- ΤΕΧΝΟΛΟΓΙΚΟ ΕΚΠΑΙΔΕΥΤΙΚΟ ΙΔΡΥΜΑ ΠΑΤΡΩΝ
  - ΣΧΟΛΗ ΔΙΟΙΚΗΣΗΣ ΟΙΚΟΝΟΜΙΑΣ
  - ΤΜΗΜΑ ΕΠΙΧΕΙΡΗΜΑΤΙΚΟΥ ΣΧΕΔΙΑΣΜΟΥ ΚΑΙ ΠΛΗΡΟΦΟΡΙΑΚΩΝ ΣΥΣΤΗΜΑΤΩΝ

ΠΤΥΧΙΑΚΗ ΕΡΓΑΣΙΑ

ΑΝΑΛΥΣΗ ΧΡΟΝΟΛΟΓΙΚΩΝ ΣΕΙΡΩΝ ΣΤΟ ΣΤΑΤΙΣΤΙΚΟ ΠΡΟΓΡΑΜΜΑ SPSS

- **ΗΛΙΑΔΗΣ ΕΥΤΥΧΙΟΣ**
- **ΤΖΟΥΜΑΝΙΚΑΣ ΔΗΜΟΣΘΕΝΗΣ**

- **ΕΠΟΠΤΕΥΩΝ ΚΑΘΗΓΗΤΗΣ**
- **ΚΟΥΝΕΤΑΣ ΚΩΝΣΤΑΝΤΙΝΟΣ**

- **ΠΑΤΡΑ-2012**

ΠΕΡΙΕΧΟΜΕΝΑ

Ευρετήριο

ΚΑΤΑΛΟΓΟΣ ΓΡΑΦΗΜΑΤΩΝ

ΚΑΤΑΛΟΓΟΣ ΠΙΝΑΚΩΝ

ΚΑΤΑΛΟΓΟΣ ΕΙΚΟΝΩΝ

ΠΡΟΛΟΓΟΣ

Ο τίτλος της συγκεκριμένης Πτυχιακής εργασίας είναι «Ανάλυση Χρονολογικών Σειρών στο Στατιτικό Πρόγραμμα SPSS» και βασίζεται στην επιστήμη της Οικονομετρίας και της Στατιστικής.

Ευχαριστούμε θερμά τον επιβλέποντα καθηγητή της συγκεκριμένης πτυχιακής εργασίας κ. Κωνσταντίνο Κουνετά για την ουσιαστική βοήθεια που μας προσέφερε σε όλη την διάρκεια εκπόνησης της, τις συμβουλές και τις διορθώσεις που μας υπέδειξε.

Επίσης θέλουμε να ευχαριστήσουμε και τους γονείς μας για την προσπάθεια και τους αγώνες που έχουν κάνει για να φτάσουμε εμείς σήμερα στο σημείο να κάνουμε αυτή την εργασία .

Ευχαριστίες θέλουμε να δώσουμε και στους φίλους μας Γιαννουδάκη Ιωάννη, Δαή Δημήτριο, Καλαμπούκα Ευάγγελο, Λουντοβίκ Νικολάου, Παπαθεοχάρη Θεοχάρη, Παπακωνσταντόπουλο Νικόλα, Στεφανογιάννη Εμμανουήλ και Φουρνιανάκη Ιωάννη για τα χρόνια που περάσαμε σαν φοιτητές.

ΠΕΡΙΛΗΨΗ

Στην πτυχιακή αυτή εργασία ασχολούμαστε με την ανάλυση χρονολογικών σειρών. Ξεκινώντας, στα πρώτα κεφάλαια γίνεται μια εισαγωγή στις χρονοσειρές, στην χρησιμότητά τους και τι μπορούν αυτές να περιγράψουν, καθώς επίσης γίνεται μια λεπτομερής ανάλυση σε βασικές έννοιες όπως διάφορα μέτρα και στασιμότητα και αναλύονται χαρακτηριστικά όπως η τάση, η περιοδικότητα κ.λ.π. Εξετάζονται συγκεκριμένες κατηγορίες χρονολογικών σειρών(στάσιμες, μη στάσιμες).Έπειτα γίνεται μια εκτενής αναφορά στα αυτοπαλίνδρομα υποδείγματα αλλά και στα υποδείγματα κινητού μέσου. Ακολουθούν τα ARMA και ARIMA και μετά η μεθοδολογία BOX-JENKINS που εκεί ουσιαστικά γίνεται η εκτίμηση των παραμέτρων. Τέλος, γίνεται μια εφαρμογή αυτών των μεθόδων στο στατιστικό πρόγραμμα SPSS και εξετάζονται οι τιμές των μετοχών(Εθνική Τράπεζα,ALPHA BANK) για ένα συγκεκριμένο χρονικό διάστημα.

ΚΕΦΑΛΑΙΟ 1

Εισαγωγή

Η περιγραφή της διαχρονικής εξελίξεως και η πρόβλεψη της μελλοντικής συμπεριφοράς μιας στοχαστικής μεταβλητής αποτελεί ένα από τα σπουδαιότερα αντικείμενα μελέτης της οικονομετρίας, της επιχειρησιακής έρευνας του Marketing, της Μετεωρολογίας και πολλών άλλων επιστημών.

Οι πιο ενδεδειγμένες μέθοδοι για να μελετηθούν τα διαχρονικά φαινόμενα θεωρούνται στις μέρες μας οι στατιστικές μέθοδοι. Ο κλάδος της Στατιστικής ο οποίος μελετά τα φαινόμενα αυτά αποδίδεται με τον όρο Ανάλυση χρονολογικών σειρών(Time Series Analysis).Χρονολογική σειρά καλείται μια συλλογή παρατηρήσεων που γίνονται σε διαδοχικές χρονικές στιγμές. Πολλές χρονολογικές σειρές συναντούμε στην Οικονομία όπως για παράδειγμα σε τιμές μετοχών σε διαδοχικές ημέρες, η αξία των εξαγωγών μιας χώρας σε συνεχόμενους μήνες, τα κέρδη μιας επιχειρήσεως σε διαδοχικά χρόνια κλπ.

Πολλοί ήταν αυτοί που βοήθησαν στην ανάπτυξη πολλών στατιστικών μεθόδων για την ανάλυση των χρονολογικών σειρών όπως οι Box-Jenkins με την θεωρία τους που θα αναφέρουμε μετέπειτα αλλά και οι Hannan, Anderson, Kendall, Chatfield , Nelson κ.α. Έτσι οι σύγχρονοι ερευνητές έχουν στην διάθεση τους όλα τα εφόδια για να μελετήσουν με επιτυχία μια χρονολογική σειρά που θα του παρουσιασθεί στην πράξη.

1.1 ΕΝΝΟΙΑ ΧΡΟΝΟΛΟΓΙΚΗΣ ΣΕΙΡΑΣ

ΟΡΙΣΜΟΣ

Χρονοσειρά(time series) είναι ένα σύνολο παρατηρήσεων $y_1, y_2, ..., y_t$ όπου ο δείκτης Τ παριστάνει ισαπέχοντα χρονικά σημεία ή διαστήματα. Τα χρονικά διαστήματα μπορεί να είναι(έτος, μήνας, ημέρα, εβδομάδα, ώρα, κ.α).

Οι παρατηρήσεις $y_1, y_2, ..., y_t$ είναι συγκεκριμένες τιμές των τυχαίων μεταβλητών Y_1, Y_2, Y_T και είναι μέρος μόνο μιας άπειρης ακολουθίας τυχαίων μεταβλητών .Οι παρατηρήσεις αναφέρονται στην έννοια του δείγματος ενώ οι τυχαίες μεταβλητές στην έννοια του πληθυσμού.

Απαραίτητη συνθήκη για τη μελέτη χρονολογικών σειρών είναι η ύπαρξη δεδομένων(data).Δεδομένα χρονολογικών σειρών συναντούμε σε πολλές επιστήμες(οικονομικές, κοινωνικές, φυσική, ιατρική κ.α) Μερικά παραδείγματα είναι οι μηνιαίες πωλήσεις μιας επιχείρησης, οι τιμές ενός αγαθού ανά τρίμηνο, οι δείκτες των μετοχών στο ΧΑΑ. Βασικό χαρακτηριστικό κάθε χρονολογικής σειράς είναι η εξάρτηση μεταξύ των διαδοχικών τιμών της.

Αντικείμενο μελέτης του κλάδου των χρονολογικών σειρών είναι η φύση της αλληλεξάρτησης που υπάρχει μεταξύ των παρατηρήσεων και χωρίζεται σε δύο μέρη. Το πρώτο περιλαμβάνει την ανάλυση των ιδιοτήτων της σειράς έτσι ώστε να προσδιοριστούν τα χαρακτηριστικά που διέπουν τη συμπεριφορά της. Αυτό γίνεται με τη χρονική όσο και τη φασματική ανάλυση (spectral analysis). Το δεύτερο μέρος με το οποίο θα ασχοληθούμε και περισσότερο αφορά τα υποδείγματα χρονολογικών σειρών (time series models).

1.2 ΧΡΟΝΟΛΟΓΙΚΕΣ ΣΕΙΡΕΣ ΚΑΙ ΠΡΟΒΛΕΨΕΙΣ

Αντικειμενικός σκοπός της μελέτης χρονολογικών σειρών είναι η χρησιμοποίησή τους στη διενέργεια προβλέψεων. Η πρόβλεψη των μελλοντικών τιμών μιας μεταβλητής μπορεί να γίνει με διάφορες μεθόδους. Οι μέθοδοι αυτοί διαφέρουν ως προς την πολυπλοκότητα, την ταχύτητα, το κόστος υπολογισμού τους καθώς επίσης και από τη διαθεσιμότητα των απαραίτητων δεδομένων.

Σε γενικές γραμμές, οι μέθοδοι πρόβλεψης(forecast methods) διαχωρίζονται σε υποκειμενικές ή ποιοτικές(subjective or qualative) και σε αντικειμενικές ή ποσοτικές(objective or quantitative).Οι υποκειμενικές μέθοδοι πρόβλεψης, οι οποίες δεν θα μας απασχολήσουν ιδιαίτερα ,γίνονται κυρίως από έμπειρους επιστημονικούς αναλυτές οι οποίοι χρησιμοποιούν τις γνώσεις τους σε συνδυασμό με την κρίση τους, για την ασφαλή διεξαγωγή συμπερασμάτων. Πρακτικά ,δεν δίνουν βάρος σε μαθηματικές ή στατιστικές μεθόδους.

Μια αρκετά διαδεδομένη ποιοτική μέθοδος είναι η μέθοδος των Δελφών, η οποία χρησιμοποιείται αρκετά στις επιχειρήσεις και βασίζεται στη συγκέντρωση πληροφοριών από ειδικές ομάδες εμπειρογνωμόνων.

Αντιθέτως ,οι αντικειμενικές μέθοδοι πρόβλεψης βασίζονται σε κάποιο μαθηματικό ή στατιστικό υπόδειγμα και σε ποσοτικά δεδομένα (model based forecasts). Τα υποδείγματα αυτά χωρίζονται σε δύο μεγάλες κατηγορίες:

1)Τα αιτιατά(causal)
2)Τα μη αιτιατά(non-causal)

ΑΙΤΙΑΤΑ ΥΠΟΔΕΓΜΑΤΑ ΠΡΟΒΛΕΨΗΣ

Με τα υποδείγματα αυτά κάνουμε προβλέψεις μιας μεταβλητής με βάση την οικονομική και στατιστική σχέση που συνδέει τη μεταβλητή μας με άλλες που σχετίζονται μαζί της. Τέτοια υποδείγματα είναι τα οικονομετρικά υποδείγματα. Για τη διενέργεια προβλέψεων με τα οικονομετρικά ακολουθούνται τα εξής βήματα:

1. Θα πρέπει να προσδιοριστεί το οικονομικό υπόδειγμα που συνδέει την εξαρτημένη μεταβλητή με τις άλλες ερμηνευτικές (ανεξάρτητες μεταβλητές).

2. Χρειάζεται να γίνει εξειδίκευση του κατάλληλου στατιστικού που εκφράζει την οικονομική σχέση των μεταβλητών και να
εκτιμάται με τις γνωστές οικονομετρικές μεθόδους.
3. Τέλος,γίνεται η εκτίμηση του επιλεγμένου υποδείγματος και
στη συνέχεια γίνονται οι προβλέψεις βάση των εκτιμήσεων.

ΜΗ ΑΙΤΙΑΤΑ ΥΠΟΔΕΙΓΜΑΤΑ ΠΡΟΒΛΕΨΗΣ ΥΠΟΔΕΙΓΜΑΤΑ ΧΡΟΝΟΛΟΓΙΚΩΝ ΣΕΙΡΩΝ

Στα υποδείγματα χρονολογικών σειρών η πρόβλεψη στηρίζεται αποκλειστικά και μόνο στις προηγούμενες τιμές της ίδιας χρονολογικής σειράς που θέλουμε να προβλέψουμε. Δηλαδή, προβλέπουμε τη μελλοντική συμπεριφορά μιας χρονοσειράς όχι σε συνάρτηση άλλων σειρών αλλά εξετάζοντας την προηγούμενη συμπεριφορά της δηλαδή το “ιστορικό’’ της.

Τα υποδείγματα χρονολογικών σειρών χωρίζονται σε καθοριστικα υποδειγματα (deterministic models) τα οποία βασίζονται σε απλές μαθηματικές μορφές (υποδείγματα κινητών μέσων όρων, εκθετικής εξομάλυνσης και τάσης) αλλά και σε στοχαστικά υποδείγματα (stochastic models) όπως το υπόδειγμα μορφής Box-Jenkins.Με κάποια από τα παραπάνω υποδείγματα θα ασχοληθούμε διεξοδικά στη συνέχεια.

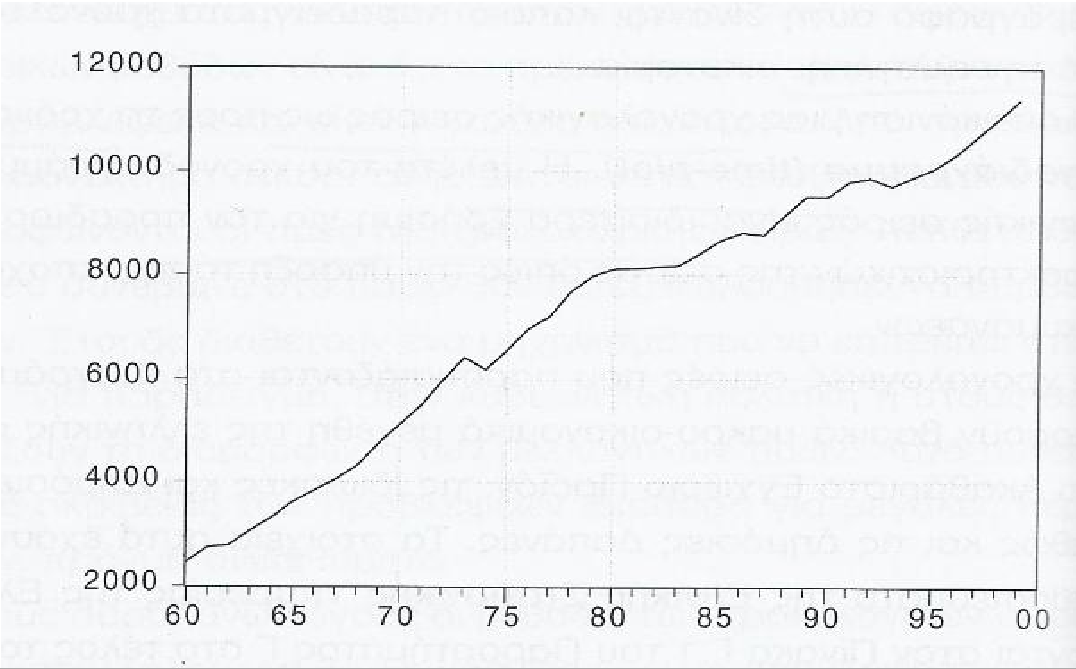
Τα πλεονεκτήματα των χρονολογικών υποδειγμάτων σε σχέση με τα οικονομετρικά είναι πως είναι έχουν χαμηλότερο κόστος διενέργειας προβλέψεων και είναι λιγότερο πολύπλοκα. Αντιθέτως, βασικό τους μειονέκτημα είναι πως δεν στηρίζονται σε κάποια θεωρία που να εξηγεί πώς διαμορφώνονται οι τιμές της χρονολογικής σειράς. Θεωρούν δηλαδή πώς αυτό που συνέβαινε στο παρελθόν θα εξακολουθήσει να συμβαίνει και στο μέλλον. Δεν υπάρχει ένας μηχανισμός που να επιτρέπει τυχόν μεταβολές στη διαμόρφωση των μελλοντικών τιμών πράγμα που συνεπάγεται μείωση της ακρίβειας των προβλέψεων ιδιαίτερα για μακροχρόνιες περιόδους.

Για όλους τους παραπάνω λόγους οι μέθοδοι των χρονολογικών σειρών κρίνονται πιο κατάλληλες για βραχυχρόνιες προβλέψεις ενώ αντίθετα οι οικονομετρικές μέθοδοι είναι καταλληλότερες στη διενέργεια μακροχρόνιων προβλέψεων.

1.3 ΠΑΡΑΔΕΙΓΜΑΤΑ ΧΡΟΝΟΛΟΓΙΚΩΝ ΣΕΙΡΩΝ

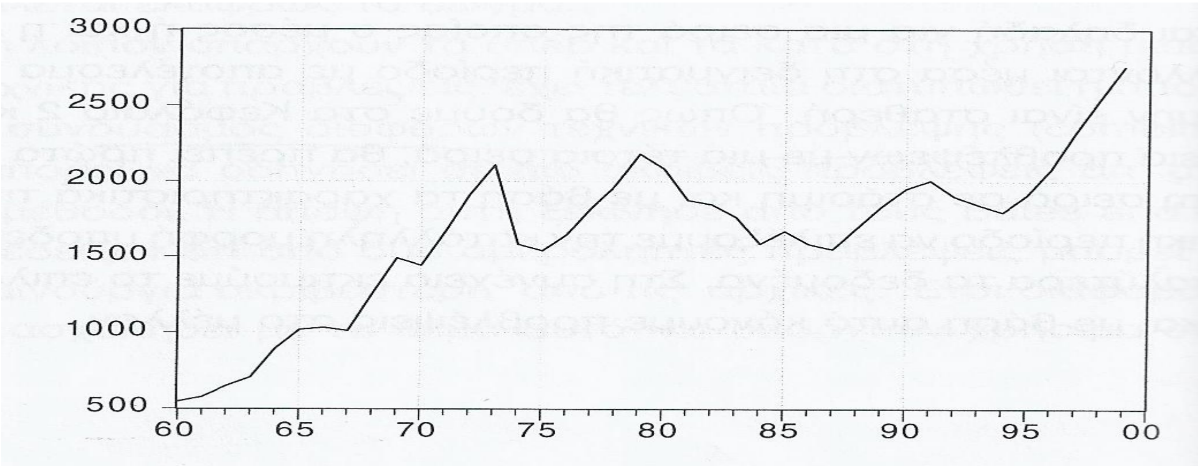
Η απεικόνιση μιας χρονολογικής σειράς ως προς το χρόνο ονομάζεται χρονοδιάγραμμα(time-plot).Η μελέτη του χρονοδιαγράμματος μιας χρονοσειράς είναι εξαιρετικά χρήσιμη για τον προσδιορισμό βασικών χαρακτηριστικών της σειράς όπως η ύπαρξη τάσης, εποχικότητας κ.α. Στο σημείο παραθέτουμε κάποια διαγράμματα χρονολογικών σειρών τα οποία αφορούν κάποια βασικά μακροοικονονομικά μεγέθη της Ελληνικής Οικονομίας όπως το ΑΕΠ(Ακαθάριστο Εγχώριο Προϊόν),τις Ιδιωτικές και Δημόσιες επενδύσεις καθώς και τις Δημόσιες Δαπάνες. Σημειώνεται πως όλα τα στοιχεία προέρχονται από δημοσιεύματα της Ελληνικής Στατιστικής Υπηρεσίας.

Στο παρακάτω διάγραμμα φανερώνεται η πορεία του ΑΕΠ της χώρας μας σε σταθερές τιμές 1988 την περίοδο 1960-1999.Παρατηρούμε πως σε γενικές γραμμές η πορεία είναι ανοδική με μικρές περιόδους μείωσης. Δηλαδή αναφερόμαστε σε μια μη στάσιμη χρονολογική σειρά(non-stationary) η οποία παρουσιάζει έντονα ανοδική τάση(trend).Με άλλα λόγια, πρόκειται για μια σειρά που ο μέσος και η διακύμανση μεταβάλλονται μέσα στη δειγματική περίοδο με αποτέλεσμα η κατανομή να μην είναι σταθερή. Όπως θα δούμε παρακάτω μετατρέπουμε πρώτα τη σειρά σε στάσιμη, έπειτα επιλέγουμε την κατάλληλη μορφή υποδείγματος που εξηγεί καλύτερα τα δεδομένα και τέλος ακολουθεί η εκτίμηση και η πρόβλεψη.



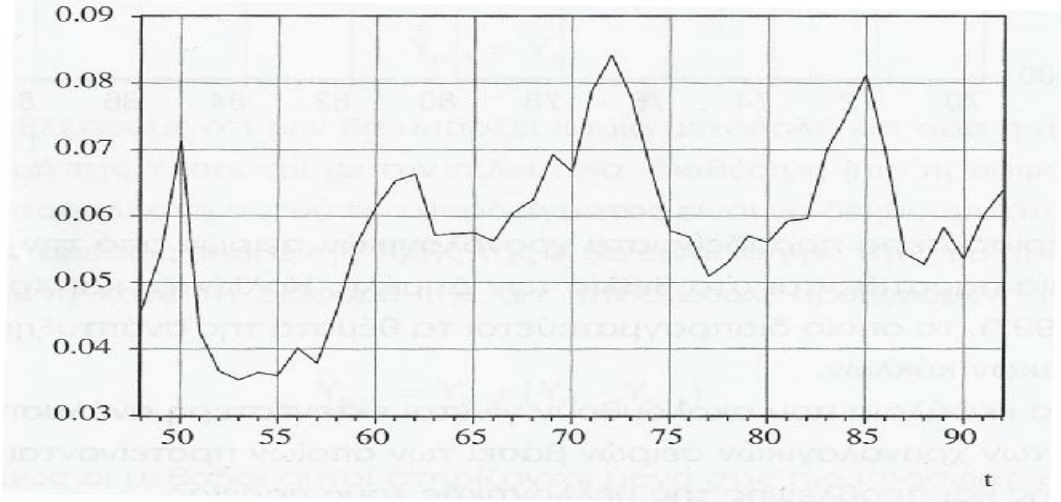
Διάγραμμα 1.1 Ακαθάριστο Εγχώριο Προϊόν (Δις.δρχ.-Σταθ.Τιμές 1988)

Στο επόμενο διάγραμμα φανερώνεται η πορεία των ιδιωτικών επενδύσεων στην Ελλάδα σε σταθερές τιμές το 1998.Η πορεία είναι μεν ανοδική αλλά όχι τόσο ομαλή όσο στο διάγραμμα του ΑΕΠ. Οι μεγαλύτερες αυξομειώσεις παρουσιάζονται στη δεκαετία 1970-1980 και οφείλονται σε διάφορες συγκυρίες της εποχής.



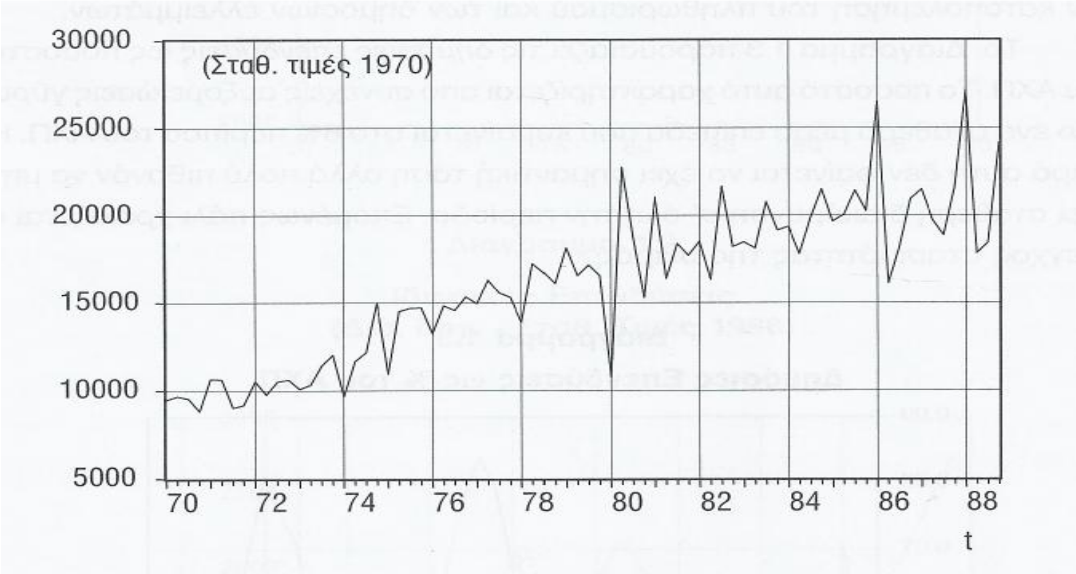
Διάγραμμα 1.2 Ιδιωτικές Επενδύσεις (Δις.δρχ-Σταθ.Τιμές 1988)

Στο επόμενο διάγραμμα παρουσιάζονται οι δημόσιες επενδύσεις ως ποσοστό του ΑΕΠ. Το ποσοστό αυτό χαρακτηρίζεται από συνεχείς αυξομειώσεις γύρω από ένα σταθερό μέσο επίπεδο που κυμαίνεται περίπου στο 6% του ΑΕΠ. Η συγκεκριμένη σειρά δεν φαίνεται να έχει σημαντική τάση ούτε σταθερή διακύμανση σε όλη την περίοδο κάτι που καθιστά απαραίτητο τον έλεγχο στασιμότητας.



Διάγραμμα 1.3 Δημόσιες Επενδύσεις ως % του ΑΕΠ

Το τελευταίο διάγραμμα απεικονίζει τις δημόσιες δαπάνες σε σταθερές τιμές 1970 της περιόδου 1970-1988.Στο συγκεκριμένο διάγραμμα υπάρχει τάση αλλά και εποχικότητα κάτι που αποτελεί δείγμα μη στασιμότητας στη συγκεκριμένη χρονολογική σειρά.



Διάγραμμα 1.4 Δημόσιες Δαπάνες

1.4 ΚΛΑΣΙΚΑ ΥΠΟΔΕΙΓΜΑΤΑ ΑΝΑΛΥΣΗΣ

Η κλασική μέθοδος ανάλυσης μιας χρονολογικής σειράς αποβλέπει στο διαχωρισμό αυτής της χρονολογικής σειράς στις επιμέρους συνιστώσες της που περιλαμβάνουν:

- Τη μακροχρόνια τάση(long-run trend)
- Την κυκλική Συνιστώσα (cyclical component)
- Την εποχιακή Συνιστώσα(seasonal component)
- Την τυχαία Συνιστώσα(random component)
- διαχωρισμός μιας χρονολογικής σειράς στις επιμέρους συνιστώσες της μπορεί να επιτευχθεί με 2 μη στοχαστικά υποδείγματα:
- Το πολλαπλασιαστικό υπόδειγμα
- Το προσθετικό υπόδειγμα

ΠΟΛΛΑΠΛΑΣΙΑΣΤΙΚΟ ΥΠΟΔΕΙΓΜΑ(MULTIPLICATIVE MODEL)

Το υπόδειγμα αυτό εκφράζεται ως:

$$Y_t = T_t C_t S_t R_t$$

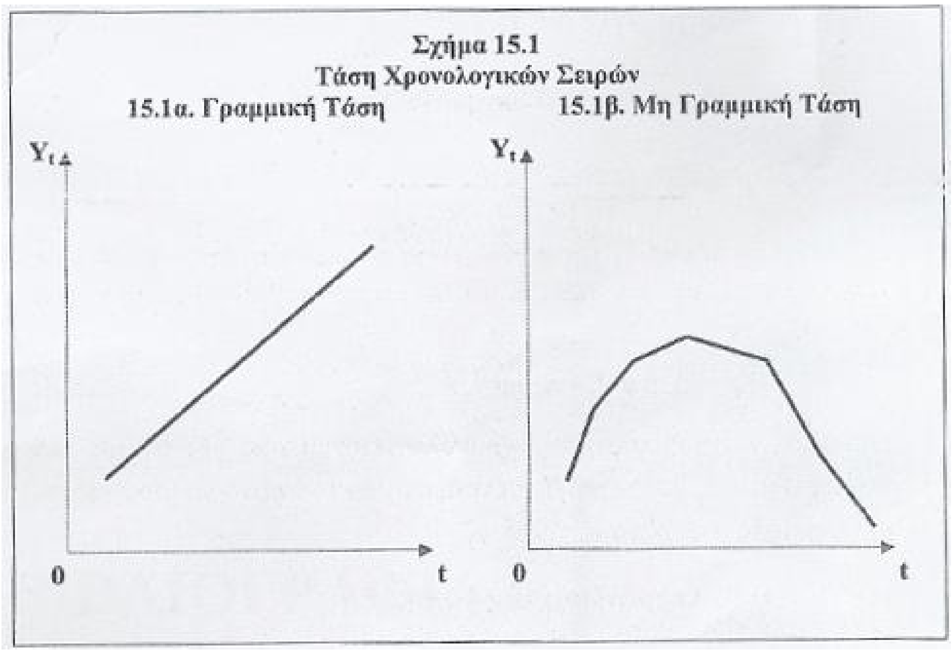
Όπου:
T=μακροχρόνια τάση
C=κυκλική συνιστώσα
S=εποχιακή συνιστώσα
R=τυχαία συνιστώσα

ΜΑΚΡΟΧΡΟΝΙΑ ΤΑΣΗ(SECULAR TREND)

Η τάση μιας χρονολογικής σειράς αναφέρεται σε μια σχετικά σταθερή συμπεριφορά ή κατεύθυνση της χρονολογικής σειράς η διάρκεια της οποίας είναι μεγαλύτερη του ενός έτους. Δηλαδή, η τάση περιγράφει την καθαρή επιρροή μακροχρόνιων παραγόντων της χρονοσειράς απαλλαγμένη από κυκλικές, εποχιακές και τυχαίες επιδράσεις .Για παράδειγμα, οι καταναλωτικές δαπάνες ενός καταναλωτή για ένα συγκεκριμένο αγαθό μπορεί μακροχρόνια να επηρεαστούν από διάφορους παράγοντες όπως εισόδημα, εποχικότητα του προϊόντος, καταναλωτικές συνήθειες, πολέμους κ.τ.λ. Όμως έχει αποδειχθεί εμπειρικά πως το εισόδημα του καταναλωτή είναι ο σημαντικότερος παράγοντας που επηρεάζει τις καταναλωτικές δαπάνες που μεταβάλλονται από περίοδο σε περίοδο με σταθερή αναλογία. Η καμπύλη που αναπαριστά την τάση μιας χρονικής σειράς δεν είναι απαραίτητα γραμμική.

ΜΕΤΡΗΣΗ ΤΗΣ ΤΑΣΗΣ

Προηγουμένως αναφερθήκαμε στο γεγονός πως η τάση μιας χρονοσειράς μπορεί να είναι γραμμική ή μη γραμμική. Ο πιο εύκολος τρόπος για να μετρήσουμε την τάση είναι το υπόδειγμα παλινδρόμησης το οποίο θα έχει εξαρτημένη μεταβλητή τις τιμές της χρονοσειράς και ανεξάρτητη μεταβλητή τη χρονική περίοδο. Η μαθηματική μορφή του υποδείγματος τάσης ποικίλει και η εκλογή της κατάλληλης μαθηματικής μορφής είναι αντικείμενο στατιστικού ελέγχου.



Διάγραμμα 1.5 Τάση Χρονολογικών Σειρών

Εμείς θα ασχοληθούμε με την εξίσωση γραμμικής τάσης (linear trend)και έχει την εξής μορφή: $Y_t = b_0 + b_1 t + e_t$

Η γραμμική εξίσωση της τάσης είναι η απλούστερη και εκφράζεται με απόλυτους αριθμούς Y_t =τάση της χρονολογικής σειράς για την περίοδο $t = (1, 2, \ldots N)$ και b_1 μετράει τη μέση μεταβολή της χρονοσειράς ανά μονάδα χρόνου .Καθαρά πληροφοριακά αξίζει να αναφέρουμε πως υπάρχουν ακόμα και κάποιες άλλες μορφές τάσεων όπως:

- Λογαριθμική τάση(Logarithmic trend)

  $Y_t = \ln b_0 + \ln b_1 t + e_t$
- Εκθετική τάση(Exponential trend)

  $Y_t = b_0 e^{b_1 t} e_t$
- Λογιστική τάση(logistic trend)

  $Y_t = \dfrac{b_0}{1 + b_1 e^{b_2 t}} e_t$
- Πολυωνυμιακή τάση(Polynomial trend)

  $Y_t = b_0 + b_1 t + b_2 t^2 + e_t$
- Gompertz τάση(Gompertz trend)

  $Y_t = b_0 b_1^{b_2^{X_t}}$

Η γραμμική εξίσωση της τάσης είναι η απλούστερη καθώς εκφράζεται σε απόλυτους αριθμούς, εκτιμάται με την OLS και έχει τη μορφή: $Y_t = b_0 + b_1 t + e_t$ όπου Y_t είναι η τάση της χρονολογικής σειράς για την χρονική περίοδο t, $t=(1,2,\ldots N)$ και το b_1 μετράει της χρονολογικής σειράς ανά μονάδα χρόνου.

ΑΠΟΜΟΝΩΣΗ ΤΗΣ ΤΑΣΗΣ

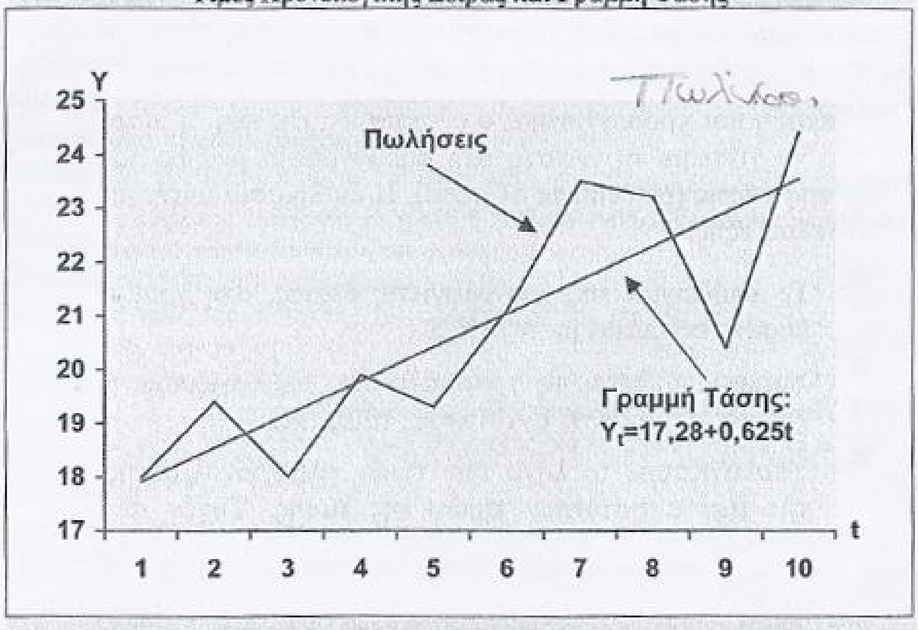
Για να εντοπίσουμε τις λοιπές συνιστώσες μιας χρονολογικής σειράς θα πρέπει η επίδραση της τάσης στις τιμές της χρονολογικής σειράς να απομακρυνθεί από τα στοιχεία του δείγματος. Η απομάκρυνση της τάσης στο πολλαπλασιαστικό υπόδειγμα επιτυγχάνεται διαιρώντας τις τιμές της χρονολογικής σειράς με τις αντίστοιχες της μακροχρόνιας τάσης.Έχουμε δηλαδή τον εξής τύπο: $\left\{\frac{Y_t}{T_t}\right\}=\left\{\frac{T_tC_tS_tR_t}{T_t}\right\}=\{C_tS_tR_t\}=Y^*$

Όπου Y^* είναι οι τιμές της χρονολογικής σειράς απαλλαγμένης της τάσης

Πίνακας 15.1
Χονδρικές Πωλήσεις Αγαθού Υ (Εκατ.Δρχ.)

Έτος	Y_t	t	Τάση: T_t	$Y^* = \{Y_t / T_t\}$
1985	18,0	1	17,91	1,01
1986	19,4	2	18,53	1,05
1987	18,0	3	19,16	0,94
1988	19,9	4	19,78	1,01
1989	19,3	5	20,41	0,95
1990	21,1	6	21,03	0,05
1991	23,5	7	21,66	1,08
1992	23,2	8	22,28	1,04
1993	20,4	9	22,91	0,89
1994	24,4	10	23,53	1,00

Διάγραμμα 15.1
Τιμές Χρονολογικής Σειράς και Γραμμή Τάσης



Διάγραμμα 1.6 Τιμές Χρονολογικής Σειράς και Γραμμή Τάσης

Στον παραπάνω πίνακα δίνονται στατιστικά που δείχνουν τις πωλήσεις ενός αγαθού την χρονική περίοδο 1985-1994.Η γραμμή τάσης είναι $Y_t = 17.28 + 0.625t$,η κλίση $b_1 = 0.625$ φανερώνει πως η μέση τιμή της τάσης αυξάνει με 0.625δρχ.το χρόνο. Οι τιμές της τάσης υπολογίζονται με τον εξής τρόπο:

$$T_{1985} = 17.28 + 0.625t = 17.28 + (0.625)(1) = 17.91$$

$$T_{1990} = 17.28 + 0.625t = 17.28 + (0.625)(6) = 21.03$$

$$T_{1993} = 17.28 + 0.625t = 17.28 + (0.625)(9) = 22.905$$

ΚΥΚΛΙΚΗ ΣΥΝΙΣΤΩΣΑ(CYCLICAL COMPONENT)

Η κυκλική συνιστώσα μιας χρονολογικής σειράς μετράει τις αποκλίσεις των τιμών γύρω από τη μακροχρόνια τάση της και έχει διάρκεια μεγαλύτερη του ενός έτους. Δεν αμφισβητείται πως η απομόνωση και η μέτρηση της κυκλικής συνιστώσας είναι αρκετά δύσκολη(με εξαίρεση τους εμπορικούς κύκλους).Ο εντοπισμός της και η μέτρησή της μπορεί να γίνει με τη μέθοδο του ποσοστού της τάσης(percentage of trend).Η διαδικασία της μεθόδου έχει ως εξής:

Το υπόδειγμα της χρονολογικής σειράς στη γραμμική του μορφή εκτιμάται με την OLS(ORDINARY LEAST SQUARES)γνωστή και ως μέθοδος ελαχίστων τετραγώνων.

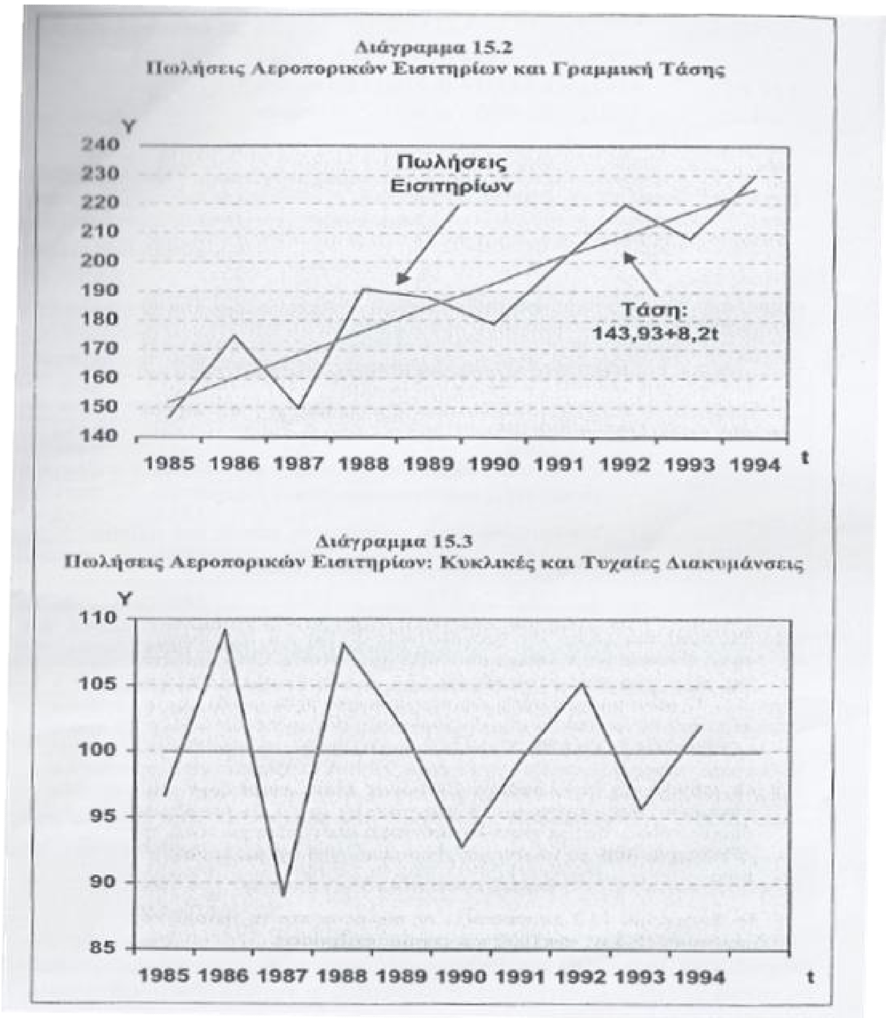
- Χρησιμοποιώντας τη γραμμή τάσης, υπολογίζουμε για κάθε χρονική περίοδο τις αντίστοιχες τιμές της.
- Υπολογίζουμε το λόγο των τιμών της χρονολογικής σειράς και των αντιστοίχων τιμών της τάσης. Αν βρεθούν τυχόν αποκλίσεις αυτής της διαίρεσης γύρω από τη μονάδα αυτό αποδίδεται στις εποχιακές και τυχαίες επιδράσεις στη χρονολογική σειρά.

Πίνακας 15.2
Πωλήσεις Αεροπορικών Εισιτηρίων της Ο.Α (χιλιάδες)

Έτος	Πωλήσεις Εισιτηρίων Y_t	Χρονική Περίοδος t	Τάση $T = 143{,}93 + 8{,}2t$	Ποσοστό $\{Y_t / T_t\}$
1985	147,0	1	152,1	96,7
1986	175,0	2	160,2	109,2
1987	150,0	3	168,4	89,1
1988	191,0	4	176,6	108,2
1989	188,0	5	184,7	101,8
1990	179,0	6	192,9	92,8
1991	200,0	7	201,8	99,5
1992	220,0	8	209,2	105,2
1993	208,0	9	217,4	95,7
1994	230,0	10	225,1	102,0

Πίνακας 1.1 Πωλήσεις Αεροπορικών Εισιτηρίων της Ο.Α σε χιλιάδες

Εκτιμούμε τη γραμμή τάσης που είναι: $Y_t = b_0 + b_1 t = 143.93 + 8.2t$ οι τιμές της τάσης για την περίοδο 1985-1994 δίνονται στην τρίτη στήλη του πίνακα. Στη συνέχεια υπολογίζουμε τις τιμές του λόγου της χρονολογικής σειράς και της τάσης που δίνονται στην τελευταία στήλη του πίνακα. Παρατηρούμε πως οι μισές τιμές της τελευταίας στήλης είναι χαμηλότερες του σημείου αναφοράς(100) και οι υπόλοιπες μισές υψηλότερες. Τέλος οι αποκλίσεις γύρω από το 100% της τάσης εντοπίζουν την ύπαρξη κυκλικών διακυμάνσεων στην χρονοσειρά κάτι το οποίο αποτελεί σημαντικό πρόβλημα ειδικότερα όταν το υπόδειγμα χρησιμοποιηθεί για μελλοντικές προβλέψεις. Τα 2 επόμενα διαγράμματα παρουσιάζουν τις πωλήσεις και τη γραμμή τάσης(1.7) και τις κυκλικές και τυχαίες επιδράσεις(1.8).



Διαγράμματα 1.7 Γραμμή Τάσης και 1.8 Κυκλικές και Τυχαίες Διακυμάνσεις

ΕΠΟΧΙΑΚΗ ΣΥΝΙΣΤΩΣΑ ΚΑΙ ΕΠΟΧΙΑΚΟΙ ΔΕΙΚΤΕΣ SEASONAL COMPONENT-SEASONAL INDICES

a.Η εποχιακή συνιστώσα θεωρείται η πηγή των εμπορικών και οικονομικών χρονολογικών σειρών. Η βασική της διαφορά με την κυκλική συνιστώσα είναι η διάρκεια(duration) καθώς μπορεί να εμφανισθεί με διάρκεια μιας εβδομάδας, ενός μήνα ή και ενός τριμήνου. Η εποχιακή συνιστώσα μετράται με τους εποχιακούς δείκτες(seasonal indices) που θα αναλύσουμε ευθύς αμέσως.
b. Οι εποχιακοί δείκτες μιας χρονοσειράς κατασκευάζονται με την παρακάτω διαδικασία:

ΒΗΜΑ 1

Το πρώτο στάδιο περιλαμβάνει την εξομάλυνση της χρονολογικής σειράς. Η εξομάλυνση επιτυγχάνεται με τον υπολογισμό των κινητών μέσων(moving averages).

Χαρακτηριστικό των κινητών μέσων στο πολλαπλασιαστικό υπόδειγμα είναι πως απομονώνουν τις εποχιακές και τυχαίες επιδράσεις από τη χρονολογική σειρά.

ΒΗΜΑ 2

Στο δεύτερο στάδιο υπολογίζουμε το λόγο των τιμών της χρονολογικής σειράς και των κινητών μέσων.Δηλαδή: $\left\{\frac{Y_t}{(KM)_t}\right\} = \left\{\frac{T_tC_tS_tR_t}{T_tC_t}\right\} = \{S_tR_t\}$

Όπου $(KM)_t = T_tC_t =$ κινητός μέσος. Ο κινητός μέσος μας δίνει την κυκλική συνιστώσα και τη συνιστώσα της τάσης(trend-cyclical component).

ΒΗΜΑ 3

Στο βήμα αυτό για κάθε χρονική περίοδο υπολογίζουμε τους αριθμητικούς μέσους των λόγων S_tR_t. Η διαδικασία αυτή απομακρύνει τη χρονολογική σειρά από τυχαίες αποκλίσεις και είναι ένα μέτρο της εποχιακής συνιστώσας. Στον παρακάτω πίνακα θα παραθέσουμε στατιστικά στοιχεία που αναφέρονται στο ποσοστό πληρότητας ξενοδοχείων της Αττικής την περίοδο 1991-1995.

Πίνακας 15.3
Ποσοστό Πληρότητας Ξενοδοχείων Αττικής (1991-1995)

Έτος	Τρίμηνο	Πληρότητα Ξενοδοχείων Y_t (%)	Κινητοί Μέσοι $\{(KM)_t\}$	Κεν/κός 4-τριμήνων Κινητός Μέσος $\{(T_t.C_t)\}$	$\{Y_t / (T_t.C_t)\} = S_t.R_t$ Δείκτης
1991:	1	0,561	...	...	...
	2	0,702	0,658	...	...
	3	0,800	0,661	0,660	1,213
	4	0,568	0,670	0,666	0,853
1992:	1	0,575	0,687	0,679	0,847
	2	0,738	0,697	0,692	1,067
	3	0,868	0,701	0,699	1,242
	4	0,605	0,701	0,666	0,853
1993:	1	0,594	0,677	0,684	0,869
	2	0,738	0,665	0,666	1,108
	3	0,729	0,672	0,669	1,090
	4	0,600	0,665	0,669	0,898
1994:	1	0,622	0,684	0,675	0,922
	2	0,708	0,692	0,688	1,156
	3	0,806	0,703	0,697	0,979
	4	0,632	0,735	0,719	0,879
1995:	1	0,665	0,752	0,743	0,895
	2	0,838	0,761	0,756	1,105
	3	0,873	...	...	...
	4	0,670	...	...	...

Πίνακας 1.2 Ποσοστό Πληρότητας Ξενοδοχείων Αττικής(1991-1995)

Βάσει των στοιχείων του πίνακα υπολογίζουμε αρχικά τους κινητούς μέσους.

$\{\text{KM}\}_1 = \{0.561\} + \{0.702\} + \{0.800\} + \{0.568\} / 4 = 0.658$

$\{\text{KM}\}_3 = \{0.800\} + \{0.568\} + \{0.575\} + \{0.738\} / 4 = 0.660$

Στη συνέχεια υπολογίζουμε τους κεντρικούς κινητούς μέσους

$$\{TC\}_1 = \{0.658\} + \{0.661\} / 2 = 0.660$$

$$\{TC\}_2 = \{0.661\} + \{0.670\} / 2 = 0.666$$

Έπειτα υπολογίζεται η εποχιακή συνιστώσα, δηλαδή διαιρούμε τις παρατηρήσεις της χρονολογικής σειράς με τις αντίστοιχες τιμές των κεντρικών κινητών μέσων

$$\{SI\}_1 = \{TC_1\} / \{Y_1\} = \{0.800\} / \{0.660\} = 1.213$$

…μέχρι το $\{SI\}_{16}$

Τέλος υπολογίζουμε τους εποχιακούς δείκτες με την εξής διαδικασία:
1)οι δείκτες SI της τελευταίας στήλης συγκεντρώνονται ανά τρίμηνο
2)υπολογίζουμε τους αριθμητικούς μέσους ανά τρίμηνο
3)αθροίζουμε τους τριμηνιαίους μέσους
4)διαιρούμε κάθε τριμηνιαίο μέσο με το σύνολο του και το αποτέλεσμα πολλαπλασιάζεται με εκείνο τον αριθμό που οδηγεί το μέσο εποχιακό δείκτη να ισούται με τη μονάδα.

Οι εποχιακοί δείκτες του παραπάνω πίνακα δίνονται στο παρακάτω διάγραμμα. Είναι εύκολα αντιληπτό πως το ποσοστό πληρότητας των ξενοδοχείων της Αττικής ήταν χαμηλότερα του ετησίου ποσοστού πληρότητας το πρώτο και το τρίτο τρίμηνο και μεγαλύτερο του ετησίου ποσοστού πληρότητας στα άλλα δύο τρίμηνα.

Πίνακας 15.4
Χρονολογικές Σειρές: Εποχιακοί Δείκτες

Έτος	Τρίμηνα				Σύνολο
	1	2	3	4	
1991	...	...	1,213	0,853	
1992	0,847	1,067	1,242	0,863	
1993	0,922	1,108	1,090	0,898	
1994	0,895	1,029	1,156	0,879	
1995	0,895	1,105	...	...	
Μέσος	0,883	1,077	1,175	0,873	4,008
Δείκτης	0,881	1,075	1,173	0,871	4,000

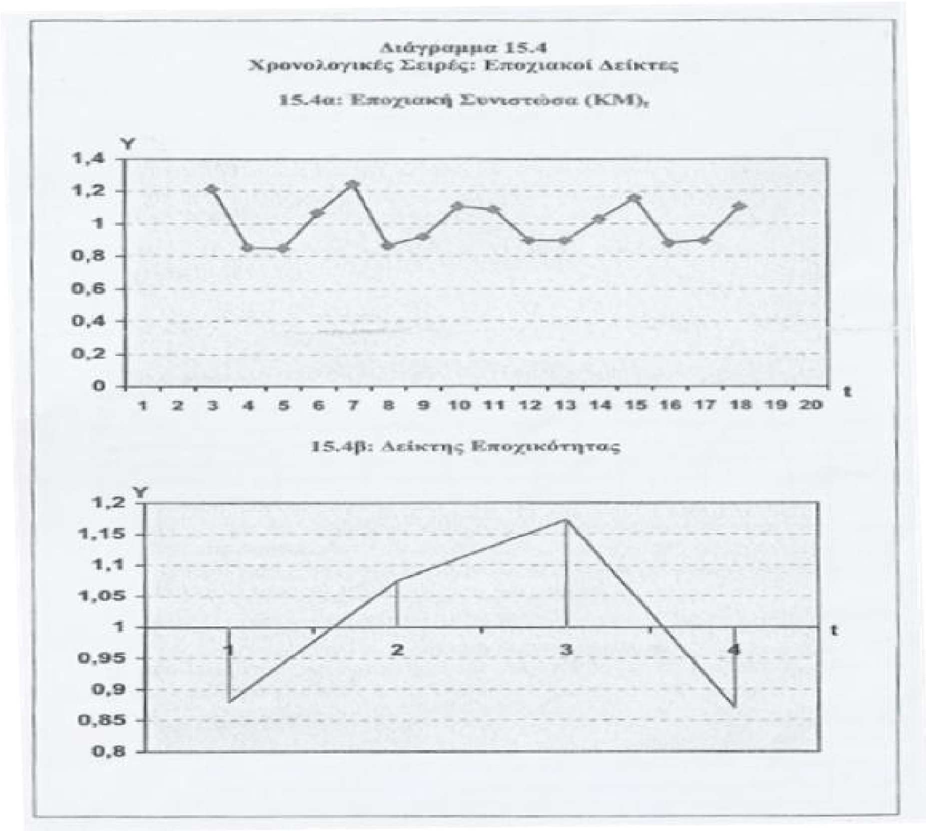
Πίνακας 1.3 Χρονολογικές Σειρές: Εποχιακοί Δείκτες

ΤΥΧΑΙΑ ΣΥΝΙΣΤΩΣΑ(RANDOM COMPONENT)

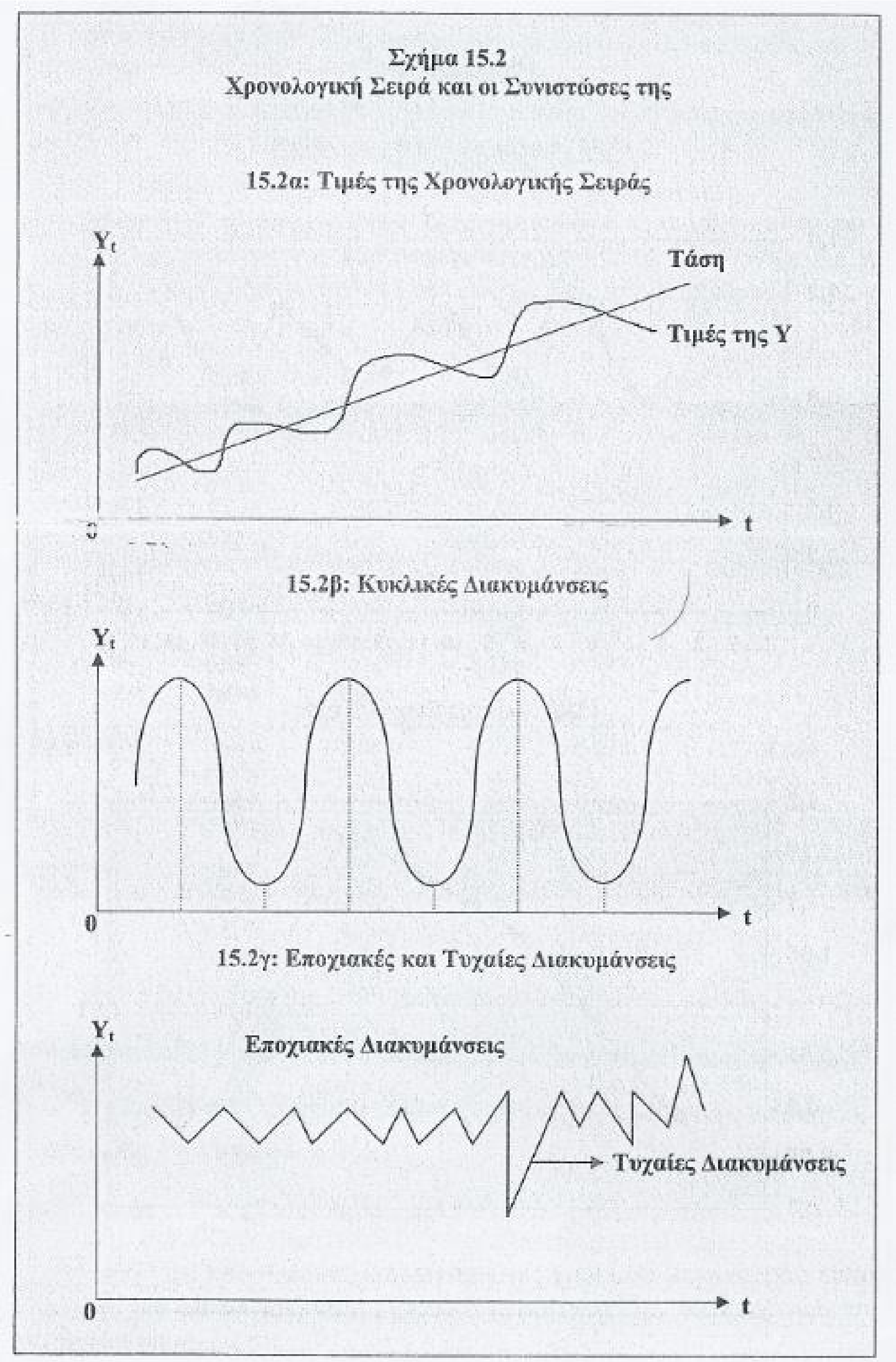
Η τυχαία συνιστώσα (απόκλιση) μιας χρονολογικής σειράς δεν επηρεάζεται από τις άλλες συνιστώσες και δεν επαναλαμβάνεται σε τακτά χρονικά διαστήματα. Υπολογίζεται ως εξής: $R_t = \left\{ \dfrac{Y_t}{C_t T_t S_t} \right\}$

ΤΟ ΠΡΟΣΘΕΤΙΚΟ ΥΠΟΔΕΙΓΜΑ(ADDITIVE MODEL)

Το προσθετικό υπόδειγμα μιας χρονολογικής σειράς εκφράζεται ως εξής: $Y_t = T_t + C_t + S_t + R_t$



Διάγραμμα 1.9 Χρονολογικές Σειρές:Εποχιακοί Δείκτες



Διάγραμμα 1.10 Χρονολογική Σειρά και οι Συνιστώσες της

ΜΕΘΟΔΟΙ ΕΞΟΜΑΛΥΝΣΗΣ ΧΡΟΝΟΛΟΓΙΚΩΝ ΣΕΙΡΩΝ

Οι μέθοδοι εξομάλυνσης(smoothing methods)μιας χρονολογικής σειράς έχουν ως σκοπό στην απομάκρυνση της τυχαίας συνιστώσας και υποβαθμίζουν την τάση αλλά και την εποχιακή και κυκλική συνιστώσα αντίστοιχα. Πρακτικά, οι επικρατέστερες μέθοδοι εξομάλυνσης χρονολογικών σειρών είναι:

- Μέθοδος κινητών μέσων(moving average method)
- Μέθοδος εκθετικής εξομάλυνσης(exponential smoothing method)

ΜΕΘΟΔΟΣ ΚΙΝΗΤΩΝ ΜΕΣΩΝ

Η διαδικασία εξομάλυνσης χρονολογικών σειρών με τη μέθοδο κινητών μέσων περιλαμβάνει τα παρακάτω στάδια.

1)Με δεδομένες τις παρατηρήσεις της χρονολογικής σειράς και γνωρίζοντας σε πόσες περιόδους επιθυμούμε τον υπολογισμό των κινητών μέσων υπολογίζουμε τα κινητά σύνολα.(moving totals)

2)Τα κινητά σύνολα διαιρούνται με τον αριθμό των παρατηρήσεων σε κάθε υποσύνολο, το πηλίκο των οποίων μας δίνει τους κινητούς μέσους. Στις χρονολογικές σειρές που έχουμε περιττό αριθμό παρατηρήσεων δεν υπολογίζονται κινητοί μέσοι για τις πρώτες και τελευταίες $\frac{(k-1)}{2}$ χρονικές περιόδους(κ= αριθμός παρατηρήσεων για κάθε κινητό σύνολο).

3)Υπολογίζουμε τη διαφορά ανάμεσα στις τιμές της χρονολογικής σειράς και των κινητών μέσων. Η διαφορά αυτή μετράει τις αποκλίσεις της χρονολογικής σειράς γύρω από τους κινητούς μέσους. Όλα αυτά φαίνονται και στον παρακάτω πίνακα.

Πίνακας 15.5
Πωλήσεις Βενζίνης: 1991Ι-1994IV

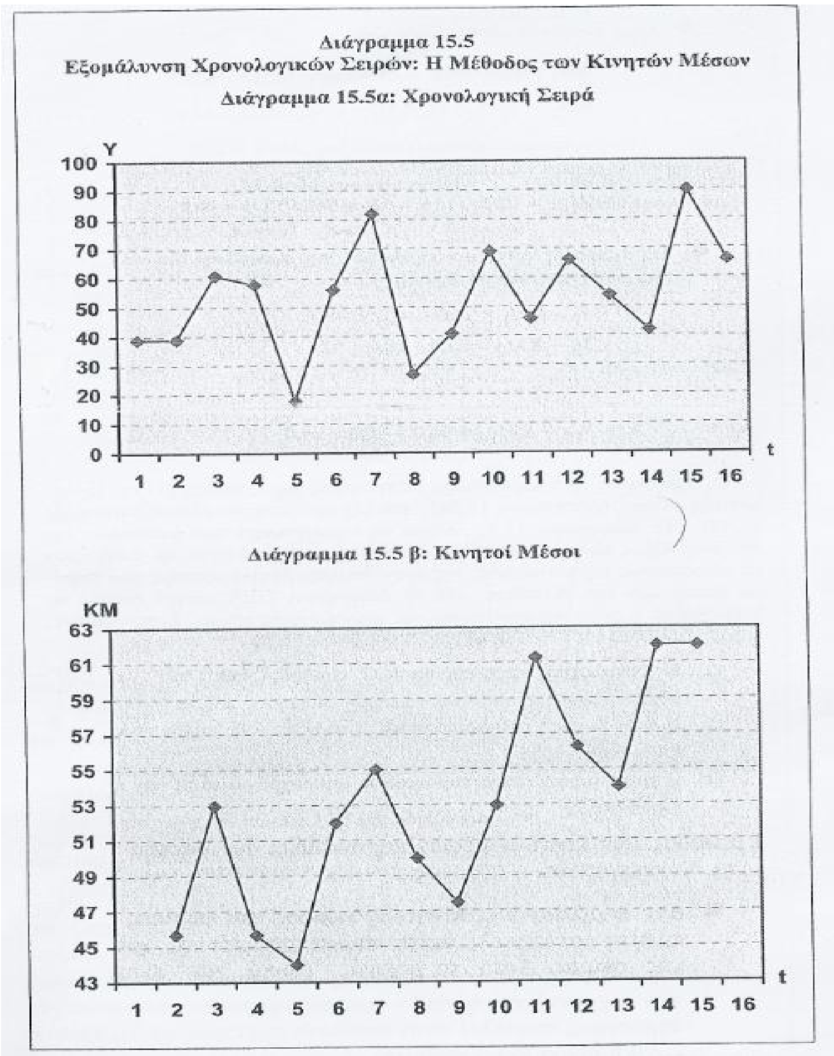
		Μέθοδος των Κινητών Μέσων			Μέθοδος Εκθετικής Εξομάλυνσης	
Έτος	Πωλήσεις Y_t	Κινητά Σύνολα ΚΣ	Κινητοί Μέσοι ΚΜ	Διαφορά (Y-KM)	S_t λ=0,2	λ=0,7
1991:1	39,0	...	...	...	39,0	39,0
2	39,0	137,0	45,7	-8,7	38,6	37,6
3	61,0	156,0	53,0	9,0	43,1	54,0
4	58,0	137,0	45,7	12,3	46,1	56,8
1992:1	18,0	132,0	44,0	-26,0	40,5	39,0
2	56,0	156,0	52,0	14,0	43,6	37,6
3	82,0	165,0	55,0	17,0	56,4	54,0
4	27,0	150,0	50,0	-23,0	46,4	56,8
1993:1	41,0	137,0	47,5	-4,7	47,3	40,6
2	69,0	159,0	53,0	16,0	50,1	60,6
3	46,0	184,0	61,3	-12,3	49,8	52,9
4	66,0	169,0	56,3	9,0	53,1	11,9
1994:1	54,0	162,0	54,0	0,0	53,3	56,4
2	42,0	186,0	62,0	-10,0	51,3	46,3
3	90,0	198,0	62,0	24,0	58,8	76,9
4	66,0	...	...	...	66,2	69,3

Πίνακας 1.4 Πωλήσεις Βενζίνης 1991-1994

Η μέθοδος των κινητών παρουσιάζει ένα σοβαρό μειονέκτημα. Πιο συγκεκριμένα δεν περιλαμβάνει κινητούς μέσους για μερικές από τις πρώτες και τελευταίες παρατηρήσεις της χρονολογικής σειράς. Το γεγονός αυτό συντελεί στην απώλεια πληροφοριών σχετικά με τη χρονολογική σειρά.

Στο πρώτο διάγραμμα φαίνεται η συμπεριφορά των πωλήσεων της βενζίνης, επειδή όμως παρουσιάζονται τυχαίες μεταβολές δεν μπορούμε να εντοπίσουμε τις συνιστώσες της χρονολογικής σειράς.

Στο δεύτερο διάγραμμα διαπιστώνουμε πως η χρονολογική σειρά παρουσιάζει κορυφές(peaks) στο τρίτο τρίμηνο κάθε έτους, πεδιάδες (valleys) πάλι στο πρώτο τρίμηνο κάθε έτους και πως γενικά η σειρά παρουσιάζει μακροχρόνια τάση των πωλήσεων της βενζίνης.



Διάγραμμα 1.11 Εξομάλυνση Χρονολογικών Σειρών: Η Μέθοδος των Κινητών Μέσων

ΕΚΘΕΤΙΚΗ ΜΕΘΟΔΟΣ ΕΞΟΜΑΛΥΝΣΗΣ

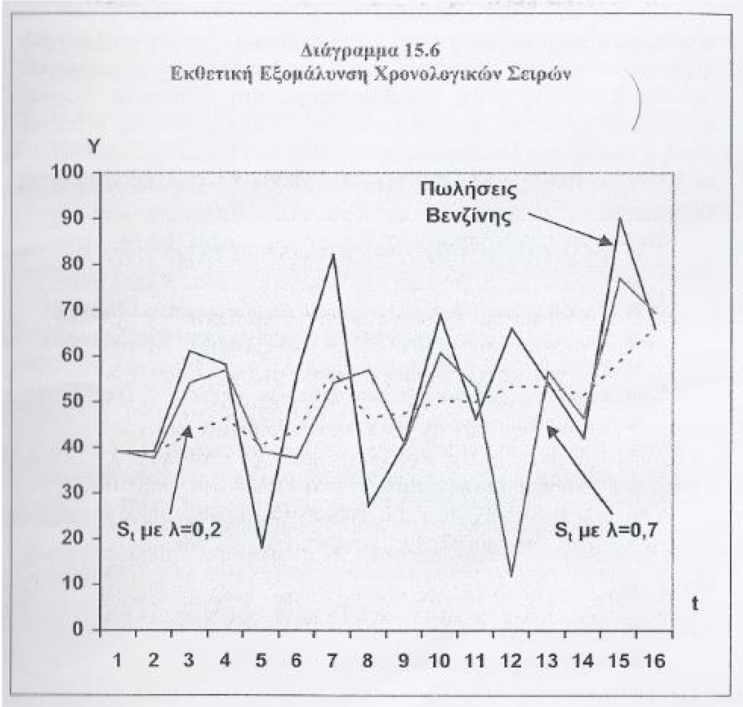
Η εκθετική μέθοδος εξομάλυνσης μιας χρονολογικής σειράς βασίζεται στην αρχή πως οι πρόσφατες τιμές της σειράς έχουν μεγαλύτερη βαρύτητα από εκείνες παλαιότερων περιόδων και δίνεται από την εξής σχέση $S_t = lY_t + (1-l)S_{t-1}$ για $0 \le l \le 1$ όπου

$Y_t =$ τιμή της χρονολογικής σειράς την περίοδο t

$S_t =$ εξομαλυσμένη τιμή της σειράς την περίοδο t

$l =$σταθμιστής (weight)

Στο τελευταίο διάγραμμα διαπιστώνεται η συμπεριφορά της σειράς στις αρχικές αλλά και στις τιμές εξομάλυνσης. Διαπιστώνουμε πως ο βαθμός εξομάλυνσης της σειράς εξαρτάται από την τιμή του συντελεστή εξομάλυνσης l (όταν $l = 0.2$ έχουμε καλύτερη εξομάλυνση από $l = 0.7$).



Διάγραμμα 1.12 Εκθετική Εξομάλυνση Χρονολογικών Σειρών

ΚΕΦΑΛΑΙΟ 2
Στοχαστικά Υποδείγματα Χρονολογικών σειρών

2.1 Εισαγωγή

Τα στοχαστικά υποδείγματα(stochastic models) βασίζονται στην ιδέα ότι μια χρονολογική σειρά της οποίας οι διαδοχικές τιμές συσχετίζονται σε μεγάλο βαθμό, μπορεί να θεωρηθεί ότι έχει παραχθεί από μια στοχαστική διαδικασία(stochastic process).Στα περισσότερα στατιστικά προβλήματα προσπαθούμε να εκτιμήσουμε τις ιδιότητες του πληθυσμού από το δείγμα. Στις χρονολογικές σειρές όμως δεν μπορούμε να έχουμε περισσότερες από μια παρατηρήσεις για κάθε μια μεταβλητή σε συγκεκριμένο χρόνο. Μπορούμε όμως να θεωρήσουμε τις παρατηρούμενες τιμές της χρονολογικής σειράς(observed time series) ως ένα δείγμα από ένα άπειρο πληθυσμό τέτοιων δειγμάτων, τα οποία θα μπορούσαν να είχαν παραχθεί από την ίδια στοχαστική διαδικασία. Η έννοια του πληθυσμού της Στατιστικής αντιστοιχεί στην έννοια της στοχαστικής διαδικασίας και η έννοια του δείγματος στην παρατηρούμενη σειρά. Ένα απλό παράδειγμα μιας στοχαστικής χρονολογικής σειράς y_t είναι να θεωρήσουμε ότι οι διαδοχικές μεταβολές των τιμών της είναι ανεξάρτητες τυχαίες μεταβλητές με μέσο το μηδέν, έχουν δηλαδή αυτή τη μορφή: $y_t - y_{t-1} = e_t$.Η μεγαλύτερη δυσκολία που αντιμετωπίζουμε στην ανάλυση χρονολογικών σειρών έχει να κάνει με την επιλογή του πιο κατάλληλου υποδείγματος. Αυτό επιτυγχάνεται με την εξέταση της δομής των ιστορικών δεδομένων της χρονολογικής σειράς με συγκεκριμένα στατιστικά μέτρα.

2.2 Στασιμότητα

Η στασιμότητα είναι πολύ σημαντική έννοια καθώς είναι απαραίτητη προϋπόθεση για τα περισσότερα εργαλεία της ανάλυσης χρονολογικών σειρών. Μια στοχαστική διαδικασία χαρακτηρίζεται ως στάσιμη όταν οι στατιστικές της ιδιότητες δεν επηρεάζονται από μια μεταβολή στην αρχή του χρόνου. Οι στατιστικές ιδιότητες των N παρατηρήσεων με αρχή $t \left(y_t, y_{t+1}, \mathbf{K}, y_{t+N-1} \right)$ είναι οι ίδιες με τις στατιστικές ιδιότητες των N παρατηρήσεων με αρχή την περίοδο $t+k \left(y_{t+k}, y_{t+k+1}, \mathbf{K}, y_{t+k+N-1} \right)$.

Μια χρονολογική σειρά θα είναι στάσιμη αν ο μέσος και η διακύμανση της δεν μεταβάλλονται με το χρόνο και η συνδιακύμανση μεταξύ των τιμών της σε δύο χρονικά σημεία εξαρτάται μόνο από την απόσταση ανάμεσα σε αυτά χρονικά σημεία και όχι από τον ίδιο το χρόνο. Για μια στάσιμη χρονολογική σειρά και για κάθε τιμή του t θα ισχύουν:

a) $E\left(y_t \right) = m_y$

b) $Var\left(y_t \right) = E\left[y_t - E\left(y_t \right) \right]^2 = s_y^2$

c) $Cov\left(y_t, y_{t+k} \right) = Cov\left(y_{t+m}, y_{t+m+k} \right) = g_k$

Οι δύο πρώτες συνθήκες δηλώνουν σταθερό μέσο και διακύμανση ενώ η τελευταία υποδηλώνει ότι η συνδιακύμανση μεταξύ δύο οποιονδήποτε τιμών y_t που απέχουν k περιόδους είναι συνάρτηση μόνο του k, δηλαδή της χρονικής υστέρησης ή προήγησης των δύο αυτών τιμών και ονομάζεται αυτοσυνδιακύμανση.

Επίσης, αν μια χρονολογική σειρά είναι στάσιμη, τότε θα έχει σταθερή κατανομή πυκνότητας πιθανότητας $f\left(y_t \right)$ για κάθε t και επομένως μια εκτίμηση του μέσου m_y και της διακύμανσης s_y^2 μπορεί να ληφθεί χρησιμοποιώντας το μέσο και τη διακύμανση αντίστοιχα του δείγματος των παρατηρήσεων της χρονολογικής σειράς:

$$\overline{y} = \frac{1}{N} \sum_{t=1}^{N} y_t \quad \text{και} \quad s_y^2 = \frac{1}{N} \sum_{t=1}^{N} \left(y_t - \overline{y} \right)^2$$

Στην οικονομία οι περισσότερες χρονολογικές σειρές που συναντάμε είναι μη στάσιμες αφού περιέχουν τάση, εποχικότητα και κυκλικές κυμάνσεις. Η ανάλυση αυτών των μη στάσιμων χρονολογικών σειρών είναι πολύ δύσκολη αλλά μπορούν με κατάλληλες τεχνικές να μετατραπούν σε στάσιμες και να μελετηθούν με τις μεθόδους ανάλυσης στάσιμων χρονολογικών σειρών που είναι πολύ απλούστερες. Για παράδειγμα, πολλές χρονολογικές σειρές μετατρέπονται σε στάσιμες , αφαιρώντας την τάση ή παίρνοντας διαδοχικά πρώτες διαφορές στα δεδομένα στοιχεία.

2.3 Αυτοσυνδιακύμανση και Αυτοσυσχέτιση

Όπως έχουμε αναφέρει η αυτοσυνδιακύμανση (autocovariance) μέτρα τη συνδιακύμανση μεταξύ δύο παρατηρήσεων της ίδιας χρονολογικής σειράς που βρίσκονται σε κάποια απόσταση μεταξύ τους. Έτσι, η αυτοσυνδιακύ-μανση μεταξύ y_t και y_{t+k} που απέχουν k χρονικές περιόδους συμβολίζεται με g_k και ορίζεται ως:

$$g_k = Cov\left(y_t, y_{t+k}\right) = E\left[y_t - E\left(y_t\right)\right]\left[y_{t+k} - E\left(y_{t+k}\right)\right]$$

Για μια στάσιμη χρονολογική σειρά, όπως είδαμε προηγουμένως θα έχουμε $E\left(y_t\right) = E\left(y_{t+k}\right) = m_y$ οπότε

α) $g_k = Cov\left(y_t, y_{t+k}\right) = E\left(y_t - m_y\right)\left(y_{t+k} - m_y\right)$

β) $g_0 = Cov\left(y_t, y_t\right) = s_y^2$, $k = 0$ (αυτοσυνδιακύμανση μηδενικής υστέρησης)

Πιο χρήσιμος είναι ο συντελεστής συσχέτισης μεταξύ δύο παρατηρήσεων y_t και y_{t+k} που απέχουν k χρονικές περιόδους και ορίζεται ως ο συντελεστής αυτοσυσχέτισης (autocorrelation coefficient):

$$r_k = \frac{Cov\left(y_t, y_{t+k}\right)}{\sqrt{Var(y_t)}\sqrt{Var(y_{t+k})}}$$

Αν μιλάμε για στάσιμη χρονολογική σειρά, τότε η διακύμανση δεν μεταβάλλεται με το χρόνο και άρα $Var\left(y_t\right) = Var(y_{t+k}) = s_y^2$

Συνεπώς και με την βοήθεια των σχέσεων α) και β) ο συντελεστής αυτοσυ-σχέτισης απλοποιείται ως:

$$r_k = \frac{Cov\left(y_t, y_{t+k}\right)}{Var\left(y_t\right)} = \frac{g_k}{g_0}$$

Οι σχέσεις g_k και r_k αναφέρονται στις θεωρητικές τιμές των αυτοσυνδια-κυμάνσεων και των συντελεστών αυτοσυσχέτισης της στοχαστικής διαδικα-σίας y_t.Στην πράξη όμως χρησιμοποιούμε ένα πεπερασμένο δείγμα παρατη- ρήσεων $y_1, y_2, \mathbf{K}\, y_N$ από το οποίο λαμβάνουμε εκτιμήσεις των αληθινών στον πληθυσμό αυτοσυνδιακυμάνσεων και αυτοσυσχετίσεων. Συμβολίζοντας με $\bar{y}$ το μέσο του δείγματος των N παρατηρήσεων θα έχουμε:

$$g_k = \frac{1}{N}\sum_{t=1}^{N-k}\left(y_t - \bar{y}\right)\left(y_{t+k} - \bar{y}\right)$$

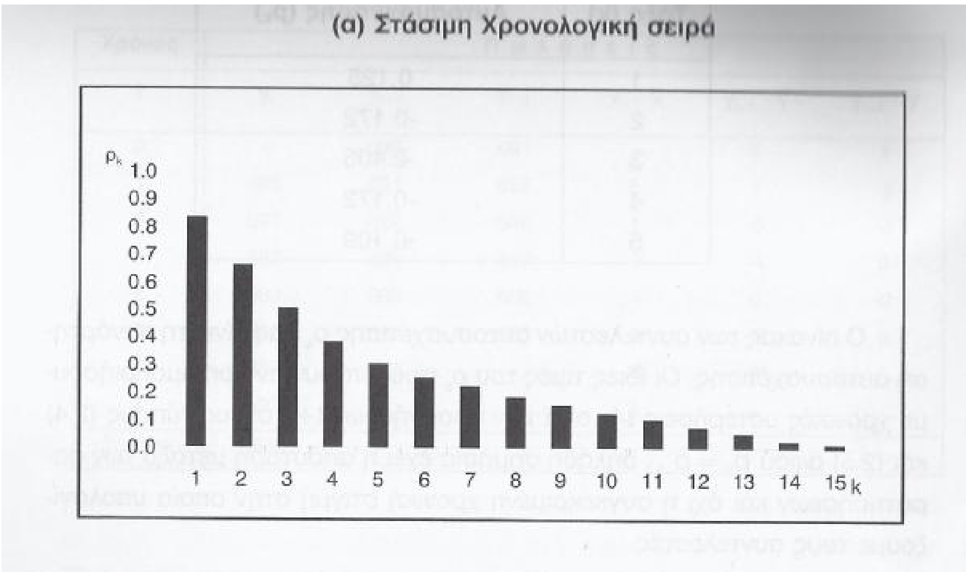
$$r_k = \frac{g_k}{g_0} = \frac{\sum_{t=1}^{N-k}\left(y_t - \bar{y}\right)\left(y_{t+k} - \bar{y}\right)}{\sum_{t=1}^{N}\left(y_t - \bar{y}\right)^2}$$

Οι εκτιμήσεις των r_k με βάση την παραπάνω σχέση αποτελούν τους δειγματικούς συντελεστές αυτοσυσχέτισης. Αν για την χρονολογική σειρά έχουμε μεγάλο αριθμό παρατηρήσεων τότε η συνάρτηση αυτοσυσχέτισης r_k θα πλησιάζει κατά προσέγγιση την αληθινή αυτοσυσχέτιση.

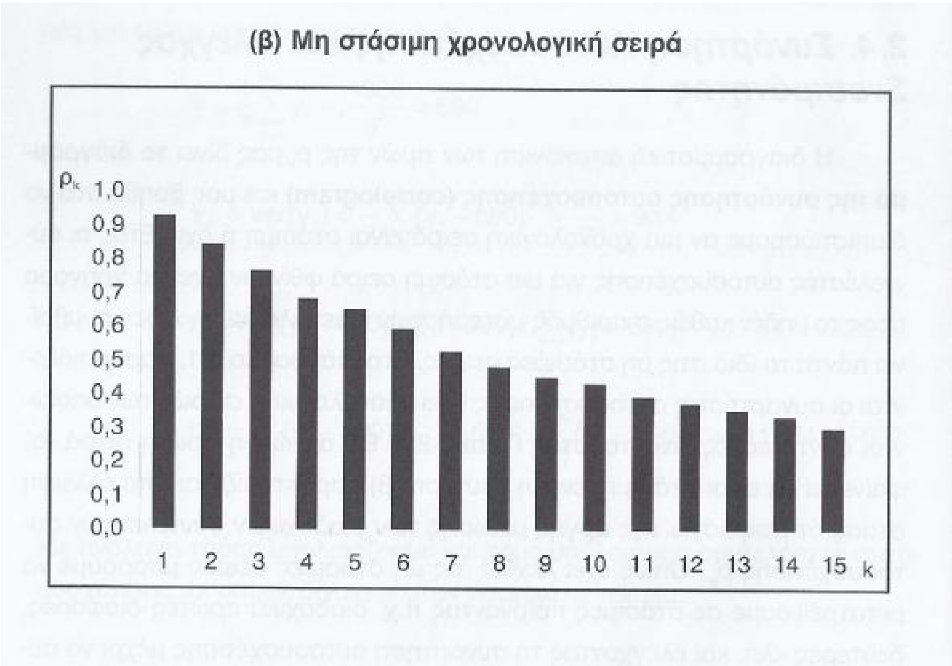
Οι τιμές που παίρνει το r_k είναι μεταξύ $-1 \leq r_k \leq 1$ και επειδή η συνάρτηση αυτοσυσχέτισης είναι συμμετρική ($r_k = r_{-k}$) εξετάζουμε μόνο τις θετικές τιμές του k. Οπότε αν έχουμε δύο παρατηρήσεις που απέχουν k χρονικές περιόδους έχουν μεγάλη σχέση μεταξύ τους τότε η τιμή του ρ_k θα είναι κοντά στη μονάδα. Η συνάρτηση αυτοσυσχέτισης r_k συμβολίζεται και ως ACF.

2.4 Συνάρτηση Αυτοσυσχέτισης και Έλεγχος Στασιμότητας

Η διαγραμματική απεικόνιση των τιμών της r_k μας δίνει το διάγραμμα συνάρτησης αυτοσυσχέτισης (correlogram) και έτσι μπορούμε να διαπιστώσουμε αν μια χρονολογική σειρά είναι στάσιμη ή όχι. Για μια στάσιμη χρονοσειρά όπως φαίνεται και παρακάτω οι συντελεστές αυτοσυσχέτισης φθίνουν γρήγορα προς το μηδέν καθώς μεγαλώνει ο αριθμός των υστερήσεων k ενώ αντίθετα δεν συμβαίνει το ίδιο στις μη στάσιμες χρονολογικές σειρές. Στο σημείο αυτό παραθέτουμε 2 διαγράμματα το ένα στασιμότητας και το άλλο μη στασιμότητας. Στο πρώτο παρατηρούμε πως η σειρά είναι στάσιμη ενώ στο δεύτερο η σειρά είναι μη στάσιμη λόγω της αργής μείωσης των διαδοχικών συντελεστών αυτοσυσχέτισης. Οι μη στάσιμες σειρές μετατρέπονται σε στάσιμες παίρνοντας πρώτες ή δεύτερες διαφορές ή με τη χρήση λογαρίθμου.



Διάγραμμα 2.1 Στάσιμη Χρονολογική Σειρά



Διάγραμμα 2.2 Μη Στάσιμη Χρονολογική Σειρά

ΕΛΕΓΧΟΣ ΤΟΥ BARTLETT(Barlett's Test)

Ο έλεγχος του Bartlett βασίζεται στην υπόθεση ότι η χρονολογική σειρά είναι στάσιμη και οι συντελεστές αυτοσυσχέτισης του δείγματος ακολουθούν προσεγγιστικά την κανονική κατανομή με μηδενικό μέσο και διακύμανση $\frac{1}{N}$ ($N =$μέγεθος του δείγματος).Επομένως για τον έλεγχο της μηδενικής υπόθεσης έχουμε: $H_0 : r_k = 0$ και η εναλλακτική είναι: $\mathrm{H}_1 : r_k \neq 0$ τα οποία συγκρίνουμε με την τιμή του στατιστικού t : $t = \frac{\hat{r}_k}{\sqrt{\frac{1}{\mathrm{N}}}} = \sqrt{\mathrm{N}}\,\hat{r}_k$.Ο έλεγχος Bartlett ισχύει για μεγάλα δείγματα και εξετάζει μεμονωμένα κάθε συντελεστή αυτοσυσχέτισης. Για να ελέγξουμε την υπόθεση ότι από κοινού ένας αριθμός συντελεστών διαφέρει ή όχι από το μηδέν χρησιμοποιούμε το στατιστικό κριτήριο Q των Box-Pierce.

Β. Q ΣΤΑΤΙΣΤΙΚΗ(Box-Pierce Test)

Ο έλεγχος στασιμότητας μιας χρονολογικής σειράς με τη στατιστική Box-Pierce χρησιμοποιείται για τον έλεγχο της συνδυαστικής υπόθεσης ότι όλοι οι συντελεστές αυτοσυσχέτισης είναι μηδέν ορίζεται ως εξής: $Q = N\sum_{k=1}^{m} \hat{r}_k^2$

Όπου $N =$ μέγεθος του δείγματος και $\mathrm{m} =$ μήκος χρονικής υστέρησης .Η στατιστική Q ακολουθεί την X^2 κατανομή με m βαθμούς ελευθερίας. Αν $Q > X^2(a,m)$ η χρονολογική σειρά δεν είναι στάσιμη και αντιστρόφως.

Γ.ΣΤΑΤΙΣΤΙΚΗ ΤΩΝ BOX-LJUNG

Η στατιστική των Box-Ljung αν και ακολουθεί την $X^2(a,m)$ δίνει καλύτερα αποτελέσματα από την Q στατιστική όταν εφαρμόζεται σε μικρά δείγματα και ορίζεται ως εξής: $LB = N(N+2)\sum_{k=1}^{m}\left(\frac{\hat{r}_k}{N-k}\right) \longrightarrow X_m^2$

Αν $LB > X^2(a,m)$ τότε η υπόθεση της στασιμότητας της χρονολογικής σειράς απορρίπτεται.

2.5 Μερική Αυτοσυσχέτιση

Μια άλλη συνάρτηση που θα εξετάσουμε και χρησιμοποιείται πολύ στη μελέτη χρονολογικών σειρών είναι η συνάρτηση μερικής αυτοσυσχέτισης. Γενικότερα, ο συντελεστής μερικής αυτοσυσχέτισης $r_{xy:z_1,\dots,z_v}$ μετρά τη συσχέτιση των δύο μεταβλητών x και y όταν έχει αφαιρεθεί η επίδραση που ασκούν άλλες μεταβλητές όπως οι $z_1,\dots,z_v$ πάνω σε αυτές. Στις χρονολογικές σειρές ο συντελεστής μερικής αυτοσυσχέτισης μεταξύ

y_t και y_{t+k} ορίζεται ως ο συντελεστής συσχέτισης μεταξύ τους όταν έχουν ληφθεί υπόψη οι συσχετίσεις όλων των ενδιάμεσων τιμών $y_{t+1}, y_{t+2}, ..., y_{t+k-1}$.

Η παλινδρόμηση μπορεί να εξηγήσει καλύτερα την έννοια της μερικής αυτοσυσχέτισης. Θεωρούμε την παλινδρόμηση της y_t πάνω στις y_{t-1} και y_{t-2} : $y_t = j_{12} y_{t-1} + j_{22} y_{t-2} + e_t$

Ο αριστερά δείκτης του j φανερώνει τη χρονική υστέρηση της μεταβλητής και αντίστοιχα ο δεξιά δείκτης δηλώνει τη μέγιστη τάξη παλινδρόμησης (ισχύουν για την y_{t-1}). Ο συντελεστής της y_{t-2} μετρά το συντελεστή μερικής αυτοσυσχέτισης δεύτερης τάξεως (j_{22}) διότι αυτός δείχνει τη συσχέτιση μεταξύ των y_t και y_{t-2} όταν έχει συμπεριληφθεί στην παλινδρόμηση η ενδιάμεση y_{t-1}. Γενικότερα, ο συντελεστής μερικής αυτοσυσχέτησης p τάξεως συμβολίζεται με το γράμμα j_{pp} και είναι ο συντελεστής του y_{t-p} στην παλινδρόμηση. Δηλαδή : $y_t = j_{1p} y_{t-1} + j_{2p} y_{t-2} + ... + j_{pp} y_{t-p} + e_t$

σημειώνουμε εδώ πως ο συντελεστής πρώτης τάξης j_{11} ταυτίζεται με τον απλό συντελεστή αυτοσυσχέτισης r_1.Πρακτικά υποδείγματα όπως τα παραπάνω τα εκτιμούμε ξεκινώντας μια χρονική υστέρηση του y_t και προσθέτοντας από μια υστέρηση κάθε φορά. Οι εκτιμήσεις ελαχίστων τετραγώνων των συντελεστών j_{ss} σε κάθε τέτοιο υπόδειγμα για $s = 1, 2, ..., p$ μας δίνει τη σειρά των μερικών αυτοσυσχετίσεων $\hat{j}_{11}, \hat{j}_{22}, \hat{j}_{33}, ..., \hat{j}_{pp}$.Οι τρεις πρώτοι συντελεστές που προκύπτουν από τις εκτιμήσεις των υποδειγμάτων είναι:

το $\hat{j}_{11}$ προκύπτει από $y_1 = j_{11} y_{t-1} + e_{1t}$

το $\hat{j}_{22}$ προκύπτει από $y_t = j_{12} y_{t-1} + j_{22} y_{t-2} + e_{2t}$

το $\hat{j}_{33}$ προκύπτει από $y_t = j_{13} y_{t-1} + j_{23} y_{t-2} + j_{33} y_{t-3} + e_{3t}$

Οι τιμές του j_{ss} για τις διάφορες τιμές του $s = 1, 2, ...$ αποτελούν τη συνάρτηση μερικής αυτοσυσχέτισης (partial autocorrelation function) και συμβολίζεται με τα γράμματα PACF.Τόσο η μορφή του PACF αλλά και του ACF μας χρησιμεύουν στον προσδιορισμό της μορφής της στοχαστικής διαδικασίας που δημιούργησε τη δεδομένη χρονολογική σειρά.

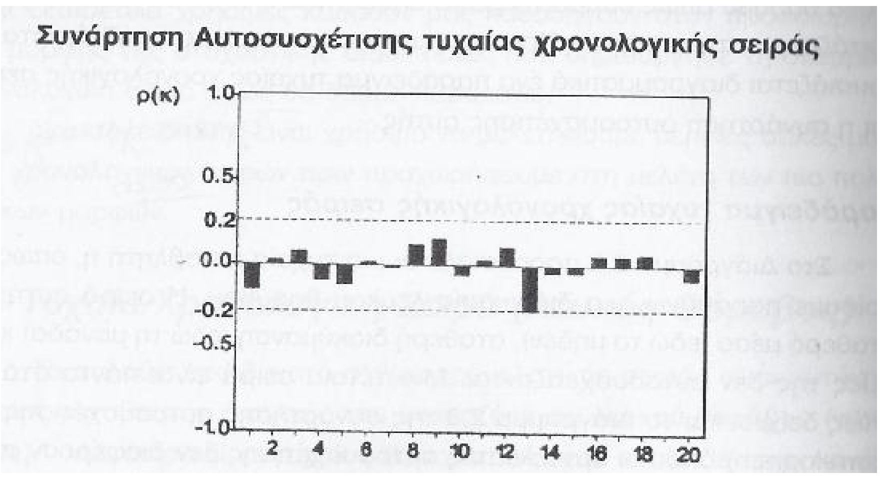
2.6 Τυχαία Χρονολογική Σειρά

Μια τυχαία χρονολογική σειρά ονομάζεται αυτή η οποία έχει τυχαία μεταβλητή ή αλλιώς λευκό θόρυβο (white noise). Μια σειρά ονομάζεται λευκός θόρυβος αν δεν έχει κανένα ευκρινές σχήμα ή πρότυπο και συμβολίζεται με ε_t.Τα χαρακτηριστικά της είναι ότι έχει σταθερό μέσο(συνήθως μηδέν), σταθερή διακύμανση και οι τιμές της δεν αυτοσυσχετίζονται. Γενικότερα για κάθε τιμή του t θα έχουμε:

$$\mathrm{E}(e_t) = 0$$
$$g_0 = \mathrm{E}(e_t^2) = s^2$$
$$g_k = \mathrm{E}(e_t e_{t-k}) = 0 \quad \text{για } k \neq 0$$

Άρα, μια τέτοια σειρά είναι πάντα στάσιμη και έχει μηδενικούς συντελεστές αυτοσυσχέτισης και μερικής αυτοσυσχέτισης.Χαρακτηριστικότερο παράδειγμα αποτελούν οι τυχεροί αριθμοί ΤΖΟΚΕΡ και ΛΟΤΤΟ που καταγράφονται κάθε εβδομάδα και είναι διαδικασία λευκού θορύβου.Σε αυτές τις περιπτώσεις οι προηγούμενες τιμές της σειράς δεν χρησιμεύουν καθόλου στην πρόβλεψη μελλοντικών τυχερών αριθμών.

Διάγραμμα 2.3 Διαδικασία Τυχαίας Χρονολογικής Σειράς (λευκός θόρυβος)



Διάγραμμα 2.4 Συνάρτηση Αυτοσυσχέτισης Τυχαίας Χρονολογικής Σειράς

2.7 Υπόδειγμα Τυχαίας Διαδρομής

Στα υποδείγματα τυχαίας διαδρομής(random walk) κάθε τιμή της χρονολογικής σειράς Y_t,προκύπτει από την αμέσως προηγούμενη Y_{t-1} με την προσθήκη ενός τυχαίου σφάλματος, δηλαδή έχουμε την εξής σχέση: $Y_t = Y_{t-1} + e_t$,t=1…,T όπου e_t είναι λευκός θόρυβος(μηδενικού μέσου).Στο υπόδειγμα τυχαίας διαδρομής βλέπουμε ότι οι διαδοχικές μεταβολές των τιμών της Y_t είναι ανεξάρτητες τυχαίες μεταβλητές.

$$\Delta y_t = y_t - y_{t-1} = e_t$$

Εξετάζοντας πρώτα πως εξελίσσεται διαχρονικά αυτή η σειρά, θεωρούμε ότι η αρχική τιμή της y_t ισούται με y_0 και κάνοντας διαδοχικές αντικαταστάσεις έχουμε:

$$y_1 = y_0 + e_1$$
$$y_2 = y_1 + e_2 = y_0 + e_1 + e_2$$

$$y_t = y_0 + e_1 + e_2 + ... + e_t = y_0 + \sum_{i=1}^{t} e_i$$

Η εξίσωση αυτή είναι και η γενική λύση της εξίσωσης διαφορών πρώτης τάξης που αντιπροσωπεύει η τυχαία διαδρομή. Καταλήγουμε δηλαδή στο συμπέρασμα πως οι διαδοχικές παρατηρήσεις της χρονολογικής σειράς y_t είναι **από κοινού εξαρτημένες τυχαίες μεταβλητές** αφού συσχετίζονται γραμμικά με τις ίδιες προηγούμενες τιμές των τυχαίων σφαλμάτων e_t.Το υπόδειγμα τυχαίας διαδρομής αποδεικνύεται μη στάσιμο καθώς ενώ έχει σταθερό μέσο $E(y_t) = y_0$,η διακύμανση αλλά και οι συνδιακυμάνσεις των τιμών δεν παραμένουν σταθερά διαχρονικά. Για την ύπαρξη στασιμότητας θα πρέπει να ισχύουν και οι τρεις συνθήκες. Άρα για το μέσο έχουμε τα εξής:

$$E(y_t) = E(y_0 + e_1 + e_2 + ... + e_t)$$
$$= y_0 + E(e_1) + E(e_2) + ... + E(e_t) = y_0$$

Εφόσον $E(e_t) = 0$.

Για τις διακυμάνσεις ισχύουν τα εξής:

$$\operatorname{var}(y_t) = E\left[y_t - E(y_t)\right]^2$$
$$= E\left(y_0 + \sum_{i=1}^{t} e_i - y_0\right) = E\left(e_1^2\right) + E\left(e_2^2\right) +$$
$$... + E\left(e_t^2\right) = ts_e^2$$

αφού ισχύει: $E\left(e_t^2\right) = s_e^2$

Όσων αφορά τη συνδιακύμανση

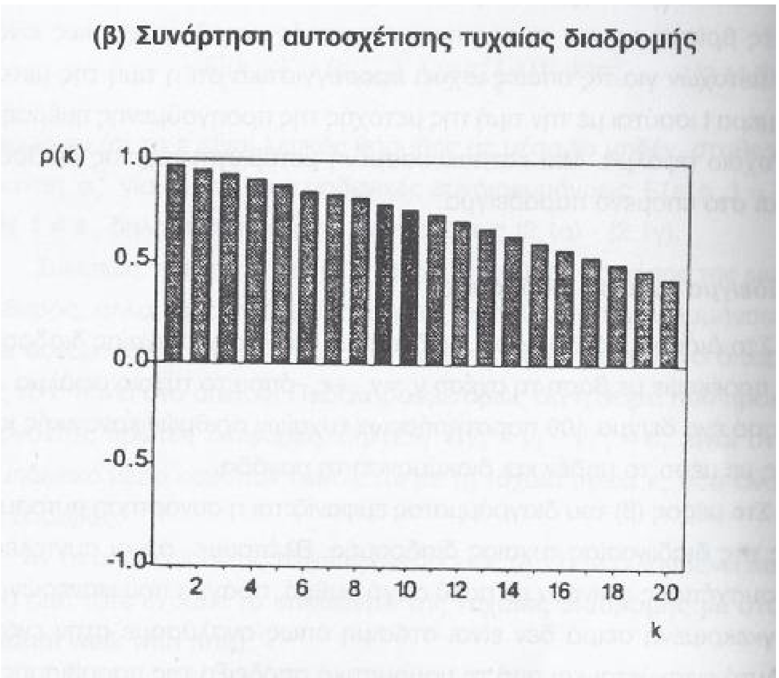
$$Cov(y_t, y_{t-k}) = E(e_t + e_{t-1} + ... + e_1)(e_{t-k} + e_{t-k-1} + e_1) =$$
$$= E\left[(e_{t-k})^2 + (e_{t-k-1})^2 + ... + e_1^2\right] =$$
$$= (t-k)s_e^2$$

έχουμε σαν δεδομένο πως το e_t είναι λευκός θόρυβος που έχει μέσο μηδέν, σταθερή διακύμανση s_e^2 για όλα τα t και οι συνδιακυμάνσεις είναι μηδενικές $\mathrm{E}(e_t e_s) = 0$ για κάθε $t \neq s$.Καταλήγουμε δηλαδή στο συμπέρασμα πως η σειρά είναι μη στάσιμη παρότι έχει σταθερό μέσο διότι τόσο η διακύμανση αλλά και οι συνδιακυμάνσεις της είναι συναρτήσεις του χρόνου t,προσεγγίζοντας μάλιστα το άπειρο καθώς το t τείνει εκεί. Παίρνοντας όμως τις πρώτες διαφορές δηλαδή $\Delta y_t = y_t - y_{t-1} = e_t$ η σειρά γίνεται στάσιμη με μηδενικό μέσο αφού ταυτίζεται με τη τυχαία σειρά e_t που είναι λευκός θόρυβος.

Αν το υπόδειγμα της τυχαίας διαδρομής περιλαμβάνει και σταθερό όρο τότε έχουμε το υπόδειγμα της **τυχαίας διαδρομής με σταθερά(random walk with drift)** το οποίο έχει αυτή τη μορφή $y_t = b + y_{t-1} + e_t$ όπου β είναι η σταθερά της εξίσωσης, παρατηρούμε και εδώ πως παίρνοντας πρώτες διαφορές η σειρά γίνεται στάσιμη και έχουμε: $\Delta y_t = b + e_t$.Αν μπούμε στη διαδικασία να συγκρίνουμε το υπόδειγμα τυχαίας διαδρομής με το αντίστοιχο της σταθεράς βλέπουμε πως στο υπόδειγμα με σταθερά οι πρώτες διαφορές είναι εν μέρει στοχαστικές και σταθερές. Χαρακτηριστικότερο παράδειγμα χρονολογικών σειρών που ακολουθούν τυχαίες διαδρομές βρίσκουμε στις χρηματιστηριακές μεταβλητές όπως είναι οι τιμές των μετοχών για τις οποίες ισχύει προσεγγιστικά ότι η τιμή της μετοχής την ημέρα t ισούται με την τιμή της μετοχής της προηγούμενης ημέρας συν ένα τυχαίο σφάλμα .Ένα τέτοιο παράδειγμα φαίνεται στο διάγραμμα.

Διάγραμμα 2.5 Υπόδειγμα Τυχαίας Διαδρομής (random walk)

Στο συγκεκριμένο διάγραμμα παρουσιάζεται μια διαδικασία τυχαίας διαδρομής που προκύπτει από την σχέση $y_t = y_{t-1} + e_t$ όπου το τυχαίο σφάλμα ελήφθη από ένα δείγμα $n = 100$ παρατηρήσεων τυχαίων αριθμών κανονικής κατανομής με μηδενικό μέσο και διακύμανση ίση με τη μονάδα.



Διάγραμμα 2.6 Συνάρτηση Αυτοσυσχέτισης Τυχαίας Διαδρομής

Στο δεύτερο μέρος του διαγράμματος παρουσιάζεται η συνάρτηση αυτοσυσχέτισης της διαδικασίας τυχαίας διαδρομής. Παρατηρούμε πως οι συντελεστές αυτοσυσχέτισης φθίνουν με αργό ρυθμό δείγμα της μη στασιμότητας της σειράς.

ΚΕΦΑΛΑΙΟ 3

ΑΥΤΟΠΑΛΙΝΔΡΟΜΑ ΥΠΟΔΕΙΓΜΑΤΑ - $AR(p)$

Ένα αυτοπαλίνδρομο υπόδειγμα $AR(p)$,στην γενική του μορφή θα διατυπώνεται ως εξής:

$$Y_t = a_0 + a_1 Y_{t-1} + a_2 Y_{t-2} + \dots a_p Y_{t-p} + e_t$$

Οι παράμετροι $a_0, a_1, a_2 \ldots a_p$ είναι σταθερές και το e_t καλείται λευκός θόρυβος (white noise) το οποίο μετράει τα τυχαία σφάλματα. Τα παραπάνω είναι ανεξάρτητες τυχαίες μεταβλητές με μέσο το μηδέν και σταθερή διακύμανση. Ο ορος αυτοπαλινδρομο έχει να κάνει στο ότι η σχέση αυτή είναι ένα υπόδειγμα παλινδρόμησης, όπου η εξαρτημένη μεταβλητή Y_t παλινδρομείται στις προηγούμενες τιμές της ίδιας της μεταβλητής Y_t.Το p υποδηλώνει την τάξη του αυτοπαλίνδρομου υποδείγματος και αναφέρεται στο μήκος της υστερήσεως ενώ τα $Y_{t-1}, Y_{t-2}, Y_{t-p}$ είναι οι τιμές της χρονοσειράς με υστέρηση.

3.1 Αυτοπαλίνδρομα Υποδείγματα Πρώτης Τάξης - $\mathrm{A}R(1)$

Η μορφή ενός υποδείγματος $\mathrm{A}R(1)$ θα είναι σύμφωνα και με τον παραπάνω τύπο η εξής:

$$Y_t = a_0 + a_1 Y_{t-1} + e_t \quad (1)$$

Τον τύπο αυτόν μπορούμε να τον αναλύσουμε πιο εύκολα εκφράζοντας το Y_t σε αποκλίσεις από το μέσο του. Αν η χρονοσειρά Y_t είναι στάσιμη τότε για κάθε χρονική στιγμή t ο μέσος(μ) θα είναι ίδιος και εφόσον $m = \mathrm{E}(Y_t)$ και $\mathrm{E}(e_t) = 0$ θα έχουμε:

$$\mathrm{E}(\mathrm{Y_t}) = \mathrm{E}(a_0 + \ a_1 \mathrm{Y_{t-1}} + e_\mathrm{t}) = \mathrm{E}(a_0) \ + \mathrm{E}(a_1 \mathrm{Y_{t-1}}) \ + \mathrm{E}(e_\mathrm{t}) \ \Leftrightarrow$$

$$\mathrm{E}(\mathrm{Y_t}) = \mathrm{E}(a_0) \ + \ a_1 \mathrm{E}(\mathrm{Y_{t-1}}) \Leftrightarrow$$

$$m = a_0 + a_1 m \Leftrightarrow m - a_1 m = a_0 \Leftrightarrow$$

a. $m = \dfrac{a_0}{1 - a_1}$

Για να είναι στάσιμη θα πρέπει $|a_1| < 1$

Αν αφαιρέσουμε από την γενική μορφή τον μέσο κατά μέλη και ο $m = 0$ τότε το $\mathrm{A}R(1)$ μπορεί να γραφτεί ως:

$$\mathrm{y_t} = a_1 \mathrm{y_{t-1}} + e_\mathrm{t} \quad (2)$$

$$\mathrm{Y_t} - m \ = a_0 + a_1 \mathrm{Y_{t-1}} + e_\mathrm{t} - m \Leftrightarrow$$

$$\mathrm{Y_t} - m = \ m - a_1 m + a_1 \mathrm{Y_{t-1}} + e_\mathrm{t} - m \ \Leftrightarrow \qquad (3)\ a_0 = m - a_1 m$$

$$\mathrm{Y_t} - m = \ a_1(\mathrm{Y_{t-2}} - m) + e_\mathrm{t} \quad \Leftrightarrow$$

$y_t = a_1 y_{t-1} + e_t \Leftrightarrow$ όπου $y_t = Y_t - m$ για κάθε $t = 1, 2, \ldots n$

Εκτός από το (a) στο υπόδειγμα $AR(1)$ ισχύουν και τα παρακάτω:

b. $g_0 = V(y_t) = \dfrac{s^2}{1 - a_1^2}$

<u>Απόδειξη</u>

Παίρνοντας την σχέση (2) αν υψώσουμε στο τετράγωνο και τα 2 μέλη θα έχουμε :

$$E(y_t)^2 = E\left[(a_1 y_{t-1}) + (e_t)\right]^2 = a_1^2 Ey^2_{t-1} + 2a_1 Ey_{t-1}e_t + e_t^2 \quad (4)$$

Το y_{t-1} εξαρτάται μόνο από το e_{t-1} οπότε $Ey_{t-1}e_t = 0$ και αφού η χρονοσειρά είναι στάσιμη το $Ey^2_t = Ey^2_{t-1} = V(y_t)$

Επομένως η σχέση (4) θα γίνει:

$$V(y_t) = a_1^2 V(Y_t) + s^2 \quad ή \quad g_0 = V(y_t) = \frac{s^2}{1 - a_1^2}$$

c. $g_s = Cov(y_t, y_{t-s})$

<u>Απόδειξη</u>

Αν πολλαπλασιάσουμε και τα δύο μέλη της σχέσης (2) με y_{t-s}

$$y_t y_{t-s} = a_1 y_{t-1} y_{t-s} + e_t y_{t-s}$$

$$E(y_t y_{t-s}) = a_1 E(y_{t-1} y_{t-s}) + E(e_t y_{t-s})$$

$g_s = a_1 g_{s-1}$ για κάθε $s > 1$ διότι $E(e_t y_{t-s}) = 0$ και $E(y_{t-1} y_{t-s}) = g_{s-1}$

Άρα γενικά θα ισχύει ότι $g_s = a_1^s g_0$ καθώς και

$$g_1 = a_1 g_0 \;, \quad g_2 = a_1 g_1 = a_1^2 g \;, \quad g_3 = a_1 g_3 = a_1^3 g_0$$

d. $r_s = \frac{g_s}{g_0} = a_1^{s}$

Θα πρέπει το $|a_1| < 1$ για να είναι η χρονοσειρά στάσιμη γιατί η διακύμανση g_0 δεν γίνεται να πάρει αρνητικές τιμές.

- Για $a_1 > 0$ η συνάρτηση αυτοσυσχέτισης αρχίζοντας από την μονάδα φθίνει γεωμετρικά και τείνει προς το μηδέν καθώς το s αυξάνει.
- Το ίδιο ισχύει και αν το $a_1 < 0$ καθώς η συνάρτηση θα φθίνει γεωμετρικά προς το μηδέν αλλά με αρνητικό πρόσημο.

3.2 Τελεστής Υστέρησης

Στην ανάλυση των χρονολογικών σειρών χρησιμοποιείται αρκετές φορές ο τελεστής υστερήσεως L (lag operator) ο οποίος μας διευκολύνει στις αλγεβρικές πράξεις. Οπότε αν έχουμε σε μια σχέση Y_{t-1} με τον τελεστή υστέρησης L μπορεί να γραφτεί ως LY_t μια Y_{t-2} ως L^2Y_t , μια Y_{t-3} ως L^3Y_t κλπ. Γενικά θα ισχύει $Y_{t-s} = L^sY_t$ δηλαδή ο εκθέτης του L δηλώνει τον αριθμό των φορών που θα πρέπει να υστερήσουμε την μεταβλητή Y .

Άρα η σχέση $y_t = a_1 y_{t-1} + e_t$ γίνεται:

$$y_t = a_1 L y_t + e_t \quad \text{ή} \quad y_t = \frac{1}{1 - a\mathrm{L}} e_t \quad \text{ή} \quad (1 - a_1\mathrm{L}) y_t = e_t \quad \text{ή} \quad y_t = (1 - a_1\mathrm{L})^{-1} e_t$$

Για $|a_1| < 1|$ ο όρος $(1 - a_1\mathrm{L})^{-1} e_t$ μπορεί να θεωρηθεί ως το όριο μιας γεωμετρικής προόδου άπειρης τάξης, οπότε θα έχουμε

$$(1-a_1\mathrm{L})^{-1} = 1+a_1\mathrm{L}+a_1{}^2\mathrm{L}^2+a_1{}^3\mathrm{L}^3+\ldots \Leftrightarrow$$

$$y_t = \sum_{j=0}^{\infty} a^j L^j e_t = \sum_{j=0}^{\infty} a^j e_{t-j}$$

3.3 Αυτοπαλίνδρομα Υποδείγματα Δεύτερης Τάξης - $AR(2)$

Τα αυτοπαλίνδρομα υποδείγματα δεύτερης (p=2) τάξεως στην γενική τους μορφή μπορούν να γραφτούν ως εξής:

$$Y_t = a_0 + a_1 Y_{t-1} + a_2 Y_{t-2} + e_t \quad (1)$$

Μια χρονολογική σειρά y_t με μέσο 0 θα είναι $y_t = a_1 y_{t-1} + a_2 y_{t-2} + e_t$ (2)

Για την $AR(2)$ προκύπτουν τα παρακάτω:

α. $m = \dfrac{a_0}{1-a_1-a_2}$

<u>Απόδειξη</u>

Από την σχέση (1) παίρνουμε τις μέσες τιμές οπότε θα έχουμε

$$\mathrm{E}(Y_t) = a_0 + a_1\mathrm{E}(Y_{t-1}) + a_2\mathrm{E}(Y_{t-2}) + \mathrm{E}(e_t), \qquad \mathrm{E}(e_t) = 0 \Leftrightarrow$$

$$m = a_0 + a_1 m + a_2 m \Leftrightarrow$$

$$m = \frac{a_0}{1-a_1-a_2}$$

β. $g_0 = Var(Y_t) = a_1 g_1 + a_2 g_2 + s^2$

<u>Απόδειξη</u>

Πολλαπλασιάζοντας και τα δύο μέλη της σχέσης (2) με y_t

$$y_t^2 = a_1 y_t y_{t-1} + a_2 y_t y_{t-2} + y_t e_t$$

Στη συνέχεια παίρνουμε τις μέσες τιμές και προκύπτει

$$\mathrm{E}(y_t^2) = Var(y_t) = a_1 \mathrm{E}(y_t y_{t-1}) + a_2 \mathrm{E}(y_t y_{t-2}) + \mathrm{E}(y_t e_t) \Leftrightarrow$$

$$g_0 = a_1 g_1 + a_2 g_2 + \mathrm{E}(y_t e_t) \Leftrightarrow$$

$$g_0 = a_1 g_1 + a_2 g_2 + s^2, \quad \text{επειδή} \quad \mathrm{E} y_t e_t = \mathrm{E}(a_1 y_{t-1} + a_2 y_{t-2} + e_t) e_t = s^2$$

γ. $g_s = Cov(Y_t, Y_{t-s}) = a_1 g_{s-1} + a_2 g_{s-2}$ για $s > 0$

<u>Απόδειξη</u>

Πολλαπλασιάζοντας και τα δύο μέλη της σχέσης $y_t = a_1 y_{t-1} + a_2 y_{t-2} + e_t$ με το y_{t-s} δηλαδή θα έχουμε $y_t y_{t-s} = a_1 y_{t-1} y_{t-s} + a_2 y_{t-2} y_{t-s} + y_{t-s} e_t$

Έπειτα παίρνουμε τις μέσες τιμές και θα προκύψει

$$\mathrm{E} y_t e_{t-s} = Cov(Y_t, Y_{t-s}) = a_1 \mathrm{E} y_{t-1} y_{t-s} + a_2 \mathrm{E} y_{t-2} y_{t-s} + \mathrm{E} y_{t-s} e_t \Leftrightarrow$$

$$g_s = Cov(Y_t, Y_{t-s}) = a_1 g_{s-1} + a_2 g_{s-2} \quad ,$$

καθώς το $\mathrm{E} y_{t-s} e_t = 0$ και τα $\mathrm{E} y_{t-1} y_{t-s} = g_{s-1}$, $\mathrm{E} y_{t-2} y_{t-s} = g_{s-2}$

δ. $r_s = a_1 r_{s-1} + a_2 r_{s-2}$ για $s > 0$

<u>Απόδειξη</u>

Διαιρούμε την σχέση (γ) με το g_0 και έχουμε:

$$\frac{g_s}{g_0} = a_1 \frac{g_{s-1}}{g_0} + a_2 \frac{g_{s-2}}{g_0} \Leftrightarrow$$

$$r_s = a_1 r_{s-1} + a_2 r_{s-2}$$

Αν χρησιμοποιήσουμε τον ορισμό του τελεστή υστέρησης το υπόδειγμα $AR(2)$ γράφεται ως

$$(1 - a_1 L - a_2 L^2) y_t = e_t \quad (3)$$

Εξισώνοντας το αριστερό μέλος με το μηδέν θα προκύψει η ομογενή μορφή της εξίσωσης διαφορών που ορίζει η σχέση (3).Η λύση της εξίσωσης αυτής εξαρτάται από της ρίζες της χαρακτηριστικής εξίσωσης

$$X^2 - a_1 X - a_2 = 0$$

Αυτή είναι μια δευτεροβάθμια εξίσωση με ρίζες λ_1και λ_2 ικανοποιούν την συνθήκη

$$(\mathrm{X} - l_1)(\mathrm{X} - l_2) = 0$$

Οι ρίζες l_1, l_2 με τις παραμέτρους α συνδέονται όπως παρακάτω

$$a_1 = l_1 + l_2$$
$$a_2 = -l_1 l_2$$

Οι ρίζες της δευτεροβάθμιας εξίσωσης υπολογίζονται από τον τύπο :

$$\lambda_1, \lambda_2 = \frac{a_1 + \sqrt{a_1^2 + 4a_2}}{2}$$

Οι ρίζες αυτές μπορεί να είναι πραγματικές αν η διακρίνουσα είναι θετική ή μιγαδικές αν είναι αρνητική. Για να υπάρχει στασιμότητα οι ρίζες l_1 και l_2 θα πρέπει να είναι μικρότερες της μονάδας σε απόλυτες τιμές.

$$|l_1| < 1|$$
$$|l_2| < 1|$$

Αν μια χρονολογική σειρά είναι της μορφής $AR(2)$ για να είναι στάσιμη θα πρέπει να ικανοποιεί τις παρακάτω προϋποθέσεις :

$$a_1 + a_2 < 1$$
$$a_2 - a_1 < 1$$
$$|a_2| < 1$$

3.4 Αυτοπαλίνδρομα Υποδείγματα Τάξεως $p - AR(p)$

Η γενική μορφή του $AR(p)$ δίνεται από τον εξής τύπο:

$$Y_t \;=\; a_0 + \; a_1 Y_{t-1} + \; a_2 Y_{t-2} + \ldots + \; a_p Y_{t-p} + \; e_t$$

ή αλλιώς με την βοήθεια του τελεστή υστέρησης όπως παρακάτω

$$(1 \;-\; a_1 L - a_2 L^2 - \ldots - \; a_p L^p) y_t \;=\; e_t$$

Οι σχέσεις που προκύπτουν στην $AR(p)$ είναι οι ακόλουθες:

α. $g_0 = Var(Y_t) = \; a_1 g_1 + a_2 g_2 + \ldots + \; a_p g_p + \; s^2$

β. $g_s = \; a_1 g_{s-1} + \; a_2 g_{s-2} + \ldots + \; a_p g_{s-p}$ για $s > 0$

γ. $r_s = \; a_1 r_{s-1} + \; a_2 r_{s-2} + \; \ldots \; + \; a_p g_{s-p}$ για $s > 0$

Από το γ προκύπτουν p εξισώσεις Yule-Walker για $s = 1, 2 \ldots p$ δηλαδή οι ακόλουθες:

Για $s = 1$, $\quad r_1 = a_1 + \; a_2 r_1 + \; a_3 r_2 + \ldots + \; a_p r_{p-1}$
Για $s = 2$, $\quad r_2 = \; a_1 r_1 + \; a_2 + a_3 r_1 + \ldots + \; a_p r_{p-2}$
Για $s = 3$, $\quad r_3 = \; a_1 r_2 + \; a_2 r_1 + \; a_3 + \ldots + \; a_p r_{p-3}$
.. .. ….
.. .. ….
Για $s = p$, $\quad r_p = \; a_1 r_{p-1} + \; a_2 r_{p-2} + \; a_3 r_{p-3} + \ldots + \; a_p$

Επομένως η μορφή της συνάρτησης αυτοσυσχέτισης ενός $AR(p)$ υποδείγματος εξαρτάται από τις τιμές των παραμέτρων του αυτοπαλίνδρομου υποδείγματος $(a_1, a_2, .., a_p)$.

Συμβολισμός σε πίνακα:

$$\begin{bmatrix} r_1 \\ r_2 \\ \ldots \\ \ldots \\ r_p \end{bmatrix} = \begin{bmatrix} 1 & r_1 \cdots \cdots r_{p-1} \\ r_1 & 1 \cdots \cdots r_{p-2} \\ \ldots & \ldots \ddots \\ \ldots & \ldots\ldots\ldots \ddots \\ r_{p-1} & r_{p-2} \cdots \cdots 1 \end{bmatrix} x \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ \vdots \\ a_p \end{bmatrix}$$

$$\mathrm{R} \quad = \quad \Pi \quad x \;\; \mathrm{A}$$

$$\Leftrightarrow \mathrm{A} = \Pi^{-1}\mathrm{R}$$

3.5 Συνάρτηση Μερικής Αυτοσυσχετίσεως για υποδείγματα μορφής - *AR*

Όλες οι αυτοπαλίνδρομες διαδικασίες έχουν συναρτήσεις αυτοσυσχέτισης οι οποίες βαίνουν φθίνουσες καθώς αυξάνει το μήκος της υστέρησης , με αποτέλεσμα να είναι πολλές φορές δύσκολο να καθοριστεί η τάξη του υποδείγματος που περιγράφει τη σειρά με βάση τη συνάρτηση αυτοσυσχέτισης. Ένα πρόσθετο κριτήριο για το σκοπό αυτό είναι η συνάρτηση της μερικής αυτοσυσχέτισης (Partial Autocorellation Function PACF). Η μερική αυτοσυσχέτιση ανάμεσα στην Y_t και την Y_{t-s} αναφέρεται στην συσχέτιση ανάμεσα στην Y_t και την Y_{t-s} όταν έχουν αφαιρεθεί οι γραμμικές επιδράσεις των ενδιάμεσων μεταβλητών $Y_{t-1}, Y_{t-2}, ..., Y_{t-(s-1)}$.Αν παραστήσουμε τον συντελεστή αυτοσυσχέτισης με r_{ss} τάξεως s δηλαδή τον συντελεστή αυτοσυσχετισης ανάμεσα στην Y_t και Y_{t-s} για $s = 1, 2..., p$, τότε το r_{ss} θα είναι ο συντελεστής μερικής παλινδρόμησης της μεταβλητής y_{t-s} στο υπόδειγμα:

$$y_t = r_{1s}y_{t-1} + r_{2s}y_{t-2} + r_{3s}y_{t-3} + \ldots + r_{ss}y_{t-s} + e_t$$

Ο μερικός συντελεστής αυτοσυσχέτισης όπως βλέπουμε έχει δυο δείκτες. Ο αριστερός δείκτης μας δείχνει την χρονική υστέρηση της μεταβλητής σύμφωνα με το $Y_{t-1}, Y_{t-2}...$ Ο δείκτης δεξιά μας δείχνει την μέγιστη τάξη της παλινδρόμησης.

Άρα η συνάρτηση μερικής αυτοσυσχέτισης είναι μηδέν για $s > p$ όταν θα μιλάμε για μια αυτοπαλίδρομη διαδικασία τάξεως p. Με άλλα λόγια ισχύουν τα εξης:

Για $AR(1)$: α) $r_{11} = r_1 = a_1$

β) $r_{ss} = 0$ για $s > 1$

Για $AR(2)$: γ) $r_{11} = r_1$

δ) $r_{22} = \frac{r_2 - r_1^2}{1 - r_1^2}$

ε) $r_{ss} = 0$ για $s > 0$

Για $AR(p)$: στ) $r_{11} = r_1$

ζ) $r_{22} \neq 0$,… , $r_{pp} \neq 0$

η) $r_{ss} = 0$ για $s > 2$

3.6 Έλεγχος Στατιστικής Σημαντικότητας Συντελεστών Αυτοσυσχέτισης και μερικού συντελεστή

Στην πράξη επειδή τόσο οι πραγματικές τιμές των αυτοσυσχετίσεων r_s όσο και των πραγματικών μερικών αυτοσυσχετίσεων r_{ss} δεν είναι γνωστές χρησιμοποιούνται οι αντίστοιχες εκτιμήσεις τους από το δείγμα.
Εάν στους τύπους που δίνουν την θεωρητική τιμή της συνάρτησης μερικής αυτοσυσχετίσεως, τα r_s θα αντικατασταθούν με τις δειγματικές τιμές τους τότε, οι τιμές των r_{ss} που προκύπτουν είναι δειγματικές.

Με βάση τις εκτιμήσεις αυτές μπορεί να γίνει έλεγχος σημαντικότητας των παραμέτρων στον πληθυσμό. Για μεγάλα δείγματα, οι εκτιμήσεις $\hat{r}_s$
των αυτοσυσχετίσεων r_s κατανέμονται κανονικά με μέση τιμή το μηδέν και διακύμανση $\frac{1}{T}$, όπου T είναι το μέγεθος του δείγματος. Το ίδιο ισχύει και
για τις εκτιμήσεις των μερικών αυτοσυσχετίσεων $\hat{r}_{ss}$ για υστερήσεις μεγαλύτερες από την τάξη p της AR διαδικασίας.

$\hat{r}_s \sim N\left(0, \frac{1}{T}\right)$

$\hat{r}_{ss} \sim N\left(0, \frac{1}{T}\right)$, για $s > p$

Στη συνέχεια θα αναπτύξουμε τον έλεγχο της σημαντικότητας του συντελεστή r_s, δηλαδή θα ελέγξουμε την υπόθεση:

$$H_0 : \hat{r}_s = 0$$
$$v\ s$$
$$H_1 : \hat{r}_s \neq 0$$

Ο έλεγχος αυτός θα γίνει με την βοήθεια της στατιστικής: $t_s = \dfrac{\hat{r}_s}{\sqrt{\frac{1}{T}}} = \hat{r}_s \sqrt{T}$

Για δεδομένο επίπεδο σημαντικότητας α=5% η μηδενική απορρίπτεται αν $|t_s| > 2$, διαφορετικά θα έχουμε αποδοχή της H_0

Το 95% διάστημα εμπιστοσύνης για αυτόν τον έλεγχο είναι :

$$\hat{r}_s - \frac{2}{\sqrt{T}} \leq r_s \leq \hat{r}_s$$

Στον έλεγχο σημαντικότητας των συντελεστών μερικής αυτοσυσχέτισης ισχύουν τα ίδια με τα παραπάνω. Δηλαδή ο r_{ss} είναι σημαντικός αν και μόνο αν $\left|\hat{r}_{ss}\sqrt{T}\right| < 2$.

Με τη βοήθεια του παραπάνω ελέγχου σημαντικότητας των συντελεστών μερικής αυτοσυσχέτισης μπορεί να καθοριστεί η τάξη μιας AR διαδικασίας. Θα επιλεγεί ως τάξη της σειράς αυτή που αντιστοιχεί στην τελευταία σημαντική τιμή του t_s.

Για παράδειγμα έστω ότι η τελευταία σημαντική τιμή του t είναι για $s = 2$, δηλαδή έστω ότι ο συντελεστής r_{22} είναι στατιστικά σημαντικός, ενώ ο συντελεστής r_{33} είναι στατιστικά μη σημαντικός. Τότε, συμπεραίνουμε ότι η τάξη p του υποδείγματος είναι 2.

ΚΕΦΑΛΑΙΟ 4

Υποδείγματα Κινητού Μέσου - $MA(q)$

Τα υποδείγματα κινητού μέσου είναι χρήσιμα για περιγραφή φαινομένων όπου τα γεγονότα παράγουν ένα άμεσο αποτέλεσμα, η επίδραση του οποίου δεν σταματά εκεί αλλά συνεχίζει, αν και το ίδιο το γεγονός παύει να υφίσταται. Τις περισσότερες φορές επηρεάζει λιγότερο και για μικρό χρονικό διάστημα τις επόμενες χρονικές στιγμές. Ένα χαρακτηριστικό παράδειγμα είναι οι απεργίες όπου επηρεάζουν την οικονομία όχι μόνο βραχυχρόνια αλλά και μακροχρόνια.

Ως προς την γενική τους μορφή, οι διαδικασίες κινητού μέσου $MA(q)$ γράφονται ως εξής:

$$Y_t = m + e_t - q_1 e_{t-1} q - q_2 e_{t-2} - \ldots - q_q e_{t-q}$$

Όπου τα q είναι σταθεροί παράμετροι και e_t ο λευκός θόρυβος. Στο υπόδειγμα $MA(q)$ υποθέτουμε ότι η χρονολογική σειρά Y_t δημιουργείται ως ένας σταθμικός μέσος των τυχαίων σφαλμάτων των q προηγούμενων περιόδων.

4.1 Υπόδειγμα Κινητού Μέσου 1ης Τάξης - $MA(1)$

Μια διαδικασία κινητού μέσου πρώτης τάξης $MA(1)$ θα έχει την εξής γενική μορφή:

$$Y_t = e_t - q_1 e_{t-1}$$

Στο υπόδειγμα αυτό ισχύουν τα παρακάτω:

α) $\mathrm{E}(Y_t) = 0$

β) $g_0 = Var(Y_t) = s_e^{\,2}\left(1 + q_1^{\,2}\right)$

γ) $g_k = Cov(Y_t, Y_{t-k}) = -q_1 s_e^{\,2}$, για $k = 1$

$g_{\mathrm{k}} = 0$, για $k > 1$

δ)

$$r_k = \frac{g_k}{g_0} = \begin{cases} \dfrac{-q_1}{1+q_1^{\,2}}, & k = 1 \\ 0, & k > 1 \end{cases}$$

<u>Αποδείξεις</u>

α) $E(Y_t) = E(e_t - q_1 e_{t-1}) = E(e_t) - E(e_{t-1}) = 0$

β) $Var(Y_t) = g_0 = E(Y_t - m)^2 = E(e_t - q_1 e_{t-1})^2 =$

$$= Ee_t^{\,2} - 2q_1 E(e_t e_{t-1}) + q_1^{\,2} Ee_{t-1}^{\,2} =$$

$$= s^2 - 2q_1 0 + q_1 s^2 = s^2 + q_1 s^2 =$$

$$= s_e^{\,2}\left(1+q_1^{\,2}\right)$$

γ) $Cov\left(Y_t, Y_{t-k}\right) = g_k = E\left(Y_t - m\right)\left(Y_{t-k} - m\right) =$

$$= E(e_t - q_1 e_{t-1})(e_{t-1} - q_1 e_{t-2}) =$$

$$= E(e_t e_{t-1}) - q_1 E(e_t e_{t-2}) - q_1 E(e_{t-1})^2 + q_1^{\,2} E(e_{t-1} e_{t-2}) =$$

$$= -q_1 E(e_{t-1})^2 = -q_1 s_e^{\,2} \qquad \text{για } k = 1$$

δ)‘

$$r_k = \frac{g_k}{g_0} = \frac{-\theta_1 \sigma^2}{\sigma^2 (1+q_1^2)} = \frac{-\theta_1}{1+q_1^2} \qquad \text{για } k = 1$$

Όπως γίνεται αντιληπτό όλες οι αυτοσυνδιακυμάνσεις και συνεπώς και οι αυτοσυσχετίσεις είναι μηδέν εκτός από την πρώτη. Αυτό σημαίνει ότι μια οποιαδήποτε παρατήρηση της Y_t σχετίζεται μόνο με την προηγούμενη ή την επομένη και δεν σχετίζεται με καμία άλλη. Για παράδειγμα η Y_3 σχετίζεται με την Y_2 ή την Y_4.

Γενικά η συνάρτηση αυτοσυσχέτισης r_k στα υποδείγματα κινητών μέσων γίνεται μηδέν με τάξη q, γίνεται μηδέν μετά από q χρονικές υστερήσεις $(k > q)$ σε αντίθεση με το ότι συμβαίνει στα AR όπου η θεωρητική συνάρτηση r_k φθίνει αλλά δεν μηδενίζεται ποτέ.

4.2 Υπόδειγμα Κινητού Μέσου 2^{ης} Τάξης - $MA(2)$

Το υπόδειγμα κινητού μέσου δεύτερης τάξης $MA(2)$ έχει την παρακάτω μορφή:

$$Y_t = e_t - q_1 e_{t-1} - q_2 e_{t-2}$$

ή αλλιώς με την βοήθεια του τελεστή υστέρησης:

$$Y_t = (1 - q_1 L - q_2 L^2) e_t$$

Οι σχέσεις που ισχύουν στο $MA(2)$ είναι οι εξής:

1. $E(y_t) = 0$

2. $Var(y_t) = g_0 = \left(1 + q_1^2 + q_2^2\right) s_e^2$

Για τις συναρτήσεις αυτοδιασποράς ισχύουν:

3. $g_1 = -q_1\left(1 - q_2\right) s_e^2$, για $k = 1$

4. $\gamma_2 = -\theta_2 \sigma_\varepsilon^2$, για $k = 2$

5. $g_k = 0$, για $k > 2$

Επίσης για τις συναρτήσεις αυτοσυσχέτισης ισχύουν τα παρακάτω:

6. $r_1 = \dfrac{-\theta_1 (1 - q_2)}{1 + q_1^2 + q_2^2}$

7. $r_2 = \dfrac{-q_2}{1 + q_1^2 + q_2^2}$

8. $r_k = 0$, για $k > 2$

<u>Αποδείξεις των παραπάνω</u>

1. $E(Y_t) = E(e_t - q_1 e_{t-1} - q_1 e_{t-2}) \; = E(e_t) - q_1 E(e_{t-1}) - q_2 E(e_{t-2}) = 0$

2. $Var(Y_t) = g_0 = \mathrm{E}(Y_t - m)^2 = \mathrm{E}(e_t - q_1 e_{t-1} - q_2 e_{t-2})^2 =$

$$= Ee_t^{\,2} - 2q_1 E(e_t e_{t-1}) + q_1^{\,2} Ee_{t-1}^{\,2} - 2q_2 E(e_t e_{t-2}) + q_2^{\,2} Ee_{t-2}^{\,2} =$$

$$= s^{\,2} - 2q_1 0 + q_1^{\,2} s^{\,2} - 2q_2 0 + q_2^{\,2} s^{\,2} = s^{\,2} + q_1^{\,2} s^{\,2} + q_2^{\,2} s^{\,2} =$$

$$= s_e^{\,2}\left(1 + q_1^{\,2} + q_2^{\,2}\right)$$

3. $Cov(Y_t, Y_{t-k}) = g_1 = E\left(Y_t - m\right)\left(Y_{t-1} - m\right) =$

$$= E(e_t - q_1 e_{t-1} - q_2 e_{t-2})(e_{t-1} - q_1 e_{t-2} - q_2 e_{t-3}) =$$

$$= -q_1 s_e^{\,2} + q_1 q_2 s^{\,2} =$$

$$= -q_1\left(1 - q_2\right) s_e^{\,2}$$

4. $g_2 = E\left(Y_t - m\right)\left(Y_{t-2} - m\right) =$

$$= E(e_t - q_1 e_{t-1} - q_2 e_{t-2})(e_{t-2} - q_1 e_{t-3} - q_2 e_{t-4}) =$$

$$= -q_2 s_e^{\,2}$$

5. Προφανώς για $s > 2$ θα ισχύει $g_s = 0$

6.

$$r_1 = \frac{g_1}{g_0} = \frac{-q_1(1 - q_2)s^{\,2}}{(1 + q_1^2 + q_2^2)s^{\,2}} = \frac{-q_1(1 - q_2)}{1 + q_1^2 + q_2^2}$$

7.

$$r_2 = \frac{g_2}{g_0} = \frac{-q_2 s^{\,2}}{(1 + q_1^2 + q_2^2)s^{\,2}} = \frac{-q_2}{1 + q_1^2 + q_2^2}$$

8. Τέλος είναι προφανές ότι για $s > 2$ θα ισχύει $r_s = 0$

Άρα το υπόδειγμα $MA(2)$ έχει "μνήμη" δυο περιόδων για το λόγο ότι το Y_t επηρεάζεται από τις τιμές Y_{t-1} και Y_{t-2} αλλά όχι και από τιμές υστέρησης μεγαλύτερης του δύο.

4.3 Υπόδειγμα Κινητού Μέσου Τάξης $q - MA(q)$

Η γενική μορφή ενός $MA(q)$ δίνεται από τον τύπο :

$$Y_t = e_t - q_1 e_{t-1} - q_2 e_{t-2} - \ldots - q_{\text{q}} e_{t-q}$$

Οι σχέσεις για το $MA(q)$ είναι οι παρακάτω:

1. $E(Y_t) = m = 0$

2. $Var(Y_t) = g_0 = (1 + q_1{}^2 + q_2{}^2 + q_3{}^2 + \ldots + q_q{}^2) s_e{}^2 = s_e{}^2$

3. $Cov(Y_t, Y_{t-k}) = g_k = (-q_k + q_1 q_{k+1} + q_2 q_{k+2} + \ldots + q_q q_{k+q}) s_e{}^2 = 0$

4.

$$r_k = \frac{g_k}{g_0} = \frac{(\text{-}\theta_k + q_1 q_{k+1} + q_2 q_{k+2} + \ldots + q_q q_{k+q})\sigma^2}{(1 + q_1^2 + q_2^2 + \ldots + q_q^2) s^2} =$$

$$= \frac{\text{-}\theta_\kappa + q_1 q_{k+1} + q_2 q_{k+2} + \ldots + q_q q_{k+q}}{1 + q_1^2 + q_2^2 + \ldots + q_q^2}$$

5. $j_{11} = r_1$

6.

$$f_{22} = \frac{\begin{pmatrix} 1 & r_1 \\ r_1 & r_2 \end{pmatrix}}{\begin{pmatrix} 1 & r_1 \\ r_1 & 1 \end{pmatrix}} = \frac{r_2 - r_1^2}{1 - r_1^2}$$

7.

$$f_{33} = \frac{r_1^3}{1 - 2 r_1^2}$$

Αντιστρεψιμότητα

Η ιδιότητα της αντιστρεψιμότητας αφορά στην μετατροπή ενός υποδείγματος $MA(q)$ σε υπόδειγμα $AR(\infty)$ (άπειρης τάξης). Αντίστοιχα θα λέμε ότι ένα υπόδειγμα $AR(p)$ είναι αντιστρέψιμο αν μπορεί να λάβει μορφή ενός υποδείγματος $MA(\infty)$.

Τα υποδείγματα $AR(p)$ θα λέμε ότι είναι αντιστρέψιμα εφόσον είναι και στάσιμα. Αντίθετα για τις διαδικασίες κινητού μέσου $MA(q)$ θα πρέπει να πληρούν κάποιες προϋποθέσεις για να είναι αντιστρέψιμα και είναι οι παρακάτω:

$MA(1)$: $|q_1| < 1$

$MA(2)$: $q_1 + q_2 < 1$

$q_2 - q_1 < 1$

$|q_2| < 1$

$MA(q)$: Θα πρέπει οι ρίζες της χαρακτηριστικής εξίσωσης του $\Theta(L)$ να είναι όλες μέσα στο μοναδιαίο κύκλο ή αλλιώς οι ρίζες του πολυωνύμου $\Theta(L) = 0$ να βρίσκονται όλες έξω από τον μοναδιαίο κύκλο.

4.4 Μεικτά υποδείγματα ARMA(p,q)

Τα μεικτά υποδείγματα ARMA είναι ένας συνδυασμός από αυτοπαλίνδρομα υποδείγματα $AR(p)$ και διαδικασίες κινητού μέσου $MA(q)$.
Αυτά παρατηρούνται όταν κατά την διερεύνηση της στασιμότητας μιας χρονοσειράς τα δεδομένα της έχουν συναρτήσεις αυτοσυσχέτισης ή μερικής αυτοσυσχέτισης που δεν φαίνονται να μηδενίζονται μετά από κάποιο σημείο αλλά φθίνουν και οι δυο με αργό ρυθμό. Για να δώσουμε την γενική μορφή των μεικτών υποδειγμάτων χρησιμοποιούμε τα υποδείγματα $AR(p)$ και $MA(q)$ οπότε καταλήγουμε στην εξής γενική μορφή για το $ARMA(p,q)$:

$$Y_t = a_0 + a_1 Y_{t-1} + a_2 Y_{t-2} + \ldots + a_p Y_{t-p} + e_t - q_1 e_{t-1} - q_2 e_{t-2} - \ldots - q_p e_{t-p}$$

ή με τους τελεστές υστέρησης θα έχουμε :

$$A(L)Y_t = \Theta(L)e_t + a_0$$

Το υπόδειγμα $ARMA(p,q)$ είναι συνδυασμός p αυτοπαλίνδρομων όρων και q όρων κινητού μέσου. Είναι προφανές ότι ένα καθαρά αυτοπαλίνδρομο υπόδειγμα ή ένα καθαρό υπόδειγμα κινητού μέσου μπορούν να θεωρηθούν ως ειδικές περιπτώσεις μιας $ARMA$ διαδικασίας. Δηλαδή, θα ισχύουν τα εξής:

$AR(p) = ARMA(p,0)$
$MA(q) = ARMA(0,q)$

Η πιο απλή μορφή μιας $ARMA$ διαδικασίας είναι η $ARMA(1,1)$. Οπότε και από την γενική μορφή, για $p=1$ και $q=1$ θα έχουμε την σχέση:

$$Y_t = a_0 + a_1 Y_{t-1} + e_t - q_1 e_{t-1} \quad (1)$$

ή $$Y_t = a_1 Y_{t-1} + e_t + q_1 e_{t-1} \quad (2)$$

Αυτό το υπόδειγμα μπορεί να γραφεί ως μια καθαρά MA διαδικασία, αλλά και ως μια αυτοπαλίνδρομη διαδικασία με άπειρους όρους $AR(\infty)$.

- Ας δούμε πρώτα πώς μπορεί να μεταμορφωθεί σε μια διαδικασία κινητού μέσου

Υστερούμε διαδοχικά τη μορφή της **(1)** και αντικαθιστώντας για $Y_{t-1}, Y_{t-2}, Y_{t-3}, \ldots$ καταλήγουμε στη μορφή:

$$Y_t = a_0 \sum_{i=0}^{\infty} a_1^i + e_t + \sum_{i=0}^{\infty} a_1^i (a_1 + q_1) e_{t-1-i}$$

Για να είναι στάσιμη αυτή η σειρά, θα πρέπει το άθροισμα $\sum_{i=0}^{\infty} a_1^i (a_1 + q_1)$ να συγκλίνει, οπότε αυτό σημαίνει ότι πρέπει το $|a_1| < 1$. Όταν ισχύει αυτό, και επομένως η σειρά είναι στάσιμη, τότε η μορφή του υποδείγματος θα είναι η:

$$Y_t = \frac{a_0}{1-a_1} + e_t + \sum_{i=0}^{\infty} a_1^i (a_1 + q_1) e_{t-1-i}$$

Αυτή η μορφή δημιουργήθηκε αντικαθιστώντας τη σειρά $\sum_{i=0}^{\infty} a_1^i = \frac{1}{1-a_1}$ από τη στιγμή που είναι άθροισμα όρων φθίνουσας γεωμετρικής προόδου. Η μορφή αυτή είναι μια MA διαδικασία με άπειρους όρους, που θα μπορούσε να προσεγγιστεί με έναν περιορισμένο αριθμό όρων, δεδομένου ότι η σημασία των συντελεστών όλο και μικραίνει. Αυτό σημαίνει ότι από κάποιο σημείο θα μπορούσαν να παραλειφθούν οι επόμενοι όροι. Μπορούμε εύκολα να καταλάβουμε ότι απαιτείται υψηλής τάξεως *MA* διαδικασίας προκειμένου να προσεγγιστεί η αντίστοιχη $ARMA(1,1)$ διαδικασία. Επομένως, είναι προφανής η οικονομία που επιτυγχάνεται με τη χρήση των μεικτών υποδειγμάτων, αφού το $ARMA(1,1)$ υπόδειγμα έχει μόνο δύο συντελεστές.

- Στη συνέχεια, θα δούμε πώς το υπόδειγμα **(1)** μπορεί να διατυπωθεί και ως AR(∞).Με διαδοχικές αντικαταστάσεις για $e_{t-1}, e_{t-2}, e_{t-3}, \ldots$ στην μορφή (1) καταλήγουμε στην σχέση:

$$Y_t = \frac{a_0}{1-q_1} + e_t + \sum_{i=0}^{\infty} q_1^{\,i}(a_1 + q_1)Y_{t-1-i}$$

Η σειρά αυτή για να γίνει αντιστρέψιμη θα πρέπει $|q_1| < 1$. Οπότε και σε αυτήν την περίπτωση επιτυγχάνεται οικονομία στους συντελεστές με την χρήση της $ARMA(1,1)$ διαδικασίας.

Για το μεικτό υπόδειγμα $ARMA(1,1)$ ισχύουν τα παρακάτω :

a) $\mathrm{E}(Y_t) = m = \dfrac{a_0}{1-a_1}$

b) $g_0 = \dfrac{1+q_1^2 + 2a_1q_1s^2}{1-a_1^2}$

c) $g_1 = a_1 g_0 + q_1 s^2$

d) $g_s = a_1 g_{s-1}$ για $s > 1$

e) $r_1 = a_1 \dfrac{q_1}{g_0} s^2 = \dfrac{(1+a_1q_1)(a_1+q_1)}{1+q_1^2+2a_1q_1}$

f) $r_s = a_1 r_{s-1}$ για $s > 1$

<u>Αποδείξεις</u>

Τώρα θα εξετάσουμε τις αποδείξεις όλων των παραπάνω σχέσεων

Στη σχέση (1) παίρνουμε τις μέσες τιμές και στα 2 μέλη:

a.

$$E\left(Y_t\right)=a_0+a_1E\left(Y_{t-1}\right)+Ee_t+q_1e_{t-1}\Leftrightarrow$$
$$\Leftrightarrow m=a_0+a_1m\Leftrightarrow$$
$$\Leftrightarrow\left(1-a_1\right)m=a_0\Leftrightarrow$$
$$\Leftrightarrow m=\frac{a_0}{1-a_1}$$

Στη (b) θα υψώσουμε τη σχέση $Y_t = a_1Y_{t-1}+e_t+q_1e_{t-1}$ (2) στο τετράγωνο και θα πάρουμε τις μέσες τιμές:

$$E\left(Y_t\right)^2=E\left(a_1Y_{t-1}+e_t+q_1e_{t-1}\right)^2\Leftrightarrow$$
$$\Leftrightarrow g_0=a_1^2g_0+s^2+q_1^2s^2+2a_1q_1s^2\Leftrightarrow$$
$$\Leftrightarrow g_0-a_1^2g_0=\left(1+q_1^2+2a_1q_1\right)s^2\Leftrightarrow$$
$$\Leftrightarrow\left(1-a_1^2\right)g_0=\left(1+q_1^2+2a_1q_1\right)s^2\Leftrightarrow$$
$$\Leftrightarrow g_0=\frac{\left(1+q_1^2+2a_1q_1\right)}{\left(1-a_1^2\right)}$$

Για την απόδειξη της (c) χρησιμοποιώντας ξανά την σχέση (2) έχουμε:

$$g_1=Cov\left(y_1,y_0\right)=E\left(y_1-m\right)\left(y_0-m\right)=$$
$$=E\left(a_1g_0+e_1+q_1e_0\right)y_0=$$
$$=E\left(a_1g_0^2\right)+E\left(e_1y_0\right)+E\left(q_1e_0y_0\right)=$$
$$=a_1g_0+q_1s^2$$

Η (d) παίρνοντας πάλι την σχέση (2) αποδεικνύεται ως:

$$g_s=Cov\left(y_s,y_{s-1}\right)=E\left(y_s-m\right)\left(y_{s-1}-m\right)=$$
$$=E\left(a_1y_{s-1}+e_s+q_1e_{s-1}\right)\left(a_1y_{s-2}+e_{s-1}+q_1e_{s-2}\right)=$$
$$=a_1^2g_{s-2}+a_1q_1s^2=a_1\left(a_1g_{s-2}+q_1s^2\right)=$$
$$=a_1g_{s-1}\qquad gia\quad s>1$$

e.

$$r_1 = \frac{g_1}{g_0} = \frac{a_1 g_0 + q_1 s^2}{g_0} = a_1 + \frac{q_1 s^2}{g_0} =$$

$$= a_1 + \frac{q_1 s^2}{\frac{1+q_1^2+2a_1 q_1}{1-a_1^2} s^2} = a_1 + \frac{q_1\left(1-a_1^2\right)}{1+q_1^2+2a_1 q_1} =$$

$$= \frac{a_1 + a_1 q_1^2 + 2a_1^2 q_1 + q_1 - q_1 a_1^2}{1+q_1^2+2a_1 q_1} = \frac{a_1 + a_1 q_1^2 + a_1^2 q_1 + q_1}{1+q_1^2+2a_1 q_1} =$$

$$= \frac{a_1\left(1+a_1 q_1\right) + q_1\left(1+a_1 q_1\right)}{1+q_1^2+2a_1 q_1} = \frac{(1+a_1 q_1)(a_1+q_1)}{1+q_1^2+2a_1 q_1}$$

f.

$$r_s = \frac{g_s}{g_0} = \frac{a_1 g_{s-1}}{g_0} = a_1 \frac{g_{s-1}}{g_0} = a_1 r_{s-1} \qquad gia \quad s > 1$$

Διακρίνουμε ότι στη συνάρτηση αυτοσυσχέτισης υπεισέρχεται ο συντελεστής από την $MA(1)$ διαδικασία, αλλά μόνο για την αυτοσυσχέτιση πρώτης τάξης r_1. Όλες οι άλλες αυτοσυσχετίσεις εξαρτώνται μόνο από το αυτοπαλίνδρομο μέρος. Η συνάρτηση αυτοσυσχέτισης για την $ARMA(1,1)$ διαδικασία φθίνει γεωμετρικά με την αύξηση του s. Η μείωση όμως, σε αντίθεση με την $AR(1)$, αρχίζει από το r_1 και όχι από το $r_0 = 1$. Η συνάρτηση μερικής αυτοσυσχέτισης για την $ARMA(1,1)$ διαδικασία φθίνει γεωμετρικά όπως στην περίπτωση της $MA(1)$ διαδικασίας.

Γενικεύοντας σε ένα $ARMA(p,q)$ οι πρώτες q αυτοσυσχετίσεις για $s \le q$, εξαρτώνται από όλους τους συντελεστές a_i και q_i του αυτοπαλίνδρομου υποδείγματος και του υποδείγματος του κινητού μέσου αντιστοίχως. Όταν το s παίρνει μεγαλύτερες τιμές από το q, οι αυτοσυνδιακυμάνσεις και οι αυτοσυσχετίσεις θα είναι παρόμοιες με αυτές μιας διαδικασίας $AR(p)$ οι οποίες θα δίνονται και αυτές από τους τύπους:

$$g_s = a_1 g_{s-1} + a_2 g_{s-2} + \mathbf{K} + a_p g_{s-p} \qquad \text{, για } s > q$$

$$r_s = a_1 r_{s-1} + a_2 r_{s-2} + \mathbf{K} + a_p r_{s-p} \qquad \text{, για } s > q$$

Η συνάρτηση αυτοσυσχέτισης μιας $ARMA(p,q)$ διαδικασίας συμπεριφέρεται όπως αυτή μιας $AR(p)$ διαδικασίας, ενώ η συνάρτηση μερικής αυτοσυσχέτισης συμπεριφέρεται όπως αυτή μιας $MA(q)$ διαδικασίας, για $s > q - p$.

Για την εκτίμηση των παραμέτρων ενός $ARMA(p,q)$, υποδείγματος μπορούν να εφαρμοστούν οι ίδιες τεχνικές που χρησιμοποιούνται και για την εκτίμηση των παραμέτρων ενός $MA(q)$ υποδείγματος. Δηλαδή μπορούν να χρησιμοποιηθούν οι σχέσεις που συνδέουν τις αυτοσυσχετίσεις με τις παραμέτρους του υποδείγματος, αλλά επίσης μπορούν να εφαρμοστούν κάποιες μη γραμμικές μέθοδοι εκτίμησης, από τη στιγμή που το υπόδειγμα είναι μη γραμμικό ως προς τις παραμέτρους.

Στην συνέχεια ακλουθούν δύο πίνακες οι όποιοι ουσιαστικά είναι μια περίληψη των όσων γράψαμε σε αυτό το κεφάλαιο.

Χαρακτηριστικά των υποδειγμάτων $ARMA(p,q)$

Υπόδειγμα	**Αυτοσυσχετίσεις ACF**	**Μερικές αυτοσυσχετίσεις PACF**
Λευκός θόρυβος	Όλες μηδέν: $r_k = 0$	Όλες μηδέν: $j_{kk} = 0$
$AR(1)$	Φθίνουν προς το μηδέν από r_1 § Ευθέως αν $a > 0$ § Με πριονωτή μορφή αν $a < 0$	Μηδέν μετά από j_{11}
$AR(2)$	Φθίνουν § Ευθέως από r_2 γιαπραγματι-κές ρίζες § Με ημιτονοειδή τρόπο για μι-γαδικές ρίζες	Μηδέν μετά από j_{22}
$AR(p)$	Φθίνουν προς το μηδέν από r_q	Μηδέν μετά το j_{pp}
$MA(1)$	Μηδέν μετά το r_1	Φθίνει σχεδόν γεωμετρικά από το j_{11}
$MA(2)$	Μηδέν μετά το r_2	Φθίνει σχεδόν γεωμετρικά από το j_{22}
$MA(q)$	Μηδέν μετά το r_q	Φθίνει σχεδόν γεωμετρικά από το j_{qq}
$ARMA(1,1)$	Φθίνει γεωμετρικά από το r_1	Φθίνει γεωμετρικά ή κυματιστά από το j_{11}
$ARMA(p,q)$	Φθίνει γεωμετρικά από το r_q	Φθίνει γεωμετρικά ή κυματιστά από το j_{pp}

Πίνακας 4.1 Χαρακτηριστικά των Υποδειγμάτων ARMA

Συνθήκες Στασιμότητας και Αντιστρεψιμότητας των υποδειγμάτων $ARMA$

Υπόδειγμα	**Εξίσωση**	**Στασιμότητα**	**Αντιστρεψιμότητα**
$AR(1)$	$y_t = a_1 y_{t-1} + e_t$	$\lvert a_1 \rvert < 1$	Αντιστρέψιμο εφόσον στάσιμο
$AR(2)$	$y_t = a_1 y_{t-1} + a_2 y_{t-2} + e_t$	$a_1 + a_2 < 1$ $a_2 - a_1 < 1$ $\lvert a_2 \rvert < 1$	Αντιστρέψιμο εφόσον στάσιμο
$AR(p)$	$\mathrm{A}(L) y_t = e_t$	Οι ρίζες του πολυωνύμου $\mathrm{A}(L)$ να κείνται εκτός του μοναδιαίου κύκλου	Αντιστρέψιμο εφόσον στάσιμο
$MA(1)$	$y_t = e_t - q_1 e_{t-1}$	Πάντα στάσιμο	$\lvert q_1 \rvert < 1$
$MA(2)$	$y_t = e_t - q_1 e_{t-1} - q_2 e_{t-2}$	Πάντα στάσιμο	$q_1 + q_2 < 1$ $q_2 - q_1 < 1$ $\lvert q_2 \rvert < 1$
$MA(q)$	$y_t = \Theta(L) e_t$	Πάντα στάσιμο	Οι ρίζες του πολυωνύμου $\Theta(L)$ να κείνται εκτός του μοναδιαίου κύκλου
$ARMA(1,1)$	$y_t = a_1 y_{t-1} + e_t - q_1 e_{t-1}$	$\lvert a_1 \rvert < 1$	$\lvert q_1 \rvert < 1$
$ARMA(p,q)$	$\mathrm{A}(L) y_t = \Theta(L) e_t$	Οι ρίζες του πολυωνύμου $\mathrm{A}(L)$ να κείνται εκτός του μοναδιαίου κύκλου	Οι ρίζες του πολυωνύμου $\Theta(L)$ να κείνται εκτός του μοναδιαίου κύκλου

Πίνακας 4.2 Συνθήκες Στασιμότητας και Αντιστρεψιμότητας των Υποδειγμάτων $ARMA$

ΚΕΦΑΛΑΙΟ 5

Μεθοδολογία Box-Jenkins και Προβλέψεις Υποδειγμάτων $ARIMA$

Η ανάλυση που κάναμε παραπάνω για τα υποδείγματα AR, MA και $ARMA$ μέχρι τώρα αναφέρονται όλα σε στάσιμες διαδικασίες που σημαίνει ότι ο μέσος, η διακύμανση και οι αυτοδιακυμάνσεις δεν εξαρτώνται από το χρόνο t, δηλαδή ο μέσος και η διακύμανση παραμένουν σταθεροί, ενώ οι αυτοδιακυμάνσεις εξαρτώνται μόνο από τη χρονική υστέρηση s. Στη συνέχεια, θα εξετάσουμε ***μη στάσιμες διαδικασίες***.

Μη στάσιμες διαδικασίες συνήθως είναι οι σειρές οι οποίες παρουσιάζουν τάση ή εποχικές διακυμάνσεις όπως είναι οι σειρές του Α.Ε.Π.(Ακαθάριστου Εγχώριου προϊόντος),των επενδύσεων και των δημόσιων δαπανών αλλά και σειρές τυχαίας διαδρομής(random walk). Αν σε μια χρονολογική σειρά ο μέσος μετατοπίζεται διαχρονικά τότε μιλάμε για σειρά μη στάσιμη ως προς το μέσο ενώ αν η διακύμανση της δεν παραμένει σταθερή τότε μιλάμε για μη στασιμότητα ως προς τη διακύμανσή της.

Βέβαια υπάρχει τρόπος μετατροπής μιας μη στάσιμης σειράς σε στάσιμη κι αυτό γίνεται με τη μεθοδολογία των Box-Jenkins.Η μεθοδολογία αυτή προτείνει την μετατροπή των σειρών σε στάσιμες με τη χρήση πρώτων, δεύτερων κ.τ.λ. διαφορών.

Όταν εξασφαλίσουμε τη στασιμότητα με τις d διαφορές τότε θα ακολουθήσουμε την ανάλυση προσαρμογής του κατάλληλου $ARMA(p,q)$ υποδείγματος στη μετασχηματισμένη σειρά. Έστω ότι οι παρατηρήσεις μιας χρονολογικής σειράς είναι $Y_1, Y_2, ..., Y_n$ τότε παίρνοντας πρώτες διαφορές έχουμε $\Delta Y_t = Y_t - Y_{t-1} = (1-L)Y_t$.

Αν αυτή η μετασχηματισμένη σειρά είναι στάσιμη τότε το υπόδειγμα που προσαρμόζουμε λέγεται ολοκληρωμένο υπόδειγμα πρώτης τάξεως $ARMA(p,q)$ ή αλλιώς
αυτοπαλίνδρομο ολοκληρωμένο υπόδειγμα κινητών μέσων(autoregressive integrated moving average) και συμβολίζεται ως $ARIMA(p,1,q)$.Γενικά, αν d είναι ο αριθμός των διαφορών που πρέπει να πάρουμε σε ένα ολοκληρωμένο υπόδειγμα προκειμένου να γίνει στάσιμο θα έχουμε $\Delta^d Y_t = (1-L)^d Y_t$ τότε το υπόδειγμα που προσαρμόζουμε στην αρχική σειρά Y_t ονομάζεται αυτοπαλίνδρομο υπόδειγμα κινητών μέσων τάξεως (p,d,q) και συμβολίζεται ως $ARIMA(p,d,q)$ με γενική μορφή:

$$A(L)(1-L)^d Y_t = d + \Theta(L)e_t$$

όπου $A(L) = 1 - a_1L - ... - a_pL^p$ και $\Theta(L) = 1 - q_1L - ... - q_qL^q$.

Ορισμός: Μια μη στάσιμη σειρά λέγεται ολοκληρωμένης τάξεως και παριστάνεται ως $I(d)$ όταν μετατρέπεται σε στάσιμη με τη χρήση d αριθμού διαφορών.

Ο όρος ολοκληρωμένη σειρά προέρχεται από τον τρόπο με τον οποίο μια μη στάσιμη διαδικασία προκύπτει από μία στάσιμη. Αυτό γίνεται ολοκληρώνοντας φορές τη μη στάσιμη σειρά για να προκύψει η στάσιμη. Μια στάσιμη σειρά, όπως ο λευκός θόρυβος, θεωρείται ολοκληρωμένη σειρά μηδενικής τάξεως $I(0)$.

5.1 ΥΠΟΔΕΙΓΜΑΤΑ $ARIMA$

Γενικά, να υπόδειγμα $ARMA(p,q)$ που εφαρμόζεται σε μια ολοκληρωμένη σειρά d τάξεως ονομάζεται αυτοπαλίνδρομο ολοκληρωμένο υπόδειγμα κινητού μέσου τάξεως

(p,d,q) και συμβολίζεται ως $ARIMA(p,d,q)$.Για να γίνει πιο κατανοητό, οι τρεις μορφές των παραμέτρων αυτού του υποδείγματος είναι: οι p παράμετροι του αυτοπαλίνδρομου υποδείγματος, ο αριθμός d των διαφορών που απαιτούνται για να γίνει η σειρά στάσιμη, και q οι παράμετροι του υποδείγματος κινητού μέσου. Για παράδειγμα, ένα υπόδειγμα που περιγράφεται ως $ARIMA(1,2,3)$ σημαίνει ότι περιέχει μία αυτοπαλίνδρομη παράμετρο, και τρεις παραμέτρους κινητού μέσου που έχουν υπολογιστεί για την προκύπτουσα σειρά των δεύτερων διαφορών. Όπως προείπαμε, μια $ARMA(p,q)$ διαδικασία στη γενική της μορφή γράφεται ως εξής:

$$Y_t = a_0 + a_1 Y_{t-1} + a_2 Y_{t-2} + ... + a_p Y_{t-p} + e_t + q_1 e_{t-1} + q_2 e_{t-2} + ... + q_q e_{t-q} .$$

Μια $ARIMA(p,d,q)$ διαδικασία μπορεί να διατυπωθεί με τρεις διαφορετικούς τρόπους και να πάρει τρεις διαφορετικές μορφές.

1. Ως συνάρτηση των παρελθουσών τιμών της και των τιμών του διαταρακτικού όρου, τρέχουσας και παρελθουσών. Η μορφή αυτή είναι γνωστή ως **εξίσωση διαφοράς** (difference equation form).

2. Ως συνάρτηση των παρελθουσών τιμών της και της τρέχουσας τιμής του διαταρακτικού όρου. Η μορφή αυτή είναι γνωστή ως η **αντίστροφη μορφή** (inverted form).

3. Ως συνάρτηση μόνο των τιμών του διαταρακτικού, τρέχουσας και παρελθουσών. Η μορφή αυτή είναι γνωστή ως **τυχαία διαταραχή** (random shock form).

5.2 ΜΕΘΟΔΟΛΟΓΙΑ BOX -JENKINS

Η προσέγγιση BOX-JENKINS στην ανάλυση χρονολογικών σειρών είναι μια μέθοδος εύρεσης ενός στατιστικού υποδείγματος $ARIMA(p,d,q)$ η οποία να παριστάνει ικανοποιητικά τη στοχαστική διαδικασία από την οποία προήλθαν τα δεδομένα, δηλαδή το δείγμα μας. Η συγκεκριμένη μέθοδος περιλαμβάνει τρία στάδια:
Την ταυτοποίηση(identification),την εκτίμηση(estimation),και το διαγνωστικό έλεγχο (diagnostic checking).

ΣΤΑΔΙΟ 1:ΤΑΥΤΟΠΟΙΗΣΗ

Στο συγκεκριμένο στάδιο γίνεται η εξειδίκευση ενός *ARIMA* υποδείγματος με βάση τις πληροφορίες που παίρνουμε από το δείγμα. Αυτό σημαίνει ότι καθορίζονται οι τιμές των p ,d και q . Με άλλα λόγια, καθορίζεται ο αριθμός d των διαφορών που απαιτούνται για να μετατραπεί η σειρά σε στάσιμη, εφ’όσον βέβαια δεν είναι, και έπειτα καθορίζεται η τάξη p της αυτοπαλίνδρομης διαδικασίας και η τάξη q της διαδικασίας κινητού μέσου. Για να διαπιστώσουμε αν η σειρά είναι στάσιμη ή όχι, θα εξετάσουμε τη συμπεριφορά της δειγματικής συνάρτησης αυτοσυσχέτισης. Αν οι αυτοσυσχετίσεις συγκλίνουν ταχύτατα προς το μηδέν σημαίνει ότι η σειρά μάλλον είναι στάσιμη. Αντιθέτως, αν οι αυτοσυσχετίσεις φθίνουν με αργό ρυθμό, είναι σοβαρή ένδειξη ότι η σειρά είναι μη στάσιμη, οπότε πρέπει να γίνει στάσιμη. Σε αυτή την περίπτωση θα χρησιμοποιήσουμε τις πρώτες ή τις δεύτερες ή κ.τ.λ. διαφορές για να μετατρέψουμε τη σειρά σε στάσιμη. Αφού η σειρά έχει γίνει στάσιμη, προσδιορίζεται στη συνέχεια η τάξη του υποδείγματος *ARIMA*, δηλαδή προσδιορίζονται οι τιμές του p και του q. Ο προσδιορισμός τους βασίζεται στις δειγματικές απλές και μερικές, αυτοσυσχετίσεις.

ΣΤΑΔΙΟ 2:ΕΚΤΙΜΗΣΗ

Το δεύτερο στάδιο ανάλυσης στη μεθοδολογία Box-Jenkins είναι η εκτίμηση των συντελεστών του υποδείγματος. Δηλαδή, εξετάζουμε την εκτίμηση των p παραμέτρων $a_1, a_2..., a_p$ της αυτοπαλίνδρομης διαδικασίας και των q παραμέτρων $q_1, q_2..., q_q$ της διαδικασίας κινητού μέσου. Αν η σειρά που εξετάζουμε είναι μόνο αυτοπαλίνδρομη, οι παράμετροί της, όπως είδαμε προηγουμένως, μπορούν να εκτιμηθούν με τη μέθοδο των ελαχίστων τετραγώνων. Αν όμως, η σειρά περιέχει και όρους κινητού μέσου τότε για την εκτίμηση των παραμέτρων του κινητού μέσου θα χρησιμοποιηθούν μη γραμμικές μέθοδοι εκτίμησης. Η κατάλληλη μέθοδος εξαρτάται από τη μορφή της στοχαστικής διαδικασίας, που διαμορφώνει τη χρονολογική σειρά και τον αριθμό των διαφορών, που απαιτούνται για τη στασιμότητα μιας χρονολογικής σειράς. Για να αναφερθούμε πιο συγκεκριμένα, σε εμπειρικές εφαρμογές η μαθηματική μορφή του υποδείγματος προσδιορίζει τη μέθοδο εκτίμησης.

ΣΤΑΔΙΟ 3:ΔΙΑΓΝΩΣΤΙΚΟΣ ΕΛΕΓΧΟΣ

Ανεξαρτήτως της μορφής του υποδείγματος, η χρησιμοποίηση του για μελλοντικές προβλέψεις θέτει σαν απαραίτητη προϋπόθεση πως το εκτιμηθέν υπόδειγμα είναι ικανοποιητικό.Ο έλεγχος της καταλληλότητας του υποδείγματος βασίζεται στα παρακάτω τέσσερα κριτήρια:

- Τη σημαντικότητα των συντελεστών του υποδείγματος
- Τη σταθερότητα των συντελεστών του υποδείγματος
- Τις ιδιότητες των καταλοίπων
- Την προβλεπτική ικανότητα του υποδείγματος

Έχοντας κάνει την εκτίμηση και την ταυτοποίηση θα πρέπει να ελέγξουμε αν το συγκεκριμένο υπόδειγμα είναι ικανοποιητικό με την έννοια του κατά πόσο καλά προσαρμόζεται στα δεδομένα(fitting). Σε γενικές γραμμές, ο πιο καλός τρόπος για να ελέγξουμε την προσαρμοστικότητα ενός υποδείγματος είναι να εξετάσουμε την προβλεπτική του ικανότητα(forecasting ability) έξω από τη δειγματοληπτική περίοδο. Συνήθως όμως επειδή δεν υπάρχει αρκετός αριθμός δεδομένων για αυτή τη διαδικασία, τόσο η ταυτοποίηση αλλά και η εκτίμηση και ο έλεγχος γίνονται με το ίδιο δείγμα παρατηρήσεων.

Θα ασχοληθούμε με τον έλεγχο των καταλοίπων. Αν ο ερευνητής θεωρεί το εκτιμηθέν υπόδειγμα κατάλληλο για τα δεδομένα μας, αν δηλαδή εκφράζει ικανοποιητικά τη διαδικασία από την οποία προέρχονται τα δεδομένα, τότε τα κατάλοιπα θα πρέπει να συμπεριφέρονται σαν μια διαδικασία λευκού θορύβου(white noise).Αυτό σημαίνει πως δεν πρέπει να υπάρχει αυτοσυσχέτιση μεταξύ των καταλοίπων. Αυτός ο έλεγχος για τα κατάλοιπα γίνεται με τη στατιστική Q των Box-Pierce, με την οποία ελέγχεται από κοινού η σημαντικότητα ενός αριθμού συντελεστών αυτοσυσχέτισης, έστω m. Η μηδενική υπόθεση τότε, θα είναι $H_0 : r_1 = r_2 = ... = r_m = 0$,όπου $r_i, i = 1, 2, ..., m$ είναι οι συντελεστές συσχέτισης των καταλοίπων. Η στατιστική Q των Box-Pierce ορίζεται ως:

$Q_{BP} = T\sum_{s=1}^{m} \hat{r}_s^{\,2}$, όπου $\hat{r}_s$ είναι δειγματικές αυτοσυσχετίσεις των καταλοίπων και T ο αριθμός των παρατηρήσεων. Ο αριθμός των αυτοσισχετίσεων ισούται με την τετραγωνική ρίζα του αριθμού των παρατηρήσεων, δηλαδή ισχύει $m = \sqrt{T}$. Η στατιστική Q_{BP} ακολουθεί την κατανομή X^2 με $m - p - q$ βαθμούς ελευθερίας. Η μηδενική υπόθεση θα απορρίπτεται αν η τιμή Q_{BP} είναι μεγαλύτερη από την κρίσιμη τιμή της κατανομής X_a^2, για δεδομένο επίπεδο σημαντικότητας a . Οπότε θα ισχύουν τα παρακάτω:

- Απόρριψη : H_0, αν $Q_{BP} > X_a^2$
- Αποδοχή : H_0,αν $Q_{BP} \le X_a^2$

Το κριτήριο Q έχει υποστεί διάφορες τροποποιήσεις από την αρχική του διατύπωση. Έτσι στις πρακτικές εφαρμογές σήμερα χρησιμοποιείται κυρίως η τροποποίηση που έκαναν οι Ljung και Box, η οποία ορίζεται ως:

$$Q_{LB} = n(n+2)\sum_{j=1}^{k} \frac{r_j^2(\hat{e})}{n-j}$$

Κι αυτή η στατιστική ακολουθεί προσεγγιστικά την κατανομή X_a^2 με $m - p - q$ βαθμούς ελευθερίας. Όπως και στο Q_{BP} έτσι και εδώ η μηδενική υπόθεση θα απορρίπτεται όταν $Q_{LB} > X_a^2$ για δεδομένο επίπεδο σημαντικότητας a . Η στατιστική

Q_{LB} χρησιμοποιείται περισσότερο για μικρά δείγματα ενώ για μεγάλα μπορούν να χρησιμοποιηθούν το ίδιο κατάλληλα και οι δύο.

5.3 Κριτήρια Επιλογής Υποδειγμάτων

Αυτός ο έλεγχος γίνεται με μια διαδικασία που ονομάζεται υπερπροσαρμογή (overfitting). Σύμφωνα με αυτή τη διαδικασία ο έλεγχος της καταλληλότητας του εκτιμημένου υποδείγματος γίνεται συγκρίνοντάς το με ένα άλλο υπόδειγμα μεγαλύτερης τάξης. Με λίγα λόγια το εκτιμημένο υπόδειγμα $ARMA(p,q)$, θα συγκριθεί με τα υποδείγματα $ARIMA(p+1,q)$ και $ARIMA(p,q+1)$ της αμέσως επόμενης τάξης. Αν το εκτιμημένο υπόδειγμα είναι τελικά το καταλληλότερο για τα δεδομένα μας, δηλαδή αν περιγράφει τη διαδικασία από την οποία παράχθηκαν τα δεδομένα, θα πρέπει οι επιπλέον συντελεστές στα μεγαλύτερα υποδείγματα να μην είναι στατιστικά διαφορετικοί από το μηδέν. Αν αυτοί οι συντελεστές δεν είναι μηδέν, τότε θα υπάρχει κάποιο άλλο υπόδειγμα που να είναι πιο κατάλληλο για τα δεδομένα μας, απ' ότι το εκτιμημένο.

Τέλος, θα αναφέρουμε κάποια κριτήρια που μας βοηθούν να επιλέξουμε το κατάλληλο υπόδειγμα. Είναι προφανές ότι αν αυξήσουμε την τάξη του υποδείγματος προσθέτοντας υστερήσεις είτε για το αυτοπαλίνδρομο τμήμα είτε για το τμήμα κινητού μέσου, θα μειώνεται το άθροισμα των τετραγώνων των καταλοίπων, αλλά ταυτόχρονα θα μειώνονται και οι βαθμοί ελευθερίας αφού εκτιμώνται περισσότερες παράμετροι. Δυο κριτήρια που χρησιμοποιούνται ευρέως στην ανάλυση χρονολογικών σειρών είναι το κριτήριο πληροφοριών Akaike (Akaike Information Criterion) ή αλλιώς AIC και το Μπαϊεσιανό κριτήριο Schwartz (Schwartz Bayesian Criterion) ή αλλιώς SBC. Τα κριτήρια αυτά ορίζονται ως εξής:

$$AIC = \ln\left(s^2\right) + \frac{2n}{T}$$

$$SBC = \ln\left(s^2\right) + n\ln\left(T\right)$$

Όπου:

s^2 =εκτίμηση της διακύμανσης των καταλοίπων
n =αριθμός εκτιμούμενων παραμέτρων υποδείγματος ($p+q+1$) όπου η μονάδα αντιστοιχεί στην σταθερά αν υπάρχει
T = αριθμός παρατηρήσεων που χρησιμοποιούνται στην παλινδρόμηση

Η προσθήκη μιας επιπλέον μεταβλητής στο υπόδειγμα μειώνει το άθροισμα των τετραγώνων των καταλοίπων άρα και τη διακύμανση s^2,αλλά ταυτόχρονα αυξάνει το n στους τύπους του AIC και SBC αντίστοιχα. Επομένως αν η προστιθέμενη μεταβλητή δεν έχει ερμηνευτική ικανότητα, τότε οι τιμές και των δύο κριτηρίων θα αυξηθούν. Η επιλογή δηλαδή των υποδειγμάτων γίνεται με βάση **τη μικρότερη τιμή των κριτηρίων.**

Με άλλα λόγια, από ένα αριθμό υποδειγμάτων με διαφορετικό αριθμό παραμέτρων επιλέγουμε εκείνο με τη μικρότερη τιμή AIC ή SBC.

Από τα δύο κριτήρια αυτά, το SBC θεωρείται ασυμπωτικά καλύτερο. Επειδή το $\ln(T) > 2$, το SBC επιβάλει μεγαλύτερη ποινή από το AIC στον επιπλέον αριθμό εκτιμούμενων παραμέτρων. Έτσι το κριτήριο SBC οδηγεί πάντα στην επιλογή ενός υποδείγματος του οποίου ο αριθμός των παραμέτρων δεν είναι σε καμία περίπτωση μεγαλύτερος από αυτόν που επιλέχτηκε με το κριτήριο AIC.

Τα δύο αυτά κριτήρια χρησιμοποιούνται στην επιλογή του καταλληλότερου υποδείγματος $ARIMA$ από πλευράς αριθμού υστερήσεων που θα πρέπει να περιληφθούν. Επίσης εφαρμόζονται και σε άλλα υποδείγματα, όπως τα υποδείγματα κατανεμημένων χρονικών υστερήσεων για την επιλογή του αριθμού των υστερήσεων των ανεξάρτητων μεταβλητών. Αν η σύγκριση γίνεται μεταξύ υποδειγμάτων με ίδιο αριθμό παραμέτρων, τότε τα κριτήρια AIC και SBC οδηγούν επιλογή του υποδείγματος με μεγαλύτερο R^2.

5.4 Προβλέψεις με AR και MA Υποδείγματα

Έχοντας διαπιστώσει το κατάλληλο υπόδειγμα μορφής $AR(p)$, $MA(q)$,ή $ARMA(p,q)$ που προσαρμόζεται καλύτερα στα δεδομένα μιας χρονολογικής σειράς και αφού έχει προηγηθεί η εκτίμηση και ο έλεγχος τώρα μπορούμε να το χρησιμοποιήσουμε για προβλέψεις. Η πρόβλεψη των τιμών μιας χρονοσειράς σε μελλοντικές περιόδους γίνεται με βάση το εκτιμημένο υπόδειγμα που προσαρμόσαμε στα δεδομένα Για παράδειγμα, έστω $Y_1, Y_{2,} ..., Y_T$,τα ιστορικά δεδομένα μιας χρονολογικής σειράς στην οποία προσαρμόσαμε ένα υπόδειγμα $ARMA(p,q)$.Η πρόβλεψη την επόμενη περίοδο $T+1$ θα είναι η υπο συνθήκη αναμενόμενη τιμή(conditional expectation) της σειράς αυτής, δηλαδή η τιμή αυτή αποτελεί την πιο πιθανή τιμή της σειράς στο μέλλον με βάση όλες τις προηγούμενες τιμές της. Μια πρόβλεψη που κάνουμε την περίοδο T ,για την επόμενη περίοδο $T+1$,συμβολίζεται με Y_{t+1} και εκφράζεται ως: $Y_{t+1} = E\frac{(Y_{t+1})}{(Y_{T,...}Y_1)} = E_T(Y_{T+1})$

όπου ο υποδείκτης T στην αναμενόμενη τιμή δηλώνει ότι στηρίζεται στις πληροφορίες μέχρι και την περίοδο T. Γενικότερα, θα συμβολίζουμε την πρόβλεψη για h περιόδους ως $Y_{T+h} = E_T(Y_{T+h})$.Το λάθος πρόβλεψης συμβολίζεται ως e_{T+h} και είναι η διαφορά μεταξύ της τιμής του Y_t την περίοδο $T+h$ και της προβλεφθείσης τιμής, δηλαδή $e_{T+h} = Y_{T+h} - Y_{T+h}$ όπου
$h = 1,2,3,...,$ περίοδοι στο μέλλον.

ΠΡΟΒΛΕΨΕΙΣ ΜΕ $AR(1)$ ΥΠΟΔΕΙΓΜΑΤΑ

Έστω το υπόδειγμα: $Y_t = a_0 + a_1 Y_{t-1} + e_t$

Για $t = T+1$ το υπόδειγμα γίνεται

$$Y_{t+1} = a_0 + a_1 Y_{t-1} + e_t$$

Αν οι παράμετροι a_0 και a_1 είναι γνωστές, τότε με βάση τις πληροφορίες που έχουμε μέχρι την περίοδο T μια πρόβλεψη για την περίοδο $T+1$ θα είναι η υπό συνθήκη προσδοκώμενη τιμή της Y_{t+1}. Αν θέσουμε ως $\hat{Y}_{t+1}$ την πρόβλεψη τότε: $\hat{Y}_{t+1} = E_T Y_{T+1}$ και γενικεύοντας θα έχουμε: $\hat{Y}_{t+h} = E_T Y_{T+h}$ που θα αποτελεί μια πρόβλεψη για h περιόδους μπροστά με βάση τις πληροφορίες που έχουμε μέχρι την περίοδο T. Αυτή η πρόβλεψη που ελαχιστοποιεί την ποσότητα $E_T\left(Y_{T+h} - \hat{Y}_{T+h}\right)^2$ δηλαδή ελαχιστοποιεί το μέσο του τετραγώνου του σφάλματος (Mean Square Error), θεωρείται **άριστη πρόβλεψη** (optimal forecast). Γενικότερα, για τις προβλέψεις που γίνονται για h περιόδους ισχύουν τα εξής:

$$\hat{Y}_{T+h} = a_0 + a_1 Y_{T+h-1}$$

$$Var\left(\hat{e}_{T+h}\right) = s_e^{\,2}\left(1 + a_1^{\,2} + a_1^{\,4} + \ldots + a_1^{\,2(h-1)}\right)$$

Από την τελευταία σχέση είναι φανερό ότι η διακύμανση του σφάλματος πρόβλεψης αυξάνει μη γραμμικά, καθώς αυξάνει η περίοδος πρόβλεψης.

ΠΡΟΒΛΕΨΕΙΣ ΜΕ $MA(1)$ ΥΠΟΔΕΙΓΜΑΤΑ

Έστω το υπόδειγμα $MA(1)$ που δίνεται από τον εξής τύπο:

$$Y_t = m + e_t + q_1 e_{t-1}$$

Η πρόβλεψη για την περίοδο $T+1$, όπως είδαμε και πριν θα είναι:

$$\hat{Y}_{T+1} = E_T\left(Y_{T+1}\right)$$

Γενικά, για προβλέψεις για $h > 1$ περιόδους μπροστά ισχύει:

$$\hat{Y}_{T+h} = m$$

$$V\left(\hat{e}_{T+h}\right) = s^{\,2}\left(1 + q_1^{\,2}\right)$$

Από την τελευταία σχέση είναι φανερό ότι ένα υπόδειγμα κινητού μέσου πρώτης τάξης είναι κατάλληλο για προβλέψεις μόνο μια περίοδο μπροστά, αφού για $h > 1$ η πρόβλεψη θα είναι πάντοτε ο μέσος.

ΠΡΟΒΛΕΨΕΙΣ ΜΕ $ARMA(1,1)$ ΥΠΟΔΕΙΓΜΑΤΑ

Έστω το υπόδειγμα $ARMA(1,1)$

$$Y_t = a_0 + a_1 Y_{t-1} + e_t + q_1 e_{t-1}$$

Για $t = T+1$ το υπόδειγμα γίνεται:

$$Y_{T+1} = a_0 + a_1 Y_T + e_{T+1} + q_1 e_T$$

Η άριστη πρόβλεψη για την επόμενη χρονική περίοδο θα είναι

$$\hat{Y}_{T+1} = a_0 + a_1 Y_T + q_1 e_T$$

Το σφάλμα πρόβλεψης και η διακύμανσή του θα δίνονται από τις παρακάτω σχέσεις:

$$\hat{e}_{T+1} = Y_{T+1} - \hat{Y}_{T+1} = e_{T+1}$$

$$V\left(\hat{e}_{T+1}\right) = V\left(e_{T+1}\right) = s^2$$

Έστω το υπόδειγμα $ARIMA(1,1,1)$:

$$W_t = a_0 + a_1 w_{t-1} + e_t \text{ όπου}$$

$$W_t = \Delta Y_t = Y_t - Y_{t-1}$$

Παίρνουμε τις πρώτες διαφορές οι οποίες είναι $AR(1)$ για να γίνει στάσιμη η μη στάσιμη η σειρά Y_t. Οπότε, θα γίνει πρόβλεψη για τη διαφορά w και στη συνέχεια θα γίνει πρόβλεψη της Y. Η άριστη πρόβλεψη για την w_{T+1} είναι:

$$\hat{w}_{T+1} = a_0 + a_1 w_T$$

Η πρόβλεψη τώρα για την αρχική σειρά Y, δηλαδή η πρόβλεψη για την Y_{T+1} θα είναι:

$$\begin{aligned}\hat{Y}_{T+1} &= Y_T + \hat{w}_{T+1} = Y_T + a_0 + a_1 w_T = \\ &= Y_T + a_0 + a_1\left(Y_T - Y_{T-1}\right) = \\ &= a_0 + \left(1 + a_1\right) Y_T - a_1 Y_{T-1}\end{aligned}$$

Γενικά, η πρόβλεψη για h περιόδους μπροστά θα είναι:

$$\hat{Y}_{T+h} = Y_T + \hat{w}_{T+1} + \hat{w}_{T+2} + \ldots + \hat{w}_{T+h}$$

Το τελευταίο βήμα στη μεθοδολογία των προβλέψεων με τη χρήση $ARMA(p,q)$ υποδείγματα έχει να κάνει με την αξιοπιστία των προβλέψεων. Έτσι κάνοντας μια πρόβλεψη δεν μας ενδιαφέρει μόνο ένα σημείο πρόβλεψης $\hat{y}_{T+h}$ αλλά και ένα μέτρο αξιοπιστίας της πρόβλεψης, δηλαδή το διάστημα εμπιστοσύνης της πρόβλεψης. Ένα $(1-a)\%$ διάστημα πρόβλεψης

θα έχει τη μορφή: $\hat{y}_{T+h} \pm z_c \begin{pmatrix} diak\acute{\upsilon}mansh \\ l\acute{a}qouV \\ pr\acute{o}bleyhV \end{pmatrix}^{1/2}$

Όπου z_c είναι η $a/2$ κριτική τιμή για την τυπική κανονική κατανομή $N(0,1)$.

5.5 ΜΕΤΡΑ ΑΞΙΟΛΟΓΗΣΗΣ ΠΡΟΒΛΕΨΕΩΝ

Για να αξιολογήσουμε την προβλεπτική ικανότητα ενός υποδείγματος θα πρέπει να συγκρίνουμε τις προβλέψεις με τα πραγματικά δεδομένα της χρονολογικής σειράς. Κάτι τέτοιο υποδηλώνει πως διαθέτουμε παρατηρήσεις ακόμα και για τις περιόδους μέσα στις οποίες κάνουμε τις προβλέψεις. Στη συγκεκριμένη περίπτωση μιλάμε για εκ των υστέρων προβλέψεις(ex-post) δηλαδή προβλέψεις που γίνονται μέσα στην περίοδο όπου έχουν ήδη πραγματοποιηθεί οι τιμές που θέλουμε να προβλέψουμε. Αντιθέτως, όταν προβλέπουμε σήμερα τις μελλοντικές τιμές της μεταβλητής τότε αναφερόμαστε για εκ των προτέρων(ex-ante) προβλέψεις, δηλαδή προβλέψεις τιμών για περιόδους που δεν έχουμε ακόμα αληθινές τιμές.

Ουσιαστικά οι προβλέψεις που κάνουμε με ένα συγκεκριμένο υπόδειγμα μπορούν να αξιολογηθούν μόνο ex-post.Για να εξυπηρετήσουμε το σκοπό αυτό, από ένα συνολικό δείγμα παρατηρήσεων που διαθέτουμε αφήνουμε τις τελευταίες παρατηρήσεις τις οποίες δεν συμπεριλαμβάνουμε στην εκτίμηση προκειμένου να τις συγκρίνουμε με τις προβλέψεις που κάνουμε στο διάστημα αυτό. Απαραίτητη προϋπόθεση είναι να έχουμε μεγάλο αριθμό παρατηρήσεων ώστε να υπάρχουν οι απαραίτητοι βαθμοί ελευθερίας για την εκτίμηση και τους ελέγχους του επιλεγμένου υποδείγματος. Η αξιολόγηση των προβλέψεων γίνεται με κάποια στατιστικά μέτρα που στηρίζονται στο μέγεθος του λάθους πρόβλεψης που κάνουμε. Έχουμε λοιπόν:

A_t =οι πραγματικές τιμές της χρονολογικής σειράς

F_t =οι προβλεφθείσες τιμές

$$e_t = F_t - A_t = \text{το λάθος πρόβλεψης}$$

Στο σημείο θα παραθέσουμε κάποιους τύπους που χρησιμοποιούνται για τη μέτρηση της ακρίβειας των ex-post προβλέψεων:

Μέσο σφάλμα τετραγώνου(Mean Squared Error)

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

Τετραγωνική ρίζα MSE (Root Mean Squared Error)

$$RMSE = \sqrt{\frac{\sum e_t^2}{N}}$$

Μέσο Απόλυτο Σφάλμα(Mean Absolute error)

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

Μέσο Απόλυτο Ποσοστιαίο Σφάλμα(Mean Absolute Percentage Error)

$$MAPE = \frac{100}{N}\sum \left|\frac{F_t - A_t}{A_t}\right|$$

Το μέσο σφάλμα τετραγώνου(MSE) αλλά και η τετραγωνική ρίζα(RMSE) δίνουν μεγαλύτερη βαρύτητα στα μεγάλα λάθη διότι τετραγωνίζονται σε αντίθεση με τα MAE και MAPE που υπολογίζουν μόνο τα απόλυτα σφάλματα. Όσο μικρότερες είναι οι τιμές των παραπάνω μεγεθών τόσο καλύτερη είναι και η προβλεπτική ικανότητα του υποδείγματος. Μειονέκτημα αποτελεί το γεγονός πως όλα τα παραπάνω μεγέθη επηρεάζονται από τις μονάδες μέτρησης των μεταβλητών κάτι που σημαίνει πως απαιτείται προσοχή κατά τη σύγκριση μεταξύ εναλλακτικών ώστε η προβλεπόμενη μεταβλητή να εκφράζεται στις ίδιες μονάδες.

Ένα άλλο μέγεθος που θα εξετάσουμε και είναι ανεξάρτητο των μονάδων μέτρησης είναι ο συντελεστής ανισότητας του Theil που συμβολίζεται με το γράμμα U και έχει τον εξής τύπο:

$$U = \frac{\sqrt{\frac{1}{N}\sum (F_t - A_t)^2}}{\sqrt{\frac{1}{N}\sum (A_t)^2}} = \frac{RMSE}{\sqrt{\frac{1}{N}\sum (A_t)^2}}$$

O Theil πρότεινε τη χρήση ποσοστιαίων μεταβολών στη θέση των A και F, αλλά ο τύπος χρησιμοποιείται και στις αρχικές τιμές. Ο συγκεκριμένος συντελεστής παίρνει την τιμή μηδέν όταν οι προβλέψεις είναι απόλυτα ακριβείς και την τιμή ένα όταν οι προβλέψεις είναι όλες μηδέν. Αντίθετα, αν ο συντελεστής υπερβαίνει τη μονάδα τότε οι προβλέψεις δεν είναι καθόλου καλές. Συμπερασματικά, όσο πιο κοντά στο μηδέν είναι ο συντελεστής U τόσο καλύτερες είναι οι προβλέψεις. Η σχέση του συντελεστή U αναλύεται σε τρεις συνιστώσες $UM + UV + UC = 1$ όπου:

$UM =$ ποσοστό μεροληψίας (bias proportion)
$UV =$ ποσοστό διακύμανσης (variance proportion)
$UC =$ ποσοστό συνδιακύμανσης (covariance proportion)

Χαρακτηριστικό είναι πως οι καλές προβλέψεις θα έχουν μικρό ποσοστό μεροληψίας και διακύμανσης και μεγάλο ποσοστό συνδιακύμανσης.

5.6 Εποχικά Υποδείγματα *SARIMA*

Τα εποχικά υποδείγματα *SARIMA* είναι μια άλλη κατηγόρια των υποδειγμάτων *ARIMA*. Γενικότερα, στοιχεία μικρότερης διάρκειας του έτους όπως μηνιαία, τριμηνιαία και υπόλοιπα στοιχεία μας φανερώνουν την εποχικότητα. Το εποχικό μέρος του υποδείγματος *ARIMA* έχει παρόμοια δομή με αυτή ενός μη εποχικού υποδείγματος και γράφεται ως $SARIMA(Sp, Sd, Sq)$ x $ARIMA(p, d, q)$. Στο εποχικό μέρος διεξάγονται πολλαπλασιασμοί της χρονικής υστέρησης S (δηλαδή τον αριθμό των περιόδων για μια εποχή) με τους συντελεστές p που είναι ο αριθμός των αυτοπαλίνδρομων εποχικών όρων (SAR), των αριθμό των εποχικών διαφορών d και το εύρος των στοιχείων των εποχικών όρων του κινητού μέσου q (SMA). Το φαινόμενο της εποχικότητας αποτελεί μια κανονική κύμανση μέσα στο χρονολογικό έτος οδηγώντας σε υψηλή συσχέτιση ανάμεσα στις τιμές τις σειράς που αντιστοιχούν στην ίδια περίοδο ανάμεσα στα διαφορετικά έτη. Την καλοκαιρινή περίοδο διαπιστώνεται ένας αυξημένος όγκος των πωλήσεων στα αναψυκτικά σε σχέση με της άλλες περιόδους του έτους λόγω της ζεστής. Αυτό είναι ένα παράδειγμα της εποχικότητας. Μια πρώτη αντιμετώπιση αφορά την αφαίρεση της εποχικότητας από την χρονολογική σειρά και την χρησιμοποίηση της μεθόδου που έχουμε μελετήσει, την γνωστή ως Box-Jenkins. Όποτε και εφαρμόζουμε της τεχνικές της ταυτοποίησης, εκτίμησης και πρόβλεψης. Μια άλλη μέθοδος η οποία είναι και η επικρατέστερη αφορά την ενσωμάτωση του εποχικού προτύπου των στοιχείων μας στο κανονικό υπόδειγμα *ARIMA* και την χρησιμοποίηση της μεθοδολογίας Box-Jenkins. Η μέθοδος αυτή καταλήγει στην εκτίμηση υποδειγμάτων με περισσότερες παραμέτρους.

ΠΡΑΚΤΙΚΟ ΜΕΡΟΣ

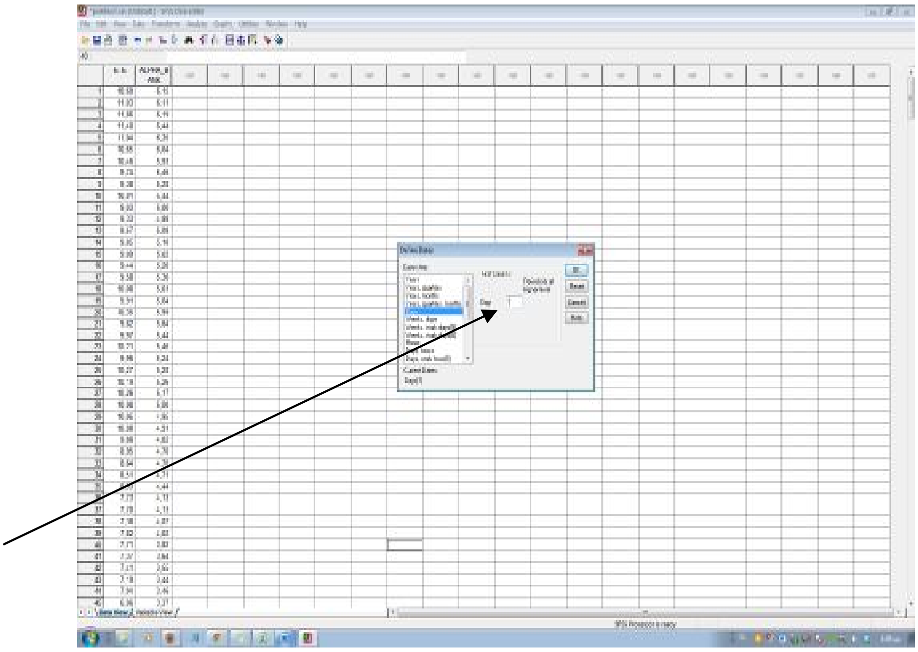
Στο σημείο αυτό θα εξεταστούν οι τιμές των μετοχών(Εθνική Τράπεζα,ALPHA BANK) για ένα συγκεκριμένο χρονικό διάστημα.Να σημειωθεί πως τα δεδομένα πάρθηκαν από το www.naftempriki.gr.

Παράδειγμα ARIMA στο SPSS για τις μετοχές της ΕΤΕ και της ALPHA-BANK

Εθνική Τράπεζα της Ελλάδος (ΕΤΕ)

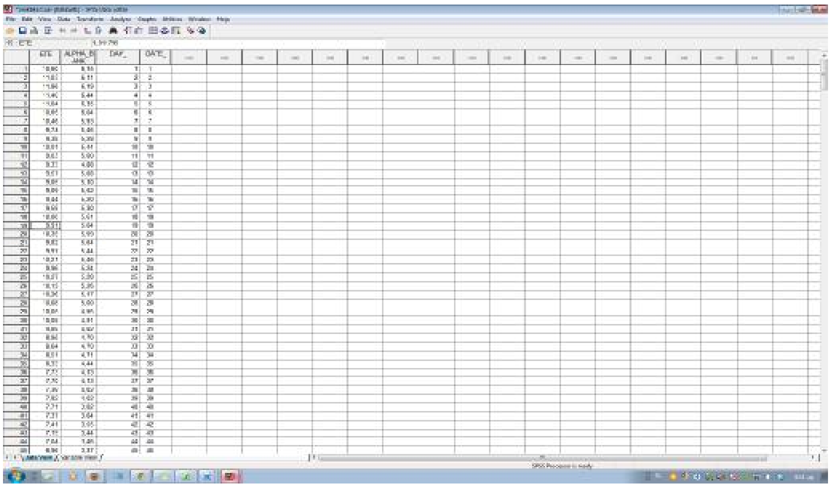
Στο συγκεκριμένο παράδειγμα έχουν περαστεί τιμές της ΕΤΕ στο SPSS για το διάστημα από 01/01/09 έως 07/07/10 δηλαδή 376 τιμές.

Εικόνα 1 (Define dates)



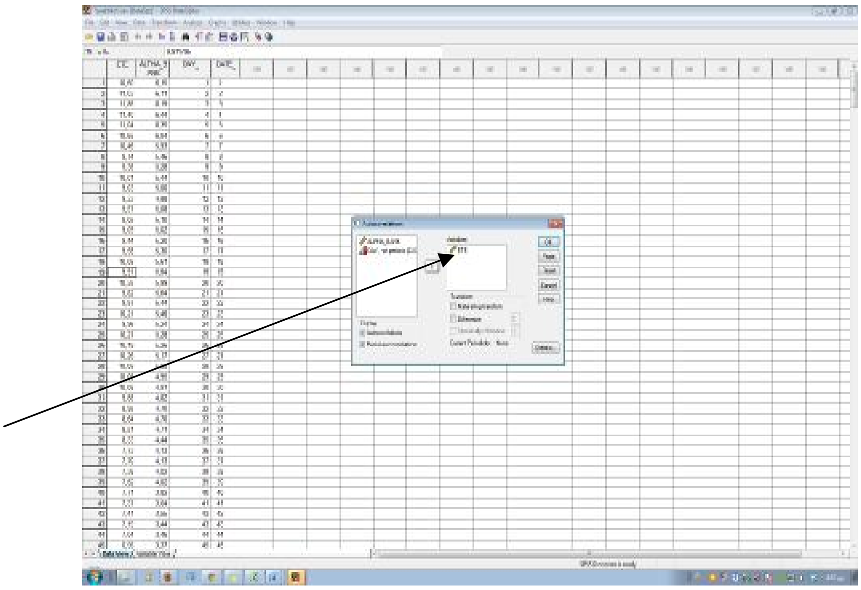
Έπειτα από το menu του SPSS και συγκριμένα από την επιλογή Data-Define Dates-Days μπαίνει Day=1 το οποίο ορίζει ότι η περιοδικότητα των δεδομένων είναι ανά ημέρα αφού πρόκειται για τιμές κλεισίματος της μετοχής.

Εικόνα 2 (Μεταβλητές DAY_,DATE_)

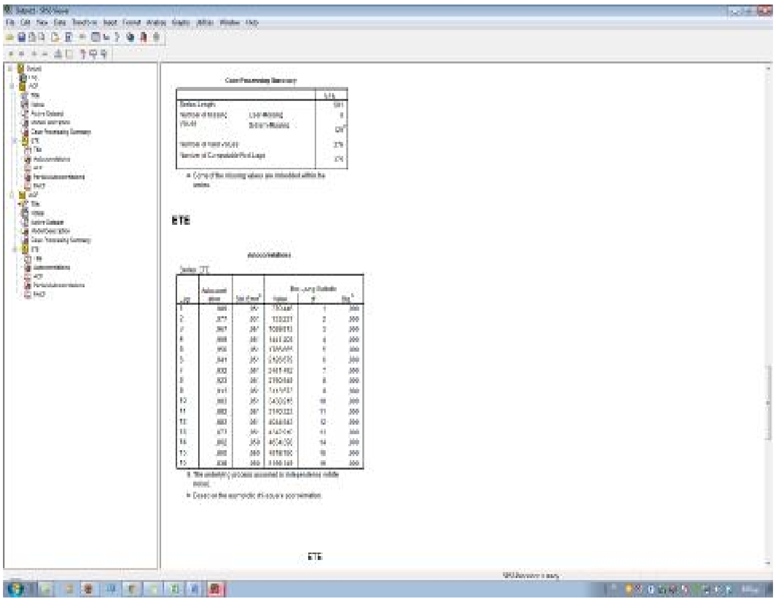


Δοκιμάζοντας από το menu του SPSS και πηγαίνοντας στο Graphs-Time Series-Autocorrelations τοποθετείται η μεταβλητή ΕΤΕ στο Variables για να γίνουν οι δοκιμές και να βρεθεί το καταλληλότερο υπόδειγμα κάνοντας τους ελέγχους στασιμότητας. Οπότε δοκιμάζεται αρχικά **χωρίς διαφορές** επιλέγεται το ok για να προκύψουν τα εξής αποτελέσματα στο output.

Εικόνα 3 (Διαδικασία για έλεγχο στασιμότητας**)**



Εικόνα 4 (Πίνακας Αυτοσυσχετίσεων sto Output)



Πίνακας 1 (Αυτοσυσχετίσεις)

Autocorrelations

Series: ETE

Lag	Autocorrelation	Std. Error[a]	Box-Ljung Statistic Value	df	Sig.[b]
1	,989	,051	370,446	1	,000
2	,977	,051	733,237	2	,000
3	,967	,051	1089,873	3	,000
4	,959	,051	1441,209	4	,000
5	,950	,051	1786,665	5	,000
6	,941	,051	2126,679	6	,000
7	,932	,051	2461,462	7	,000
8	,923	,051	2790,549	8	,000
9	,913	,051	3113,573	9	,000
10	,903	,051	3430,215	10	,000
11	,892	,051	3740,323	11	,000
12	,883	,051	4044,543	12	,000
13	,873	,051	4342,910	13	,000
14	,862	,050	4634,398	14	,000
15	,850	,050	4918,750	15	,000
16	,838	,050	5196,149	16	,000

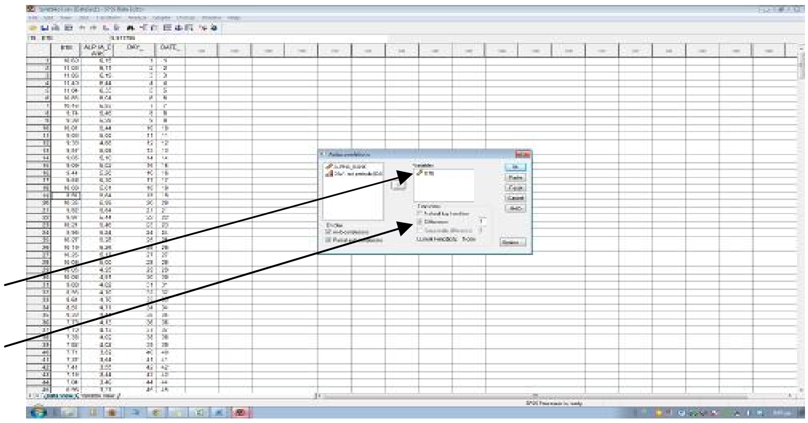
a. The underlying process assumed is independence (white noise).

b. Based on the asymptotic chi-square approximation.

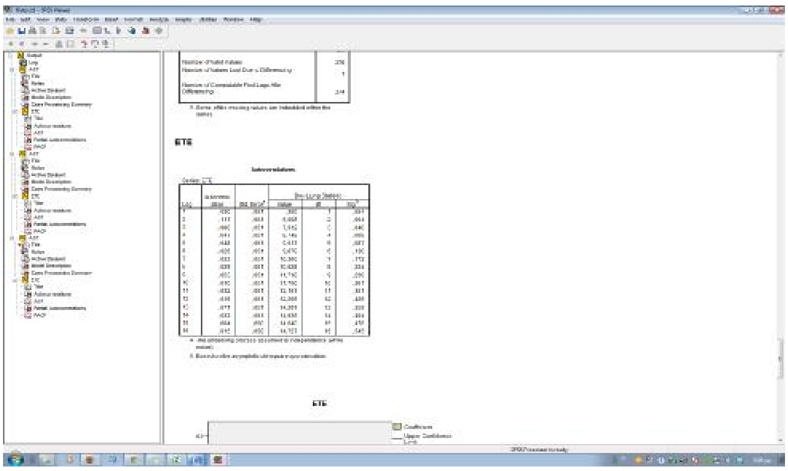
- Χρονολογικές υστερήσεις(Lag),
- Τιμή του συντελεστή αυτοσυσχέτισης(Autocorrelation),
- Τυπικό σφάλμα(Std Error),
- Τιμή της στατιστικής συνάρτησης Box-Ljung(Value) ,
- Βαθμοί ελευθερίας(df),
- Τιμή πιθανότητας της στατιστικής συνάρτησης(Sig)

Από τον παραπάνω πίνακα είναι ξεκάθαρο ότι για όλα τα lags υπάρχει αυτοσυσχέτιση γατί όλα τα ρ είναι μηδέν οπότε δεν είναι στάσιμη χωρίς διαφορές και θα δοκιμαστεί με **πρώτες διαφορές.** Πατώντας πάλι οκ εμφανίζονται τα εξής στον πίνακα Autocorrelations.

Εικόνα 5 (Διαδικασία για έλεγχο στασιμότητας)



Εικόνα 6 (Πίνακας Αυτοσυσχετίσεων sto Output)



Πίνακας 2 (Αυτοσυσχετίσεις)

Autocorrelations

Series: ETE

Lag	Autocorrelation	Std. Error [a]	Box-Ljung Statistic Value	df	Sig. [b]
1	,030	,051	,350	1	,554
2	-,117	,051	5,508	2	,064
3	-,080	,051	7,912	3	,048
4	,047	,051	8,748	4	,068
5	-,048	,051	9,617	5	,087
6	-,026	,051	9,878	6	,130
7	,033	,051	10,306	7	,172
8	,029	,051	10,628	8	,224
9	,053	,051	11,718	9	,230
10	,010	,051	11,760	10	,301
11	-,032	,051	12,167	11	,351
12	,015	,051	12,255	12	,425
13	,071	,051	14,207	13	,359
14	,033	,051	14,635	14	,404
15	,004	,050	14,640	15	,478
16	,015	,050	14,727	16	,545

a. The underlying process assumed is independence (white noise).

b. Based on the asymptotic chi-square approximation.

Έλεγχος στασιμότητας

$H_0 : r = 0$

vs

$H_1 : r \neq 0$

Με την υπόθεση H_0 δεν υπάρχει αυτοσυσχέτιση, άρα είναι στάσιμη ενώ με την H_1 υπάρχει αυτοσυσχέτιση και δεν προκύπτει στασιμότητα.

1) H_o: $\rho_1=0$ $\rho_1=0.554>\alpha=0.05$ Αποδοχή H_0 (δεν υπάρχει αυτοσυσχέτιση, άρα στάσιμη)

vs

H_1: $\rho_1\neq0$

2) H_o: $\rho_2=0$ $\rho_2=0.064>\alpha=0.05$ Αποδοχή H_0 (δεν υπάρχει αυτοσυσχέτιση, άρα στάσιμη)

vs

H_1: $\rho_2\neq0$

3) H_o: $\rho_3=0$ $\rho_3=0.048<\alpha=0.05$ Απόρριψη H_0 (υπάρχει αυτοσυσχέτιση)

vs

H_1: $\rho_3\neq0$

4) H_0: $\rho_4=0$ $\rho_4=0.068>\alpha=0.05$ Αποδοχή H_0 (δεν υπάρχει αυτοσυσχέτιση,

vs άρα στάσιμη)

H_1: $\rho_4\neq0$

5) H_o: $\rho_5=0$ $\rho_5=0.087>\alpha=0.05$ Αποδοχή H_0 (δεν υπάρχει αυτοσυσχέτιση,

vs άρα στάσιμη)

H_1: $\rho_5\neq0$

6) H_o: $\rho_6=0$ $\rho_6=0.13>\alpha=0.05$ Αποδοχή H_0 (δεν υπάρχει αυτοσυσχέτιση,

vs άρα στάσιμη)

H_1: $\rho_6\neq0$

7) H_o: $\rho_7=0$ $\rho_7=0.172>\alpha=0.05$ Αποδοχή H_0 (δεν υπάρχει αυτοσυσχέτιση,

vs άρα στάσιμη)

H_1: $\rho_7\neq0$

8) H_o: $\rho_8=0$ $\rho_8=0.224>\alpha=0.05$ Αποδοχή H_0 (δεν υπάρχει αυτοσυσχέτιση,

vs άρα στάσιμη)

H_1: $\rho_8\neq0$

…..

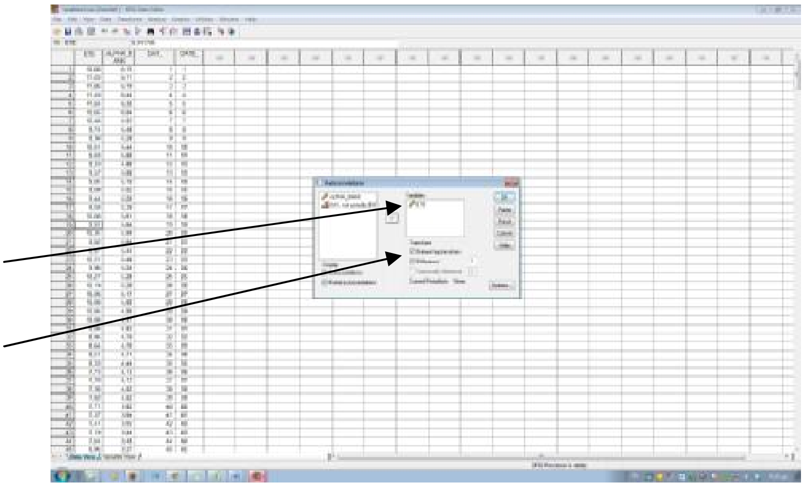
16) H_o: $\rho_{16}=0$ $\rho_1=0.545>\alpha=0.05$ Αποδοχή H_0 (δεν υπάρχει αυτοσυσχέτιση,

vs άρα στάσιμη)

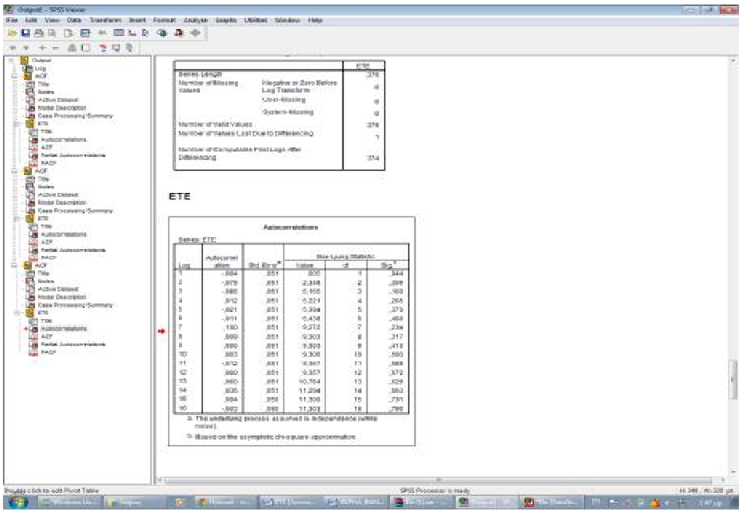
H_1: $\rho_{16}\neq0$

Επειδή το ρ_3 εμφανίζει αυτοσυσχέτιση θα ήταν καλύτερα να γίνει δοκιμή με **πρώτες διαφορές και με λογάριθμο** για να ελεγχθεί αν υπάρχουν καλύτερα αποτελέσματα.

Εικόνα 7 (Διαδικασία για έλεγχο στασιμότητας**)**



Εικόνα 8 (Πίνακας Αυτοσυσχετίσεων στο Output)



Πίνακας 3 (Αυτοσυσχετίσεις)

Autocorrelations

Series: ETE

Lag	Autocorrelation	Std. Error [a]	Box-Ljung Statistic Value	df	Sig. [b]
1	-,004	,051	,005	1	,944
2	-,079	,051	2,348	2	,309
3	-,086	,051	5,165	3	,160
4	,012	,051	5,221	4	,265
5	-,021	,051	5,394	5	,370
6	-,011	,051	5,438	6	,489
7	,100	,051	9,272	7	,234
8	,009	,051	9,303	8	,317
9	,000	,051	9,303	9	,410
10	,003	,051	9,306	10	,503
11	-,012	,051	9,357	11	,589
12	,000	,051	9,357	12	,672
13	,060	,051	10,784	13	,629
14	,036	,051	11,294	14	,663
15	,004	,050	11,300	15	,731
16	-,003	,050	11,303	16	,790

a. The underlying process assumed is independence (white noise).

b. Based on the asymptotic chi-square approximation.

Οπότε στο output εμφανίζεται ο πίνακας Αυτοσυσχετίσεων και γίνεται ο έλεγχος στασιμότητας.

Έλεγχος στασιμότητας

$H_0 : r = 0$

vs

$H_1 : r \neq 0$

1) H_o: $\rho_1=0$ $\quad$ $\rho_1=0.944>\alpha=0.05$ $\quad$ Αποδοχή H_0 (δεν υπάρχει αυτοσυσχέτιση,

vs $\quad$ άρα στάσιμη)

H_1: $\rho_1 \neq 0$

2) H_o: $\rho_2=0$ $\quad$ $\rho_2=0.309>\alpha=0.05$ $\quad$ Αποδοχή H_0 (δεν υπάρχει αυτοσυσχέτιση,

vs $\quad$ άρα στάσιμη)

H_1: $\rho_2 \neq 0$

3) H_o: $\rho_3=0$ $\quad$ $\rho_3=0.16>\alpha=0.05$ $\quad$ Αποδοχή H_0 (δεν υπάρχει αυτοσυσχέτιση,

vs $\quad$ άρα στάσιμη)

H_1: $\rho_3 \neq 0$

4) H_o: $\rho_4=0$ $\quad$ $\rho_4=0.265>\alpha=0.05$ $\quad$ Αποδοχή H_0 (δεν υπάρχει αυτοσυσχέτιση,

vs άρα στάσιμη)

$H_{1:}$ $\rho_4 \neq 0$

5) H_o: $\rho_5=0$ $\rho_5=0.37>\alpha=0.05$ Αποδοχή H_0 (δεν υπάρχει αυτοσυσχέτιση,

vs άρα στάσιμη)

$H_{1:}$ $\rho_5 \neq 0$

6) H_o: $\rho_6=0$ $\rho_6=0.489>\alpha=0.05$ Αποδοχή H_0 (δεν υπάρχει αυτοσυσχέτιση,

vs άρα στάσιμη)

$H_{1:}$ $\rho_6 \neq 0$

7) H_o: $\rho_7=0$ $\rho_7=0.234>\alpha=0.05$ Αποδοχή H_0 (δεν υπάρχει αυτοσυσχέτιση,

vs άρα στάσιμη)

$H_{1:}$ $\rho_7 \neq 0$

8) H_o: $\rho_8=0$ $\rho_8=0.317>\alpha=0.05$ Αποδοχή H_0 (δεν υπάρχει αυτοσυσχέτιση,

vs άρα στάσιμη)

$H_{1:}$ $\rho_8 \neq 0$

….

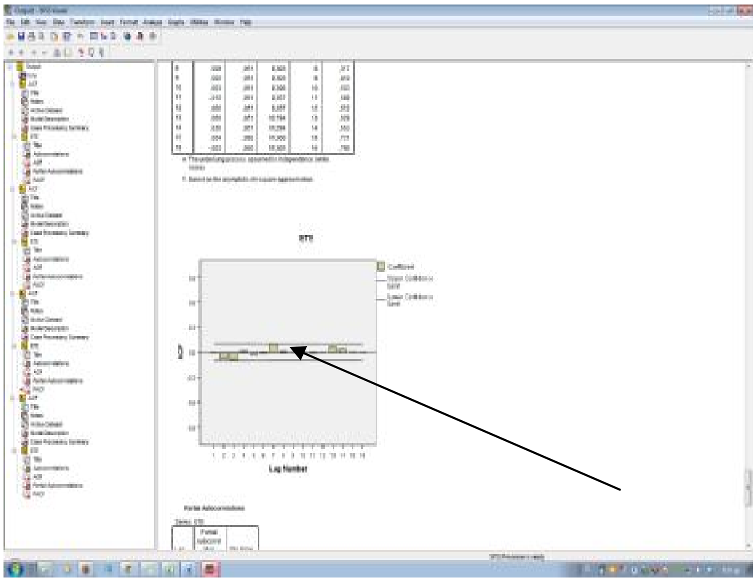
16) H_o: $\rho_{16}=0$ $\rho_{16}=0.79>\alpha=0.05$ Αποδοχή H_0 (δεν υπάρχει αυτοσυσχέτιση,

vs άρα στάσιμη)

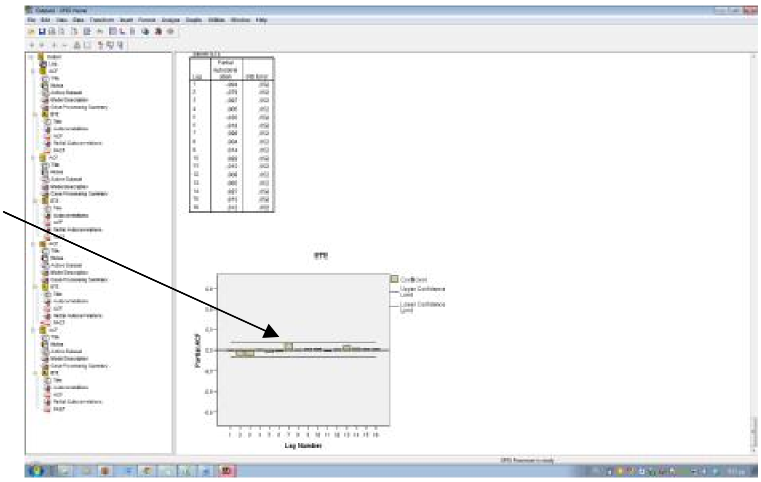
$H_{1:}$ $\rho_{16} \neq 0$

Πράγματι με πρώτες διαφορές και λογάριθμο σε όλα τα p-value υπάρχει στασιμότητα. Τα αποτελέσματα που φαίνονται να είναι καλύτερα από ότι με **πρώτες διαφορές** και χωρίς λογάριθμο. Οπότε βάση του έλεγχου κανονικότητας φαίνεται ότι η διαδικασία είναι στάσιμη με **πρώτες διαφορές και λογάριθμο.** Έπειτα, με τα διαγράμματα που εμφανίζονται στο output των ACF και Partial ACF εντοπίζονται τα q και τα p αντίστοιχα.

Εικόνα 9 (Εύρεση τάξεως q)



Εικόνα 10 (Εύρεση τάξεως p)

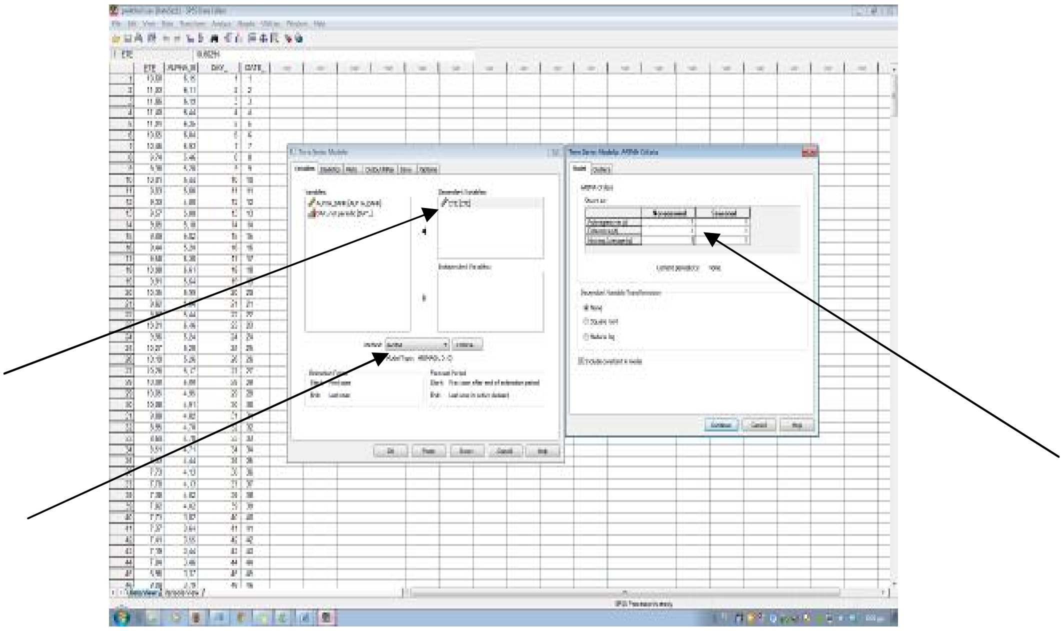


Από τα διαγράμματα ACF και PACF προκύπτει q=1 και p=1. Το υπόδειγμα είναι της μορφής ARIMA(p,d,q)-ARIMA(1,1,1) (d=1 λόγω πρώτων διαφορών). Η θεωρητική μορφή του υποδείγματος είναι η εξής:

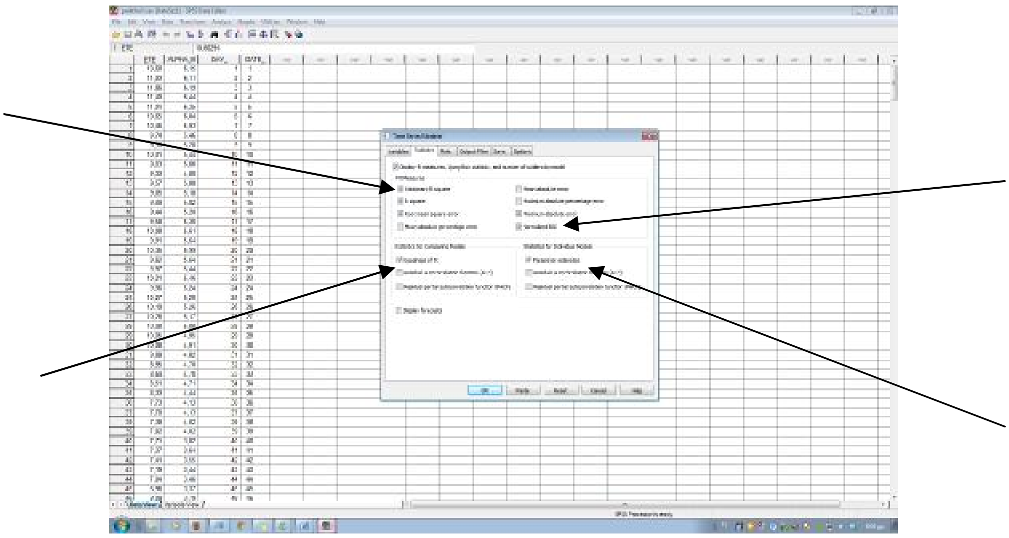
$$\mathrm{Y}_{t} = a_{0} + a_{1}y_{t-1} + q_{1}e_{t-1} + e_{t}$$

Έπειτα πηγαίνοντας στο SPSS στα δεδομένα ακολουθούνται οι παρακάτω διαδικασίες. Από το menu επιλέγεται Analyze-Time Series-Create Models.Στο Dependent Variables μπαίνει η μεταβλητή της ΕΤΕ και στο Method επιλέγεται ARIMA.Πατώντας το Criteria εμφανίζεται ο πίνακας ARIMA Criteria και στο Nonseasonal μπαίνουν οι τιμές των p,d,q που έχουν εντοπιστεί 1,1,1 αντίστοιχα. Πατώντας το Continue και πηγαίνοντας στο Statistics γίνονται οι παρακάτω επιλογές και στη συνέχεια ok.

Εικόνα 11 (Καθορισμός υποδείγματος ARIMA(1,1,1) στο SPSS)

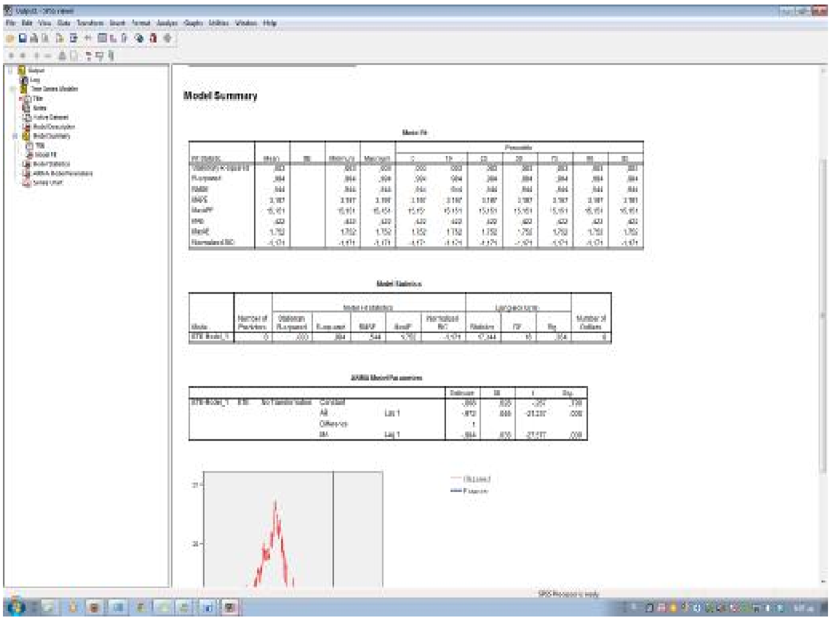


Εικόνα 12 (Διαδικασία για έλεγχο στατιστικής σημαντικότητας)



Στη συνέχεια στο Output βγαίνει ο πίνακας ARIMA Model Parameters από τον οποίο μελετώνται οι παραμέτροι που είναι στατιστικά σημαντικοί κάνοντας τον έλεγχο στατιστικής σημαντικότητας για το υπόδειγμα ARIMA(1,1,1).

Εικόνα 13 (Πίνακας ARIMA Model Parameters στο Output)



Πίνακας 4 (ARIMA Model Parameters)

ARIMA Model Parameters

					Estimate	SE	t	Sig.
ETE-Mod	ETE	No Transforn	Constant		-,008	,028	-,267	,790
			AR	Lag 1	-,972	,046	21,237	,000
			Difference		1			
			MA	Lag 1	-,984	,036	27,577	,000

Έλεγχος στατιστικής σημαντικότητας p-value των εκτιμητών του υποδείγματος

$H_0 : r = 0$

vs

$H_1 : r \neq 0$

Με την υπόθεση H_0 ο εκτιμητής μας είναι στατιστικά μη σημαντικός, ενώ με την H_1 ο εκτιμητής μας είναι στατιστικά σημαντικός.

1) H_o: $\rho_0=0$	$\rho_0=0.79>\alpha=0.05$	Αποδοχή H_o άρα ο εκτιμητής είναι στατιστικά
vs		μη σημαντικός (Σ.Μ.Σ.)
H_1: $\rho_0\neq 0$		
2) H_o: $\rho_1=0$	$\rho_1=0<\alpha=0.05$	Απόρριψη H_o άρα ο εκτιμητής είναι στατιστικά
vs		σημαντικός (Σ.Σ.)
H_1: $\rho_1\neq 0$		
3) H_o: $\rho_2=0$	$\rho_2=0<\alpha=0.05$	Απόρριψη H_o άρα ο εκτιμητής είναι στατιστικά
vs		σημαντικός (Σ.Σ.)
H_1: $\rho_2\neq 0$		

Αφού βρέθηκαν όλοι εκτιμητές στατιστικά σημαντικοί (εκτός του σταθερού όρου ο οποίος δεν επηρεάζει σε κάτι) και γνωρίζοντας τις παραμέτρους του υποδείγματος ARIMA(1,1,1) θεωρείται ως το καταλληλότερο. Οι εκτιμώμενοι παράμετροι a_0, a_1, q_1 θα είναι ίσοι με:

$a_0 = -0.008,\ a_1 = -0.972,\ q_1 = -0.984$

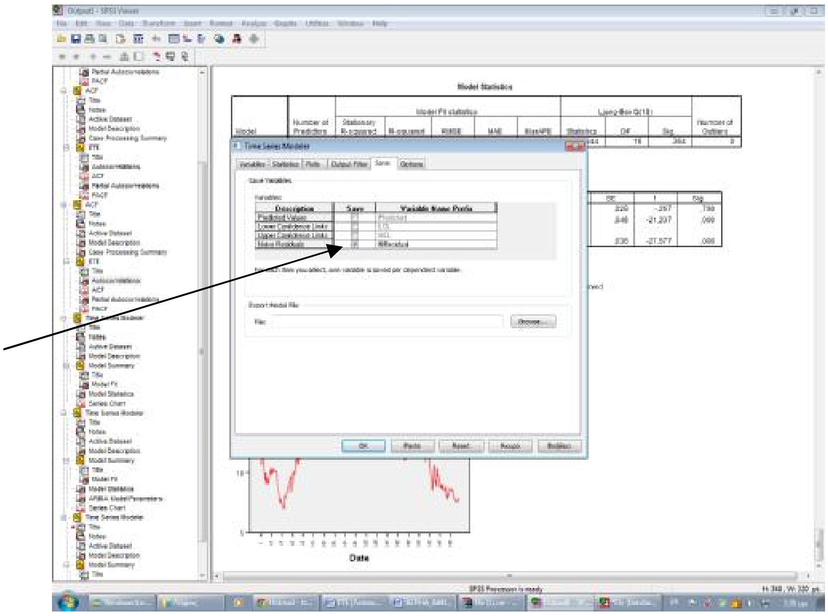
Η μαθηματική μορφή του υποδείγματος είναι:

$$Y_t = -0.008 - 0.972 y_{t-1} - 0.984 e_{t-1} + e_t$$

Για να είναι το υπόδειγμα Y_t αποδεκτό, για να θεωρηθεί δηλαδή ότι περιγράφει τη διαδικασία που παρήγαγε τα δεδομένα, θα πρέπει τα κατάλοιπα να είναι λευκός θόρυβος, δηλαδή δεν θα πρέπει να αυτοσυσχετίζονται. Για αυτό το λόγο θα γίνει έλεγχος στασιμότητας στα κατάλοιπα (residuals) επαληθεύοντας έτσι αν τα αποτελέσματα είναι αληθή για το ARIMA(1,1,1).

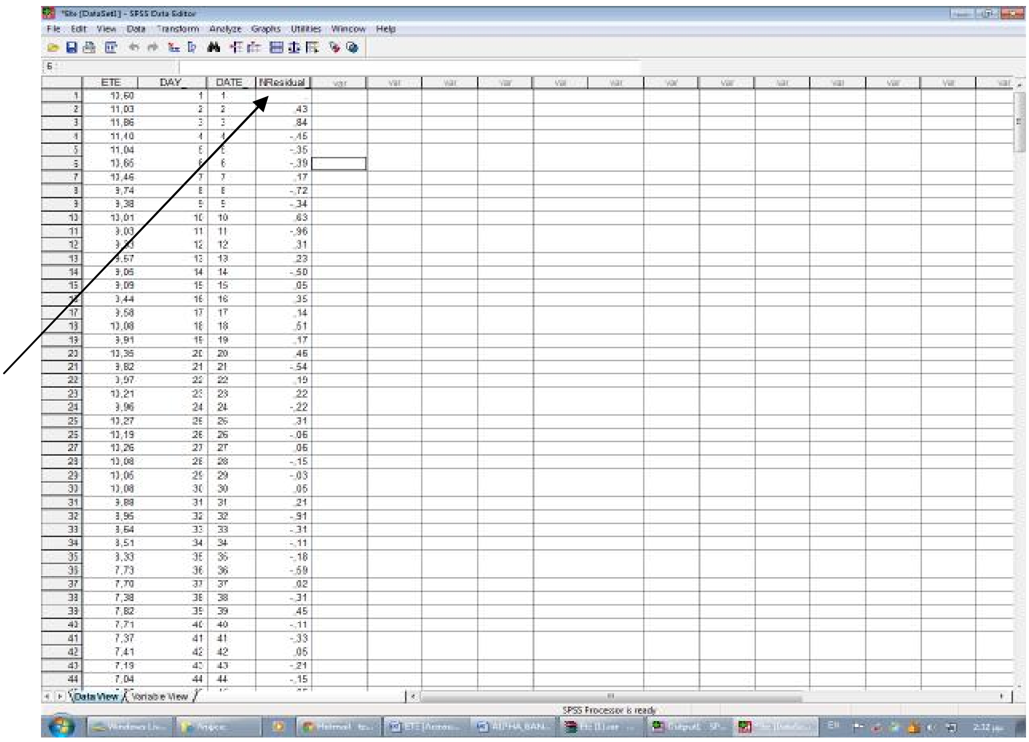
Από το menu του SPSS λοιπόν πηγαίνοντας Analyze-Time Series-Create Models και στο Time Series Modeler και συγκεκριμένα στο Save επιλέγεται το Noise Residuals.

Εικόνα 14 (Διαδικασία για την εύρεση των Καταλοίπων)



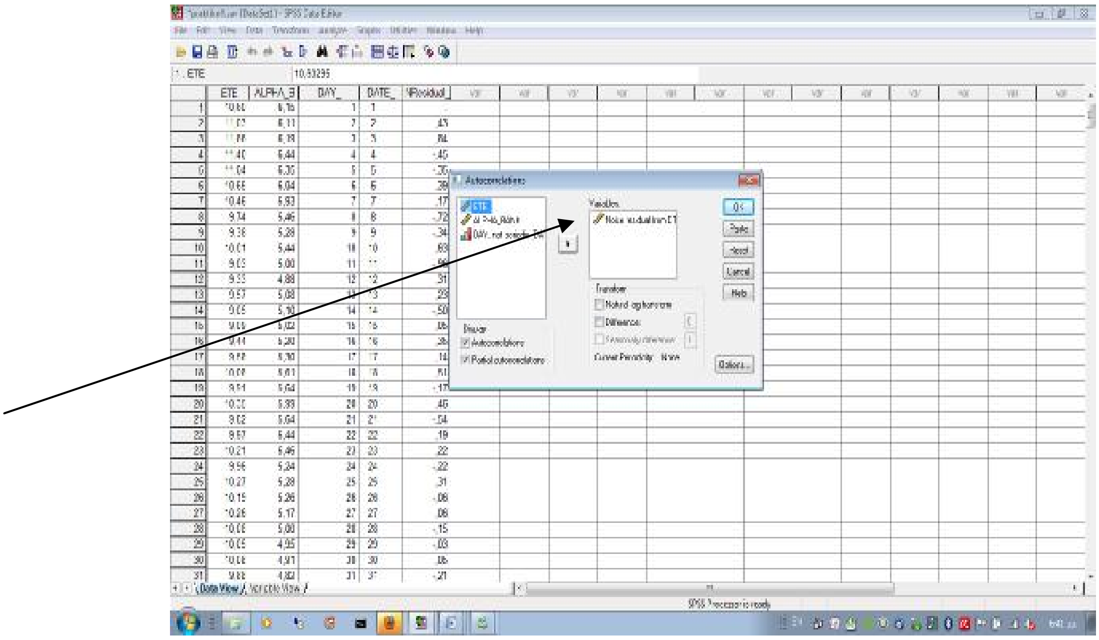
Πατώντας οκ εμφανίζεται η μεταβλητή Noise Residual ETE στο SPSS που δείχνει τα υπόλοιπα(residuals).

Εικόνα 15 (Τιμές των Καταλοίπων)



Από το menu του SPSS επιλέγοντας Graphs-Time Series-Autocorrelations η μεταβλητή NResidual μπαίνει στο Variables και στη συνέχεια ok για να βγει ο πίνακας αυτοσυσχέτισης στο output και να γίνει έλεγχος στασιμότητας.

Εικόνα 16 (Διαδικασία για έλεγχο στασιμότητας των Καταλοίπων)



Πίνακας 5 (Αυτοσυσχετίσεις)

Autocorrelations

Series: Noise residual from ETE-Model_1

Lag	Autocorrelation	Std. Error [a]	Box-Ljung Statistic Value	df	Sig. [b]
1	,020	,051	,158	1	,691
2	-,106	,051	4,419	2	,110
3	-,088	,051	7,360	3	,061
4	,055	,051	8,528	4	,074
5	-,056	,051	9,724	5	,083
6	-,017	,051	9,840	6	,132
7	,025	,051	10,078	7	,184
8	,037	,051	10,592	8	,226
9	,045	,051	11,370	9	,251
10	,018	,051	11,492	10	,321
11	-,039	,051	12,090	11	,357
12	,022	,051	12,276	12	,424
13	,063	,051	13,843	13	,385
14	,039	,051	14,442	14	,417
15	-,003	,050	14,445	15	,492
16	,021	,050	14,612	16	,553

a. The underlying process assumed is independence (white noise).

b. Based on the asymptotic chi-square approximation.

Έλεγχος στασιμότητας

$H_0 : r = 0$
vs
$H_1 : r \neq 0$

Με την υπόθεση H_0 δεν υπάρχει αυτοσυσχέτιση, άρα είναι στάσιμη ενώ με την H_1 υπάρχει αυτοσυσχέτιση και δεν προκύπτει στασιμότητα.

1) H_o: $\rho_1=0$ vs H_1: $\rho_1 \neq 0$ — $\rho_1=0.691>\alpha=0.05$ — Αποδοχή H_0 (δεν υπάρχει αυτοσυσχέτιση, άρα στάσιμη)

2) H_o: $\rho_2=0$ vs H_1: $\rho_2 \neq 0$ — $\rho_2=0.11>\alpha=0.05$ — Αποδοχή H_0 (δεν υπάρχει αυτοσυσχέτιση, άρα στάσιμη)

3) H_o: $\rho_3=0$ vs H_1: $\rho_3 \neq 0$ — $\rho_3=0.061>\alpha=0.05$ — Αποδοχή H_0 (δεν υπάρχει αυτοσυσχέτιση, άρα στάσιμη)

4) H_o: $\rho_4=0$ vs H_1: $\rho_4 \neq 0$ — $\rho_4=0.74>\alpha=0.05$ — Αποδοχή H_0 (δεν υπάρχει αυτοσυσχέτιση, άρα στάσιμη)

5) H_o: $\rho_5=0$ vs H_1: $\rho_5 \neq 0$ — $\rho_5=0.83>\alpha=0.05$ — Αποδοχή H_0 (δεν υπάρχει αυτοσυσχέτιση, άρα στάσιμη)

6) H_o: $\rho_6=0$ vs H_1: $\rho_6 \neq 0$ — $\rho_6=0.132>\alpha=0.05$ — Αποδοχή H_0 (δεν υπάρχει αυτοσυσχέτιση, άρα στάσιμη)

7) H_o: $\rho_7=0$ $\rho_7=0.184>\alpha=0.05$ Αποδοχή H_0 (δεν υπάρχει αυτοσυσχέτιση, άρα στάσιμη)

vs

H_1: $\rho_7 \neq 0$

8) H_o: $\rho_8=0$ $\rho_8=0.226>\alpha=0.05$ Αποδοχή H_0 (δεν υπάρχει αυτοσυσχέτιση, άρα στάσιμη)

vs

H_1: $\rho_8 \neq 0$

…

16) H_o: $\rho_{16}=0$ $\rho_{16}=0.553>\alpha=0.05$ Αποδοχή H_0 (δεν υπάρχει αυτοσυσχέτιση, άρα στάσιμη)

vs

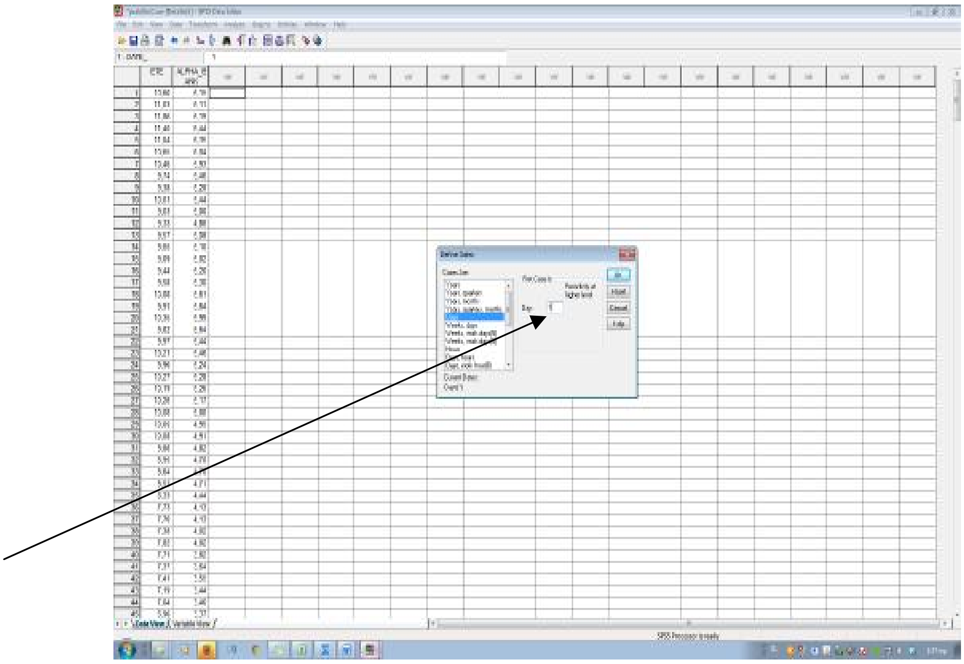
H_1: $\rho_{16} \neq 0$

Άρα από τον έλεγχο καταλοίπων φαίνεται ότι η διαδικασία είναι στάσιμη καθώς για όλες της χρονολογικές υστερήσεις τα ρ είναι στάσιμα. Έτσι το υπόδειγμα ARIMA(1,1,1) επιβεβαιώνεται ότι είναι το πλέον κατάλληλο.

ALPHA_BANK

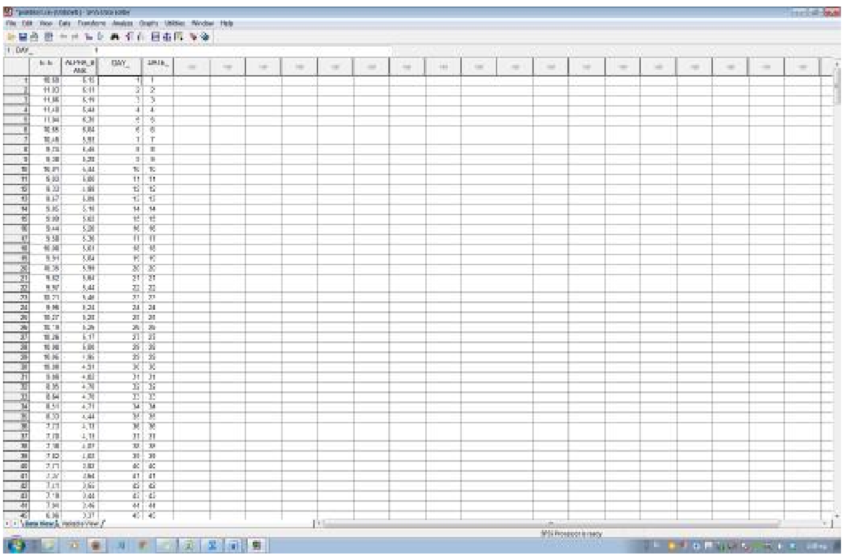
Στο παρακάτω παράδειγμα έχουν συλλεχθεί στο SPSS μετοχές από 01/01/09 εώς 31/12/10(συνολικά 501 τιμές).

Εικόνα 17 (Define dates)



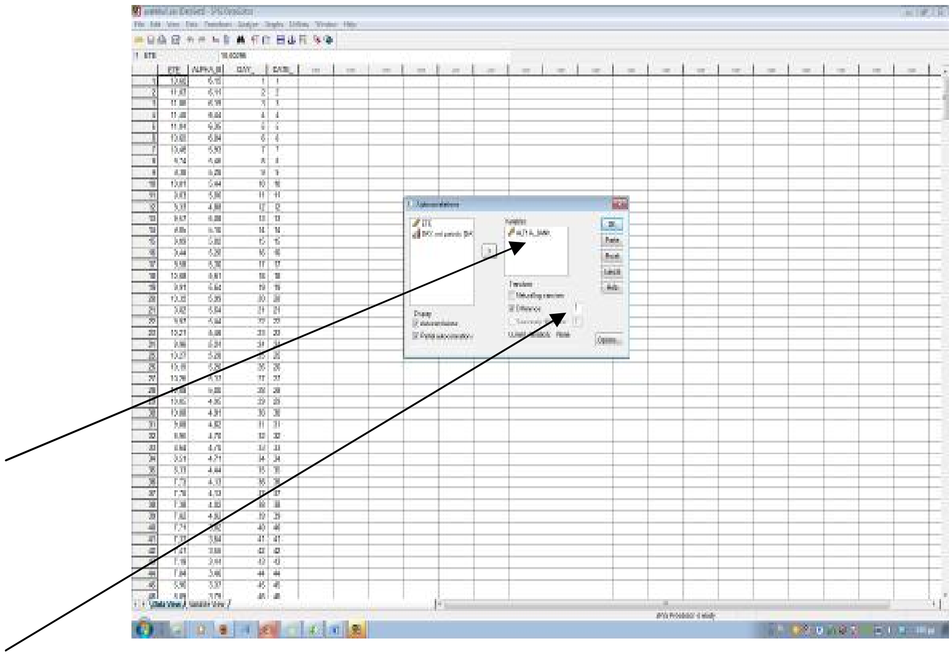
Στη συνέχεια από το menu του SPSS επιλέγονται οι εντολές Data-Define Dates-Days και στο Days τοποθετήθηκε Day =1 καταλήγοντας στο παρακάτω

Εικόνα 18 (Μεταβλητές DAY_,DATE_)

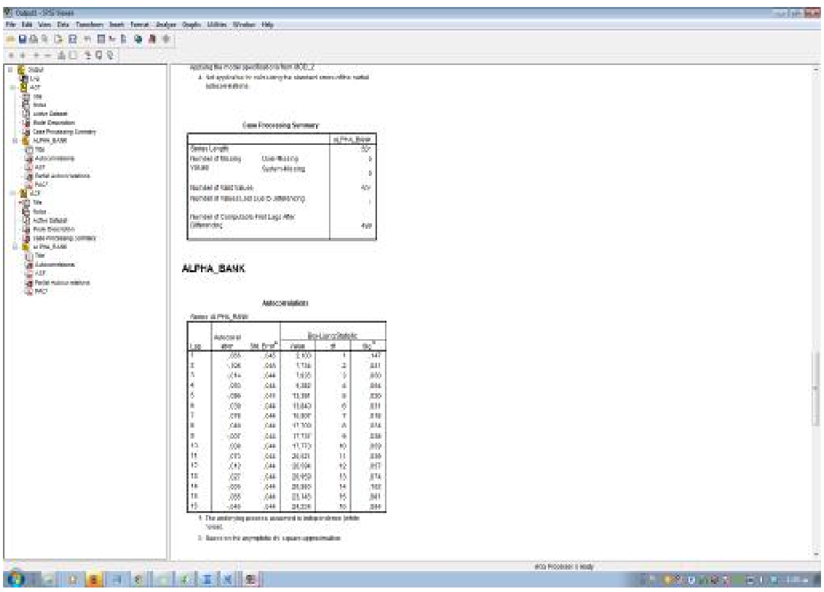


Έπειτα από το menu του SPSS και συγκεκριμένα από το Graphs-Time Series-Autocorrelations μπαίνει η μεταβλητή ALPHA BANK στο Variables και γίνονται δοκιμές για να βρεθεί το καταλληλότερο υπόδειγμα. Για την διαδικασία αυτή απαιτείται ελέγχος στασιμότητας(Πίνακας 1). Αρχικά χρειάζεται με **πρώτες διαφορές** επιλέγεται το ok για να προκύψουν τα εξής αποτελέσματα στο output.

Εικόνα 19 (Διαδικασία για έλεγχο στασιμότητας**)**



Εικόνα 20 (Πίνακας Αυτοσυσχετίσεων sto Output)



Πίνακας 6 (Αυτοσυσχετίσεις)

Autocorrelations

Series: ALPHA_BANK

Lag	Autocorrelation	Std. Error[a]	Box-Ljung Statistic Value	df	Sig.[b]
1	,065	,045	2,100	1	,147
2	-,106	,045	7,734	2	,021
3	-,014	,044	7,835	3	,050
4	,053	,044	9,282	4	,054
5	-,090	,044	13,391	5	,020
6	,030	,044	13,840	6	,031
7	,078	,044	16,907	7	,018
8	,040	,044	17,709	8	,024
9	-,007	,044	17,737	9	,038
10	,008	,044	17,770	10	,059
11	,073	,044	20,521	11	,039
12	,012	,044	20,594	12	,057
13	,027	,044	20,959	13	,074
14	-,006	,044	20,980	14	,102
15	,065	,044	23,145	15	,081
16	-,046	,044	24,254	16	,084

a. The underlying process assumed is independence (white noise).

b. Based on the asymptotic chi-square approximation.

- Χρονολογικές υστερήσεις(Lag),
- Τιμή του συντελεστή αυτοσυσχέτισης(Autocorrelation),
- Τυπικό σφάλμα(Std Error),
- Τιμή της στατιστικής συνάρτησης Box-Ljung(Value) ,
- Βαθμοί ελευθερίας(df),
- Τιμή πιθανότητας της στατιστικής συνάρτησης(Sig)

Έλεγχος στασιμότητας

$H_0 : r = 0$

vs

$H_1 : r \neq 0$

Με την υπόθεση H_0 δεν υπάρχει αυτοσυσχέτιση, άρα είναι στάσιμη ενώ με την H_1 υπάρχει αυτοσυσχέτιση και δεν προκύπτει στασιμότητα.

1) H_o: $\rho_1=0$ $\rho_1=0.147>\alpha=0.05$ Αποδοχή H_0 (δεν υπάρχει αυτοσυσχέτιση, άρα στάσιμη)

vs

H_1: $\rho_1\neq0$

2) H_o: $\rho_2=0$ $\rho_2=0.021<\alpha=0.05$ Απόρριψη H_0 (υπάρχει αυτοσυσχέτιση)

vs

H_1: $\rho_2\neq0$

3) H_o: $\rho_3=0$ $\rho_3=0.05=\alpha$ Απόρριψη H_0 (υπάρχει αυτοσυσχέτιση)

vs

H_1: $\rho_3\neq0$

4) H_o: $\rho_4=0$ $\rho_4=0.054>\alpha=0.05$ Αποδοχή H_0 (δεν υπάρχει αυτοσυσχέτιση, άρα στάσιμη)

vs

H_1: $\rho_4\neq0$

5) H_o: $\rho_5=0$ $\rho_5=0.02<\alpha=0.05$ Απόρριψη H_0 (υπάρχει αυτοσυσχέτιση)

vs

H_1: $\rho_5\neq0$

6) H_o: $\rho_6=0$ $\rho_6=0.031<\alpha=0.05$ Απόρριψη H_0 (υπάρχει αυτοσυσχέτιση)

vs

H_1: $\rho_6\neq0$

7) H_o: $\rho_7=0$ $\rho_7=0.018<\alpha=0.05$ Απόρριψη H_0 (υπάρχει αυτοσυσχέτιση)

vs

H_1: $\rho_7\neq0$

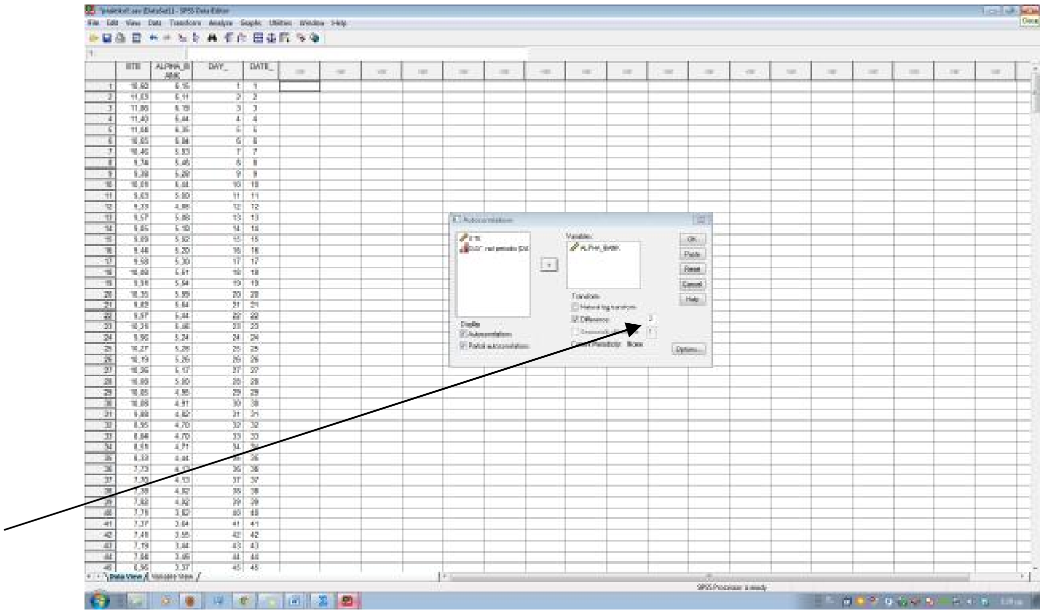
8) H_o: $\rho_8=0$ $\rho_8=0.024<\alpha=0.05$ Απόρριψη H_0 (υπάρχει αυτοσυσχέτιση)

vs

H_1: $\rho_8\neq0$

Από ότι φαίνεται από τους ελέγχους των υποθέσεων με πρώτες διαφορές υπάρχει αυτοσυσχέτιση στα περισσότερα από τα πρώτα 8 και έτσι δεν είναι στάσιμη η διαδικασία. Οπότε θα δοκιμαστεί με **δεύτερες διαφορές.**

Εικόνα 21 (Διαδικασία για έλεγχο στασιμότητας)



Πίνακας 7 (Αυτοσυσχετίσεις)

Autocorrelations

Series: ALPHA_BANK

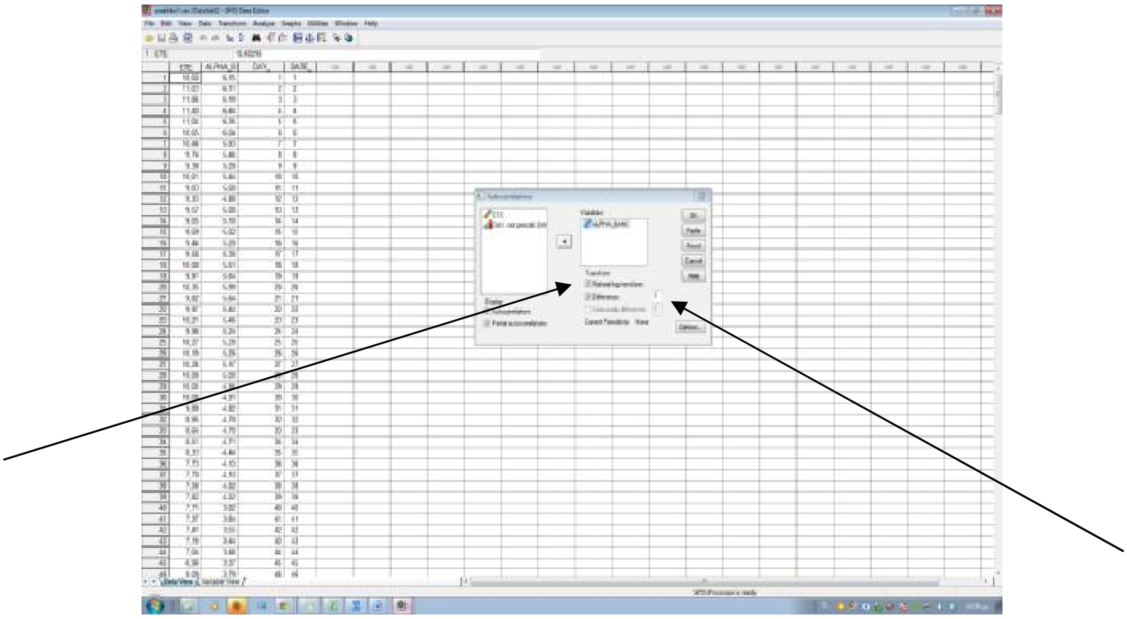
Lag	Autocorrelation	Std. Error[a]	Box-Ljung Statistic Value	df	Sig.[b]
1	-,409	,045	83,954	1	,000
2	-,140	,045	93,774	2	,000
3	,013	,045	93,853	3	,000
4	,113	,044	100,255	4	,000
5	-,140	,044	110,242	5	,000
6	,038	,044	110,982	6	,000
7	,046	,044	112,054	7	,000
8	,005	,044	112,066	8	,000
9	-,033	,044	112,633	9	,000
10	-,027	,044	112,992	10	,000
11	,068	,044	115,344	11	,000
12	-,041	,044	116,192	12	,000
13	,025	,044	116,520	13	,000
14	-,055	,044	118,107	14	,000
15	,097	,044	123,013	15	,000
16	-,113	,044	129,613	16	,000

a. The underlying process assumed is independence (white noise).

b. Based on the asymptotic chi-square approximation.

Από τον παραπάνω πίνακα είναι ξεκάθαρο ότι για όλα τα lags υπάρχει αυτοσυσχέτιση γατί όλα τα p είναι μηδέν οπότε δεν είναι στάσιμη με δεύτερες διαφορές και δοκιμάζεται με **πρώτες διαφορές και λογάριθμο.** Πατώντας πάλι οκ και εμφανίζονται τα εξής αποτελέσματα στον πίνακα Autocorrelations.

Εικόνα 22 (Διαδικασία για έλεγχο στασιμότητας)



Πίνακας 8 (Αυτοσυσχετίσεις)

Autocorrelations

Series: ALPHA_BANK

Lag	Autocorrelation	Std. Error [a]	Box-Ljung Statistic Value	df	Sig. [b]
1	,072	,045	2,632	1	,105
2	-,088	,045	6,540	2	,038
3	-,017	,044	6,691	3	,082
4	,019	,044	6,878	4	,143
5	-,043	,044	7,825	5	,166
6	,048	,044	9,005	6	,173
7	,119	,044	16,270	7	,023
8	,006	,044	16,286	8	,038
9	-,027	,044	16,652	9	,054
10	-,001	,044	16,653	10	,082
11	,044	,044	17,645	11	,090
12	,047	,044	18,776	12	,094
13	,041	,044	19,642	13	,105
14	,013	,044	19,730	14	,139
15	,035	,044	20,349	15	,159
16	-,048	,044	21,523	16	,159

a. The underlying process assumed is independence (white noise).

b. Based on the asymptotic chi-square approximation.

Έλεγχος στασιμότητας

$H_0 : r = 0$
vs
$H_1 : r \neq 0$

Με την υπόθεση H_0 δεν υπάρχει αυτοσυσχέτιση, άρα είναι στάσιμη ενώ με την H_1 υπάρχει αυτοσυσχέτιση και δεν έχουμε στασιμότητα.

1) H_o: $\rho_1=0$ vs H_1: $\rho_1 \neq 0$ — $\rho_1=0.105>\alpha=0.05$ Αποδοχή H_0 (δεν υπάρχει αυτοσυσχέτιση, άρα στάσιμη)

2) H_o: $\rho_2=0$ vs H_1: $\rho_2 \neq 0$ — $\rho_2=0.038<\alpha=0.05$ Απόρριψη H_0 (υπάρχει αυτοσυσχέτιση)

3) H_o: $\rho_3=0$ vs H_1: $\rho_3 \neq 0$ — $\rho_3=0.082>\alpha=0.05$ Αποδοχή H_0 (δεν υπάρχει αυτοσυσχέτιση, άρα στάσιμη)

4) H_o: $\rho_4=0$ vs H_1: $\rho_4 \neq 0$ — $\rho_4=0.143>\alpha=0.05$ Αποδοχή H_0 (δεν υπάρχει αυτοσυσχέτιση, άρα στάσιμη)

5) H_o: $\rho_5=0$ vs H_1: $\rho_5 \neq 0$ — $\rho_5=0.166>\alpha=0.05$ Αποδοχή H_0 (δεν υπάρχει αυτοσυσχέτιση, άρα στάσιμη)

6) H_o: $\rho_6=0$ vs H_1: $\rho_6 \neq 0$ — $\rho_6=0.173>\alpha=0.05$ Αποδοχή H_0 (δεν υπάρχει αυτοσυσχέτιση, άρα στάσιμη)

7) H_o: $\rho_7=0$ $\rho_7=0.023<\alpha=0.05$ Απόρριψη H_0 (υπάρχει αυτοσυσχέτιση)

vs

H_1: $\rho_7\neq0$

8) H_o: $\rho_8=0$ $\rho_8=0.038<\alpha=0.05$ Απόρριψη H_0 (υπάρχει αυτοσυσχέτιση)

vs

H_1: $\rho_8\neq0$

…….

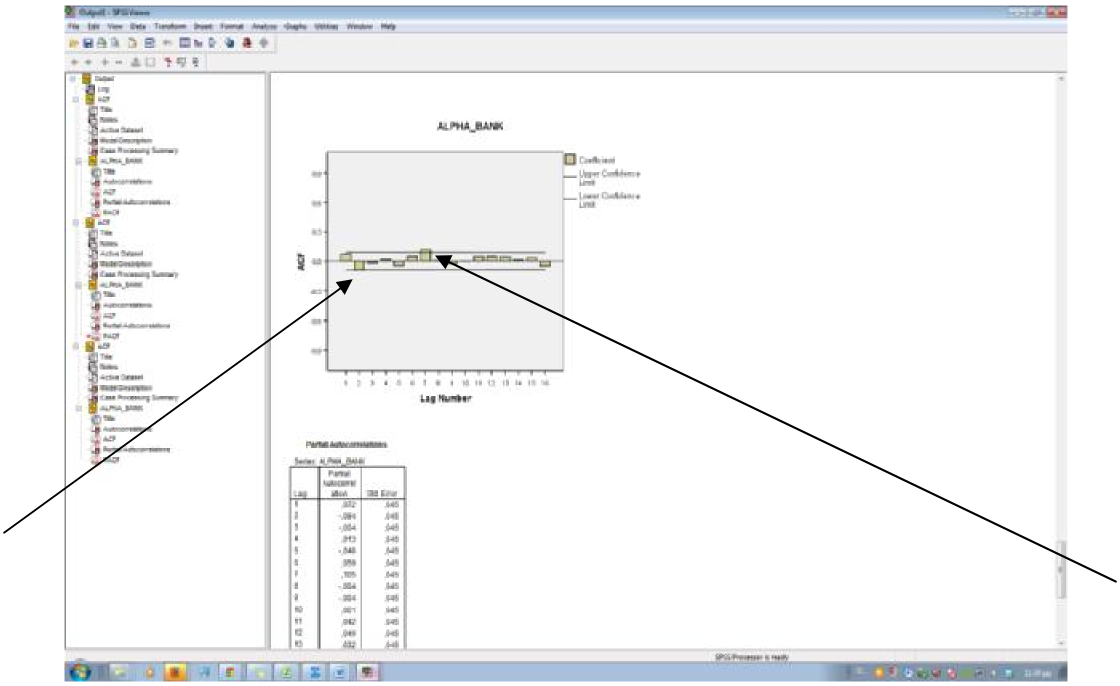
16) H_o: $\rho_{16}=0$ $\rho_{16}=0.159>\alpha=0.05$ Αποδοχή H_0 (δεν υπάρχει αυτοσυσχέτιση, άρα στάσιμη)

vs

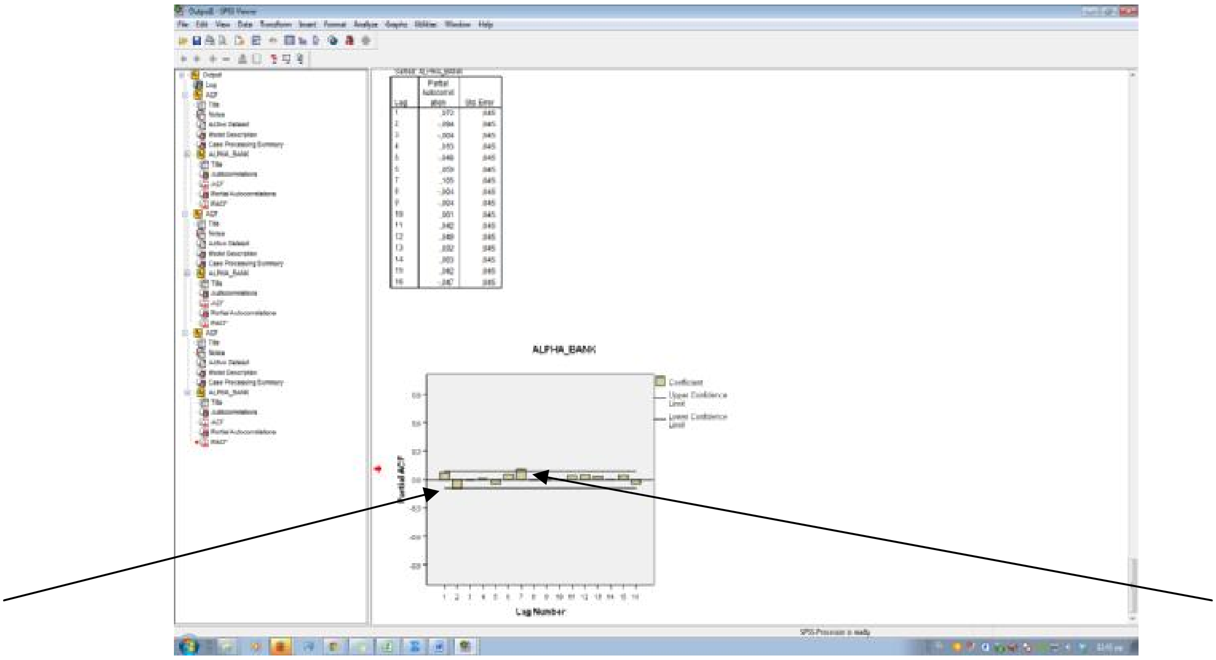
H_1: $\rho_{16}\neq0$

Οπότε βάση του έλεγχου στασιμότητας που έγινε φαίνεται ότι η διαδικασία αυτή είναι στάσιμη με **πρώτες διαφορές και λογάριθμο.** Ύστερα με τα διαγράμματα που εμφανίζονται στο output των ACF και Partial ACF εντοπίζονται τα q και τα p αντίστοιχα.

Εικόνα 23 (Εύρεση τάξεως q)



Εικόνα 24 (Εύρεση τάξεως p)



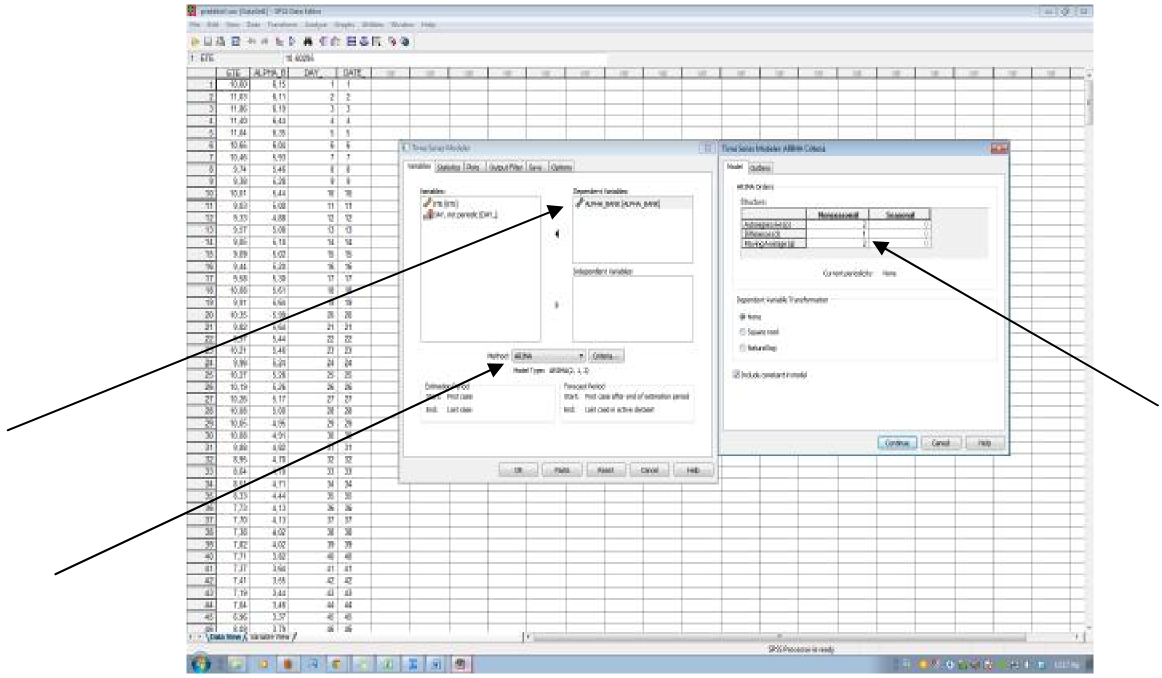
Άρα q=2 και p=2. Το υπόδειγμα είναι της μορφής ARIMA(p,d,q)-ARIMA(2,1,2)

Η θεωρητική μορφή του υποδείγματος είναι η εξής:

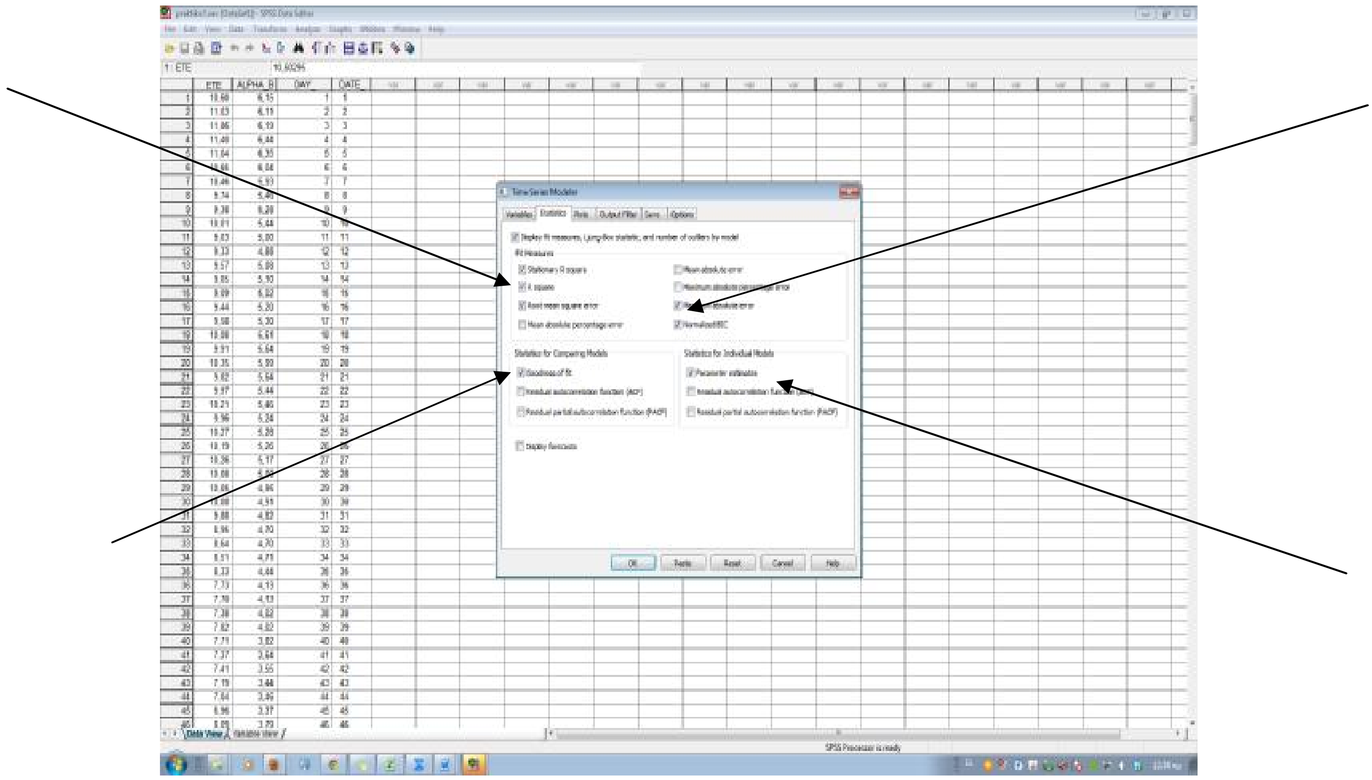
$$Y_t = a_0 + a_1 y_{t-1} + a_2 y_{t-2} + q_1 e_{t-1} + q_2 e_{t-2} + e_t$$

Πηγαίνοντας στο SPSS που υπάρχουν τα δεδομένα ακολουθείται η παρακάτω διαδικασία. Από το menu επιλέγεται Analyze-Time Series-Create Models.Στο Dependent Variables μπαίνει η μεταβλητή ALPHA_BANK και στο Method επιλέγεται το ARIMA.Πατώντας το Criteria εμφανίζεται ο πίνακας ARIMA Criteria και στο Nonseasonal μπαίνουν οι τιμές των p,d,q δηλαδή 2,1,2 αντιστοίχως. Επιλέγοντας Continue, και πηγαίνοντας στο Statistics επιλέγονται τα παρακάτω πατώντας ok.

Εικόνα 25 (Καθορισμός υποδείγματος ARIMA(2,1,2) στο SPSS)

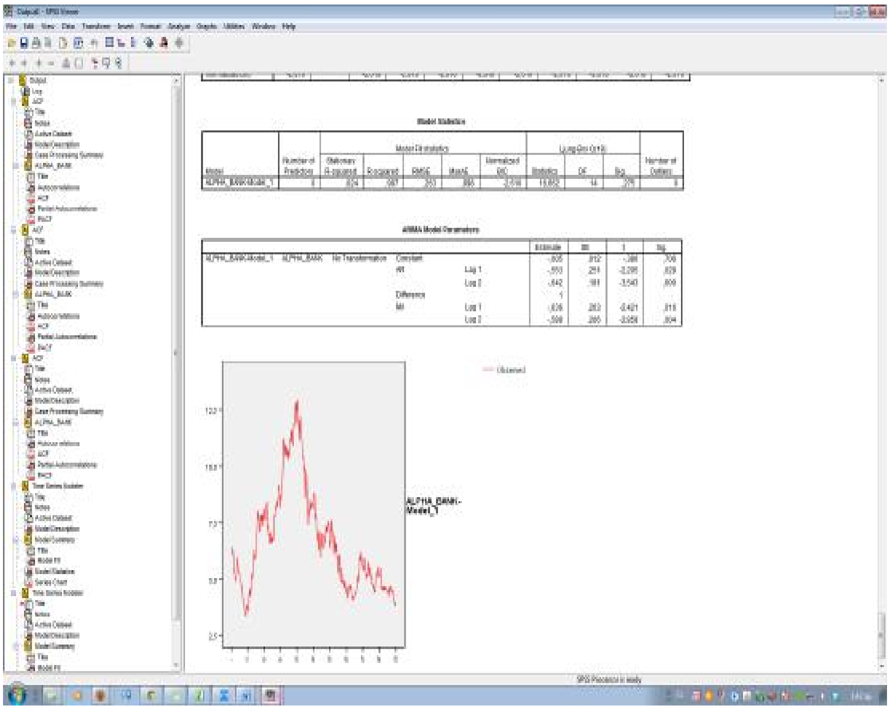


Εικόνα 26 (Διαδικασία για έλεγχο στατιστικής σημαντικότητας)



Στη συνέχεια στο Output βγαίνει ο πίνακας ARIMA Model Parameters ο οποίος χρειάζεται για τον έλεγχο στατιστικής σημαντικότητας στο υπόδειγμα ARIMA(2,1,2).

Εικόνα 27 (Πίνακας ARIMA Model Parameters στο Output)



Πίνακας 9 (ARIMA Model Parameters)

ARIMA Model Parameters

					Estimate	SE	t	Sig.
ALPHA_BANK-M(	ALPHA_BA	No Transforma	Constant		-,005	,012	-,386	,700
			AR	Lag 1	-,553	,251	-2,205	,028
				Lag 2	-,642	,181	-3,543	,000
			Difference		1			
			MA	Lag 1	-,636	,263	-2,421	,016
				Lag 2	-,588	,206	-2,858	,004

Έλεγχος στατιστικής σημαντικότητας p-value των εκτιμητών του υποδείγματος

$H_0 : r = 0$

vs

$H_1 : r \neq 0$

Με την υπόθεση H_0 ο εκτιμητής είναι στατιστικά μη σημαντικός, ενώ με την H_1 ο εκτιμητής είναι στατιστικά σημαντικός.

1) H_o: $\rho_0=0$ $\rho_0=0.7>\alpha=0.05$ Αποδοχή H_o άρα ο εκτιμητής είναι στατιστικά μη σημαντικός (Σ.Μ.Σ.)

vs

H_1: $\rho_0\neq0$

2) H_o: $\rho_1=0$ $\rho_1=0.028<\alpha=0.05$ Απόρριψη H_o άρα ο εκτιμητής είναι στατιστικά σημαντικός (Σ.Σ.)

vs

H_1: $\rho_1\neq0$

3) H_o: $\rho_2=0$ $\rho_2=0<\alpha=0.05$ Απόρριψη H_o άρα ο εκτιμητής είναι στατιστικά σημαντικός (Σ.Σ.)

vs

H_1: $\rho_2\neq0$

4) H_o: $\rho_3=0$ $\rho_3=0.016<\alpha=0.05$ Απόρριψη H_o άρα ο εκτιμητής είναι στατιστικά σημαντικός (Σ.Σ.)

vs

H_1: $\rho_3\neq0$

5) H_o: $\rho_4=0$ $\rho_4=0.004<\alpha=0.05$ Απόρριψη H_o άρα ο εκτιμητής είναι στατιστικά σημαντικός (Σ.Σ.)

vs

H_1: $\rho_4\neq0$

Αφού βρέθηκαν όλοι οι εκτιμητές στατιστικά σημαντικοί (εκτός του σταθερού όρου ο οποίος δεν επηρεάζει το υπόδειγμα) και γνωρίζοντας τις παραμέτρους του υποδείγματος ARIMA(2,1,2) θεωρείται ως το καταλληλότερο. Οι εκτιμώμενοι παράμετροι a_0, a_1, a_2, q_1, q_2 θα είναι ίσοι με:

$$a_0 = -0.005,\ a_1 = -0.553,\ a_2 = -0.642,\ q_1 = -0.636,\ q_2 = -0.588$$

Το a_0 είναι ο σταθερός όρος, τα a_1, a_2 εκτιμώμενοι παράμετροι αυτοπαλίνδρομης διαδικασίας AR και τα q_1, q_2 εκτιμώμενοι παράμετροι κινητού μέσου MA.

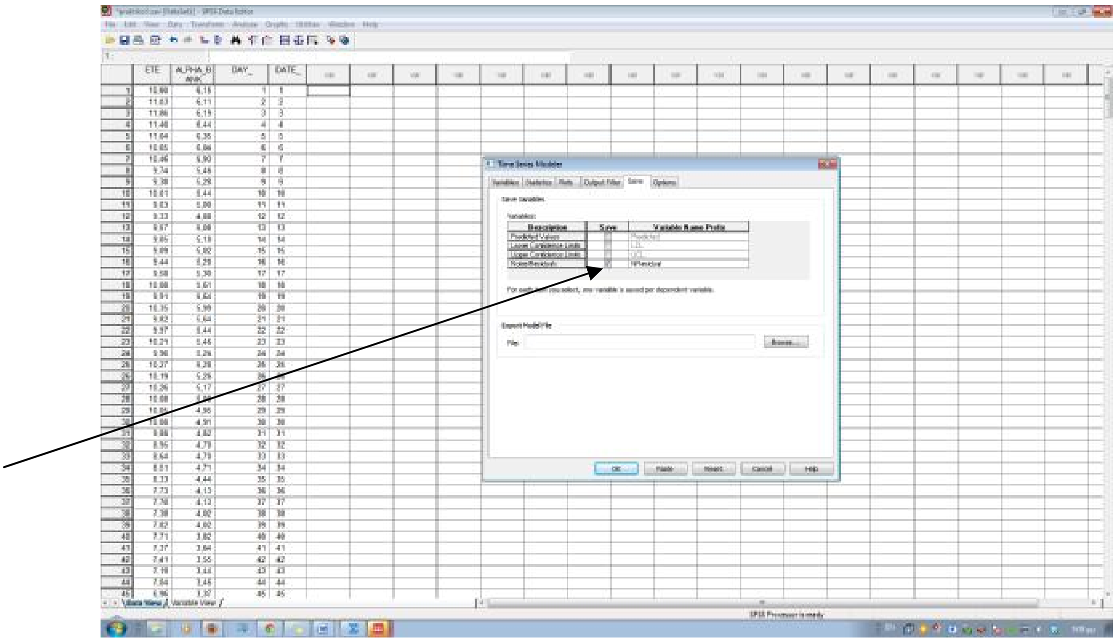
και θα έχει την εξής μαθηματική μορφή:

$$Y_t = -0.005 - 0.553y_{t-1} - 0.642y_{t-2} - 0.636e_{t-1} - 0.588e_{t-2} + e_t$$

Για να είναι το υπόδειγμα Y_t αποδεκτό, για να θεωρηθεί δηλαδή ότι περιγράφει τη διαδικασία που παρήγαγε τα δεδομένα, θα πρέπει τα κατάλοιπα να είναι λευκός θόρυβος, οπότε δεν θα πρέπει να αυτοσυσχετίζονται. Για αυτό το λόγο θα γίνει έλεγχος στασιμότητας στα κατάλοιπα (residuals) επαληθεύοντας έτσι αν τα αποτελέσματα είναι αληθή για το ARIMA(2,1,2).

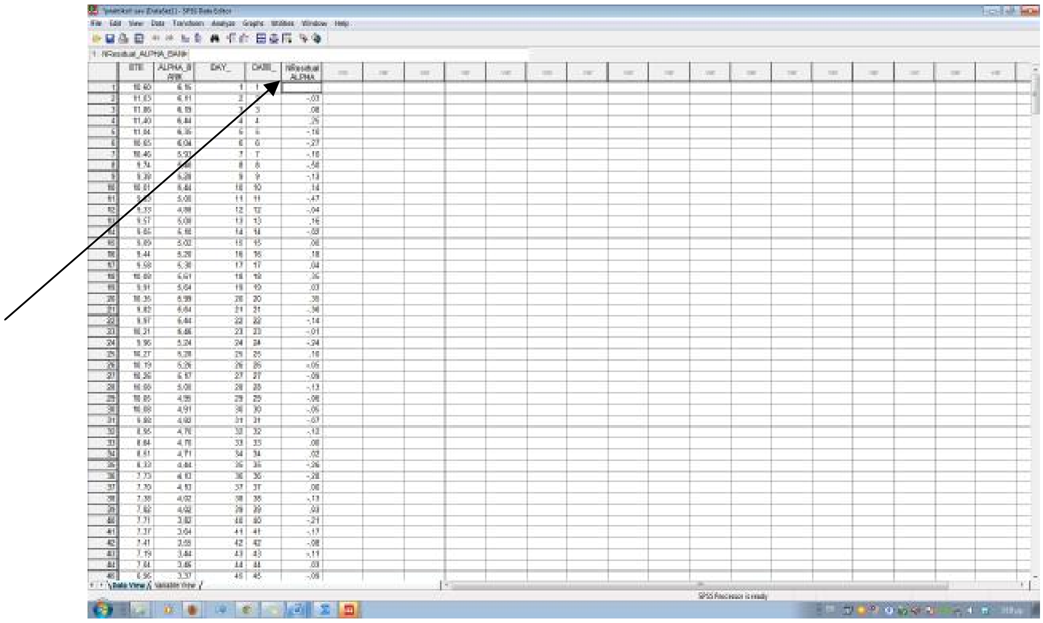
Από το menu του SPSS λοιπόν πηγαίνοντας Analyze-Time Series-Create Models και στο Time Series Modeler και συγκεκριμένα στο Save επιλέγεται το Noise Residuals.

Εικόνα 28 (Διαδικασία για την εύρεση των Καταλοίπων)



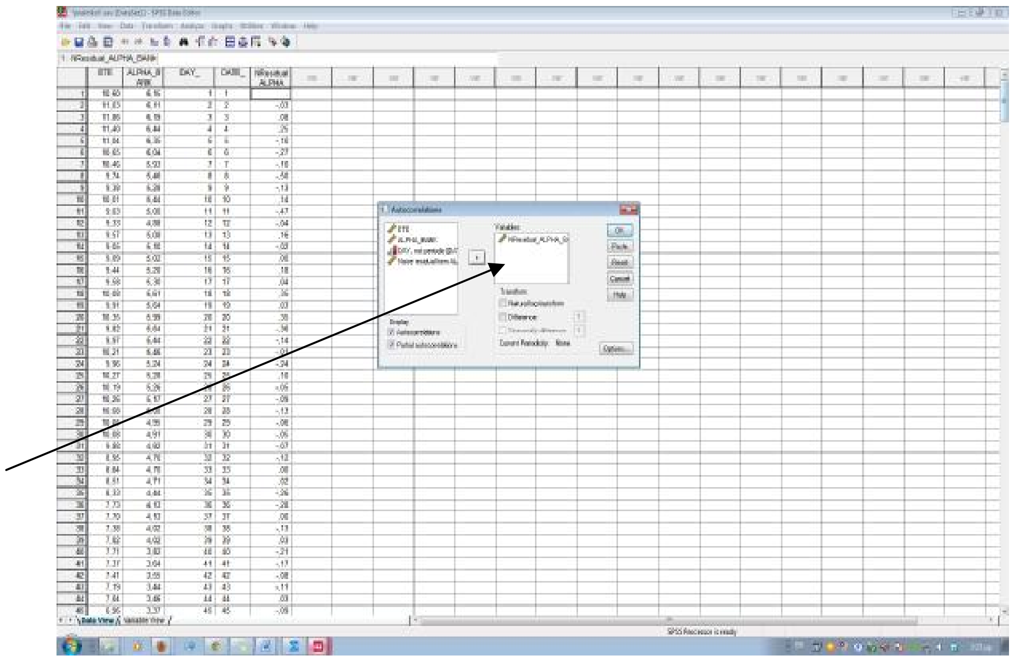
Πατώντας οκ εμφανίζεται η μεταβλητή Noise Residual ALPHA στο SPSS που δείχνει τα υπόλοιπα(residuals).

Εικόνα 29 (Τιμές των Καταλοίπων)



Από το menu του SPSS επιλέγοντας Graphs-Time Series-Autocorrelations η μεταβλητή NResidual μπαίνει στο Variables και στη συνέχεια ok για να βγει ο πίνακας αυτοσυσχέτισης στο output και να γίνει έλεγχος στασιμότητας.

Εικόνα 30 (Διαδικασία για έλεγχο στασιμότητας των Καταλοίπων)



Πίνακας 10 (Αυτοσυσχετίσεις**)**

Autocorrelations

Series: Noise residual from ALPHA_BANK-Model_1

Lag	Autocorrelation	Std. Error[a]	Box-Ljung Statistic Value	df	Sig.[b]
1	-,006	,045	,018	1	,894
2	,003	,045	,023	2	,989
3	-,034	,044	,594	3	,898
4	,000	,044	,594	4	,964
5	-,054	,044	2,068	5	,840
6	,056	,044	3,648	6	,724
7	,030	,044	4,119	7	,766
8	,047	,044	5,252	8	,730
9	,019	,044	5,446	9	,794
10	-,010	,044	5,497	10	,856
11	,070	,044	7,992	11	,714
12	,015	,044	8,114	12	,776
13	,036	,044	8,788	13	,789
14	-,022	,044	9,046	14	,828
15	,077	,044	12,107	15	,671
16	-,057	,044	13,769	16	,616

a. The underlying process assumed is independence (white noise).

b. Based on the asymptotic chi-square approximation.

Έλεγχος στασιμότητας

$H_0 : r = 0$

vs

$H_1 : r \neq 0$

Με την υπόθεση H_0 δεν υπάρχει αυτοσυσχέτιση, άρα είναι στάσιμη ενώ με την H_1 υπάρχει αυτοσυσχέτιση και δεν έχουμε στασιμότητα.

1) H_o: $\rho_1=0$ $\rho_1=0.894>\alpha=0.05$ Αποδοχή H_0 (δεν υπάρχει αυτοσυσχέτιση,

vs άρα στάσιμη)

H_1: $\rho_1 \neq 0$

2) H_o: $\rho_2=0$ $\rho_2=0.989>\alpha=0.05$ Αποδοχή H_0 (δεν υπάρχει αυτοσυσχέτιση,

vs άρα στάσιμη)

H_1: $\rho_2 \neq 0$

3) H_o: $\rho_3=0$ $\rho_3=0.898>\alpha=0.05$ Αποδοχή H_0 (δεν υπάρχει αυτοσυσχέτιση,

vs άρα στάσιμη)

H_1: $\rho_3\neq 0$

4) H_o: $\rho_4=0$ $\rho_4=0.964>\alpha=0.05$ Αποδοχή H_0 (δεν υπάρχει αυτοσυσχέτιση,

vs άρα στάσιμη)

H_1: $\rho_4\neq 0$

5) H_o: $\rho_5=0$ $\rho_5=0.84>\alpha=0.05$ Αποδοχή H_0 (δεν υπάρχει αυτοσυσχέτιση,

vs άρα στάσιμη)

H_1: $\rho_5\neq 0$

6) H_o: $\rho_6=0$ $\rho_6=0.724>\alpha=0.05$ Αποδοχή H_0 (δεν υπάρχει αυτοσυσχέτιση,

vs άρα στάσιμη)

H_1: $\rho_6\neq 0$

7) H_o: $\rho_7=0$ $\rho_7=0.766>\alpha=0.05$ Αποδοχή H_0 (δεν υπάρχει αυτοσυσχέτιση,

vs άρα στάσιμη)

H_1: $\rho_7\neq 0$

8) H_o: $\rho_8=0$ $\rho_8=0.73>\alpha=0.05$ Αποδοχή H_0 (δεν υπάρχει αυτοσυσχέτιση,

vs άρα στάσιμη)

H_1: $\rho_8\neq 0$

…

16) H_o: $\rho_{16}=0$ $\rho_{16}=0.616>\alpha=0.05$ Αποδοχή H_0 (δεν υπάρχει αυτοσυσχέτιση,

vs άρα στάσιμη)

H_1: $\rho_{16}\neq 0$

Άρα από τον έλεγχο καταλοίπων φαίνεται ότι η διαδικασία είναι στάσιμη καθώς για όλες της χρονολογικές υστερήσεις τα ρ είναι στάσιμα. Έτσι το υπόδειγμα ARIMA(2,1,2) επιβεβαιώνεται ότι είναι το πλέον κατάλληλο.

ΒΙΒΛΙΟΓΡΑΦΙΑ

1. Σύγχρονες μέθοδοι ανάλυσης χρονολογικών σειρών(Σοφία Δημελή-Καθηγήτρια Τμήματος Πληροφορικής Οικονομικού Πανεπιστημίου Αθηνών)
2. Κοινωνική Στατιστική με το SPSS(Δρ.Βασίλειος Δαφέρμος-Τμήμα Πολιτικής Επιστήμης Πανεπιστημίου Κρήτης)
3. Ανάλυση Χρονολογικών Σειρών(Διπλωματική Εργασία Χαράλαμπου Καρβέλη)
4. Πολλές πληροφορίες σχετικά με την εργασία μας αντλήσαμε και από το Διαδίκτυο
5. Εισαγωγή στην Οικονομετρία(Γεώργιος Χρήστου)
6. www.naftemporiki.gr