόβλεψης είναι 1952884. Το κατώτερο όριο που ανήκει στο 95% διάστημα εμπιστοσύνης είναι 1929031 και το ανώτερο 1976737.

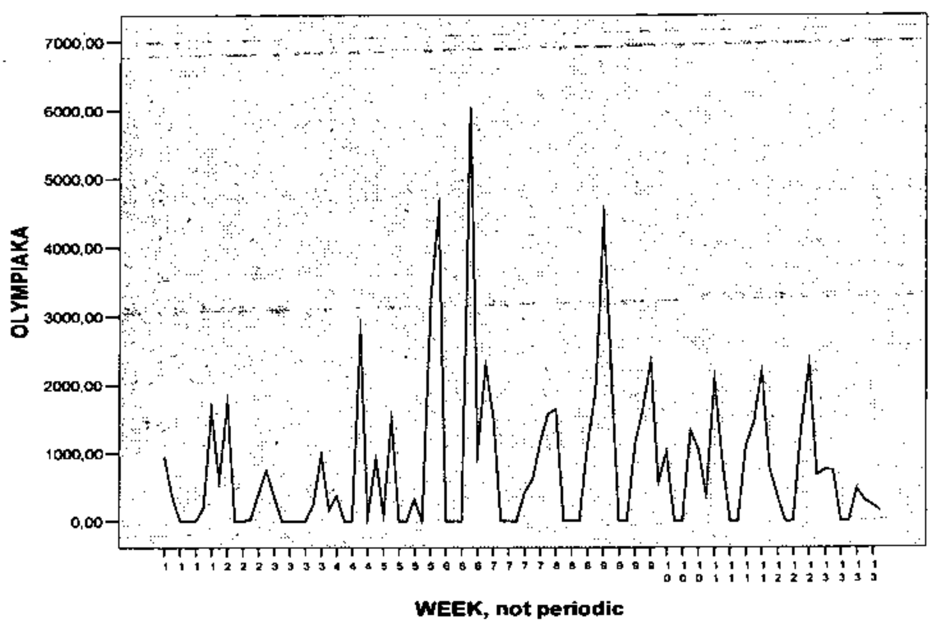
Το έτος 2004 η τιμή της πρόβλεψης είναι1950832. Το κατώτερο όριο που ανήκει στο 95% διάστημα εμπιστοσύνης είναι1909165 και το ανώτερο1992499.

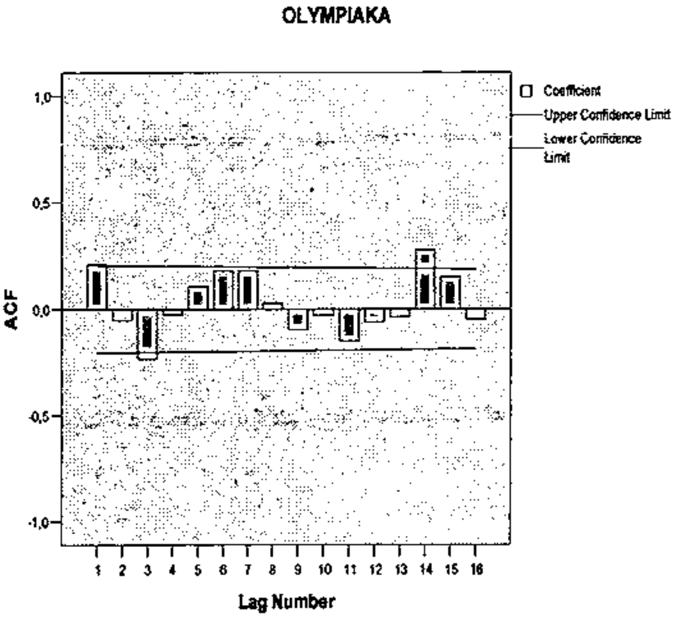
Το έτος 2005 η τιμή της πρόβλεψης είναι 1946077. Το κατώτερο όριο που ανήκει στο 95% διάστημα εμπιστοσύνης είναι 1884853 και το ανώτερο 2007300.

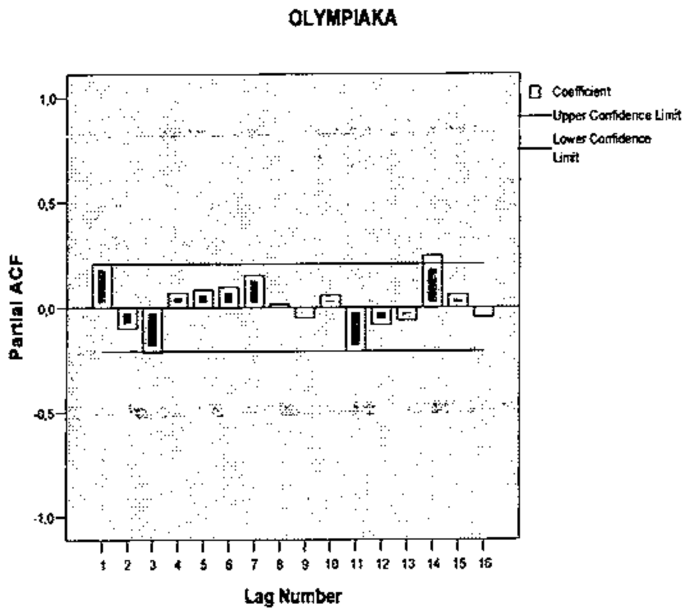
Από την πρόβλεψη των τριών αυτών χρόνων βλέπουμε ότι το 2003 συνεχίζεται η σταδιακή αύξηση των ασφαλισμένων του ΙΚΑ κατά ένα μικρό ποσοστό ενώ τα δύο τελευταία χρόνια παρατηρείται μια μικρή μείωση του αριθμού των ασφαλισμένων. Οπότε τα υπόλοιπα ικανοποιούν τις προϋποθέσεις.

6.5 <u>Εφαρμογή - Ολυμπιακά</u>

Τα δεδομένα που χρησιμοποιούνται στο μοντέλο αφορούν τις πωλήσεις των ολυμπιακών προϊόντων από τα ΕΛΤΑ των καταστημάτων της Πάτρας. Οι τιμές είναι ημερήσιες από το διάστημα 1/6/2004 έως 31/8/2004. Υπάρχει και μια δεύτερη ανεξάρτητη μεταβλητή στο μοντέλο η οποία αναφέρεται στις ημέρες που η Ελλάδα κέρδιζε κάποιο ολυμπιακό μετάλλιο. Οι τιμές που λαμβάνει αυτή μεταβλητή είναι 0=όχι μετάλλιο και 1=μετάλλιο. Θα προσαρμόσουμε ένα μοντέλο χωρίς αυτή τη μεταβλητή και στη συνέχεια θα θεωρήσουμε και αυτή τη μεταβλητή.







ΕΡΜΗΝΕΙΑ

Παρατηρώντας κανείς τη γραφική παράσταση sequence βλέπει ότι η χρονοσειρά είναι στάσιμη.

Στο διάγραμμα των αυτοσυσχετίσεων (ACF) βλέπουμε ότι οι παρατηρήσεις φθίνουν γρήγορα επομένως η χρονοσειρά είναι στάσιμη. Ακόμα έχουμε έναν στατιστικά σημαντικό συντελεστή αυτοσυσχέτισης γιατί η πρώτη υστέρηση βγαίνει έξω από τα όρια και συγκεκριμένα από το ανώτατο όριο. Άρα η τιμή του q=1.

Από το PACF διάγραμμα επιλέγουμε p=1 γιατί η πρώτη υστέρηση βρίσκεται στα όρια.

Επομένως έχουμε ένα ARIMA (1,0,1) μοντέλο. Στην συνέχεια βλέπουμε τα αποτελέσματα.

Residual Diagnostics

Number of Residuals	92
Number of Parameters	2
Residual df	89
Adjusted Residual Sum of Squares	10548341 9,520
Residual Sum of Squares	10749775 9,859
Residual Variance	1184580,9 54
Model Std. Error	1088,385
Log-Likelihood	-772,371
Akaike's Information Criterion (AIC)	1550,742
Schwarz's Bayesian Criterion (BIC)	1558,308

Parameter Estimates

		Estimates	Std Error	t	Approx Sig
Non-Seasonal Lags	AR1	,083	,486	,170	,866
	MA1	-,138	,483	-,285	,776
Constant		796,826	140,395	5,676	,000

Melard's algorithm was used for estimation.

Από τον πίνακα "Parameter Estimates" έχουμε τα εξής αποτελέσματα:

- $\hat{\phi}=0,083$ δηλαδή ο συντελεστής του AR(1) με αντίστοιχο p-value=0,866. Συκρίνοντάς το με το 0,05 επίπεδο σημαντικότητας 0,866>0,05 βλέπουμε ότι είναι μεγαλύτερο. Άρα το $\hat{\phi}$ δεν είναι στατιστικά σημαντικός συντελεστής.
- $\hat{\theta}=-0,138$ δηλαδή ο συντελεστής του MA(1) με p-value=0,776. Το τελευταίο σε σχέση με το 0,05 επίπεδο σημαντικότητας 0,776>0,05 είναι μεγαλύτερο. Άρα το $\hat{\theta}$ δεν είναι στατιστικά σημαντικός συντελεστής.
- $\hat{a}=796,826$ δηλαδή η σταθερά με αντίστοιχο p-value=0. Σε σχέση με το 0,05 επίπεδο σημαντικότητας 0<0,05 είναι μικρότερο. Άρα το $\hat{a}$ είναι στατιστικά σημαντικός συντελεστής.

Θα επαναλάβουμε την ίδια διαδικασία αφαιρώντας τον συντελεστή $\hat{\phi}$ από το μοντέλο επειδή δεν είναι στατιστικά σημαντικός αλλά και γιατί έχει μεγαλύτερη τιμή από τον $\hat{\theta}$, που και αυτός δεν είναι στατιστικά σημαντικός.

Τα αποτελέσματα μετά την εξάλειψη του συντελεστή $\hat{\phi}$ τα βλέπουμε στους δύο πίνακες που ακολουθούν.

Residual Diagnostics

Number of Residuals	92
Number of Parameters	1
Residual df	90
Adjusted Residual Sum of Squares	105551861,575
Residual Sum of Squares	105562647,769
Residual Variance	1172223,091
Model Std. Error	1082,693
Log-Likelihood	-772,387
Akaike's Information Criterion (AIC)	1548,775
Schwarz's Bayesian Criterion (BIC)	1553,818

Parameter Estimates

		Estimates	Std Error	t	Approx Sig
Non-Seasonal Lags	MA1	-,210	,103	-2,036	,045
Constant		797,053	136,337	5,846	,000

Melard's algorithm was used for estimation.

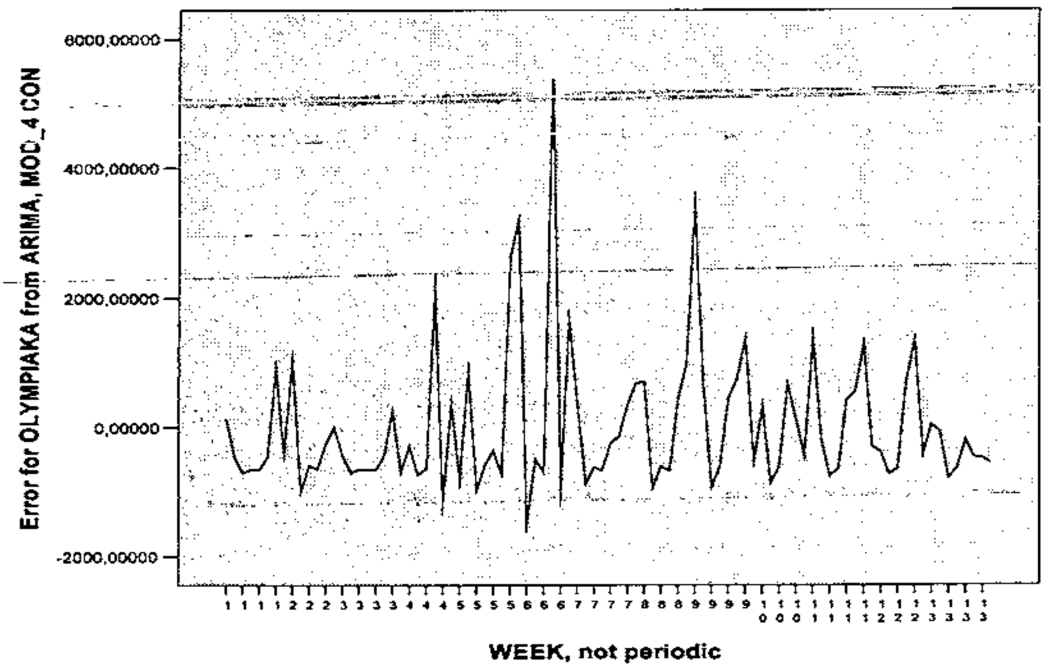
Από τον πίνακα "Parameter Estimates" έχουμε τα εξής αποτελέσματα:

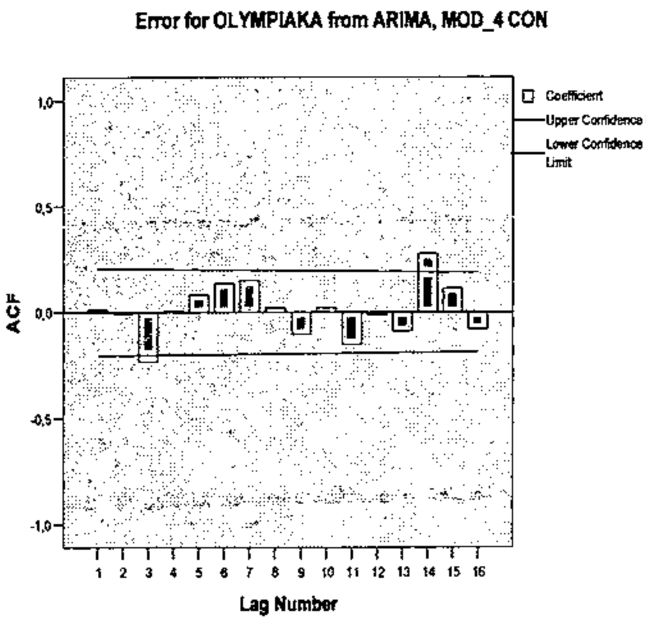
- $\hat{\theta} = -0,210$ δηλαδή ο συντελεστής του MA(1) με p-value=0,045. Το τελευταίο σε σχέση με το 0,05 επίπεδο σημαντικότητας 0,045<0,05 είναι μικρότερο. Άρα το $\hat{\theta}$ είναι στατιστικά σημαντικός συντελεστής.
- $\hat{a} = 797,053$ δηλαδή η σταθερά με αντίστοιχο p-value=0. Σε σχέση με το 0,05 επίπεδο σημαντικότητας 0<0,05 είναι μικρότερο. Άρα το $\hat{a}$ είναι στατιστικά σημαντικός συντελεστής.

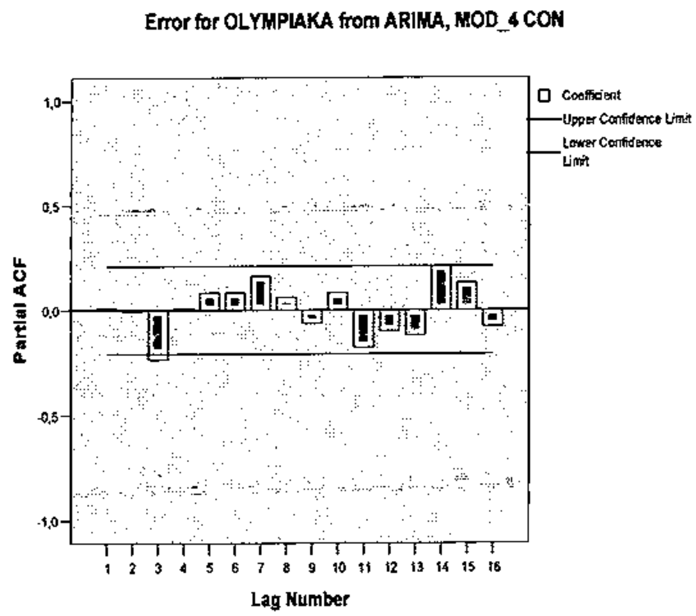
Το μοντέλο μας είναι ARIMA(0,0,1) και έχει την μορφή:

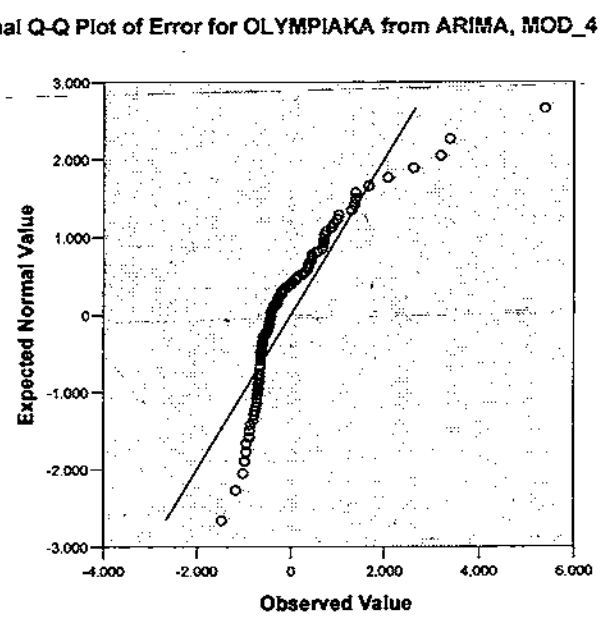
$$X_t = (1 + 0,210)e_t + 797,053$$

Στη συνέχεια παρουσιάζονται τα διαγράμματα όσο αναφορά τα υπόλοιπα.









Από το πρώτο διάγραμμα παρατηρούμε ότι τα σφάλματα είναι κατανεμημένα κανονικά γιατί έχουμε ένα διασκορπισμένο ορθογώνιο γύρω από το μηδενικό οριζόντιο επίπεδο χωρίς πιθανές τάσεις.

Επίσης από τα διαγράμματα ACF και PACF παρατηρούμε ότι όλες οι τιμές σχεδόν είναι εντός των ορίων.

Εξετάζοντας το Q-Q διάγραμμα βλέπουμε ότι τα υπόλοιπα είναι κανονικά κατανεμημένα γιατί τα περισσότερα από αυτά πέφτουν επάνω στην διαγώνιο. Άρα τα δεδομένα μας ακολουθούν κανονική κατανομή.

Κάνουμε πάλι τώρα την ίδια διαδικασία ARIMA με μοντέλο αυτό στο οποίο καταλήξαμε νωρίτερα, δηλαδή το ARIMA(0,0,1) προσθέτοντας την ανεξάρτητη μεταβλητή μετάλλιο με τιμές 0 και 1 έχουμε:

Residual Diagnostics

Number of Residuals	92
Number of Parameters	1
Residual df	89
Adjusted Residual Sum of Squares	10400057 2,550
Residual Sum of Squares	10400167 0,755
Residual Variance	1168013,9 98
Model Std. Error	1080,747
Log-Likelihood	-771,720
Akaike's Information Criterion (AIC)	1549,440
Schwarz's Bayesian Criterion (BIC)	1557,006

Parameter Estimates

		Estimates	Std Error	t	Approx Sig
Non-Seasonal Lags	MA1	-,203	,104	-1,947	,055
Regression Coefficients	METALIO	449,682	389,427	1,155	,251
Constant		743,502	142,982	5,200	,000

Melard's algorithm was used for estimation.

Από τον πίνακα "Parameter Estimates" έχουμε τα εξής αποτελέσματα:

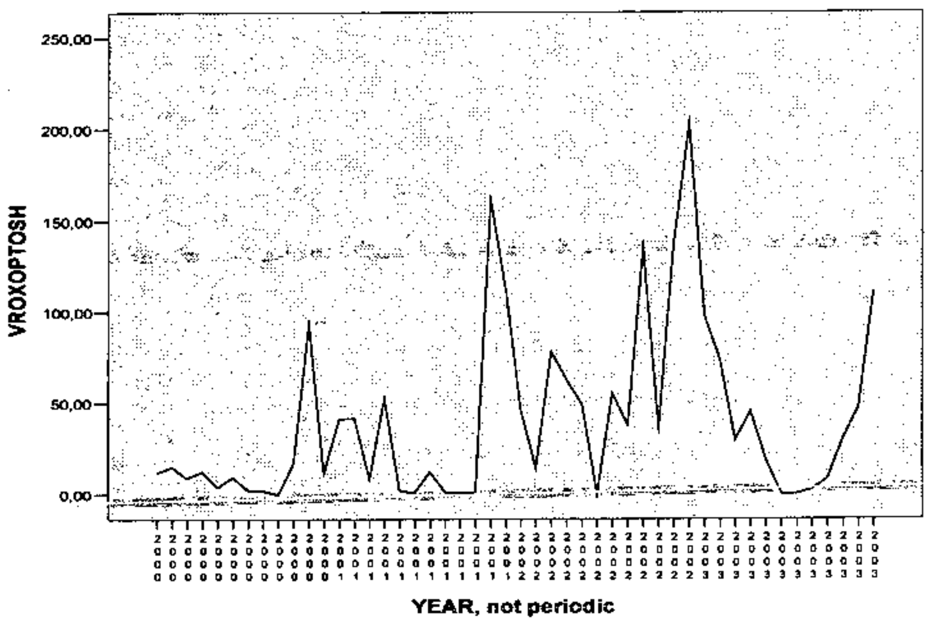
- $\hat{\theta}=-0,203$ δηλαδή ο συντελεστής του MA(1) με p-value=0,055. Το τελευταίο σε σχέση με το 0,05 επίπεδο σημαντικότητας 0,055>0,05 είναι μεγαλύτερο. Άρα το $\hat{\theta}$ δεν είναι στατιστικά σημαντικός συντελεστής.
- $\hat{a}=743,502$ δηλαδή η σταθερά με αντίστοιχο p-value=0. Σε σχέση με το 0,05 επίπεδο σημαντικότητας 0<0,05 είναι μικρότερο. Άρα το $\hat{a}$ είναι στατιστικά σημαντικός συντελεστής.

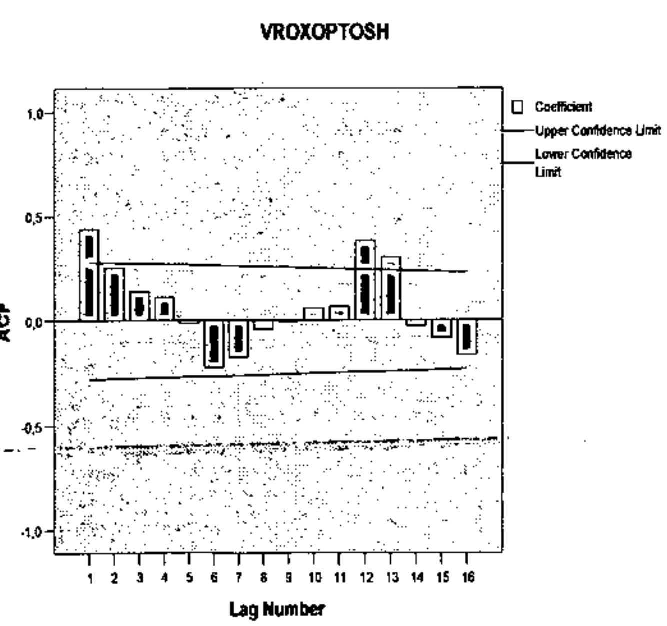
- Η εκτίμηση της ανεξάρτητης μεταβλητής είναι ίση με 449,682 με p-value=0,251. Συγκρίνοντας το με το 0,05 επίπεδο σημαντικότητας 0,251>0,05 είναι μεγαλύτερο. Επομένως δεν είναι στατιστικά σημαντική. Το συμπέρασμα είναι ότι η ανεξάρτητη μεταβλητή δεν επηρεάζει το μοντέλο μας. Με άλλα λόγια τα μετάλλια τις ημέρες που κέρδιζε η Ελλάδα δεν επηρέαζαν τις πωλήσεις των ολυμπιακών προϊόντων των επόμενων ημερών. Δηλαδή οι πωλήσεις εκείνες τις ημέρες δεν αυξάνονταν.

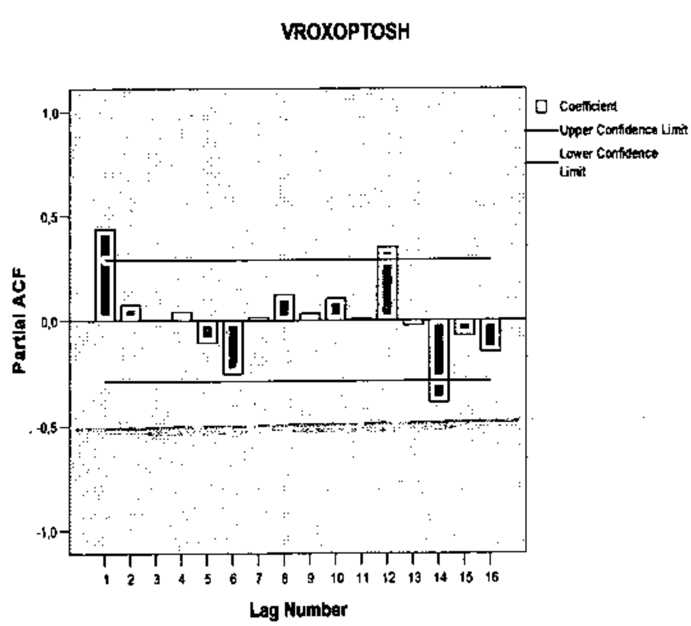
Από τα παραπάνω αποτελέσματα κρίνουμε ότι το μοντέλο που ισχύει είναι το ARIMA(0,0,1).

6.6 Εφαρμογή - Βροχόπτωση

Τα δεδομένα που χρησιμοποιούνται στο μοντέλο αναφέρονται στη βροχόπτωση στην Ελλάδα και οι τιμές είναι μηνιαίες καλύπτοντας το διάστημα 2000 έως 2003.





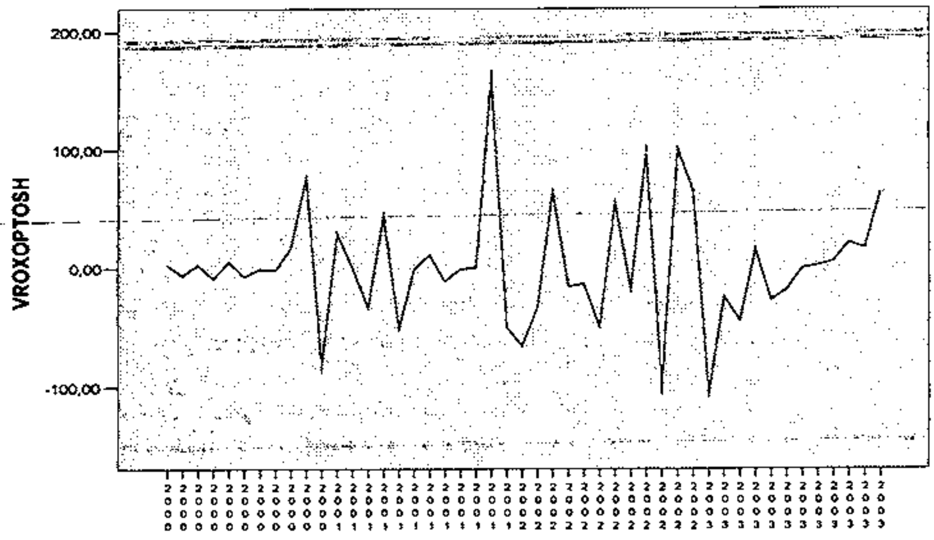


ΕΡΜΗΝΕΙΑ

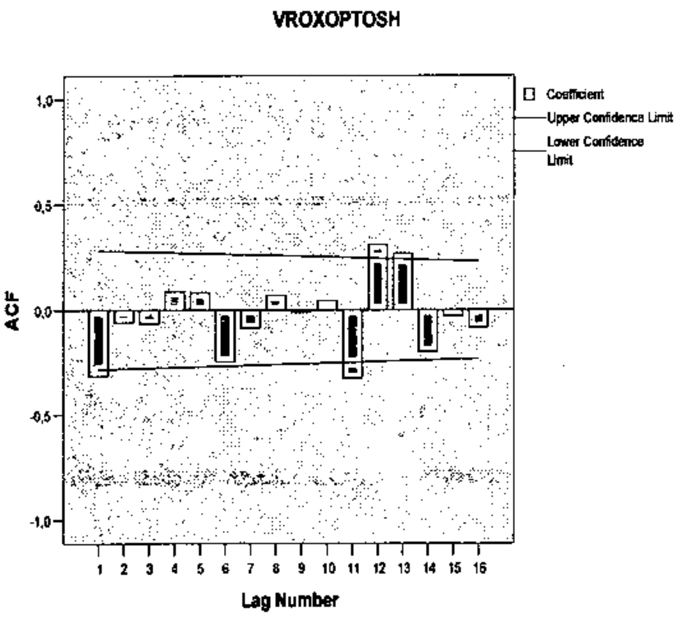
Παρατηρώντας κανείς τη γραφική παράσταση sequence βλέπει ότι η χρονοσειρά δεν είναι στάσιμη.

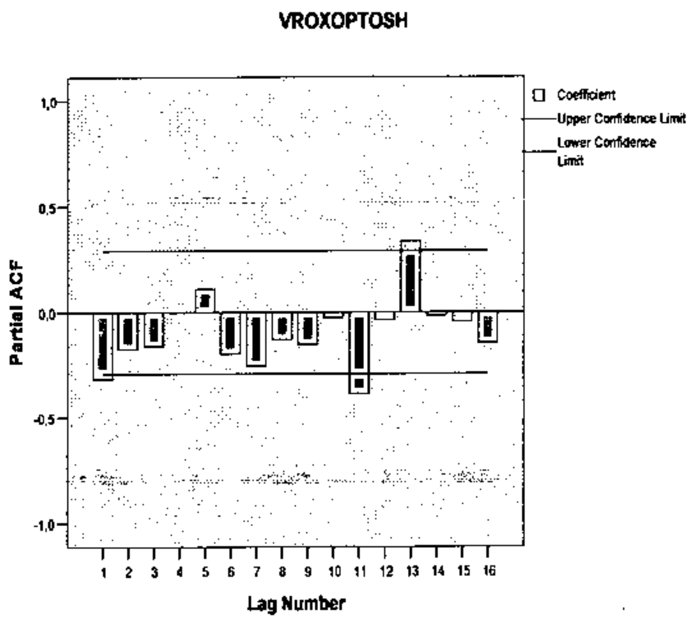
Στο διάγραμμα των αυτοσυσχετίσεων (ACF) και των μερικών αυτοσυσχετίσεων (PACF) αντίστοιχα βλέπουμε ότι οι παρατηρήσεις φθίνουν αργά επομένως η χρονοσειρά δεν είναι στάσιμη.

Για τον λόγο αυτό, επειδή δηλαδή η χρονοσειρά δεν είναι στάσιμη, θα την διαφορίσουμε. Τα αποτελέσματα φαίνονται στους παρακάτω πίνακες.



Transforms: difference(1)





Από το διάγραμμα sequence είναι φανερό ότι η χρονοσειρά "μαζεύτηκε" αρκετά και προς τα επάνω σε σχέση με το αρχικό διάγραμμα. Τα όριά του από το 250 κατέβηκε στο 200 και οι κορυφές του κυμαίνονται σχεδόν στο ίδιο ύψος. Άρα το διάγραμμα μας δείχνει ότι η διαδικασία αυτή είναι στάσιμη γιατί έχουμε ένα διασκορπισμένο ορθογώνιο γύρω από το μηδενικό οριζόντιο επίπεδο χωρίς πιθανές τάσεις.

Εν συνεχεία το διάγραμμα των αυτοσυχετίσεων (ACF) φθίνει γρήγορα πράγμα που δείχνει την στασιμότητα της χρονοσειράς. Επίσης επιλέγουμε q=1 γιατί η πρώτη υστέρηση βγαίνει έξω από τα όρια.

Από το PACF διάγραμμα επιλέγουμε p=1 γιατί η πρώτη υστέρηση βγαίνει έξω από τα όρια.

Επομένως έχουμε ένα ARIMA (1,1,1) μοντέλο με τα εξής παρακάτω αποτελέσματα.

Residual Diagnostics

Number of Residuals	47
Number of Parameters	2
Residual df	44
Adjusted Residual Sum of Squares	88712,556
Residual Sum of Squares	98601,335
Residual Variance	1947,126
Model Std. Error	44,126
Log-Likelihood	-243,962
Akaike's Information Criterion (AIC)	493,924
Schwarz's Bayesian Criterion (BIC)	499,475

Parameter Estimates

		Estimates	Std Error	t	Approx Sig
Non-Seasonal Lags	AR1	,417	,176	2,369	,022
	MA1	,957	,142	6,734	,000
Constant		1,145	,894	1,281	,207

Melard's algorithm was used for estimation.

Από τον πίνακα "Parameter Estimates" έχουμε τα εξής αποτελέσματα:

- $\hat{\phi}=0,417$ δηλαδή ο συντελεστής του AR(1) με αντίστοιχο p-value=0,022. Συκρίνοντάς το με το 0,05 επίπεδο σημαντικότητας 0,022<0,05 βλέπουμε ότι είναι μικρότερο. Άρα το $\hat{\phi}$ είναι στατιστικά σημαντικός συντελεστής.
- $\hat{\theta}=0,957$ δηλαδή ο συντελεστής του MA(1) με p-value=0. Το τελευταίο σε σχέση με το 0,05 επίπεδο σημαντικότητας 0<0,05 είναι μικρότερο. Άρα το $\hat{\theta}$ είναι στατιστικά σημαντικός συντελεστής.
- $\hat{a}=1,145$ δηλαδή η σταθερά με αντίστοιχο p-value=0,207. Σε σχέση με το 0,05 επίπεδο σημαντικότητας 0,207>0,05 είναι μεγαλύτερο. Άρα το $\hat{a}$ δεν είναι στατιστικά σημαντικός συντελεστής.

Άρα θα επαναλάβουμε την ίδια διαδικασία αφαιρώντας τη σταθερά από το μοντέλο επειδή δεν είναι στατιστικά σημαντική.

Residual Diagnostics

Number of Residuals	47
Number of Parameters	2
Residual df	45
Adjusted Residual Sum of Squares	91695,078
Residual Sum of Squares	98956,856
Residual Variance	1984,830
Model Std. Error	44,551
Log-Likelihood	-244,731
Akaike's Information Criterion (AIC)	493,463
Schwarz's Bayesian Criterion (BIC)	497,163

Parameter Estimates

		Estimates	Std Error	t	Approx Sig
Non-Seasonal Lags	AR1	,435	,164	2,644	,011
	MA1	,935	,082	11,422	,000

Melard's algorithm was used for estimation.

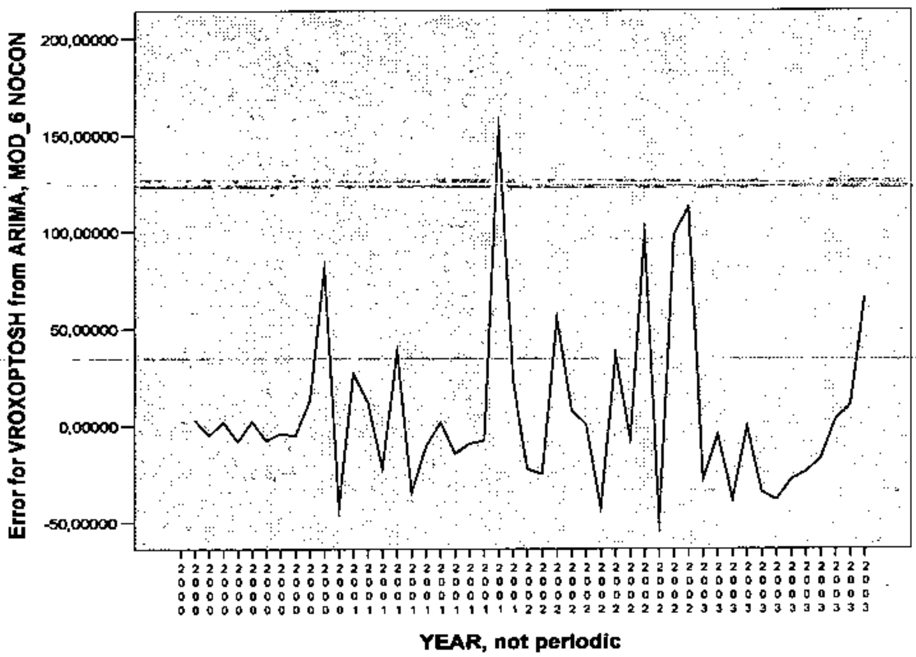
Από τον πίνακα "Parameter Estimates" έχουμε τα εξής αποτελέσματα:

- $\hat{\phi}=0,435$ δηλαδή ο συντελεστής του MA(1) με p-value=0,011. Συγκρίνοντας το τελευταίο με το 0,05 επίπεδο σημαντικότητας 0,011<0,05 είναι μικρότερο. Άρα το $\hat{\phi}$ είναι στατιστικά σημαντικός συντελεστής.
- $\hat{\theta}=0,935$ δηλαδή ο συντελεστής του MA(1) με p-value=0. Συγκρίνοντας το τελευταίο με το 0,05 επίπεδο σημαντικότητας 0<0,05 είναι μικρότερο. Άρα το $\hat{\theta}$ είναι στατιστικά σημαντικός συντελεστής.

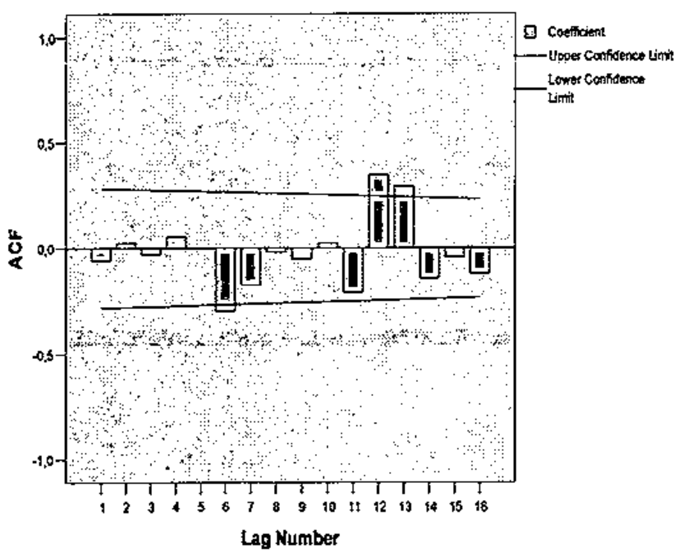
Το μοντέλο μας είναι ARIMA(1,1,1) και έχει την μορφή:

$$(1-B)(1-0,435B)X_t=(1-0,935B)e_t$$

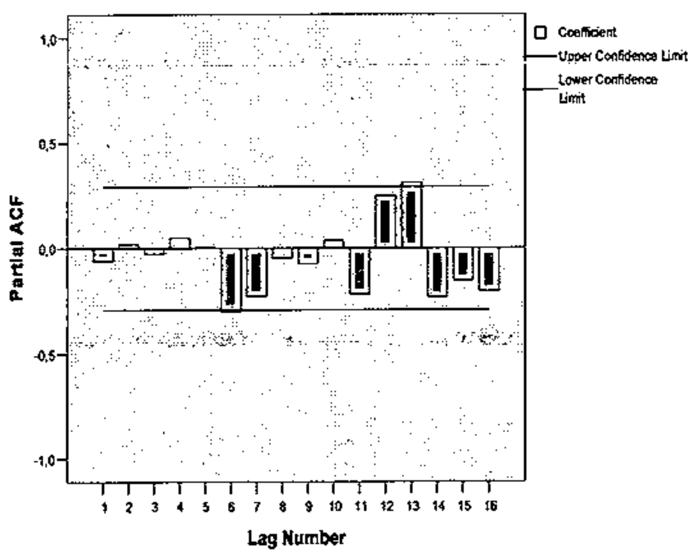
Εν συνεχεία εξετάζουμε τα υπόλοιπα.



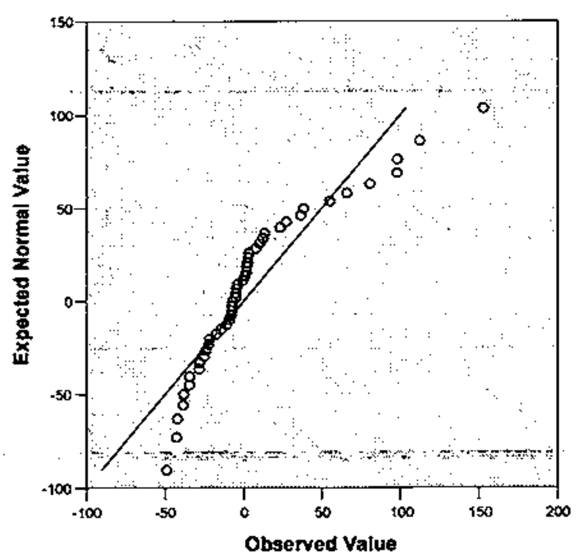
Error for VROXOPTOSH from ARIMA, MOD_6 NOCON



Error for VROXOPTOSH from ARIMA, MOD_6 NOCON



Normal Q-Q Plot of Error for VROXOPTOSH from ARIMA, MOD_6 NOCON



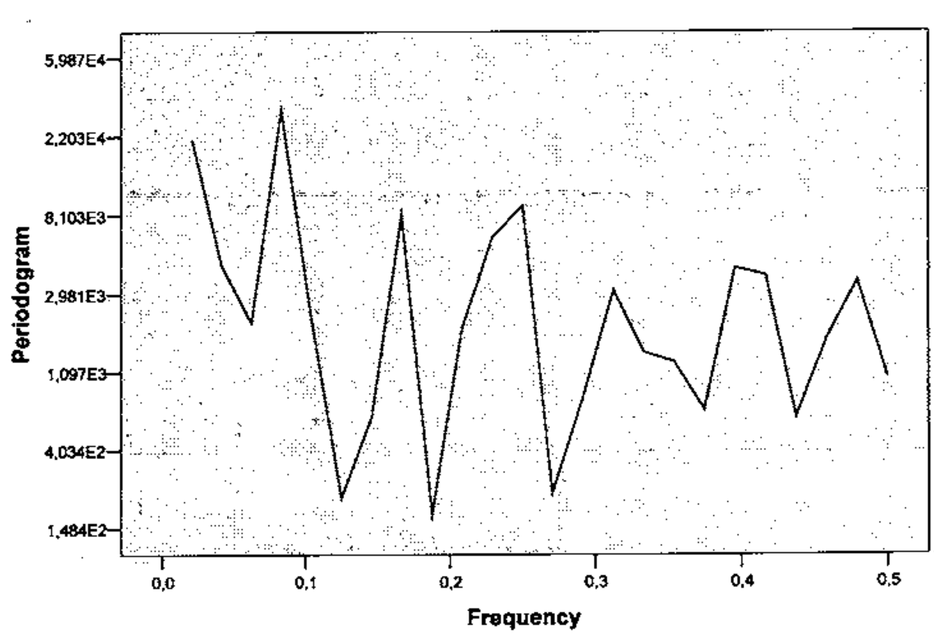
Από το πρώτο διάγραμμα παρατηρούμε ότι τα υπόλοιπα είναι κατανεμημένα κανονικά γιατί έχουμε ένα διασκορπισμένο ορθογώνιο γύρω από το μηδενικό οριζόντιο επίπεδο χωρίς πιθανές τάσεις.

Επίσης από τα διαγράμματα ACF και PACF παρατηρούμε ότι όλες οι τιμές σχεδόν είναι εντός των ορίων.

Εξετάζοντας τώρα το Q-Q διάγραμμα βλέπουμε ότι τα υπόλοιπα είναι κανονικά κατανεμημένα γιατί τα περισσότερα από αυτά πέφτουν επάνω στην διαγώνιο. Άρα τα δεδομένα μας ακολουθούν κανονική κατανομή.

Με το μοντέλο που βρήκαμε πιο πάνω ARIMA(1,1,1) θα συνεχίσουμε την διαδικασία πάνω στην χρονοσειρά μας στο πεδίο των συχνοτήτων. Έχοντας πάλι ως μεταβλητή μας την "Βροχόπτωση" θα φτιάξουμε το περιοδόγραμμα και το διάγραμμα της φασματικής πυκνότητας.

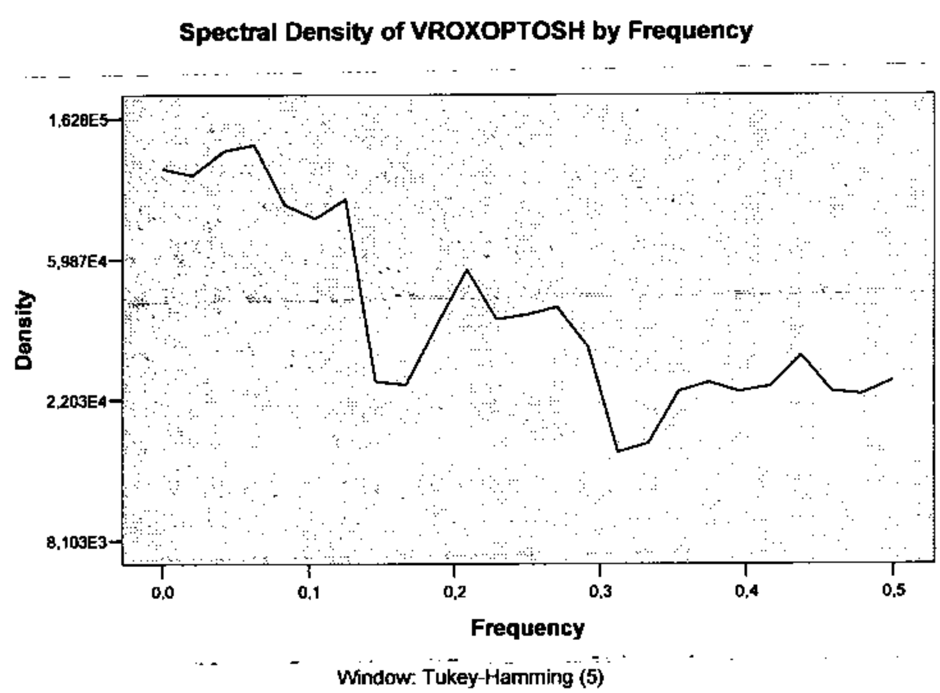
Periodogram of VROXOPTOSH by Frequency



Στο περιοδόγραμμα η οριζόντια γραμμή δείχνει την συχνότητα και η κάθετη γραμμή δείχνει την σχετική τιμή ή σημαντικότητα κάθε που αντιστοιχεί σε κάθε συχνότητα. Στο διάγραμμα λοιπόν παρατηρούμε δύο υψηλές τιμές και πολλές άλλες χαμηλότερες σε σχέση με τις προηγούμενες. Ο τύπος της συχνότητας είναι:

$$Frequency_j = \frac{j}{N}$$

Όπου j είναι το πόσες φορές επαναλαμβάνεται ο κύκλος μέσα στο δείγμα και N είναι ο αριθμός των παρατηρήσεων. Άρα στο δείγμα μας έχουμε σύνολο 48 παρατηρήσεις και ο κύκλος αφού επαναλαμβάνεται μηνιαίως τότε το $j = 11$. Η συχνότητα είναι 11/48=0,23.



Στη συνέχεια η εκτίμηση της "Βροχόπτωσης" φαίνεται από το διάγραμμα της φασματικής πυκνότητας. Παρατηρούμε ότι οι περισσότερες από τις κορυφές έχουν μετακινηθεί. Βέβαια υπάρχουν δύο φανερές κορυφές μέσα στο έτος 2000 (οι οποίες αντιστοιχούν στο να σκιαγραφεί συχνότητες γύρω στο 1/11, ή 0,1). Οι κορυφές λοιπόν αυτές έχουν γίνει πιο ομαλές ή έχουν σταθεροποιηθεί με αποτέλεσμα να είναι πιο πλατιές σε σχέση με αυτές που παρουσιάζει το περιοδόγραμμα. Το παράθυρο που χρησιμοποιήσαμε στην φασματική πυκνότητα είναι αυτό των Tukey-Hamming με άνοιγμα 5. Το παράθυρο αναφέρεται στο μοντέλο των βαρών εφαρμοσμένο για την ερμηνεία του κινητού μέσου. Αυτά τα βάρη γύρω από το κεντρικό σημείο. Το άνοιγμα του παραθύρου πιθανόν το επηρεάζει.

Το άνοιγμα των παρατηρήσεων δείχνει τον αριθμό των σημείων που περιλαμβάνει ο κινητός μέσος.

ΚΕΦΑΛΑΙΟ 7: ΣΥΜΠΕΡΑΣΜΑ

Η παρούσα εργασία βασίστηκε στην ανάλυση των χρονοσειρών που μπορεί να γίνει είτε στο πεδίο του χρόνου είτε στο πεδίο των συχνοτήτων, προκειμένου να προσδιοριστούν τα βασικά χαρακτηριστικά που διέπουν την συμπεριφορά τους. Η ανάλυση αυτή έγινε με την εξέταση βασικών μεθόδων χρονοσειρών που διαφέρουν ως προς την πολυπλοκότητα, την ευχρηστία και την ακρίβεια των προβλέψεων τους.

Οι περιγραφικές μέθοδοι ανάλυσης χρονοσειρών θεωρούνται απλές και δεν απαιτούν σύνθετα μαθηματικά μοντέλα. Μελετώνται κάτω από το πρίσμα της συμπεριφοράς της χρονοσειράς στο παρελθόν σαν βάση για να γίνει η πρόβλεψη της στο μέλλον. Αυτές οι μέθοδοι αναφέρονται σε μη στοχαστικά μοντέλα και μπορούμε να τις χωρίσουμε στις μεθόδους εξομάλυνσης και στις μεθόδους ανάλυσης και διαχωρισμού των χρονοσειρών. Οι μεν πρώτες παρέχουν ικανοποιητικές προβλέψεις στον βραχυπρόθεσμο ορίζοντα, ενώ οι άλλες και στον μακροπρόθεσμο.

Οι μέθοδοι ανάλυσης χρονοσειρών στο πεδίο του χρόνου αναφέρονται σε στοχαστικά μοντέλα. Αυτά τα μοντέλα γνωστά και ως ARIMA προτάθηκαν από τους Box και Jenkins με σκοπό την επιλογή του καλυτέρου μοντέλου για την διαδικασία της πρόβλεψης. Η πρόβλεψη σε αυτά τα μοντέλα βασίζεται στις παλαιότερες τιμές της χρονοσειρας όσο και από τα εμφανισθέντα πρότυπα συμπεριφοράς που παρουσιάζονται. Τα μοντέλα αυτά αποτελούν μια σύνθετη τεχνική που απαιτεί καλή γνώση από τον χρήστη. Η σωστή τους χρήση δίνει ικανοποιητικές προβλέψεις τόσο στον βραχυπρόθεσμο όσο και στο μεσοπρόθεσμο ορίζοντα.

Οι οικονομετρικές μέθοδοι βασίζονται πάνω στην εξέταση μοντέλων με χρονικές υστερήσεις που εκφράζουν την δυναμική του οικονομικού συστήματος. Τέτοια μοντέλα είναι τα δυναμικά ή γνωστά ως VAR. Το πεδίο εφαρμογής τους αφορά την βραχυχρόνια και μεσοπρόθεσμη πρόβλεψη και θεωρούνται απλά στην χρήση τους.

Οι φασματικές μέθοδοι βασίζονται στην ανάλυση της χρονοσειράς στο πεδίο των συχνοτήτων που είναι συμπληρωματική της ανάλυσης στο πεδίο του χρόνου. Με αυτές τις μεθόδους μπορούν να εξεταστούν χαρακτηριστικά της χρονοσειράς που δεν θα μπορούσαν εύκολα να εντοπιστούν με άλλες μεθόδους. Οι φασματικές μέθοδοι βασίζονται στην μελέτη της φασματικής πυκνότητας και της αυτοσυσχέτισης έχοντας ως αποτέλεσμα την γνωστή ως φασματική ανάλυση.

Το βασικό ερώτημα που προκύπτει μετά την ανάλυση διάφορων μεθόδων χρονοσειρών είναι: *ποια μέθοδο πρέπει να ακολουθήσει κανείς για να πετύχει τις πιο άριστες προβλέψεις για τα μεγέθη που τον ενδιαφέρουν;*

Στο τέλος της εργασίας έγινε μια προσπάθεια με τις εφαρμογές που παρουσιαστήκαν να αποσαφηνιστεί ο τρόπος επιλογής των καλύτερων μοντέλων. Ανάλογα με τα δεδομένα που υπήρχαν, έγινε προσαρμογή σε κάποιο μοντέλο με σκοπό να περιγράφουν όσον το δυνατό καλυτέρα οι ιδιαιτερότητες του καθενός ξεχωριστά. Πέρα από αυτό έγινε προσπάθεια να γίνει πρόβλεψη της μελλοντικής συμπεριφοράς των μοντέλων αυτών με την μέγιστη ακρίβεια.

Στην πράξη η εφαρμογή της καταλληλότερης μεθόδου είναι κυρίως θέμα επιλογής του κάθε αναλυτή. Συχνά, δεν είναι δίκαιη η σύγκριση διαφορετικών μεθόδων χρονοσειρών με μεγάλα δείγματα. Εκείνο που έχει πρωτεύουσα σημασία είναι η ικανότητα με την οποία η μέθοδος κάνει τις προβλέψεις.

ΒΙΒΛΙΟΓΡΑΦΙΑ

Ελληνική Βιβλιογραφία

1. Ανδρικόπουλος Ανδρέας Α. «Οικονομετρία: Θεωρία και Εμπειρικές Εφαρμογές 1998».
2. Δημελή Σοφία «Σύγχρονες Μέθοδοι Ανάλυσης Χρονολογικών σειρών», Εκδόσεις Κριτική 2002.
3. Καρυώτη Βασιλική «Σημειώσεις από το μάθημα Τεχνικές Προβλέψεων» 2004.
4. Κασκαρέλης Α. Ιωάννης « Ένδεκα Μαθήματα Οικονομετρίας» 2η Έκδοση Αναθεωρημένη, Εκδόσεις Gutenberg 1999.
5. Κεντρικό Ταχυδρομείο Πάτρας, «Στοιχεία Έρευνας, 2004».
6. Κιντής Α. Ανδρέας, «Στατιστικές και Οικονομετρικές Μέθοδοι», Εκδόσεις Gutenberg.»
7. Μπόρα-Σέντα Ευθυμία, Λέκτορας Α.Π.Θ., Χ. Μωυσιάδης, Επίκουρος Καθηγητής Α.Π.Θ., «Εφαρμοσμένη Στατιστική, Πολλαπλή Παλινδρόμηση, Ανάλυση Διασποράς, Χρονοσειρές» εκδόσεις Gutenberg.
8. Ξενάκης Σ. Ανδρέας «Ανάλυση Χρονολογικών Σειρών και Προβλέψεις - Ανάλυση στο Πεδίο του Χρόνου» πανεπιστημιακές παραδόσεις Αθήνα 1998.
9. Παπαδήμας Όθωνας M.Sc Μαθηματικού – Στατιστικού, «Στατιστική ΙΙ», Μακεδονικές Εκδόσεις 1994.

Ξένη Βιβλιογραφία

1. Banerjee, A., R.L. Lumsdaine and J.H. Stock (1986). «Recursive and Sequential Tests of the Unit-Root and Trend-Break Hypothesis: Theory and International Evidence».
2. Bartlett, M.S. (1946). «On the Theoretical Specification of Sampling Properties of Autocorrelated Time Series».
3. Box, G.E.P. and Jenkins, G.M. (1976). «Time Series Analysis, Forecasting and Control» Oakland, California, Holden day.

4. Box, G.E.P. and Pierce, D.A. (1970). «Distribution of residual autocorrelations in ARIMA models» Oakland, California, Holden day.
5. Engle, R.F., and C.W.J. Granger (1987). «Co-integration and Error Correction: Representation, Estimation and Testing».
6. Granger, C.W.J. (1969). «Investigating Causal Relations by econometric Models and Cross-Spectral Methods».
7. Granger, C.W.J. (1981). «Some Properties of Time Series Data and Their Use in Econometric Model Specification».
8. Granger, C.W.J. (1983). «Co-integrated Variables and Error Correcting Models», University of California, San Diego.
9. Jarrett Jeffrey «Μέθοδοι προβλέψεων Για Οικονομικές – Επιχειρηματικές Αποφάσεις», Εκδόσεις Gutenberg 2002.
10. Karioti Vassiliki Department of Mathematics «Master Time Series Analysis of Peak Expiratory Flow», September 1997.
11. Ljung, G.M. and Box, G.E.P. (1978). «On a Measure of Lack of Fit in Time Series Models», Biometrica.
12. Makridakis/Wheelright/McGee, «Forecasting/Methods and Applications/», 2nd Ed.

Ηλεκτρονική Βιβλιογραφία

1. http://fsu.ece.ntua.gr/old/gr/texnikes_2004.html
2. http://www.itl.nist.gov
3. http://meteo.chi.civil.ntua.gr/el/jointtime.php?mode=custom
4. http://www.users.auth.gr
5. http://home.ubalt.edu/
6. http://www.cas.lancs.ac.uk/
7. http://www.foreteler.gr/exom.html
8. http://webmail.ucy.ac.cy/~fokianos/bookts.htm
9. http://users.auth.gr/~dkugiu/Teach/TimeSeries/
10. http://www.stasoftinc.com
11. http://www.dap-papei.gr/downloads/ode/4%20eksamino/tada.doc
12. http://www.ekato.gr

ΠΑΡΑΡΤΗΜΑ

Δεδομένα ετήσιων ατυχημάτων των ασφαλισμένων στην Ελλάδα.

ΗΜΕΡΟΜΗΝΙΑ	ΑΤΥΧΗΜΑΤΑ
1977	46597
1978	47115
1979	46976
1980	44950
1981	45493
1982	41327
1983	38828
1984	38658
1985	38836
1986	36913
1987	36590
1988	32192
1989	29847
1990	27846
1991	25185
1992	25063
1993	23959
1994	22608
1995	21540
1996	21255
1997	20046
1998	18615
1999	17658
2000	16822
2001	16483
2002	16031

ΠΗΓΗ: http://www.ekato.gr

Δεδομένα ετήσιων τροχαίων στην Ελλάδα.

ΗΜΕΡΟΜΗΝΙΑ	ΤΡΟΧΑΙΑ
1992	22006
1993	22165
1994	22222
1995	22798
1996	23775
1997	24295
1998	24819
1999	24231
2000	23001
2001	19671
2002	16809
2003	15751

ΠΗΓΗ:http://www.ekato.gr

Δεδομένα ετήσιων ασφαλισμένων του ΙΚΑ στην Ελλάδα.

ΗΜΕΡΟΜΗΝΙΑ	ΑΣΦΑΛΙΣΜΕΝΟΙ
1977	1222000
1978	1280000
1979	1356000
1980	1431000
1981	1508000
1982	1546000
1983	1589000
1984	1646000
1985	1683000
1986	1719000
1987	1744000
1988	1766000
1989	1795000
1990	1812000
1991	1830000
1992	1849000
1993	1861000
1994	1874000
1995	1884000
1996	1889000
1997	1902000
1998	1908000
1999	1935000
2000	1941265
2001	1949495
2002	1952232

ΠΗΓΗ: http://www.ekato.gr

Δεδομένα πωληθέντων ολυμπιακών προϊόντων από τα ΕΛΤΑ Πατρών και οι ημέρες νίκης μεταλλίων της Ελλάδας.

ΟΛΥΜΠΙΑΚΑ			ΜΕΤΑΛΛΙΟ		
942,97	320,2	561,82	0	0	0
382,05	0	1008,03	0	0	0
0	3261,4	0	0	0	0
0	4540,16	0	0	0	0
200,7	0	1338,32	0	0	0
1618,71	0	1059,89	0	0	0
594,55	0	385,13	0	0	0
1737,15	6042,08	2087,28	0	0	0
0	1049,34	903,49	0	0	0
0	2284,62	0	0	0	0
27,3	1527,14	0	0	0	0
400,52	0	1098,03	0	0	0
742,9	0	1446,03	0	0	0
371,5	0	2204,29	0	0	0
0	417,7	777,1	0	1	0
0	612,48	364,75	0	1	0
0	1139,48	0	0	1	0
0	1570,66	0	0	0	0
231,02	1644,92	1374,09	0	1	0
966,59	0	2303,34	0	1	0
162	0	671,32	0	1	0
368,83	0	758,06	0	0	0
0	1106,84	748,24	0	1	0
0	1882,87	0	0	1	0
2738,13	4390,2	0	0	1	0
67,08	2213,44	474,78	0	1	0
912,71	0	307,15	0	1	0
83,27	0	229,13	0	0	0
1497,93	1118,37	148,3	0	0	0
0	1648,99		0	0	
0	2336,02		0	0	

ΑΠΟ 1/6/2004 ΕΩΣ 31/8/2004

ΠΗΓΗ: ΕΛΤΑ ΠΑΤΡΑΣ

Δεδομένα μηνιαία για την βροχόπτωση στην Ελλάδα.

ΒΡΟΧΟΠΤΩΣΗ

11,80	46,60
15	14
9	78,20
12,20	62,60
3,80	49,20
9,40	0
2,20	55
1,80	37,80
0	135
17	36,60
93,20	137,60
11,20	202,60
41	97,40
41,40	73,40
8,40	29,20
52,40	45,20
1,80	18,20
0,80	0
12,20	0
1,20	2,20
0,40	8,20
1	30
161,20	47
112	110,8

ΠΗΓΗ: http: //meteo.chi.civil.ntua.gr/el/jointtime.php?mode=custom