ΤΕΧΝΟΛΟΓΙΚΟ ΕΚΠΑΙΔΕΥΤΙΚΟ ΙΔΡΥΜΑ ΔΥΤΙΚΗΣ ΕΛΛΑΔΑΣ

ΣΧΟΛΗ ΔΙΟΙΚΗΣΗΣ ΚΑΙ ΟΙΚΟΝΟΜΙΑΣ

ΤΜΗΜΑ ΔΙΟΙΚΗΣΗΣ ΕΠΙΧΕΙΡΗΣΕΩΝ

ΠΤΥΧΙΑΚΗ ΕΡΓΑΣΙΑ

ΤΕΧΝΙΚΕΣ ΠΡΟΒΛΕΨΕΩΝ ΚΑΙ ΕΛΕΓΧΟΥ ΣΕ ΧΡΟΝΟΛΟΓΙΚΕΣ ΣΕΙΡΕΣ

ΦΟΙΤΗΤΗΣ:ΚΡΙΤΣΙΩΝΗΣ ΒΛΑΣΙΟΣ

ΕΠΟΠΤΕΥΩΝ ΚΑΘΗΓΗΤΗΣ:ΠΑΠΑΔΟΠΟΥΛΟΣ ΔΗΜΗΤΡΗΣ

ΠΑΤΡΑ 2016

ΠΕΡΙΕΧΟΜΕΝΑ

ΠΕΡΙΛΗΨΗ

Βασικός σκοπός της εργασίας είναι η μελέτη και πρόβλεψη χρονολογικών σειρών και τα ωφέλη που προσφέρει η συγκεκριμένη διαδικασία.Παρακάτω θα παρουσιαστούν τρόποι ανάλυσης χρονολογικών σειρών καθώς και βασικές έννοιες οι οποίες είναι σημαντικές για τα δεδομένα που εκπονούνται από την ανάλυση.Τέλος μετά την ανάλυση των χρονολογικών σειρών και την επεξεργασία δεδομένων θα μπορεί κάποιος να καταλήξει σε πρόβλεψη μελλοντικών δεδομένων.Η πρόβλεψη θα πραγματοποιηθεί με πολλούς τρόπους οι οποίοι θα παρουσιαστούν στη συνέχεια της εργασίας.Με λίγα λόγια θα χρησιμοποιηθούν παρελθοντικές μεταβλητές για την πρόβλεψη νέων.

ΕΙΣΑΓΩΓΗ

Αρχικά θα αναλυθεί το τι σημαίνει χρονολική σειρά και γιατί είναι ένα ιδιαίτερα σημαντικό εργαλείο.

Χρονολογική σειρά ονομάζεται ένα σύνολο από καταχωρήσεις δεδομένων μέσα σε τακτά χρονικά διαστήματα, η οποία χρησιμοποιείται για να μελετηθούν συστήματα,διαδικασίες και πρότυπα που εξελίσσονται χρονικά.Επίσης χρησιμοποιείται για την πρόβλεψη και την μελλοντική εξέλιξη των παρατηρήσεων.

Για παράδειγμα: Μια χρονολογική σειρά μπορεί να θεωρηθεί το σύνολο των μηνιαίων καταγραφών εσόδων μιας επιχείρησης μέσα στο έτος.Ακόμα και η καταγραφή πληθυσμού σε μια χώρα αναπαριστά μια χρονολογική σειρά.

Πολλές φορές στον 21^{ο} αιώνα και παλίοτερα επιχειρήσεις και οργανισμοί προσπαθούν να επιτύχουν τις πιο ακριβείς προβλέψεις για την μελλοντική εξέλιξη των χρονοσειρών που τους αφορούν, ένα αρκετά εύστοχο παράδειγμα είναι οι τιμές των ακινήτων.Από τη στιγμή που υπάρχουν οι τιμές των ακινήτων στο κέντρο της Αθήνας τα τελευταία 4 χρόνια ο χρήστης μπορεί να τρέξει στο PASW STATISTICS την εντολή για την πρόβλεψη των τιμών για το 2017.Έχοντας λοιπόν τις προβλέψεις για το 2017 θα πραγματοποιηθεί αγορά για το ακίνητο όταν εκτιμάται πως θα φτάσει στην χαμηλότερη τιμή.Φυσικά δεν είναι όλες οι χρονοσειρές έγκυρες για πρόβλεψη και οι προβλέψεις δεν είναι πάντα 100% ακριβείς.Τα κριτήρια για την εγκυρότητα πρόβλεψης θα αναλυθούν στη συνέχεια της εργασίας,όπως και τους τρόπους πρόβλεψης χρονοσειρών τόσο θεωρητικά όσο και πρακτικά.

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 σήμερα 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 φανερώνοτανι στην Οικονομία όπως για παράδειγμα σε τιμές μετοχών σε διαδοχικές ημέρες, η αξία των εξαγωγών μιας χώρας σε συνεχόμενους μήνες, τα κέρδη μιας επιχείρησης 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 η οποία θα αναλυθεί στη συνέχεια. Έτσι οι σύγχρονοι ερευνητές έχουν στην διάθεση τους όλα τα εφόδια για να μελετήσουν με επιτυχία μια χρονολογική σειρά που θα του παρουσιαστεί στην πράξη.(ΑΝΑΛΥΣΗ ΧΡΟΝΟΛΟΓΙΚΩΝ ΣΕΙΡΩΝ(ΚΟΥΝΕΤΑΣ e-class)

Επίσης θα παρουσιαστούν τα αιτιατά υποδείγματα πρόβλεψης. Με τα υποδείγματα αυτά πραγματοποιηθούν προβλέψεις μιας μεταβλητής με βάση την οικονομική και στατιστική σχέση που συνδέει τη παραπάνω μεταβλητή 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.

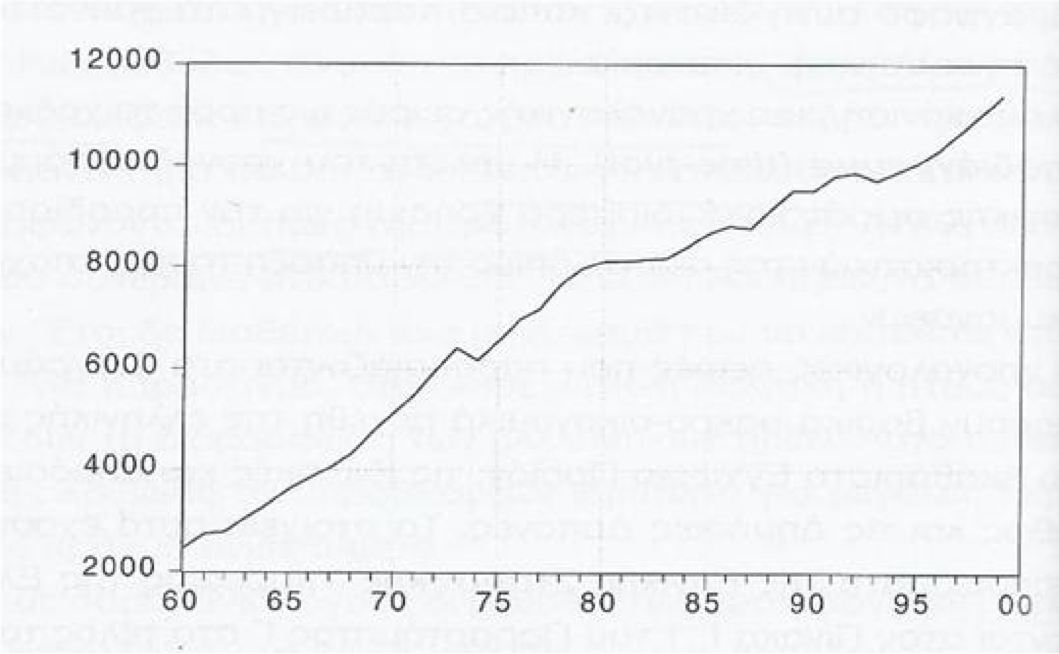
3) Τέλος, γίνεται η εκτίμηση του επιλεγμένου υποδείγματος και στη συνέχεια γίνονται οι προβλέψεις βάση των εκτιμήσεων.

Υπάρχουν όμως και τα μη αιτιατά υποδείγματα πρόβλεψης.Στα υποδείγματα χρονολογικών σειρών η πρόβλεψη στηρίζεται αποκλειστικά και μόνο στις προηγούμενες τιμές της ίδιας χρονολογικής σειράς στην οποία πρέπει να γίνει πρόβλεψη. Δηλαδή, προβλέπεται ημελλοντική συμπεριφορά μιας χρονοσειράς όχι σε συνάρτηση άλλων σειρών αλλά εξετάζοντας την προηγούμενη συμπεριφορά της δηλαδή το "ιστορικό'' της.Τα υποδείγματα χρονολογικών σειρών χωρίζονται σε καθοριστικά υποδείγματα (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. Κάποια από τα παραπάνω υποδείγματα θα αναλυθούν και θα σχολιαστούν 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.(Γεωργίου Κ. Χρήστου, (2007) Εισαγωγή στην Οικονομετρία, Β΄ Τόμος, Γ΄ Έκδοση, Αθηνά, Gutenberg.)

Πριν αναλυθούν οι βασικές έννοιες θα παρουσιαστούν γραφικά κάποιες χρονολογικές σειρές.

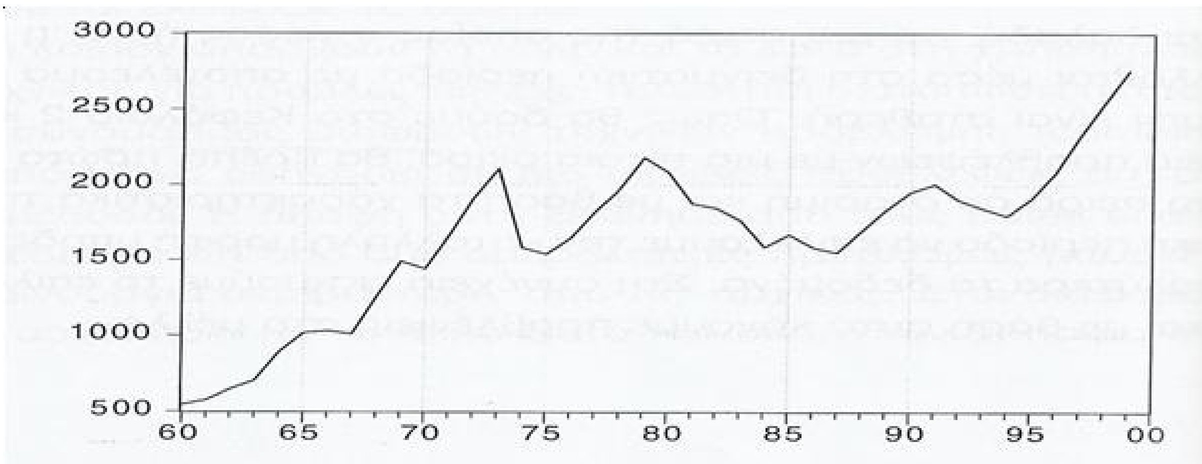
Η απεικόνιση μιας χρονολογικής σειράς ως προς το χρόνο ονομάζεται χρονοδιάγραμμα(time-plot).Η μελέτη του χρονοδιαγράμματος μιας χρονοσειράς είναι εξαιρετικά χρήσιμη για τον προσδιορισμό βασικών χαρακτηριστικών της σειράς όπως η ύπαρξη τάσης, εποχικότητας κ.α. Στο σημείο αυτό παραθέτονται κάποια διαγράμματα χρονολογικών σειρών τα οποία αφορούν κάποια βασικά μακροοικονομικά 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 σε σταθερές τιμές 1988 την περίοδο 1960-1999.Παρατηρείται πως σε γενικές γραμμές η πορεία είναι ανοδική με μικρές περιόδους μείωσης. Δηλαδή γίνεται αναφορά σε μια μη στάσιμη χρονολογική σειρά (non-stationary) η οποία παρουσιάζει έντονα ανοδική τάση (trend).Με άλλα λόγια, πρόκειται γιαμια σειρά που ο μέσος και η διακύμανση μεταβάλλονται μέσα στη δειγματική περίοδο με αποτέλεσμα η κατανομή να μην είναι σταθερή. Όπως θα παρουσιαστεί παρακάτω μετατρέπεται πρώτα τη σειρά σε στάσιμη, έπειτα επιλέγεται η πιο κατάλληλη μορφή υποδείγματος που εξηγεί καλύτερα τα δεδομένα και τέλος ακολουθεί η εκτίμηση και η πρόβλεψη.

Ακαθάριστο Εγχώριο Προϊόν (Δις.δρχ.-Σταθ τιμές 1988)



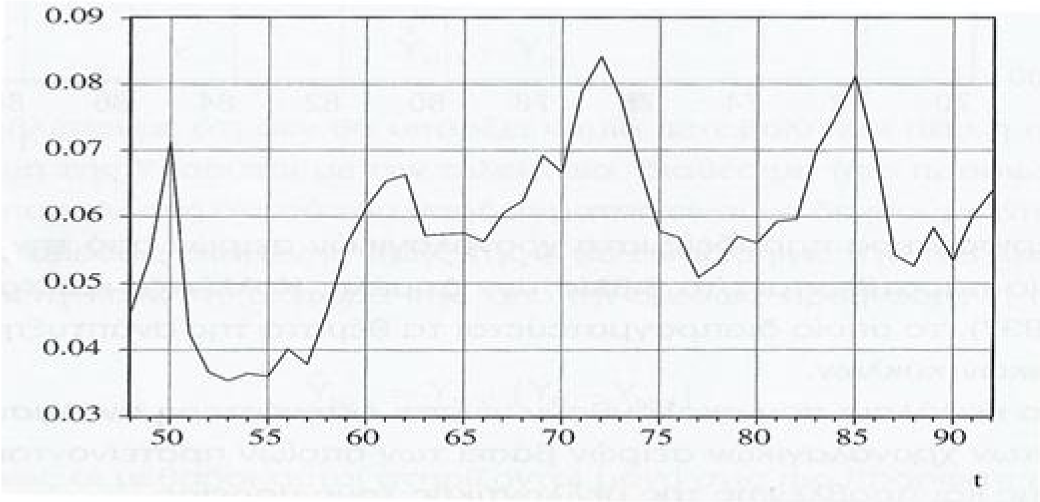
Στο επόμενο διάγραμμα φανερώνεται η πορεία των ιδιωτικών επενδύσεων στην Ελλάδα σε σταθερές τιμές το 1998.Η πορεία είναι μεν ανοδική αλλά όχι τόσο ομαλή όσο στο διάγραμμα του ΑΕΠ.Οι μεγαλύτερες αυξομειώσεις παρουσιάζονται στη δεκαετία 1970-1980 και οφείλονται σε διάφορες συγκυρίες της εποχής.

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



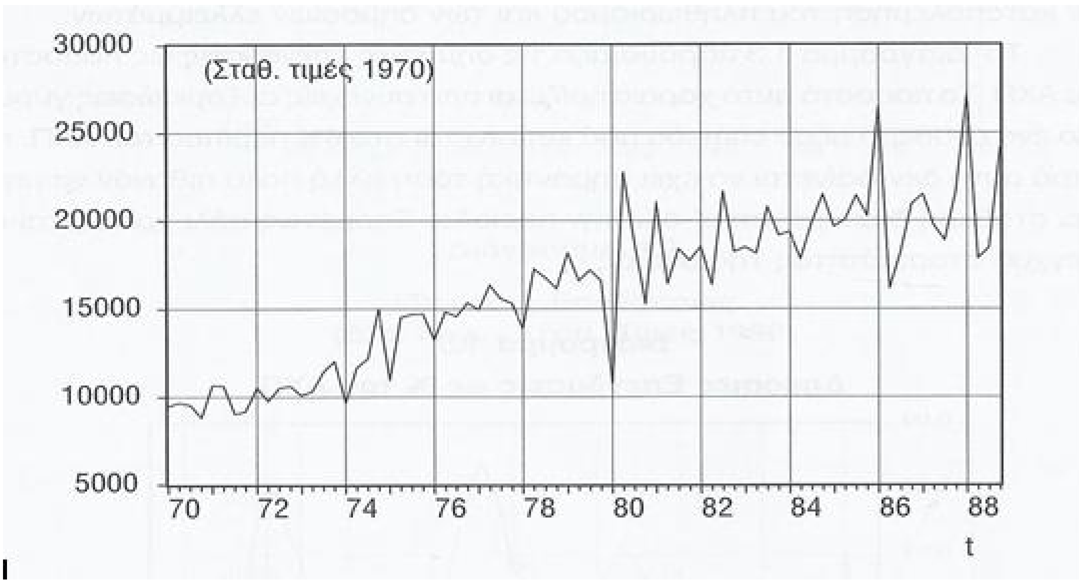
Στο επόμενο διάγραμμα παρουσιάζονται οι δημόσιες επενδύσεις ως ποσοστό του ΑΕΠ. Το ποσοστό αυτό χαρακτηρίζεται από συνεχείς αυξομειώσεις γύρω από ένα σταθερό μέσο επίπεδο που κυμαίνεται περίπου στο 6% του ΑΕΠ. Η συγκεκριμένη σειρά δεν φαίνεται να έχει σημαντική τάση ούτε σταθερή διακύμανση σε όλη την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

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

Δημόσιες Δαπάνες 1



ΒΑΣΙΚΕΣ ΕΝΝΟΙΕΣ

Σαν βασικές έννοιες θα αναλυθούν παρακάτω κάποια θεωρητικά εργαλεία τα οποία χρειάζονται για την κατανόηση των τρόπων πρόβλεψης διάφορων χρονολογικών σειρών.

Ξεκινώντας με τον όρο «στοχαστική διαδικασία»,σε μία χρονολογική σειρά υπάρχουν οι δείκτες y1,y2,…,yT οι οποίοι είναι ισαπέχοντες χρονικά.Οι παραπάνω δείκτες αποτελούν συγκεκριμένες τιμές όπου αποτελούν τυχαίες μεταβλητές(Y1,Y2,…,YT)οι οποίες είναι μέρος μόνο μιας άπειρης ακολουθίας.Αυτό ονομάζεται στοχαστική διαδικασία.Για παράδειγμα τα υποδείγματα Box-Jenkins αποτελούν μέρος στοχαστικών διαδικασιών και στα υποδείγματα αυτά ο τυχαίος παράγοντας αποτελεί το μηχανισμό μέσα από τον οποίοι δημιουργείται η χρονολογική σειρά.Με λίγα λόγια οι υπεύθυνοι για την πρόβλεψη των χρονοσειρών πρέπει να μελετήσουν την στοχαστική διαδικασία να επεξεργαστούν τα δεδομένα και τέλος να προχωρήσουν στην πρόβλεψη.

Απαραίτητη συνθήκη για τη μελέτη χρονολογικών σειρών είναι η ύπαρξη δεδομένων (data). Δεδομένα χρονολογικών σειρών συναντόνται 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).Με το δεύτερο ασχολείται πιο πολύς κόσμος και αφορά περισσότερα τα υποδείγματα χρονολογικών σειρών (time series models). Στα συγκεκριμένα υποδείγματα απώτερος σκοπός είναι η δημιουργία προβλέψεων και συνεπακόλουθα η μείωση της αβεβαιότητας και η καλύτερη εκτίμηση διαφόρων γεγονότων.

Στη συνέχεια πρέπει πάντα μια στοχαστική διαδικασία να είναι στάσιμη ώστε να είναι έγκυρη για πρόβλεψη. Μια στοχαστική διαδικασία μπορεί να θεώρηθεί στάσιμη ή μή στάσιμη.Στάσιμη θεωρείται όταν διατηρεί τις στατιστικές της ιδιότητες σταθερές,επίσης όταν:ισχύει ο ίδιος μέσος(μy) για όλα τα t,ίση διακύμανση($σ^2$y) για όλα τα t και σταθερή συνδιακύμανση(γk) για όλα τα t.(ΧΡΟΝΟΛΟΓΙΚΕΣ ΣΕΙΡΕΣ(Φ. ΚΟΥΝΤΟΥΡΗ)).

<u>Συνδιακύμανση</u>:Είναι συνάρτηση δύο τιμών της yT που απέχουν k περιόδους (χρονική υστέρηση). Ωστόσο,όταν υπάρχει εποχικότητα και τάση τότε η διαδικασία δεν είναι στάσιμη!

Επιπλέον σε κάθε στοχαστική διαδικασία υπάρχει αυτοσυσχέτιση μεταξύ των τιμών η οποία είναι αρκετά σημαντική στη διαμόρφωση των χρονοσειρών. Θέλωντας να ερευνηθεί κατά πόσο οι τιμές της διαδικασίας επηρεάζουν η μία την άλλη χρησιμοποιοείται η συνάρτηση αυτοσυσχέτισης.Η συνάρτηση αυτοσυσχέτισης (ACF) δείχνει την σχέση μετάξύ του συντελεστή αυτοσυσχέτισης και του χρονικού διαστήματος(k).Ο συντελεστής αυτοσυσχέτισης(ρκ) είναι η συσχέτιση μεταξύ δύο παρατηρήσεων σε k περιόδους.Όταν οι τιμές των συντελεστών αυτών φθίνουν γρήγορα προςτο 0 είναι ένδειξη στασιμότητας,ενώ όταν παίρνουν τιμές υψηλές κοντά στη μονάδα[-1,1]τότε η παραπάνω χρονοσειρά δεν είναι στάσιμη.

Τέλος υπάρχει και η μερική αυτοσυσχέτιση, η οποία χρησιμοποιείται για για μελέτη χαρακτηριστικών μιας χρονολογικής σειράς.Η λειτουργία του συντελεστή μερικής αυτοσυσχέτισης είναι να αφαιρεί άλλες επιδράσεις που ασκούνται πάνω στην χρονολογική σειρά, δηλαδή από τη στιγμή που έχουν ληφθεί υπόψη οι συσχετίσεις των ενδιάμεσων τιμών είναι ο νεός συντελεστής συσχέτισης.Με λίγα λόγια ο συντελεστής μερικής αυτοσυσχέτισης(PACF) συμβολίζεται με το (φρρ) όπου το ρ είναι η τάξη του συντελεστή και είναι ο συντελεστής του yt-ρ στην παλινδρόμηση.ΧΡΟΝΟΛΟΓΙΚΕΣ ΣΕΙΡΕΣ(Φ. ΚΟΥΝΤΟΥΡΗ)).

$Yt=\varphi_{1\rho}y_{t-1}+\varphi_{2\rho}y_{t-2}+\ldots\varphi_{\rho\rho}y_{t-\rho}+e_t$

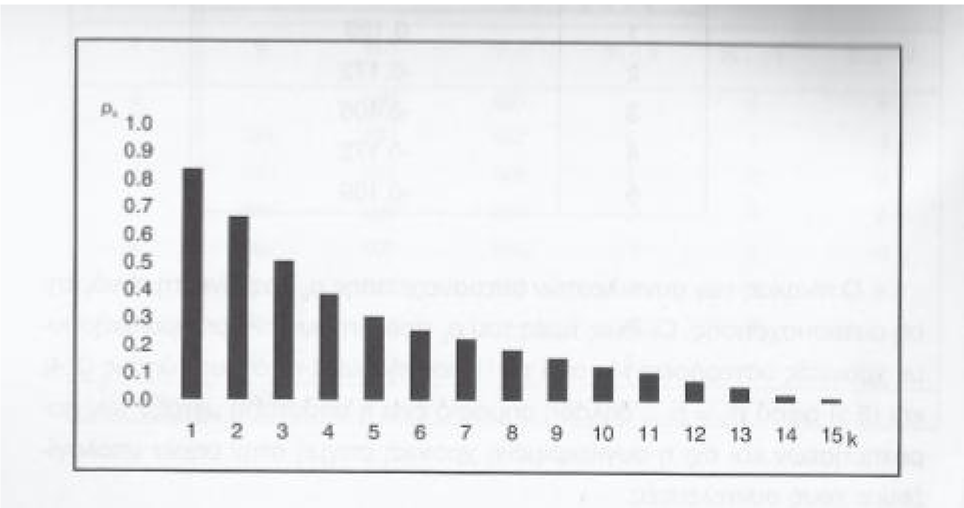
Ο μερικός συντελεστής αυτοσυσχέτισης όπως φαίνεται έχει δυο δείκτες. Ο αριστερός δείκτης δείχνει την χρονική υστέρηση της μεταβλητής σύμφωνα με το Yt-1, Yt-2... Ο δείκτης δεξιά δείχνει την μέγιστη τάξη της παλινδρόμησης. Άρα η συνάρτηση μερικής αυτοσυσχέτισης είναι μηδέν για s > p όταν θα μιλάμε για μια αυτοπαλίνδρομη διαδικασία τάξεως p. Με άλλα λόγια ισχύουν τα εξής: Για AR(1): α) ρ11=ρ1=α1 β) ρss= 0 για s > 1 Για AR(2): γ) ρ11=ρ1 δ) ε) ρss = 0 για s > 0 Για AR(p): στ) ρ11= ρ1 ζ) ,… , η) ρss = 0 για s > 2

Τα AR(p) θα αναλυθεί πιο μετά στην εργασία και οι μερικές αυτοσυσχετίσεις θα γίνουν πιο ξεκάθαρες στην πράξη.

Η διαγραμματική απεικόνιση των τιμών της ρκ με λίγα λόγια δίνει το διάγραμμα συνάρτησης αυτοσυσχέτισης (correlogram) και έτσι μπορεί να διαπιστωθεί 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 παραθέτονται 2 διαγράμματα το ένα στασιμότητας και το άλλο μη στασιμότητας. Στο πρώτο παρατηρείται 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

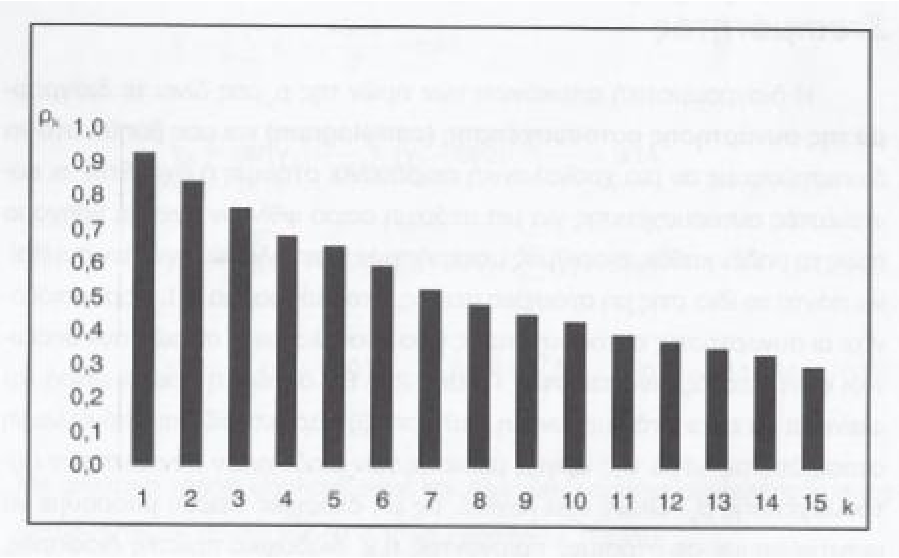
1)Στάσιμη χρονολογική σειρά

ΣΤΑΣΙΜΟΤΗΤΑ 1



2)μη στάσιμη χρονολογική σειρά

ΣΤΑΣΙΜΟΤΗΤΑ 2



Είναι σημαντικό να γίνει αναφορά στον έλεγχο σημαντικότητας των συντελεστών αυτοσυσχέτισης.

Ο έλεγχος του Bartlett (Bartlett Test) στηρίζεται στην ακόλουθη υπόθεση: Αν η χρονολογική σειρά είναι στάσιμη τότε οι συντελεστές αυτοσυσχέτισης ρ s του δείγματος ακολουθούν προσεγγιστικά την κανονική κατανομή με μέσο μηδέν και διακύμανση 1/ N (N το μέγεθος του δείγματος). Για μεγάλα δείγματα οι συντελεστές αυτοσυσχέτισης ρs του δείγματος ακολουθούν την κανονική κατανομή με μέσο μηδέν και διακύμανση 1/ N (N το μέγεθος του δείγματος)]. Θεωρούνται οι εξής υποθέσεις:

H0: ρ_s=0 (Η χρονολογική σειρά είναι στάσιμη)

H1: ρ_s#0 (Η χρονολογική σειρά δεν είναι στάσιμη) (ΑΝΑΛΥΣΗ ΧΡΟΝΟΛΟΓΙΚΩΝ ΣΕΙΡΩΝ(ΚΟΥΝΕΤΑΣ e-class).

ΣΥΝΙΣΤΩΣΕΣ ΧΡΟΝΟΛΟΓΙΚΩΝ ΣΕΙΡΩΝ

Οι συνιστώσες των χρονολογικών σειρών είναι τέσσερις,οι οποίες είναι εξίσου σημαντικές.Αρχίζοντας από την Τάση(T trend) κάποιες φορές μια χρονοσειρά δείχνει μια σταθερή τάση αύξησης ή μείωσης με την πάροδο του χρόνου.Μία τέτοια συμπεριφορά αποκαλείται τάση.Όταν γίνεται ένα διάγραμμα παρατηρήσεων προς τον χρόνο μπορεί να παρατηρηθεί ότι μία ευθεία γραμμή περιγράφει την αύξηση ή την μείωση στις σειρές καθώς κυλάει ο χρόνος.Επειδή πρέπει να πραγματοποιηθεί πρόβλεψη στην κίνηση μιας χρονοσειράς μέσα στον χρόνο χρησιμοποιείται η τάση σε συνδυασμό με το υπόδειγμα της απλής γραμμικής παλινδρόμησης.(Theil , H., (1967) Applied Economic Forecasting, Rand McNally).

$yt=\beta 0+\beta 1_t+\alpha_t$ όπου:

Το β0 είναι σταθερός αριθμός που μένει για όλα τα t.

Το β1 είναι η κλίση της ευθείας στο σημείο t.

Το α είναι τυχόν σφάλμα στο σημείο t.

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. Προηγουμένως έγινε αναφορά στο γεγονός πως η τάση μιας χρονοσειράς μπορεί να είναι γραμμική ή μη γραμμική. Ο πιο εύκολος τρόπος μέτρησης της τάσης 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. Σημαντικότερη όμως ειναι η εξίσωση γραμμικής τάσης. (Theil, H., (1967) Applied Economic Forecasting, Rand McNally).

Προηγουμένως έγινε αναφορά στο γεγονός πως η τάση μιας χρονοσειράς μπορεί να είναι γραμμική ή μη γραμμική. Ο πιο εύκολος τρόπος για να μετρηθεί η 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.

Παρακάτω αναγράφεται η εξίσωση γραμμικής τάσης (linear trend) η οποία έχει την εξής μορφή: $Y_t=b0+b1_t+e_t$.

Η γραμμική εξίσωση της τάσης είναι η απλούστερη καθώς εκφράζεται σε απόλυτους αριθμούς, εκτιμάται με την μέθοδο ελαχίστων τετραγώνων OLS και έχει τη μορφή $Y_t=b0+b1_t+e_t$ όπου Υτ είναι η τάση της χρονολογικής σειράς στην περίοδο t=(1,2…N) και το b_1 μέτρο της χρονολογικής σειράς ανά μονάδα χρόνου

Για να εντοπίστούν οι 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

Στην συνέχεια υπάρχει ο κύκλος(C circle) ο οποίος παρουσιάζει επαναλαμβανόμενες κυμάνσεις που διαρκούν πάνω από ένα έτος και εκφράζουν τις βραχυχρόνιες μεταβολές γύρω από την μακροχρόνια πορεία της σειράς. Ο κύκλος θα παρουσιαστεί καλύτερα στα γραφήματα αργότερα. 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 γνωστή και ως μέθοδος ελαχίστων τετραγώνων. Χρησιμοποιώντας τη γραμμή τάσης, υπολογίζονται για κάθε χρονική περίοδο οι αντίστοιχες τιμές της. Υπολογίζονται ο λόγος των τιμών της χρονολογικής σειράς και οι αντίστοιχες τιμές της τάσης. Αν βρεθούν τυχόν αποκλίσεις αυτής της διαίρεσης γύρω από τη μονάδα αυτό αποδίδεται στις εποχιακές και τυχαίες επιδράσεις στη χρονολογική σειρά.

Όσον αφορά την εποχικότητα(S seasonality) παρουσιάζεται αρκετά συχνά στις μηνιαίες παρατηρήσεις. εποχικότητα.Η εποχικότητα ακολουθεί έναν πλήρη κύκλο κατά τη διάρκεια ενός χρόνου.Ένα κλασσικό παράδειγμα είναι οι πωλήσεις των παγωτών,η ζήτηση των οποίων το καλοκαίρι είναι υψηλή ενώ τους άλλους μήνες χαμηλότερη.Με τη χρήση των εποχικών δεικτών μπορεί να εξαληφθεί η εποχικότητα και να υπολογιστεί η μακροχρόνια τάση.(ΑΝΑΛΥΣΗ ΧΡΟΝΟΛΟΙΚΩΝ ΣΕΙΡΩΝ(Χρυσανθη Παπαθανασοπουλου e-class).

Οι εποχικόι δείκτες που αναφέρθηκαν παραπάνω εμφανίζονται μια φορά για κάθε τρίμηνο του έτους και είναι ίδιοι για όλα τα έτη της χρονοσειράς. Επίσης ενδεικνύουν κατά πόσο οι τιμές του y είναι μικρότερες ή μεγαλύτερες απο τον μέσο όρο των τιμών μέσα στο έτος.(Οι εποχικοί δείκτες θα αναλυθούν καλύτερα στο πρακτικό κομμάτι της εργασίας). 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. Οι εποχιακοί δείκτες μιας χρονοσειράς κατασκευάζονται με την παρακάτω διαδικασία:

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.

Στο δεύτερο στάδιο υπολογίζεται ο λόγος των τιμών της χρονολογικής σειράς και των κινητών μέσων. Ο κινητός μέσος δίνει την κυκλική συνιστώσα και τη συνιστώσα της τάσης.

Στη συνέχεια για κάθε χρονική περίοδο υπολογίζεται ο λόγος των yt και των κινητών μέσων. Η διαδικασία αυτή απομακρύνει τη χρονολογική σειρά από τυχαίες αποκλίσεις και είναι ένα μέτρο της εποχιακής συνιστώσας.Έπειτα υπολογίζεται η εποχιακή συνιστώσα, δηλαδή διαιρούνται οι παρατηρήσεις της χρονολογικής σειράς με τις αντίστοιχες τιμές των κινητών μέσων.

Τέλος υπολογίζονται οι εποχιακοί δείκτες με την εξής διαδικασία:

1)υπολογίζονται οι αριθμητικοί μέσοι ανά τρίμηνο

2)αθροίζονται οι τριμηνιαίοι μέσοι

3)διαιρείται κάθε τριμηνιαίος μέσος με το σύνολο του και το αποτέλεσμα πολλαπλασιάζεται με εκείνο τον αριθμό που οδηγεί το μέσο εποχιακό δείκτη να ισούται με τη μονάδα.

Τέλος υπάρχει ο τυχαίος παράγοντας.Ο τυχαίος παράγοντας εμφανίζει τις βραχυχρόνιες αλλαγές των τιμών που οφείλονται σε μη προβλέψιμες καταστάσεις(όπως για παράδειγμα τα capital control).Είναι σημαντικό να σημειωθεί πως όταν γίνονται αυξομειώσεις των τιμών χωρίς να προκαλούνται από τον κύκλο την εποχικότητα και την τάση είναι πάντα υπέθυνος ο τυχαίος παράγοντας.Κατά την διάσπαση της χρονοσειράς στους 4 παραπάνω παράγοντες, δεν μπορεί να γίνει προβλέψη του τυχαίου παράγοντα ο οποίος ορισμένες φορές ονομάζεται κατάλοιπο(residual).(ΣΤΑΤΙΣΤΙΚΗ ΣΚΕΨΗ ΣΤΟΝ ΚΟΣΜΟ ΤΩΝ ΕΠΙΧΕΙΡΗΣΕΩΝ(ACZEL D. AMIR)).

Για την διάσπαση της χρονοσειράς χρησιμοποιούνται 2 μεθόδοι ή μοντέλα.

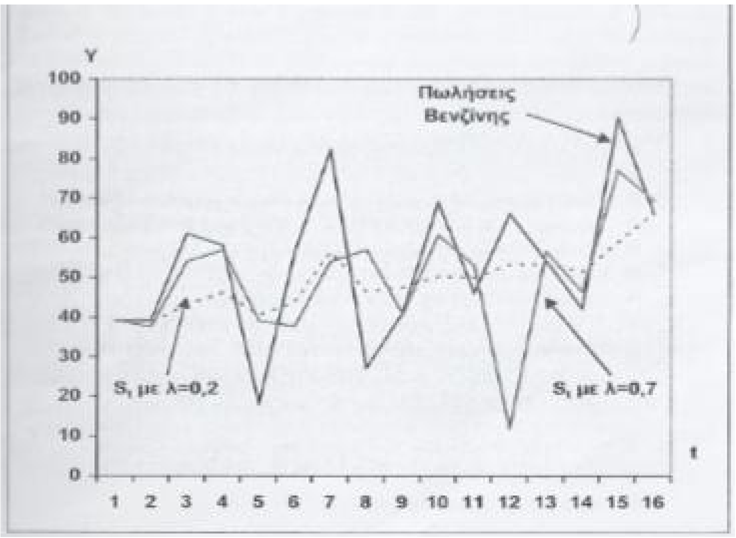
Το προσθετικό, το οποίο χρησιμοποιείται όταν οι συνιστώσες είναι ανεξάρτητες μεταξύ τους

$Y_t = T_t + C_t + S_t + e_t$ (T είναι η τάση,C ο κύκλος,S η εποχικότητα και e τα κατάλοιπα) το t ορίζεται από 1…n.

Το πολλαπλασιαστικό χρησιμοποιείται όταν οι τιμές των παρατηρήσεων είναι ακραίες και μη κανονικές.

$Y_t = T_t * C_t * S_t * e_t.$

Στα παραπάνω μοντέλα χρησιμοποιούνται συχνά τεχνικές εξομάλυνσεις οι οποίες σχεδόν εκμηδενίζουν τον τυχαία παράγονται, οι πιο διαδεδομένες είναι η εκθετική και η μέθοδος κινητών μέσων.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 $S_t=\lambda Y_t+(1-\lambda)S_{t-1}$ για $0<=\lambda<=1$ όπου Y_t η τιμή της χρονολογικής σειράς την περίοδο t, S_t η εξομαλυσμένη τιμή της σειράς την περίοδο t και λ= σταθμιστής (weight).Στο παρακάτω διάγραμμα διαπιστώνεται η συμπεριφορά της σειράς στις αρχικές αλλά και στις τιμές εξομάλυνσης. Διαπιστώνεται πως ο βαθμός εξομάλυνσης της σειράς εξαρτάται από την τιμή του συντελεστή εξομάλυνσης λ (όταν λ=0,2 υπάρχει καλύτερη εξομάλυνση απ ότι λ=0,7)



ΚΙΝΗΤΟΙ ΜΕΣΟΙ ΠΙΟ ΑΝΑΛΥΤΙΚΑ:

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

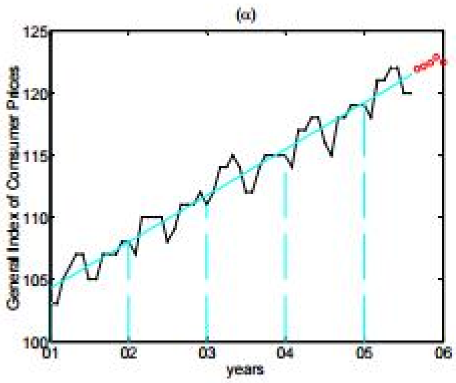
1) Με δεδομένες τις παρατηρήσεις της χρονολογικής σειράς και γνωρίζοντας σε πόσες περιόδους επιθυμείται ο υπολογισμός των κινητών μέσων υπολογίζονται τα κινητά σύνολα (moving totals).

2) Τα κινητά σύνολα διαιρούνται με τον αριθμό των παρατηρήσεων σε κάθε υποσύνολο, το πηλίκο των οποίων δίνει τους κινητούς μέσους. Στις χρονολογικές σειρές που υπάρχει περιττός αριθμός παρατηρήσεων δεν υπολογίζονται κινητοί μέσοι για τις πρώτες και τελευταίες (κ-1)/2 χρονικές περιόδους(κ= αριθμός παρατηρήσεων για κάθε κινητό σύνολο).

3) Υπολογίζεται η 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. 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.(Box, G. E. P. and Pierce, D.A (1970) Distribution of residual auto-correlation in autoregressive-integrated moving average time-series models. J. Amer. Statist. Ass).

Τέλος στον έλεγχο που πραγματοποιοείται πρέπει να γίνεται και μία μέτρηση ακριβείας για να γνωρίζει ο χειριστής πόσο αξιόπιστο είναι το μοντέλο.Αυτό γίνεται με την χρήση της μέσης απόλυτης απόκλισης (MEAN ABSOLUTE DEVIATION) MAD και του αθροίσματος τετραγώνων σφαλμάτων πρόβλεψης(SUM OF SQUARES OF FORECAST ERRORS) SSE.Παρακάτω φαίνεται ος τύπος του MAD:($\Sigma_{I=1}^{v}|yt-yi|$)/n ,όσο πιο κοντά στο 0 είναι η τιμή της MAD τόσο λιγότερες αποκλίσεις υπάρχουν άρα το μοντέλο είναι καλό για πρόβλεψη.Στη συνέχεια αναλύεται ο τύπος του SSE:$\Sigma_{I=1}^{v}(yt-y^{t})^{2}$,στον οποίο όσο πιο κοντά στο 0 βρίσκεται η τιμή του SSE τόσο καλύτερα γιατί έτσι υφίστανται λίγες ακραίες αποκλίσεις.

Παράδειγμα: Απαλοιφή τάσης και εποχικότητας στη χρονοσειρά GICP. Η διαδικασία απαλοιφής τάσης και εποχικότητας θα παρουσιαστεί μέσα από ένα πραγματικό παράδειγμα. Στο παρακάτω σχήμα, δίνεται ο γενικός δείκτης τιμών καταναλωτή (general index for consumer price, GICP) σε μηνιαίες τιμές από τον Ιανουάριο 2001 ως τον Αύγουστο 2005 (οι ανοιχτοί κύκλοι στο σχήμα είναι προβλέψεις που φαίνεται παρακάτω). Η χρονοσειρά έχει μήκος n = 56 .



Φαίνεται καθαρά, πως ο GICP παρουσιάζει σταθερή πληθωριστική τάση για όλη την περίοδο παρατήρησης αλλά και ασθενή ετήσια εποχικότητα. Η πληθωριστική τάση μπορεί να εξηγηθεί ικανοποιητικά από γραμμική συνάρτηση μt του χρόνου t (μήνα). Η προσαρμογή απλού γραμμικού μοντέλου παλινδρόμησης του Tt ως προς το t, θεωρώντας τις παρατηρήσεις yt για t από 1 μέχρι 56 ως τιμές του Tt , έδωσε 103.9 + 0.31t. Αφαιρώντας αυτή τη γραμμική τάση από τη χρονοσειρά GICP εμφανίζεται η χρονοσειρά απαλλαγμένη από τάση (detrended) y'=yt-Tt. Αυτή είναι η κατάλληλη χρονοσειρά για ανάλυση αν χρειαστεί να γίνει μελέτης της μεταβολής του GICP απαλλαγμένη από τον πληθωρισμό. Εδώ φαίνεται καλύτερα η ετήσια εποχικότητα του GICP αλλά δεν είναι προφανές ποια είναι η αναλυτική μορφή της περιοδικής συνάρτησης του χρόνου του ετήσιου κύκλου. Μπορεί όμως να εκτιμηθεί ο ετήσιος κύκλο (annual cycle) του GICP από τις μέσες τιμές του κάθε μήνα για τα έτη 2001 – 2005 (ως το 2004 για τους μήνες Σεπτέμβριο – Δεκέμβριο). Υπολογίζεται για τον Ιανουάριο ο μέσος όρος από τις τιμές Ιανουαρίου για τα 5 χρόνια και η τιμή αυτή αφαιρείται από τις 5 παρατηρήσεις του Ιανουαρίου (για τα έτη 2001 – 2005). Το ίδιο γίνεται και για τους άλλους μήνες. Η εποχική χρονοσειρά St (του επαναλαμβανόμενου ετήσιου κύκλου) έχει περίοδο d =12 . Αν χρειαστεί να γίνει εξέταση για το αν μπορούν να αντληθούνε κάποια επιπλέον πληροφορία από τη χρονοσειρά GICP πέρα από τη γραμμική τάση, που απαλείφθηκε, και τον ετήσιο κύκλο που εκτιμήθηκε, τότε από τη χρονοσειρά χωρίς τάση Tt αφαιρείται η εποχική χρονοσειρά St. Η χρονοσειρά των υπολοίπων et δίνεται ως et=Y’t-St=yt-Tt-St. Παρατηρείται ότι αυτή η χρονοσειρά είναι απαλλαγμένη από τάση και εποχικότητα και δε φαίνεται να έχει κάποια κανονικότητα ή δομή. Αν η χρονοσειρά et είναι εντελώς τυχαία, δηλαδή δεν έχει να δώσει καμιά πληροφορία, τότε η ανάλυση σταματάει εδώ και περιγράφεται μόνο η χρονική μεταβολή του GICP ως μια πληθωριστική γραμμική τάση σε συνδυασμό με έναν ετήσιο κύκλο. Έτσι χρειαστεί να γίνει πρόβλεψη του GICP για το μήνα Σεπτέμβριο θα επεκταθεί (extrapolate) η συνάρτηση τάσης για το μήνα Σεπτέμβριο και θα προστεθεί σε αυτήν την τιμή του ετήσιου κύκλου για το μήνα Σεπτέμβριο. Αναλυτικά, έχοντας τις τιμές GICP από τον Ιανουάριο 2001 ως τον Αύγουστο 2005 (n = 56), η τάση επεκτείνεται για το Σεπτέμβριο 2005 ως Tt+1= 103.9 + 0.31*(n+1)=>T57= 103.9 + 0.31*57=121.70. Η τιμή του ετήσιου κύκλου για το Σεπτέμβριο βρέθηκε ως η μέση τιμή Σεπτεμβρίου για τα έτη 2001 – 2004, 9s = 0.16 , και προστίθεται στην τιμή του T57 για να δώσει την πρόβλεψη του GICP για το μήνα Σεπτέμβριο y57=121.86.Όμοια μπορούν να πραγματοποιηθούν προβλέψεις και για τους επόμενους μήνες.

ΣΤΟΧΑΣΤΙΚΑ ΥΠΟΔΕΙΓΜΑΤΑ

Στην συνέχεια της εργασίας θα παρουσιαστούν τρεις διαφορετικοί τύποι στοχαστικών υποδειγμάτων χρονολογικών σειρών.Η πρώτη κατηγορία είναι τα Αυτοπαλίνδρομα υποδείγματα Ar, στη συνέχεια ακολουθούν τα υποδείγματα κινητών μέσων MA, και τέλος τα μικτά υποδείγματα ARMA που είναι ο συνδυασμός των παραπάνω.Καλό για να ειναι όλα πιο ξεκάθαρα στη συνέχεια πρέπει να σημειωθεί πως υπάρχουν χρονοσειρές πάνω στις οποίες δεν γίνεται πρόβλεψη Πχ(τα τυχερά παιχνίδια τζόκερ κλπ).Αυτές ονομάζονται τυχαίες χρονολογικές σειρές η αλλιώς (λευκός θόρυβος).(Nelson, C. R., (1973) Applied Time Series Analysis for Managerial Forecasting, Holden – Day).

Αν συμβολιστεί με εt, μια τέτοια σειρά, τότε θα θεωρείται λευκός θόρυβος, αν έχει σταθερό μέσο (συνήθως μηδέν) E(εt)=0 ,σταθερή διακύμανση, γ0=E(εt 2)=σ2 , και οι τιμές της δεν αυτοσυσχετίζονται, γκ=E(εt,εt-κ)=0, για κ≠0.

Οι χρονολογικές σειρές τέτοιου είδους είναι πάντα στάσιμες και έχουν μηδενικούς συντελεστές αυτοσυσχέτισης και μερικής αυτοσυσχέτισης.

Από αυτές τις χρονοσειρές βγαίνουν και τα υποδείγματα τυχαίας διαδρομής, αυτό ισχύει γιατί κάθε τιμή της τυχαίας διαδρομής είναι απλά ένα επιπλέον σφάλμα et.

Για να φανεί καλύτερα στην πράξη έστω ότι η πρώτη τιμή της τυχαίας διαδρομής είναι το y0 άρα:

Y1=y0+e1

Y2=y1+e2 όπου ισούται με y2=y0+e1+e2 και πάει λέγοντας...

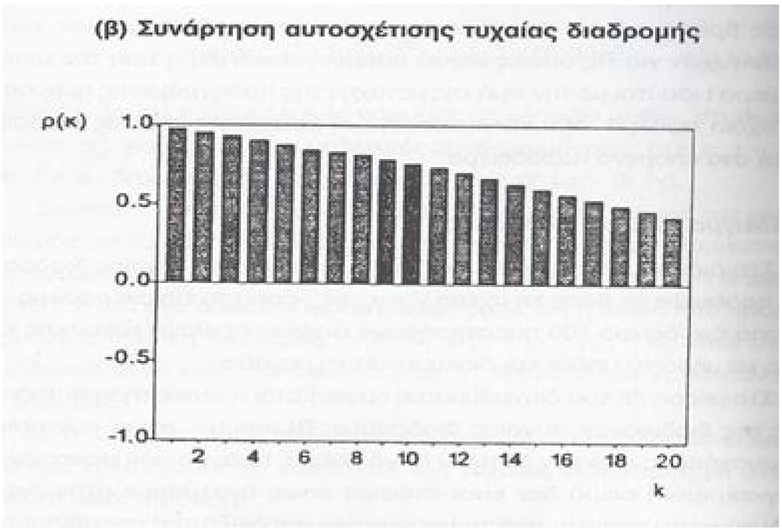
Η παραπάνω εξίσωση αντιπροσωπεύει την τυχαία διαδρομή άραν εν τέλει οι διαδοχικές τιμές της χρονολογικής σειράς έιναι από κοινού τυχαίες εξαρτημένες μεταβλητές αφού συσχετίζονται με τις ίδιες προηγούμενες τιμές των τυχαίων σφαλμάτων et.

Αν το υπόδειγμα της τυχαίας διαδρομής περιλαμβάνει και σταθερό όρο τότε μετονομάζεται σε υπόδειγμα της **τυχαίας διαδρομής με σταθερά(random walk with drift)** το οποίο έχει αυτή τη μορφή yt=β+y_{t-1}-εt όπου β είναι η σταθερά της εξίσωσης. Αν ξεκινήσει η διαδικασία σύγκρισης το υπόδειγμα τυχαίας διαδρομής με το αντίστοιχο της σταθεράς, φαίνεται πως στο υπόδειγμα με σταθερά οι πρώτες διαφορές είναι εν μέρει στοχαστικές και σταθερές(οι πρώτες διαφορές θα αναλυθούν λίγο αργότερα στην εργασία). Χαρακτηριστικότερο παράδειγμα χρονολογικών σειρών που ακολουθούν τυχαίες διαδρομές βρίσκεται 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.

τυχαία διαδρομή

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

συνάρτηση αυτοσυσχέτισης



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

ΑΥΤΟΠΑΛΙΝΔΡΟΜΑ ΥΠΟΔΕΙΓΜΑΤΑ

Ξεκινώντας με τα αυτοπαλίνδρομα υποδείγματα,η γενική μορφή του αυτοπαλίνδρομου υποδείγματος εκφράζεται ως εξής:

$Yt=\alpha0+\alpha1Y_{t-1}+\alpha2Y_{t-2}+\alpha pY_{t-p}+e_t$.

Στην παραπάνω εξίσωση το α0 είναι σταθερός όρος ενώ α1 και α2 ονομάζονται αυτοπαλινδρούμενοι παράμετροι και το et είναι ο τυχαίος παράγοντας στον οποίο έχει γίνει αναφορά παραπάνω.Επιπλέον το p δείχνει την τάξη του αυτοπαλίνδρομου υποδείγματος και το μήκος της.

Το υπόδειγμα αυτό ονομάζεται αυτοπαλίνδρομο για τον εξής λόγο:Η τιμή της Yt όπου είναι η εξαρτημένη μεταβλητή παλινδρομείται στις προηγούμενες τιμές της και όχι σε άλλες ανεξάρτητες μεταβλητές.Το υπόδειγμα είναι τάξεως p και το συναντάμε ως (Autoregressive model)ή για συντομία:Ar(p).(ΣΗΜΕΙΩΣΕΙΣ ΟΙΚΟΝΟΜΕΤΡΙΑΣ(e-class ΑΤΕΙ ΠΑΤΡΩΝ).

Τα αυτοπαλίνδρομα υποδείγματα είναι γενικότερα χρήσιμα διότι μπορεί να πραγματοποιηθεί πρόβλεψη πάνω σε όλων των ειδών τα δεδομένα, αλλά για να γίνει αυτό πρέπει πρώτα να βρεθεί η τάξη του μοντέλου.

Στη συνέχεια αναγράφεται ένα αυτοπαλίνδρομο υπόδειγμα πρώτης τάξης Ar(1)

$Yt=\alpha+\alpha1Y_{t-1}+e_t$.

Όπως παρατηρήθηκε στις βασικές έννοιες πρέπει ένα μοντέλο τέτοιου τύπου να είναι στάσιμο ώστε να γίνει σωστή πρόβλεψη, για να είναι στάσιμο το μοντέλο πρέπει να ισχύει το εξής:

$|\alpha 1|<1$.

Στο μοντέλο πρώτης τάξης γίνεται μια φορά η χρονική υστέρηση, επειδή αρχίζει από τη μονάδα και φθίνει προς το 0 και υπάρχει μόνο ένας στατιστικά σημαντικός συντελεστής μερικής αυτοσυσχέτισης.

Για α1 η συνάρτηση αυτοσυσχέτισης αρχίζοντας από την μονάδα φθίνει γεωμετρικά και τείνει προς το μηδέν καθώς το p αυξάνει, το ίδιο ισχύει και αν το α1<0 καθώς η συνάρτηση θα φθίνει γεωμετρικά προς το μηδέν αλλά με αρνητικό πρόσημο.

Για το Ar(1) ο συντελεστής αυτοσυσχέτισης είναι: ρ1=α1

Βέβαια είναι καλό σε αυτό το σημείο να εξηγηθεί και ο τελεστής υστέρησης (Lag).

Στην ανάλυση των χρονολογικών σειρών χρησιμοποιείται αρκετές φορές ο τελεστής υστερήσεως L (lag operator) ο οποίος διευκολύνει γενικότερα στις αλγεβρικές πράξεις. Οπότε αν υπάρχει σε μια σχέση Yt-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 υστερηθεί η μεταβλητή Y .

Στη συνέχεια αναλύονται τα αυτοπαλίνδρομα υποδείγματα 2ης τάξης AR(2).

Ο τύπος για τα υποδείγματα 2ης τάξης είναι ο εξής:

$Yt=α0+α1Y_{t-1}+α2Y_{t-2}+e_t$

Επομένως οι συνθήκες στασιμότητας είναι οι εξής:

α1+α2<1

α2-α1<1

-1<α1<1

Και ο συντελεστής αυτοσυσχέτισης είναι για $ρ1=α1/(1-α2)$, $ρ2=α2+(α1^2/(1-α2))$

Χρειάζονται πάντα οι συντελεστές αυτοσυσχέτισης γιατί τα αυτοπαλίνδρομα υποδείγματα προσαρμόζονται μόνο όταν οι συντελεστές φθίνουν γρήγορα προς το 0 αλλά ποτέ δεν μηδενίζονται.

Στη συνέχεια όσον αφορά τους συντελεστές μερικής αυτοσυσχέτισης πρέπει να είναι στατιστικά σημαντικοί ανάλογα με την τάξη του μοντέλου, δηλαδή αν το μοντέλο είναι 1ης τάξης πρέπει να έχει έναν σημαντικό συντελεστή μερικής αυτοσυσχέτισης και όλοι οι υπόλοιποι μηδενίζονται.

Σε αυτό το σημείο πρεπει είναι καλό να δούμε πως οι συντελεστές μερικής αυτοσυσχέτισης υπολογίζονται με τον κανόνα του Cramer ως εξής:

$ρ_{11}=ρ_1$

$$ρ_{22}=\frac{\begin{matrix}1 & \rho 1\\ \rho 1 & \rho 2\end{matrix}}{\begin{matrix}1 & \rho 1\\ \rho 1 & 1\end{matrix}}=\frac{\rho 2-(\rho 1*\rho 1)}{1-(\rho 1*\rho 1)}$$

$$ρ_{33}=\frac{\begin{matrix}1 & P1 & \rho 1\\ \rho 1 & 1 & \rho 2\\ \rho 2 & \rho 1 & \rho 3\end{matrix}}{\begin{matrix}1 & \rho 1 & \rho 2\\ \rho 1 & 1 & \rho 1\\ \rho 2 & \rho 1 & 1\end{matrix}}$$

Οι συντελεστές $ρ_{ss}$, για διάφορες τιμές του s, είναι η συνάρτηση μερικής αυτοσυσχέτισης. Είναι προφανές ότι, για μία αυτοπαλίνδρομη διαδικασία p τάξης, η συνάρτηση μερικής αυτοσυσχέτισης είναι μηδέν για $s > p$, σύμφωνα με τον ορισμό της μερικής αυτοσυσχέτισης.

Τέλος πρέπει να πραγματοποιείται πάντα έλεγχος σημαντικότητας στους συντελεστές αυτοσυσχέτισης και μερικής αυτοσυσχέτισης.Στην πράξη επειδή τόσο οι πραγματικές τιμές των αυτοσυσχετίσεων $ρ_s$ όσο και των πραγματικών μερικών αυτοσυσχετίσεων $ρ_{ss}$ δεν είναι γνωστές χρησιμοποιούνται οι αντίστοιχες εκτιμήσεις τους από το δείγμα. Εάν στους τύπους που δίνουν την θεωρητική τιμή της συνάρτησης μερικής αυτοσυσχετίσεως, τα $ρ_s$ θα αντικατασταθούν με τις δειγματικές τιμές τους τότε, οι τιμές των $ρ_{ss}$ που προκύπτουν είναι δειγματικές. Με βάση τις εκτιμήσεις αυτές μπορεί να γίνει έλεγχος σημαντικότητας των παραμέτρων στον πληθυσμό. Για μεγάλα δείγματα, οι εκτιμήσεις $\hat{ρ}_s$ των αυτοσυσχετίσεων $ρ_s$ κατανέμονται κανονικά με μέση τιμή το μηδέν και διακύμανση 1/T, όπου T είναι το μέγεθος του δείγματος. Το ίδιο ισχύει και για τις εκτιμήσεις των μερικών αυτοσυσχετίσεων ss $\hat{ρ}_{ss}$ για υστερήσεις μεγαλύτερες από την τάξη p της AR διαδικασίας.(ΣΗΜΕΙΩΣΕΙΣ ΟΙΚΟΝΟΜΕΤΡΙΑΣ(e-class ΑΤΕΙ ΠΑΤΡΩΝ)).

Στη συνέχεια θα αναπτυχθεί ο έλεγχος της σημαντικότητας του συντελεστή $ρ_s$, δηλαδή θα ελεγχθεί η υπόθεση:

$H0:\hat{ρ}_s=0$ και $H1:\hat{ρ}_s=1$

Ο έλεγχος αυτός θα γίνει με την βοήθεια της στατιστικής: $t_s=\hat{\rho}_s/\sqrt{(1/t)}=\hat{\rho}_s\sqrt{T}$. Για δεδομένο επίπεδο σημαντικότητας α=5% η μηδενική απορρίπτεται αν $|ts|>2$, διαφορετικά θα υπάρξει αποδοχή της H0 . Το 95% διάστημα εμπιστοσύνης για αυτόν τον έλεγχο είναι : $\hat{\rho}_s-2/\sqrt{T}<=\rho_s<=\hat{\rho}_s$. Στον έλεγχο σημαντικότητας των συντελεστών μερικής αυτοσυσχέτισης ισχύουν τα ίδια με τα παραπάνω. Δηλαδή ο ρ_{ss} είναι σημαντικός αν και μόνο αν I $\rho_{ss}\sqrt{T}$ I<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 ρ_{22} είναι στατιστικά σημαντικός, ενώ ο συντελεστής ρ_{33}είναι στατιστικά μη σημαντικός. Τότε, συμπεραίνεται ότι η τάξη p του υποδείγματος είναι 2.(ΤΕΧΝΙΚΕΣ ΟΙΚΟΝΟΜΕΤΡΙΑΣ ΚΑΙ ΑΝΑΛΥΣΗΣ ΧΡΟΝΟΛΟΓΙΚΩΝ ΣΕΙΡΩΝ(ΘΕΟΔΩΡΟΣ ΣΙΜΟΣ).

ΥΠΟΔΕΙΓΜΑΤΑ ΚΙΝΗΤΩΝ ΜΕΣΩΝ

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 στάσεις εργασίας όπου επηρεάζουν την οικονομία όχι μόνο βραχυχρόνια αλλά και μακροχρόνια. Σε αντίθεση με τα αυτοπαλίνδρομα υποδείγματα που κάνουν πρόβλεψη μέσα από την παλινδρόμηση,τα υποδείγματα κινητών μέσων MA(q) (Moving Average) χρησιμοποιούν έναν συνδυασμό σφαλμάτων πρόβλεψης, τρεχόντων και παρελθοντικών όπως θα φανεί από την γενική μορφή:

$Yt=\mu+\varepsilon t-\theta 1\varepsilon t-1-\theta 2\varepsilon t-2\ldots-\theta q\varepsilon t-q$

Όπου εt: είναι τα σφάλματα και θ είναι η σταθερά

Επίσης κάθε υπόδειγμα κινητών μέσων είναι πάντα στάσιμο γιατί πληρεί όλες τις συνθήκες στασιμότητας.

Παρακάτω αναλύεται η συνάρτηση για q=1 δηλαδή πρώτης τάξης υπόδειγμα MA(1)

$Yt=\mu+\varepsilon 1-\theta 1\varepsilon t-1$

Επίσης για το MA(1) ισχύουν τα παρακάτω:

α) $E(Y_t) = 0$

β) $y_0 = Var(Y_t) = \sigma^2(1+\theta_1^2)$

γ) $y_k = Cov(Y_t, Y_{t-k}) = -\theta_1\sigma_\varepsilon^2$, για $k=1$

Το γεγονός ότι οι αυτοσυσχετίσεις μεγαλύτερης από πρώτης τάξης είναι όλες μηδενικές σημαίνει πρακτικά ότι η MA(1) χρονοσειρά έχει μνήμη μόνο μιας περιόδου, δηλαδή μια οποιαδήποτε παρατήρηση της Y_t συσχετίζεται με την προηγούμενη ή την επόμενη παρατήρηση, αλλά δεν συσχετίζεται με καμία άλλη.

Ακόμη, θετική τιμή της παραμέτρου θ σημαίνει θετική αυτοσυσχέτιση πρώτης τάξης ρ_1, ενώ αντιστοίχως αρνητική τιμή σημαίνει αρνητική αυτοσυσχέτιση. Επίσης, η μεγαλύτερη τιμή που μπορεί να λάβει η ρ_1 είναι 0.5 για θ = 1, ενώ η μικρότερη τιμή είναι -0.5 για θ = -1.

Το υπόδειγμα μηδενίζεται μετά από μία υστέρηση γιατί έχει μόνο ένα στατιστικά σημαντικό συντελεστή αυτοσυσχέτισης, αυτό μπορεί να το δει κάποιος κι από την συνάρτηση αυτοσυσχέτισης ρκ.(ΣΤΑΤΙΣΤΙΚΗ ΣΚΕΨΗ ΣΤΟΝ ΚΟΣΜΟ ΤΩΝ ΕΠΙΧΕΙΡΗΣΕΩΝ(ACZEL D. AMIR)).

:$\rho\kappa=-\theta/1+\theta^2$, για κ=1

Με πιο απλά λόγια η Yt εξαρτάται άμεσα από την Yt-1 αλλά επειδή μιλάμε για διαστήματα μίας περιόδου,η πρόβλεψη πραγματοποιείται μόνο για την επόμενη περίοδο.Παρακάτω φαίνεται το υπόδειγμα κινητών μέσων δεύτερης τάξης MA(2) με τον εξής τύπο:

$Yt=\mu+\varepsilon 1-\theta 1\varepsilon t-1+\varepsilon 2-\theta 2\varepsilon t-2$

Όπως έχει προαναφερθεί το υπόδειγμα είναι στάσιμο και τα Yt δεν επηρεάζονται μεταξύ τους πάνω από ένα διάστημα δύο περιόδων.

Στη συνέχεια για τα υποδείγματα κινητών μέσων η συνάρτηση αυτοσυσχέτισης ρκ γίνεται μηδενική μετά την τελευταία υστέρηση q σε αντίθεση με τα αυτοπαλίνδρομα στα οπόια δεν μηδενίζεται.Τέλος θα γίνει συζήτηση για την αντιστρεψιμότητα πάνω στα υποδείγματα κινητών μέσων.

Η ιδιότητα της αντιστρεψιμότητας αφορά στην μετατροπή ενός υποδείγματος MA (q) σε υπόδειγμα AR(∞) (άπειρης τάξης). Αντίστοιχα θα ισχύει ότι ένα υπόδειγμα AR (p) είναι αντιστρέψιμο αν μπορεί να λάβει μορφή ενός υποδείγματος MA(∞) .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): Ιθ1Ι<1

MA(2):θ1+θ2<1

θ2-θ1<1

Ιθ2Ι<1

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

Για παράδειγμα ας εξεταστελι το υπόδειγμα MA(2).

Για το υπόδειγμα MA(2) ισχύουν τα παρακάτω:

$Y0=(1+θ1^2+θ2^2)σ^2$

$Y1=(θ1+θ2θ1)σ^2$ $ρ1=(θ1+θ2θ1)/(1+θ1^2+θ2^2)$

$Y2=θ2σ^2$ $ρ2=θ2/(1+θ1^2+θ2^2)$

Yς=0 για ς>2 ρς=0 για ς>2

Γενικά, οι αυτοσυνδιακυμάνσεις και συνεπώς η συνάρτηση αυτοσυσχέτισης είναι μηδέν μετά από q υστερήσεις, δηλαδή για $s > q$. Η συμπεριφορά της συνάρτησης μερικής αυτοσυσχέτισης προσομοιάζει με αυτή της συνάρτησης αυτοσυσχέτισης μίας AR διαδικασίας.

Ας σημειωθεί ότι, γενικά, ενώ η συνάρτηση αυτοσυσχέτισης μίας AR(p) διαδικασίας μπορεί να εκτείνεται στο άπειρο, η συνάρτηση αυτοσυσχέτισης μίας MA(q) διαδικασίας μηδενίζεται μετά από q υστερήσεις. Με άλλα λόγια, η “μνήμη” της εξαντλείται σε q περιόδους. Αντίθετα, η συνάρτηση μερικής αυτοσυσχέτισης μίας AR(p) διαδικασίας μηδενίζεται μετά από p υστερήσεις, ενώ η συνάρτηση μερικής αυτοσυσχέτισης μίας MA(q) διαδικασίας εκτείνεται στο άπειρο.

Οι μερικές αυτοσυσχετίσεις (ρςς) μπορούν να εκφραστούν ως συνάρτηση των αυτοσυσχετίσεων (ρς) με τον ίδιο ακριβώς τρόπο όπως και για τις AR διαδικασίες. Για παράδειγμα, για την MA(1) διαδικασία έχουν:

$ρ1=θ1/(1+θ1^2)$ και ρ2=0.

Οι συντελεστές μερικής αυτοσυσχέτισης είναι οι εξής:
Ο ρ11=ρ1=θ1/(1+θ1^2)

$$\rho_{22} = \frac{\begin{vmatrix} 1 & \rho_1 \\ \rho_1 & \rho_2 \end{vmatrix}}{\begin{vmatrix} 1 & \rho_1 \\ \rho_1 & 1 \end{vmatrix}} = \frac{\begin{vmatrix} 1 & \rho_1 \\ \rho_1 & 0 \end{vmatrix}}{\begin{vmatrix} 1 & \rho_1 \\ \rho_1 & 1 \end{vmatrix}} = \frac{-\rho_1^2}{1-\rho_1^2} = -\frac{\left(\frac{\theta_1}{1+\theta_1^2}\right)^2}{1-\left(\frac{\theta_1}{1+\theta_1^2}\right)^2} = -\frac{\theta_1^2}{1+\theta_1^2+\theta_1^4}$$

$$\rho_{33} = \frac{\begin{vmatrix} 1 & \rho_1 & \rho_1 \\ \rho_1 & 1 & \rho_2 \\ \rho_2 & \rho_1 & \rho_3 \end{vmatrix}}{\begin{vmatrix} 1 & \rho_1 & \rho_2 \\ \rho_1 & 1 & \rho_1 \\ \rho_2 & \rho_1 & 1 \end{vmatrix}} = \frac{\begin{vmatrix} 1 & \rho_1 & \rho_1 \\ \rho_1 & 1 & 0 \\ 0 & \rho_1 & 0 \end{vmatrix}}{\begin{vmatrix} 1 & \rho_1 & 0 \\ \rho_1 & 1 & \rho_1 \\ 0 & \rho_1 & 1 \end{vmatrix}} = \frac{\rho_1^3}{1-2\rho_1^2}$$

Και πάει λέγοντας για τους υπόλοιπους συντελεστές.

ΜΕΙΚΤΑ ΥΠΟΔΕΙΓΜΑΤΑ ARMA

Η γενική μορφή ARMA υποδειγμάτων είναι η εξής:

$Yt=δ+α1yt-1+α2yt-2+…αpyt-p+εt+θ1εt-1+θ2εt-2+…θqεt-q$

Πολλές στάσιμες χρονοσειρές δεν μπορούν να μοντελοποιηθούν αποκλειστικά ως MA ή AR χρονοσειρές, μιας και παρουσιάζουν ιδιότητες και από τις δύο κατηγορίες.Έτσι, οι χρονοσειρές που εκφράζονται ως συνδυασμός χρονοσειρών κινητού μέσου και αυτοπαλινρομικών χρονοσειρών τάξης (p,q) και συμβολίζεται ως **ARMA(p,q)**.

Όπως είναι ανιληπτό πρόκειται για έναν συνδυασμό p αυτοπαλίνδρομων όρων και q όρων κινητών μέσων ARMA(p,q).Επιπλέον θεωρείται πως ένα υπόδειγμα AR(1) είναι ίσο με ένα υπόδειγμα ARMA(1,0) από τη στιγμή που δεν υπάρχουν q όροι κινητών μέσων,επίσης ισχύει το ίδιο και για το MA(1) το οποίο ισούται με ARIMA(0,1).

Παρακάτω παρουσιάζεται ένα μεικτό υπόδειγμα πρώτης τάξης ARMA(1,1).

$Yt=δ+α1yt-1+εt+θ1εt-1$

Για να είναι μια τέτοια σειρά στάσιμη θα πρέπει το Ια1Ι < 1 όπως και στα αυτοπαλίνδρομα υποδείγματα AR(p).

Παρακάτω φαίνεται ο συντελεστής συσχέτισης για το μοντέλο ARMA(1,1).

$P1=α1+(θ1/y0)*σ^2$

Στην παραπάνω σχέση φαίνεται πως στον συντελεστή αυτοσυσχέτισης παίζουν εξίσου σημαντικό ρόλο η τάξη του μοντέλου κινητών μέσων και η τάξη του αυτοπαλίνδρομου υποδείγματος, αλλά αυτό ισχύει μόνο για μοντέλα ARMA $1^{ης}$ τάξης καθώς για τα μοντέλα μεγαλύτερης τάξης ο συντελεστής αυτοσυσχέτισης υπολογίζεται μόνο από το αυτοπαλίνδρομο μέρος.

Συντελεστής αυτοσυσχέτισης για s>1:

$Ps=α1ρ_{s-1}$

Το S συμβολίζει την τάξη μεικτών υποδειγμάτων.

Η συνάρτηση αυτοσυσχέτισης για την ARMA(1,1) διαδικασία φθίνει γεωμετρικά καθώς αυξάνεται το s, η μείωση αρχίζει από το ρ1 και όχι από τη μονάδα (ρ0=1) όπως στην περίπτωση της AR(1) διαδικασίας. Η συνάρτηση μερικής αυτοσυσχέτισης, συμπεριφέρεται όπως στην περίπτωση της MA(1) διαδικασίας, δηλαδή φθίνει γεωμετρικά. Αυτό το υπόδειγμα μπορεί να γραφεί ως μια καθαρά MA διαδικασία, αλλά και ως μια αυτοπαλίνδρομη διαδικασία με άπειρους όρους AR(∞).(ΑΝΑΛΥΣΗ ΧΡΟΝΟΛΟΓΙΚΩΝ ΣΕΙΡΩΝ(ΚΟΥΝΕΤΑΣ e-class)).

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.

Σε αυτό το σημείο θα παρουσιαστούν μερικά θεωρητικά παραδείγματα για την καλύτερη κατανόηση των ARMA υποδειγμάτων

1)Έστω ότι η μεταβολή των καθαρών κερδών μιας επιχείρησης ακολουθεί το υπόδειγμα $P_t=\varepsilon_t-0.21e_{t-1}+0.155\varepsilon_{t-2}-0.512P_{t-1}$ όπου P_t η μεταβολή των κερδών (*χωρίς να έχουν χρησιμοποιηθεί διαφορές*). Ποιο είναι το παραπάνω υπόδειγμα και πως ερμηνέυεται; Αποτελεί στάσιμη και αντιστρέψιμη χρονοσειρά;

Απάντηση)$P_t=\varepsilon_t-0.21e_{t-1}+0.155\varepsilon_{t-2}-0.512P_{t-1}$=ARMA(1,2).Για να αποδειχθεί πως ένα υπόδειγμα ARMA(1,2) είναι στάσιμο φτάνει να ισχύει η συνθήκη στασιμότητας για το υπόδειγμα AR(1). Άρα Ια1Ι<1 δηλαδή Ι0.512Ι<1 το οποίο ισχύει επομένως είναι στάσιμη χρονοσειρά.Η αντιστρεψιμότητα ενός υποδείγματος ARMA βασίζεται στο MA μέρος. Επομένως,

Θ1+θ2<1 -0.21+0.155<1 = -0.055<1 το οποίο ισχύει

Θ2-θ1 <1 0.155+0.23<1 = 0.365<1 το οποίο ισχύει

Θ2<1 0.155<1 το οποίο ισχύει

Άρα η χρονοσειρά είναι αντιστρέψιμη όσον αφορά τις προυποθέσεις.

2) Για την παρακάτω χρονοσειρά $Yt=0.22 + \varepsilon t - 0.068^{\varepsilon}{}_{t-1}$, $\sigma^2=4$

1. Είναι η παραπάνω χρονοσειρά στάσιμη και αντιστρέψιμη;

2. Να βρεθούν ο μέσος, η διακύμανση και οι αυτοσυσχετίσεις ρ1 ρ2 ρ22.

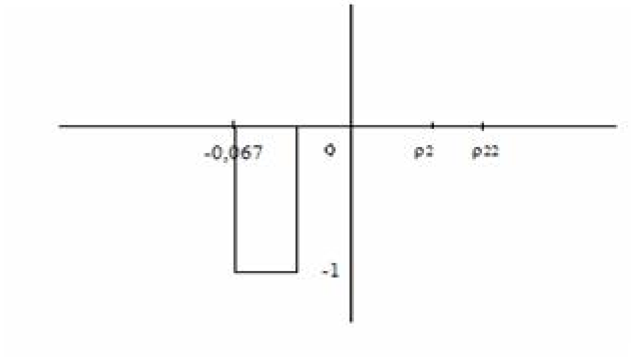
3. Να γίνει το διάγραμμα αυτοσυσχέτισης.

4. Να διατυπωθεί το υπόδειγμα με τον συμβολισμό του τελεστή υστέρησης L

Απάντηση)

1) $Yt=0.22 + \varepsilon_t - 0.068^{\varepsilon}{}_{t-1}$, $\sigma^2=4$, MA(1) Για να είναι αντιστρέψιμη θα πρέπει: Ιθ1Ι<1 δηλαδή Ι0.068Ι<1 0.068<1, το οποίο ισχύει. Άρα η χρονοσειρά είναι αντιστρέψιμη.Το υπόδειγμα είναι στάσιμο, διότι κάθε υπόδειγμα κινητού μέσου είναι στάσιμο.
2) $Y_t=\varepsilon_t - \theta_1\varepsilon_{t-1}$
   Yt=Yt-μ
   Yt=Yt-μ= $\varepsilon t-\theta 1\varepsilon_{t-1}$
   Μέσος E(Yt)=0 $Yt=\varepsilon t-\theta 1\varepsilon_{\tau-1} + \mu$, $0.22+\varepsilon t - 0.068\varepsilon_{t-1} = \varepsilon t-\theta 1\varepsilon_{\tau-1} + \mu$, μ=0.22
   $Y0 =\sigma_\varepsilon^2(1+\theta^2)$ $y_0 = 4(1+(-0.068)^2)$ y0= 4.016
   $Y1=-\theta\sigma_\varepsilon^2$ y1=-0.068*4 y1=-0.272

   $P\kappa=\frac{yk}{y0}$
   Για κ=1 ισχύει $y\kappa= \frac{-\theta}{1+\theta*\theta}$
   Για κ>0 ισχύει yκ=0
   Για κ=1 , $\rho1=\frac{-\theta}{1+\theta*\theta}=\frac{-0.068}{1+0.068*0.068}$=-0,067
   3)



4)

$L\varepsilon_t=\varepsilon_{t-1}$ $\quad$ $yt=0.22+\varepsilon t-0.068\varepsilon_{t-1}$

$L^2\varepsilon_t=\varepsilon_{t-2}$ $\quad$ $Yt=\varepsilon_t -0.068L\varepsilon_t$

$Yt=\varepsilon_t\ (1-0.068)$

Τέλος απάντησης.

ΜΗ ΣΤΑΣΙΜΕΣ ΔΙΑΔΙΚΑΣΙΕΣ,ΥΠΟΔΕΙΓΜΑΤΑ ARIMA

Τα ολοκληρωμένα αυτοπαλινδρομικά μοντέλα κινητού μέσου όρου (AutoRegressive-Integrated-Moving Average) ARIMA(p,d,q) είναι στοχαστικά μαθηματικά μοντέλα με γίνεται η προσπάθεια περιγραφής της διαχρονικής εξέλιξης κάποιου φυσικού μεγέθους. Δεδομένου ότι για την πλειοψηφία των φυσικών μεγεθών είναι αδύνατη η πλήρης γνώση και καταγραφή όλων των παραγόντων που επηρεάζουν την εξέλιξη τους στο χρόνο, είναι πολύ δύσκολη η διαχρονική περιγραφή του μεγέθους από ένα ντετερμινιστικό μοντέλο. Από την άλλη μεριά, η εξάρτηση τέτοιων μεγεθών από μη ντετερμινιστικούς παράγοντες (π.χ. καιρός, τυχαία γεγονότα) καθιστά δυνατή την περιγραφή της διαχρονικής τους εξέλιξης από ένα στοχαστικό μοντέλο, με το οποίο μπορεί να υπολογιστεί η πιθανότητα με την οποία η τιμή του μεγέθους βρίσκεται σε κάποιο διάστημα.(Γεωργίου Κ. Χρήστου, (2007) Εισαγωγή στην Οικονομετρία, Β΄ Τόμος, Γ΄ Έκδοση, Αθηνά, Gutenberg).

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

Από την αρχή της εργασίας έχουν αναλυθεί στοχαστικές διαδικασίες και υποδείγματα τα οποία είναι πάντα στάσιμα, αυτό σημαίνει πως ο μέσος, η διακύμανση και οι αυτοσυνδιακυμάνσεις δεν εξαρτώνται από τον χρόνο(επίσης πραγματοποιείται έλεγχος για εποχικότητα ή τάση).Παρακάτω θα εξετασθούν διαδικασίες οι οποίες είναι μη στάσιμες, και θα γίνει ανάλυση της διαδικασίας μετατροπής σε στάσιμες.

Στην οικονομετρία θεωρείται σημαντικό να είναι στάστιμες οι χρονοσειρές, έτσι ώστε να γίνονται σωστές προβλέψεις χωρίς πολλά σφάλματα και για να αποφεύγονται άλλα πρβλήματα.

Για να μετατραπεί μια χρονοσειρά σε στάσιμη χρησιμοποιούνται οι πρώτες ή και δεύτερες διαφορές.

Τι είναι οι διαφορές;

Οι διαφορές συμβολίζονται με Δ και χρησιμοποιούνται πάνω σε μη στάσιμες χρονολογικές σειρές με την εξής συνάρτηση:

$\Delta Yt=yt-yt-1$

Όταν μια χρονοσειρά μετατρέπεται σε στάσιμη με τις πρώτες διαφορές η σειρά λέγεται πλέον πρώτης τάξης I(1). Σε γενικές γραμμές, αν d είναι ο αριθμός των διαφορών που μετατρέπει μία σειρά σε στάσιμη, η σειρά καλείται ολοκληρωμένη d τάξης και συμβολίζεται ως I(d).

Ένα κλασσικό παράδειγμα μη στάσιμης χρονοσειράς είναι η τυχαία διαδρομή.

Αναγράφεται λοιπόν μια αυτοπαλίνδρομη διαδικασία πρώτης τάξης:

$Yt=\alpha 1yt-1+et$ όπου et είναι λευκός θόρυβος.

Αν α=1 τότε ισχύει:

$Yt=Yt-1+et$

Το παραπάνω υπόδειγμα ονομάζεται τυχαίος περίπατος ή τυχαία διαδρομή.Μία στοχαστική διαδικασία που ακολουθεί τυχαία διαδρομή δεν είναι ποτέ στάσιμη.Από τη στιγμή όμως που χρησιμοποιούνται οι πρώτες διαφορές η Yt γίνεται αμέσως στάσιμη.

ΔYt=yt-yt-1

Όταν υπάρχει σταθερός όρος α,δηλαδή : $Y_t=α+Y_{t-1+}ε_t$, το υπόδειγμα λέγεται τυχαία διαδρομή με περιπλάνηση.Μια στοχαστική διαδικασία που ακολουθεί την τυχαία διαδρομή δεν είναι στάσιμη.

Οι περισσότερες χρονολογικές σειρές δεν έχουν τα χαρακτηριστικά στάσιμων διαδικασιών.Όπως στην περίπτωση του τυχαίου περιπάτου , που οι διαφορές πρώτης τάξης οδήγησαν σε στάσιμη διαδικασία (διαδικασία λευκού θορύβου) , έτσι και πολλές άλλες περιπτώσεις χρονολογικών σειρών μετατρέπονται σε στάσιμες διαδικασίες με τη βοήθεια διαφορών.

Οι σειρές είναι χρήσιμο να είναι στάσιμες γιατί έτσι αποφεύγονται διάφορα προβλήματα , όπως για παράδειγμα το πρόβλημα της φαινομενικής παλινδρόμησης .Τέτοιες μη-στάσιμες χρονοσειρές αποκαλούνται **ολοκληρωμένες τάξης d ,**όπου η τάξη d είναι ο αριθμός των διαφορών που απαιτούνται για να μετατρέπει μια σειρά σε στάσιμη και παριστάνεται ως I(d) .

Πιο συγκεκριμένα ,αν η Y_t είναι μια μη-στάσιμη ολοκληρωμένη χρονοσειρά πρώτης τάξης ,η νέα χρονοσειρά $W_{t:}$

$W_t = ΔY_t = Y_τ – Y_{t-1}$

Η οποία προκύπτει χρησιμοποιώντας τον τελεστή υστέρησης πρώτων διαφορών Δ =(1-L) ,θα είναι στάσιμη .Αντιστοίχως ,αν η Y_t είναι ολοκληρωμένη χρονοσειρά δεύτερης τάξης , τότε χρησιμοποιώντας τον τελεστή υστέρησης δεύτερων διαφορών $Δ^2 =(1-L)^2$,προκύπτει η στάσιμη χρονοσειρά W_t :

$W_t = Δ^2Y_t =(1-L)^2 = (1-2L+L^2)Y_t$

$= Y_t – 2LY_t + L^2Y_t = Y_t – 2Y_{t-1} + Y_{t-2}$

ή με απευθείας εφαρμογή ,θα είναι :

$W_t = Δ^2Y_t = Δ(Y_t – Y_{t-1}) = ΔY_t - ΔY_{t-1}$

$= Y_t – Y_{t-1} – (Y_{t-1} – Y_{t-2})$

$= Y_t – Y_{t-1} – Y_{t-1} + Y_{t-2}$

$= Y_t – 2Y_{t-1} + Y_{t-2}$

Ανάλογα αποτελέσματα προκύπτουν χρησιμοποιώντας τελεστές διαφορών τρίτης ,τέταρτης έως και d τάξης .

Με τον τρόπο αυτό, αν κανείς μελετά μια μη-στάσιμη χρονοσειρά, μπορεί εύκολα να την μετατρέψει σε στάσιμη και να εφαρμόσει τεχνικές ανάλυσης στη νέα στάσιμη χρονοσειρά. Το υπόδειγμα τυχαίας διαδρομής, είναι σειρά ολοκληρωμένη πρώτης τάξεως, αφού μετατρέπεται σε στάσιμη με τις πρώτες διαφορές. Προφανώς, μια στάσιμη σειρά, όπως ο λευκός θόρυβος, θεωρείται ολοκληρωμένη μηδενικής τάξεως.

Πλέον πρέπει να η ταυτοποίηση των μοντέλων με την χρήση των αυτοπαλίνδρομων υποδειγμάτων, των κινητών μέσων και των διαφορών να είναι εύκολη.

Εδώ έρχεται να βοηθήσει η μεθοδολογία BOX-JENKINS.Η συγκεκριμένη μέθοδος βοηθάει στο να βρεθεί ένα μοντέλο ARIMA(p,d,q) από το οποίο εξάγονται δεδομένα για την πρόβλεψη και αποτελείται από 3 μέλη 1) ταυτοποίηση 2)εκτίμηση 3)διαγνωστικός έλεγχος.

Το πρώτο στάδιο είναι η ταυτοποίηση του μοντέλου, δηλαδή η εκτίμηση των (p,d,q). Η ταυτοποίηση ξεικινάει από τις διαφορές, δηλαδή γίνεται έλεγχος για το αν η χρονοσειρά είναι στάσιμη (Επαλήθευση για το αν οι αυτοσυσχετίσεις φθίνουν γρήγορα προς το 0).Από τη στιγμή που είναι στάσιμη το d παίρνει την τιμή 0, στη περίπτωση που δεν έιναι στάσιμη χρησιμοποιούνται οι πρώτες και δεύτερες διαφορές και ορίζεται το d αναλόγως.Έχοντας καταλήξει σε μια στάσιμη χρονοσειρά ορίζεται το p από τις αυτοσυσχετίσεις και το q από τις μερικές αυτοσυσχετίσεις.(ΑΝΑΛΥΣΗ ΧΡΟΝΟΛΟΓΙΚΩΝ ΣΕΙΡΩΝ(Χρυσανθη Παπαθανασοπούλου e-class)).

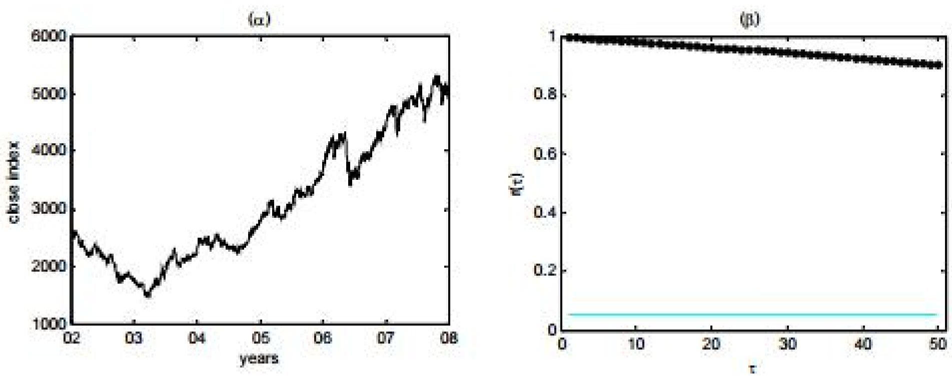
Το δεύτερο στάδιο είναι η εκτίμηση των p παραμέτρων της αυτοπαλίνδρομης διαδικασίας α1,α2….αp και των παραμέτρων των κινητών μέσων θ1,θ2,…θq.

Το τελευταίο στάδιο είναι ο διαγνωστικός έλεγχος ο οποίος αναφέρεται στο κατά πόσο καλή είναι η εγκυρότητα του υποδείγματος, πιο συγκεκριμένα πραγματοποιεί στατιστικούς ελέγχους και τσεκάρει την σημαντικότητα των συντελεστών καθώς και την συμπεριφορά των καταλοίπων et.

Τα κατάλοιπα δεν πρέπει να αυτοσυσχετίζονται, δηλαδή πρέπει να ακολουθούν μια διαδικασία λευκού θορύβου, ένας τρόπος με τον οποίο γίνεται έλεγχοσ είναι με την μέθοδο BOX-PIERCE.Η παραπάνω μέθοδος λειτουργεί ως εξής: Χρησιμοποιείται η μηδενική υπόθεση Η0:ρ1=ρ2=ρ3=ρ4…=ρm=0 και εφαρμόζεται ο τύπος: $Qbp=T\Sigma^{m}_{s-1}\rho^{2}_{s}$ όπου ρ_s είναι οι δειγματικές αυτοσυσχετίσεις των καταλοίπων και Τ ο αριθμός των παρατηρήσεων.Η μέθοδος BOX-PIERCE ακολουθεί την κατανομή X^2 αυτό σημαίνει πως όταν η Qbp είναι μεγαλύτερη της κρίσιμης τιμής X^2 τότε απορρίπτεται η Η0 άρα υπάρχουν ισχυρές αυτοσυσχετίσεις.Είναι σημαντικό να προσθεθεί πως υπάρχει και η μέθοδος LJUNG-BOX η οποία είναι παρόμοια με την BOX-PIERCE, η μόνη διαφορά είναι πως η LJUNG-BOX αναφέρεται σε μεγαλύτερα δείγματα δηλαδή για Τ>50.(ΣΗΜΕΙΩΣΕΙΣ ΟΙΚΟΝΟΜΕΤΡΙΑΣ(e-class ΑΤΕΙ ΠΑΤΡΩΝ)).

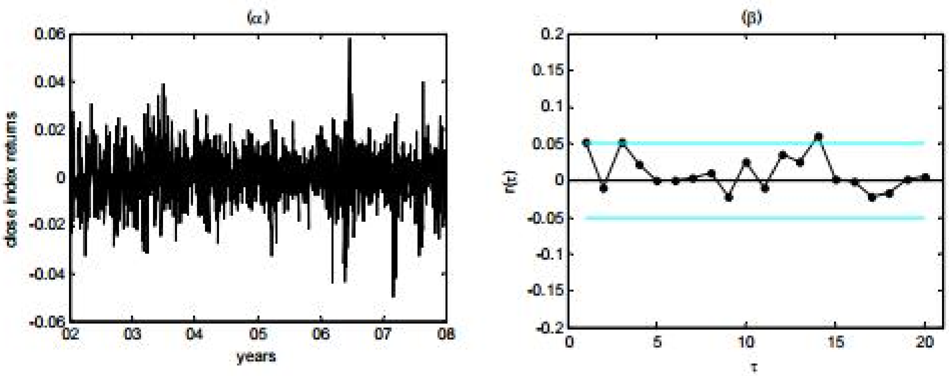
Παρακάτω παρουσιάζεται ένα ένα παράδειγμα της διαδικασίας box-jenkins.

Παράδειγμα: Παρατηρώντας τη διαδικασία Box-Jenkins στην πράξη στη χρονοσειρά του ημερήσιου γενικού δείκτη χρηματιστηρίου αξιών Αθηνών (ΧΑΑ). Επιλέγεται μια αρκετά μεγάλη χρονοσειρά με μήκος n =1496 που καλύπτει την περίοδο 1/1/2002 – 31/12/2007. Η χρονοσειρά {yt} για t από 1 έως n σε αυτήν την περίοδο φαίνεται να έχει ομοσκεδαστικότητα, δηλαδή η διακύμανση των τιμών της δε διαφέρει σημαντικά σε διαφορετικές περιόδους. Φυσικά περιέχει τάσεις όπως εύκολα φαίνεται από το διάγραμμα της χρονοσειράς και της αυτοσυσχέτισης, η οποία φθίνει πολύ αργά από τη μονάδα προς το μηδέν (βλέπε παρακάτω σχήμα)



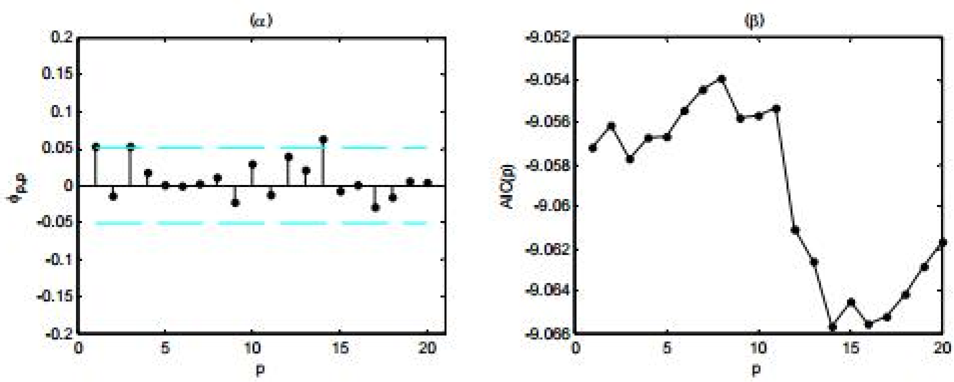
Σχήμα (α) Η χρονοσειρά του γενικού δείκτη ΧΑΑ την περίοδο 1/1/2002 – 31/13/2007. (β) Η αυτοσυσχέτιση της χρονοσειράς στο (α).

Ως πρώτο βήμα της ανάλυσης γίνεται χρήση των πρώτων διαφορών του γενικού δείκτη, ή αντίστοιχα τις αποδόσεις ή τις λογαριθμικές αποδόσεις του γενικού δείκτη. Η χρονοσειρά των αποδόσεων {xt} φαίνεται να είναι στάσιμη(βλέπε παρακάτω σχήμα). Το διάγραμμα της αυτοσυσχέτισης στο Σχήμα β δε δείχνει να υπάρχουν στατιστικά σημαντικές αυτοσυσχετίσεις, παρά μόνο οριακά για υστερήσεις 1, 3 και 14. Το αποτέλεσμα αυτά συνιστούν πολύ ασθενείς συσχετίσεις στα δεδομένα, που μπορεί απλά να οφείλονται σε κάποια μικρή ετεροσκεδαστικότητα ή και ασθενείς τάσεις (π.χ. όπως στην περίοδο του 2003 στο Σχήμα α. Μπορεί όμως και να οφείλονται στη δυναμική του συστήματος του ελληνικού χρηματιστηρίου, που φυσικά χρειάζεται εξήγηση. Στα πλαίσια της γραμμικής ανάλυσης η περιγραφή της δυναμικής γίνεται με μοντέλα τύπου ARMA που περιγράφουν την εξάρτηση της απόδοσης του γενικού δείκτη από τις αποδόσεις σε προηγούμενες μέρες.



Σχήμα (α) Η χρονοσειρά των αποδόσεων του γενικού δείκτη ΧΑΑ την περίοδο 1/1/2002 – 31/12/2007. (β) Η αυτοσυσχέτιση της χρονοσειράς στο (α). Τα όρια της στατιστικής σημαντικότητας δίνονται με τις οριζόντιες γκρίζες γραμμές.

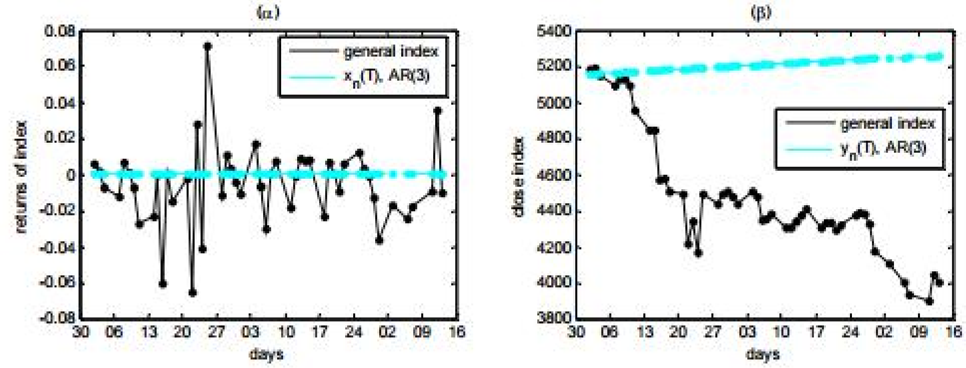
Επειδή πρέπει να γίνουν προβλεψεις προτιμάται η χρήση ενός AR μοντέλου, παρά ARMA, στη χρονοσειρά των αποδόσεων {χ1....χ1495}Για τον προσδιορισμό της τάξης του AR μοντέλου υπολογίζεται η μερική αυτοσυσχέτιση και η συνάρτηση AIC για τάξεις p=1….20(βλέπε παρακάτω σχήμα)



Σχήμα (α) Η μερική αυτοσυσχέτιση για τη χρονοσειρά των αποδόσεων του γενικού δείκτη ΧΑΑ την περίοδο 1/1/2002 – 31/13/2007. Τα όρια της στατιστικής σημαντικότητας δίνονται με τις οριζόντιες γκρίζες γραμμές (β) Το AIC κριτήριο για την ίδια χρονοσειρά και AR μοντέλα τάξης p=1…20

Η μερική αυτοσυσχέτιση έχει την ίδια μορφή με την αυτοσυσχέτιση και γενικά δε παρέχει ασφαλή εκτίμηση της κατάλληλης τάξης. Η συνάρτηση AIC φαίνεται να έχι τοπικά ελάχιστα στο 3 και στο 9 και το σφαιρικό ελάχιστο είναι για p =14 . Αυτή η εκτίμηση της τάξης είναι ύποπτα υψηλή για δεδομένα που οριακά φαίνονται να έχουν συσχετίσεις. Τα κριτήρια της μερικής αυτοσυσχέτισης και AIC δεν φαίνεται λοιπόν να προσδιορίζουν με κάποια ασφάλεια την πιο κατάλληλη τάξη του AR μοντέλου.

Δοκιμάστηκαν όλα τα μοντέλα AR, p =1,…20 για να πραγματοποιηθούν αυθεντικές προβλέψεις σε τιμές αποδόσεων του γενικού δείκτη πέρα από τη χρονική περίοδο που καλύπτει η χρονοσειρά, και συγκεκριμένα για τις επόμενες μέρες και μέχρι τις 13/3/2008. Με αφετηρία την 2/1/2008 προβλέφθηκαν οι τιμές απόδοσης του ημερήσιου γενικού δείκτη για αυξανόμενο ορίζοντα πρόβλεψης T από 2/1/2008 ως 13/3/2008 (δηλαδή τα βήματα πρόβλεψης είναι T=1….49 και οι προβλέψεις είναι y293(1)….y293(16)) Οι προβλέψεις με όλα τα AR μοντέλα συνέκλιναν γρήγορα με το βήμα T στη μέση απόδοση του γενικού δείκτη (περίπου 0). Στο Σχήμα α φαίνεται αυτό για το AR(3). Αντίστοιχα η πρόβλεψη του γενικού δείκτη τείνει προς τη μέση τιμή του δείκτη και δεν καταφέρνει να προσεγγίσει την πορεία τιμών του δείκτη ΧΑΑ(βλέπε παρακάτω σχήμα).

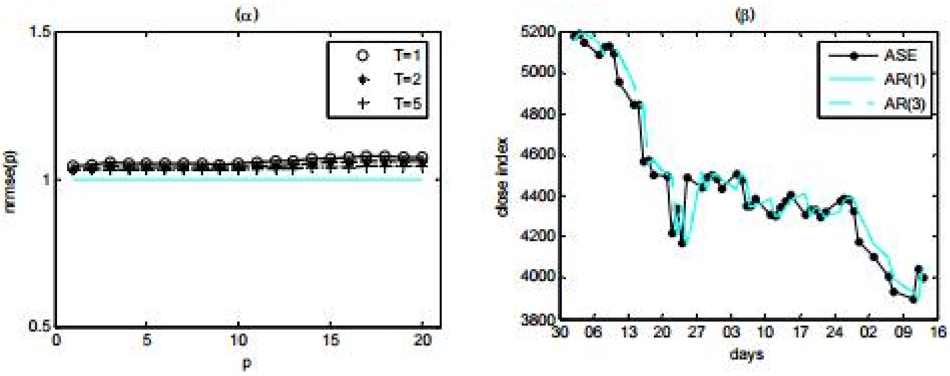


Σχήμα (α) Πρόβλεψη με AR(3) των αποδόσεων του γενικού δείκτη με αφετηρία την 31/12/2007 και για 1 μέχρι 49 μέρες μπροστά. Ο τύπος γραμμών για τις πραγματικές και προβλεπόμενες τιμές δίνονται στο ένθετο. Στον οριζόντιο άξονα δίνονται οι μέρες του μήνα. (β) Το ίδιο όπως στο (α) αλλά για τις τιμές του γενικού δείκτη ΧΑΑ.

Υπολογίσστηκαν σε αυτήν την περίοδο και για μικρά βήματα (T = 1,...5) το nrmse, ως στατιστικό μέτρο για το σφάλμα πρόβλεψης, και βρέθηκε να είναι λίγο μεγαλύτερο του 1 για όλα τα βήματα T, δηλαδή η πρόβλεψη είναι λίγο χειρότερη από την πρόβλεψη με τη μέση απόδοση, όπως φαίνεται στο παρακάτω σχήμα α. Οι προβλέψεις των αποδόσεων με AR μοντέλα είναι κοντά στη μέση απόδοση, που είναι σχεδόν 0. Η πρόβλεψη με τη μέση απόδοση αντιστοιχεί στη διατήρηση της τελευταίας τιμής του γενικού δείκτη για την πρόβλεψη μελλοντικών τιμών του γενικού δείκτη. Αυτό φαίνεται από τη σχέση απόδοσης και τιμής.

$Xt=\frac{yt-yt-1}{yt-1}$ και $yt=y_{t-1}(1+x_t)$

αν θεωρηθεί ότι $x \cong 0$. Πράγματι παρατηρείται (Σχήμα β) ότι οι προβλέψεις του γενικού δείκτη για την επόμενη μέρα (T=1) με μοντέλο των αποδόσεων AR(1) και AR(3) ακολουθούν τις τιμές του γενικού δείκτη με υστέρηση 1, δηλαδή yn(1)=yn.



Σχήμα (α) Σφάλμα nrmse για την πρόβλεψη ενός, δύο και πέντε βημάτων μπροστά του ημερήσιου γενικού δείκτη ΧΑΑ για την περίοδο 2/1/2007 – 13/3/2007 με μοντέλα AR(p) για p = 1,…20 . (β) Πρόβλεψη ενός βήματος για την ίδια περίοδο (στον οριζόντια άξονα δίνονται οι μέρες του μήνα. Οι προβλέψεις έγιναν στις αποδόσεις με AR(1) και AR(3) μοντέλο.

Το συμπέρασμα είναι ότι η πρόβλεψη του ημερήσιου γενικού δείκτη ΧΑΑ ακολουθώντας τη διαδικασία Box – Jenkins αποτυγχάνει. Πράγματι η χρονοσειρά του γενικού δείκτη ΧΑΑ φαίνεται να είναι συνεπής με την υπόθεση του τυχαίου περιπάτου. Τα αποτελέσματα αυτά αφήνουν δύο ανοιχτά ενδεχόμενα: (α) δεν υπάρχει κάποια δυναμική της χρηματαγοράς που μπορεί να εκτιμηθεί από τις τιμές του γενικού δείκτη, (β) υπάρχει αλλά η γραμμική ανάλυση και τα γραμμικά μοντέλα δεν καταφέρνουν να αναδείξουν.

Να σημειωθεί πως πρέπει να υπάρχουν κάποια κριτήρια για το ποιο μοντέλο είναι καλύτερο για πρόβλεψη της χρονοσειράς.

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 παραπάνω δεδομένα, 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, απ' ότι το εκτιμημένο.

Τέλος, θα γίνει αναφορά σε κάποια κριτήρια τα οποία βοηθούν στην επιλογή του κατάλληλου υποδείγματος. Είναι προφανές ότι αν αυξηθεί η 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: $\ln(s^2)+2n/T$

SBC: $\ln(s^2)+n\ln(T)$

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 $σ^2$,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 θα γίνει επιλογή εκείνου 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.

Έχοντας λοιπόν επιλέξει το τελικό μοντέλο πρόβλεψης. Για να αξιολογηθεί η προβλεπτική ικανότητα ενός υποδείγματος θα πρέπει να συγκριθούν οιπροβλέψεις με τα πραγματικά δεδομένα της χρονολογικής σειράς. Κάτι τέτοιο υποδηλώνει πως διαθέτονται παρατηρήσεις ακόμα και για τις περιόδους μέσα στις οποίες πραγματοποιούνται προβλέψεις. Στη συγκεκριμένη περίπτωση γίνεται συζήτηση για εκ των υστέρων προβλέψεις δηλαδή προβλέψεις που γίνονται μέσα στην περίοδο όπου έχουν ήδη πραγματοποιηθεί οι τιμές που χρειάζονται πρόβλεψη. Αντιθέτως, όταν προβλέπονται σήμερα οι μελλοντικές τιμές της μεταβλητής τότε γίνεται αναφορά για εκ των προτέρων προβλέψεις, δηλαδή προβλέψεις τιμών για περιόδους που δεν έχουν ακόμα αληθινές τιμές.

Ουσιαστικά οι προβλέψεις που πραγματοποιούνται με ένα συγκεκριμένο υπόδειγμα μπορούν να αξιολογηθούν μόνο εκ των υστέρων. Για να εξυπηρετηθεί ο σκοπός αυτός, από ένα συνολικό δείγμα παρατηρήσεων που διαθέτεται δεν χρησιμοποιούνται οι τελευταίες παρατηρήσεις οι οποίες δεν συμπεριλαμβάνονται στην εκτίμηση προκειμένου να συγκριθούν με τις προβλέψεις που γίνονται στο διάστημα αυτό. Απαραίτητη προϋπόθεση είναι να υπάρχει μεγάλος αριθμός 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 πραγματοποιείται. Φαίνεται λοιπόν πως:

A =οι πραγματικές τιμές της χρονολογικής σειράς

F_t =οι προβλεφθείσες τιμές

$e_t = F_t - A_t$=το λάθος πρόβλεψης

Στο σημείο θα παρατεθούν κάποιοι τύποι που χρησιμοποιούνται για τη μέτρηση της ακρίβειας των εκ των υστέρων προβλέψεων:

Μέσο σφάλμα τετραγώνου (Mean Squared Error) $MSE=1/N\sum_{t=1}^{n} e^2_t$

Τετραγωνική ρίζα MSE (Root Mean Squared Error) $RMSE=\sqrt{\sum e^2_t}/\sqrt{N}$

Μέσο Απόλυτο Σφάλμα(Mean Absolute error)

$MAE=1/N*\sum^{n}_{t=1}|e_t|$

Μέσο Απόλυτο Ποσοστιαίο Σφάλμα(Mean Absolute Percentage Error) MAPE= $100/N*\sum|F_t-A_t|/|A_t|$

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 εξετσθεί και είναι ανεξάρτητο των μονάδων μέτρησης είναι ο συντελεστής ανισότητας του Theil που συμβολίζεται με το γράμμα U και έχει τον εξής τύπο: $RMSE/(\sqrt{1/N*\sum(A_t)^2})$

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(Theil, H., (1967) Applied Economic Forecasting, Rand McNally).

όπου:

UM = ποσοστό μεροληψίας (bias proportion)

UV=ποσοστό διακύμανσης (variance proportion)

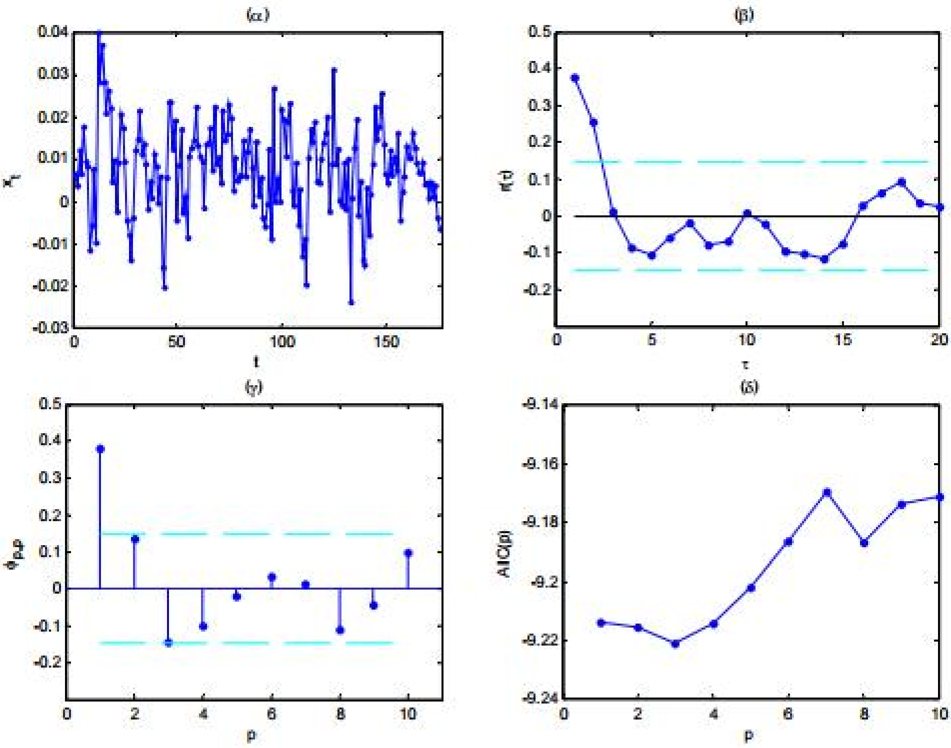
UC = ποσοστό συνδιακύμανσης (covariance proportion)Χαρακτηριστικό είναι πως οι καλές προβλέψεις θα έχουν μικρό ποσοστό μεροληψίας και διακύμανσης και μεγάλο ποσοστό συνδιακύμανσης.

Σε αυτό το σημείο θα παρουσιαστούν κάποια παραδείγματα (εφαρμογές απο αυτά που εμφανίστηκαν παραπάνω.

Παράδειγμα 1:

(Προσαρμογή μοντέλου AR στη χρονοσειρά του ΑΕΠ στις ΗΠΑ.)

Στη συνέχεια θα παρουσιαστεί η εφαρμογή της διαδικασίας προσαρμογής σε ένα πραγματικό παράδειγμα, το ρυθμό μεταβολής του ακαθάριστου εθνικού προϊόντος (ΑΕΠ) των ΗΠΑ που δίνεται σε τετραμηνιαίες τιμές από το 2ο τετράμηνο 1947 ως το 1ο τετράμηνο 1991. Η χρονοσειρά δίνεται στο Σχήμα α, όπου η εποχικότητα έχει διορθωθεί (αφαιρώντας τον εποχικό κύκλο). Η αυτοσυσχέτιση της χρονοσειράς που δίνεται στο Σχήμα β υποδηλώνει σημαντικές αυτοσυσχετίσεις για τις δύο πρώτες υστερήσεις (τα όρια που φαίνονται στο σχήμα υπολογίζονται ως ±2 / n κάτω από την υπόθεση ότι η χρονοσειρά είναι στάσιμη). Εκτιμήθηκαν τα AR μοντέλα για τάξεις p =1, ,10 … και στη συνέχεια υπολογίστηκαν η μερική αυτοσυσχέτιση και το AIC κριτήριο για κάθε τάξη p. Από τη μερική αυτοσυσχέτιση στο Σχήμα γ δεν μπορεί να γίνει εκτίμηση με ασφάλεια την τάξη. Φαίνεται πως μόνο το φ11 είναι στατιστικά μη-μηδενικό, ενώ τα φ22 και φ33 βρίσκονται στα όρια στατιστικής σημαντικότητας. Το κριτήριο AIC στο Σχήμα δ έχει ελάχιστο για p = 3 . Άρα η επιλογή της τάξης 3 υποστηρίζεται από το κριτήριο AIC αλλά όχι από τη μερική αυτοσυσχέτιση. Αυτό φυσικά δεν αποτελεί ένδειξη ότι η μια μέθοδος προσδιορισμού τάξης είναι καλύτερη από την άλλη, αλλά απλά δείχνει ότι διαφορετικές μέθοδοι μπορεί να δώσουν διαφορετικά αποτελέσματα σε πραγματικές εφαρμογές



Σχήμα (α) χρονοσειρά του ρυθμού μεταβολής του ΑΕΠ των ΗΠΑ. (β) Η αυτοσυσχέτιση της χρονοσειράς. (γ) Η μερική αυτοσυσχέτιση της χρονοσειράς. (δ) Το κριτήριο AIC για την ίδια χρονοσειρά. Στο (β) και (γ) οι οριζόντιες διακεκομμένες γραμμές δηλώνουν τα όρια σημαντικότητας σε στάθμη α = 0.05 .

Προσαρμόζεται στη χρονοσειρά μοντέλο AR(3) και το εκτιμημένο μοντέλο είναι $Yt=0,0047+0,35y_{t-1}+0,18y_{t-2}-0,14y_{t-3}+et$ με εκτιμώμενη τυπική απόκλιση του λευκού θορύβου s=0.0098.

Για την εκτίμηση, αφαιρέθηκε πρώτα η μέση τιμή των 176 παρατηρήσεων από κάθε παρατήρηση (βρέθηκε $\hat{\mu} = 0.0077$) και σε αυτήν τη χρονοσειρά εκτιμήθηκαν με τη μέθοδο των ελαχίστων τετραγώνων τους συντελεστές $\varphi1 = 0.35$, $\varphi2 = 0.18$ και $\varphi3 = -0.14$, και ο σταθερός όρος υπολογίστηκε ως $\varphi0 = \mu(1-\varphi1-\varphi2-\varphi3) = 0.0047$. Από το μοντέλο αυτό υπολογίστηκε για κάθε χρονική στιγμή t = 3,...,176 ο εκτιμώμενος ρυθμός μεταβολής

$Yt=0,0047+0,35y_{t-1}+0,18y_{t-2}-0,14y_{t-3}$

και στη συνέχεια τα σφάλματα εκτίμησης et . Η διασπορά αυτών των σφαλμάτων είναι η εκτίμηση της διασποράς του λευκού θορύβου στο μοντέλο AR(3) s=0.0098.

Έστω τώρα ότι έχει εκτιμηθεί ένα μοντέλο AR(p) σε μια στάσιμη χρονοσειρά. Για ένα AR(p) μοντέλο θεωρείται πως η πραγματική τιμή της χρονοσειράς για χρόνο t +1 δίνεται $Y_{t+1}=\varphi0+\varphi1_{yt}+....+\varphi_n\varphi_{n+1-p}+e_{t+1}$, (εδώ σημειώνεται ότι η χρονοσειρά έχει μη-μηδενική μέση τιμή). Η καλύτερη πρόβλεψη του Y_{t+1} γνωρίζοντας τις τιμές yt,yt_{-1} και με βάση το μοντέλο AR(p) είναι σύμφωνα με τον κανόνα της δεσμευμένης μέσης τιμής δηλαδή η πρόβλεψη δίνεται από το γραμμικό συνδυασμό των p τελευταίων (γνωστών) παρατηρήσεων, που είναι το αιτιοκρατικό μέρος του AR(p). Το σφάλμα πρόβλεψης (forecast error) είναι $et=yt-yt_{-1}$ και η διασπορά του σφάλματος πρόβλεψης ενός βήματος είναι $Var(en(1))=\sigma^2$. Αν θεωρηθεί ότι ο λευκός θόρυβος στο μοντέλο ακολουθεί κανονική κατανομή, τότε μπορεί να ορισθεί 95% διάστημα πρόβλεψης για το ένα βήμα μπροστά ως $yt=+-1.96\sigma$. Στα οικονομικά το σφάλμα πρόβλεψης et_{+1} αναφέρεται ως το σοκ (shock) της χρονοσειράς στη χρονική στιγμή n +1. Στην πράξη οι συντελεστές του AR(p) μοντέλου, καθώς και η διασπορά του λευκού θορύβου εκτιμώνται από τη χρονοσειρά και οι εκτιμήσεις τους χρησιμοποιούνται για τη σημειακή πρόβλεψη και το διάστημα πρόβλεψης. Έχοντας υπολογίσει την πρόβλεψη yt(1) για το yt+1 , μπορεί να γίνει αντικατάστασή της στη σχέση που δίνεται στην πρόβλεψη για το yt+2 Y_{t+1}=φ0+φ1$_{yt}$+....+$φ_n φ_{n+2-p}$+e_{t+1} ,Το σφάλμα πρόβλεψης για δύο βήματα μπροστά είναι et=yt$_{-1}$-yt$_{-2}$.και η διασπορά του σφάλματος πρόβλεψης δύο βημάτων είναι Var(en(2))=1+φ1^2)σ^2. Το διάστημα πρόβλεψης υπολογίζεται όπως και για ένα βήμα μπροστά. Αυτή η επαναληπτική διαδικασία συνεχίζεται για να δώσει την πρόβλεψη για 3 βήματα μπροστά και γενικά για ορίζοντα πρόβλεψης T, όπου κάθε φορά αντικαθιστώνται οι παρατηρήσεις σε χρόνους μεγαλύτερους της αφετηρίας n (και μικρότερους φυσικά του n T+) με τις προβλέψεις που ήδη έγιναν σε προηγούμενα βήματα. Μπορεί να δειχθεί ότι για στάσιμο AR(p) μοντέλο καθώς ο ορίζοντας πρόβλεψης αυξάνει (T → ∞), η πρόβλεψη Y(t) τείνει στη μέση τιμή της χρονοσειράς.

Παρακάτω παρουσιάζεται ένα παράδειγμα στο οποίο θα συγκριθούν 3 ειδών υποδείγματα για να βρεθεί ο καλύτερο υπόδειγμα πρόβλεψης.

Έστω ότι παρέχονται οι παρακάτω πληροφορίες:

	AR(1)	ARIMA (1,1,1)	ARIMA (2,1,1)
Root Mean Squared Error	2.151	2.625	2.964
Mean Absolute Error	1.857	2.245	2.210
Theil Inequality	0.195	0.198	0.214

Σύμφωνα με όσα έχουν παρουσιαστεί παραπάνω, έτσι ώστε να βγει ένα συμπέρασμα για το ποιό από τα υποδείγματα είναι το κατάλληλο χρησιμοποιούνται τα εξής: η ρίζα μέσου τετραγωνικού σφάλματος (RMSe) και το μέσο απόλυτο σφάλμα (MAE). Όσο πιο μικρές είναι οι τιμές των παραπάνω μεγεθών τόσο καλύτερο είναι το υπόδειγμα. Όμως επειδή επηρεάζονται από τις μονάδες μέτρησης των μεταβλητών κοιτάμε ένα άλλο μέγεθος το οποίο είναι ανεξάρτητο από τις μονάδες μέτρησης. Αυτό είναι το theil, δηλαδή ο συντελεστής ανισότητας. Αν ο συντελεστής υπερβεί την μονάδα, τότε οι προβλέψεις είναι πολύ κακές. Όσο πιο κοντά είναι στο μηδέν ο συντελεστής του theil, τόσο καλύτερες είναι οι προβλέψεις. Επομένως, το καλύτερο υπόδειγμά είναι το AR(1) με theil 0<0.

ΥΠΟΔΕΙΓΜΑ SARIMA

Τα εποχικά υποδείγματα SARIMA είναι μια άλλη κατηγόρια των υποδειγμάτων ARIMA. Γενικότερα, στοιχεία μικρότερης διάρκειας του έτους όπως μηνιαία, τριμηνιαία και υπόλοιπα στοιχεία 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 χρήση της μεθόδου που έχει μελετηθεί, την γνωστή ως Box-Jenkins. Όποτε και εφαρμόζονται οι τεχνικές της ταυτοποίησης, εκτίμησης και πρόβλεψης. Μια άλλη μέθοδος η οποία είναι και η επικρατέστερη αφορά την ενσωμάτωση του εποχικού προτύπου των στοιχείων στο κανονικό υπόδειγμα ARIMA και την χρήση της μεθοδολογίας Box-Jenkins. Η μέθοδος αυτή καταλήγει στην εκτίμηση υποδειγμάτων με περισσότερες παραμέτρους. (ΑΝΑΛΥΣΗ ΧΡΟΝΟΛΟΓΙΚΩΝ ΣΕΙΡΩΝ(ΚΟΥΝΕΤΑΣ e-class)).

ΠΡΑΚΤΙΚΟ ΜΕΡΟΣ

Σε αυτό το μέρος της εργασίας θα αναλυθεί με τη βοήθεια του PASW-statistics πως από την θεωρία γίνεται η μεταφορά στην πράξη δηλαδή θα αναλυθούν κάποιες χρονολογικές σειρές και ανάλογα με το πρόβλημα που θα παρουσιάζεται θα εφαρμοστούν συγκεκριμένες μέθοδοι για να επιλυθεί. Αρχικά θα αναλυθούν πιο απλές χρονοσειρές και προβλήματα και στη συνέχεια θα αντιμετωπιστούν πιο περίπλοκα ζητήματα.

Ξεκινώντας, έστω ότι υπάρχει ένα ζήτημα πρόβλεψης, πιο συγκεκριμένα στην περίπτωση του παρακάτω παραδείγματος παρουσιάζεται ένας πωλητή ελαστικών ο οποίος έχει καταγράψει τις 3μηνιαίες πωλήσεις ελαστικών για τα τελευταία 4 χρόνια και θέλει να πραγματοποιήσει μελλοντικές προβλέψεις.

Εκτελώντας το SPSS 17 προκύπτει ένα κενό data file, στο οποίο εισάγονται τα δεδομένα της χρονοσειράς. Στη συνέχεια, μπορεί να επιλεχθεί να δημιουργηθεί μια καινούρια μεταβλητή που να περιέχει την ημερομηνία που αντιστοιχεί η κάθε παρατήρηση της χρονοσειράς. Για να πραγματοποιηθεί αυτό επιλέγεται Data/Define Dates και καθορίζεται το χρονικό διάστημα αναφοράς της κάθε παρατήρησης της χρονοσειράς. Στην περίπτωσή της χρονοσειράς που εξετάζεται επιλέγεται το έτος, δηλαδή years, αφού οι παρατηρήσεις της είναι σε ετήσια βάση, στο First case is εισάγεται το έτος που αντιστοιχεί στην πρώτη παρατήρηση και πατιέται OK.

Επιπλέον πατώντας το κουμπί Variable View μπορούν να ονομαστούν οι μεταβλητές και να καθοριστεί ο αριθμός των δεκαδικών ψηφίων που θα περιλαμβάνει κάθε παρατήρηση. Χρησιμοποιείται το όνομα Pwlhseis για τη χρονοσειρά του παραδείγματος. Πατώντας το κουμπί Data View γίνεται επιστροφή στην προηγούμενη οθόνη.

Εδώ αναγράφονται οι πωλήσεις με τη χρήση του PASW statistics:

ΠΩΛΗΣΕΙΣ 1

	PWLHSEIS	YEAR_	QUARTER_	DATE_
1	39	1	1	Q1 1
2	37	1	2	Q2 1
3	62	1	3	Q3 1
4	58	1	4	Q4 1
5	15	2	1	Q1 2
6	54	2	2	Q2 2
7	89	2	3	Q3 2
8	23	2	4	Q4 2
9	43	3	1	Q1 3
10	70	3	2	Q2 3
11	45	3	3	Q3 3
12	68	3	4	Q4 3
13	55	4	1	Q1 4
14	50	4	2	Q2 4
15	67	4	3	Q3 4
16	56	4	4	Q4 4

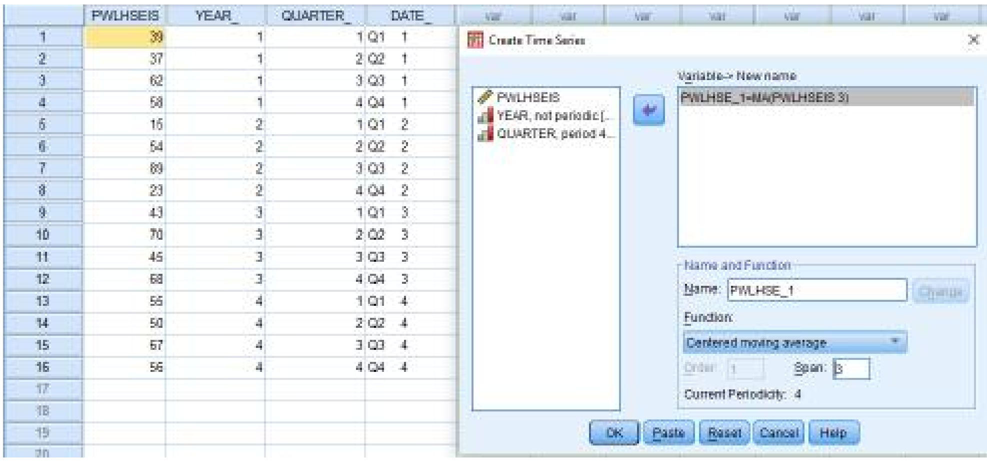
Πολλές φορές χρειάζεται να εκτιμηθεί κάποιο υπόδειγμα με διαφορετικό αριθμό παρατηρήσεων, π.χ. με μια λιγότερη παρατήρηση. Για παράδειγμα έστω ότι επιθυμείται στην ανάλυσή, να μη συμπεριλαμβάνεται το έτος 1, δηλαδή η παρατήρηση της χρονοσειράς που αναφέρεται στο πρώτο έτος. Στην περίπτωση αυτή επιλέγεται το Data/Select Cases. Στο πλαίσιο διαλόγου που ανοίγει επιλέγεται το Select Based on time or case range και η επιλογή Range. Στη συνέχεια, πληκτρολογείται το πρώτο έτος που χρειάζεται να περιλαμβάνει το δείγμα, δηλαδή το 2, και το τελευταίο, δηλαδή το 3 και επιλέγεται το Continue και Ok. Στον Data Editor του SPSS διαγράφεται ο αριθμός του δείγματος που αντιστοιχεί στην πρώτη παρατήρηση και το SPSS στη μετέπειτα ανάλυση θα αγνοεί το πρώτο έτος. Αν χρειάζεται να ξαναχρησιμοποιηθεί όλο το δείγμα επαναλαμβάνεται η προηγούμενη διαδικασία επιλέγοντας All Cases.

Σε αυτό το σημείο αυτό που πρέπει να γίνει είναι η απομονώση της τάσης και η κατάληξη στο αν η χρονοσειρά ακολουθεί ανοδική πορεία ή καθοδική πορεία καθώς και αν υπάρχει ένδειξη εποχικότητας.

Στη συνέχεια θα χρησιμοποιηθεί η μέθοδος κινητών μέσων για να προσδιορισθεί καλύτερα τη τάση της χρονοσειράς.Προχωρώντας επιλέγεται να η καταχώρηση τριών περιόδων στον κυλιόμενο μέσο (span=3) αυτό σημαίνει πως το πρόγραμμα παίρνει ανά 3 τις παρατηρήσεις τις προσθέτει και τις διαιρεί με το 3.

Γενικότερα μία χρήσιμη δυνατότητα του SPSS είναι η κατασκευή μετασχηματισμένων χρονοσειρών από τις αρχικές τιμές μιας χρονοσειράς. Τέτοιοι μετασχηματισμοί μπορεί να είναι η εφαρμογή διαφορών στην αρχική χρονοσειρά, οι χρονικές υστερήσεις, η εφαρμογή κάποιου κινητού μέσου κ.τ.λ. Για παράδειγμα, για να υπολογιστούν οι πρώτες διαφορές της χρονοσειράς SeriesA επιλέγεται Transform/Create Time Series και στο πλαίσιο διαλόγου που προκύπτει επιλέγεται μέσα στο Function η επιλογή Difference και στο Order καταχωρείται ο αριθμός 1. Στη συνέχεια, επιλέγεται η μεταβλητή Pwlhseis και εισάγεται στο πλαίσιο New Variable(s). Πατώντας το κουμπί OK δημιουργείται στον Data Editor η καινούρια μεταβλητή pwlhseis_1 που αποτελείται από τους κυλιόμενους μέσους της σειράς Pwlhseis

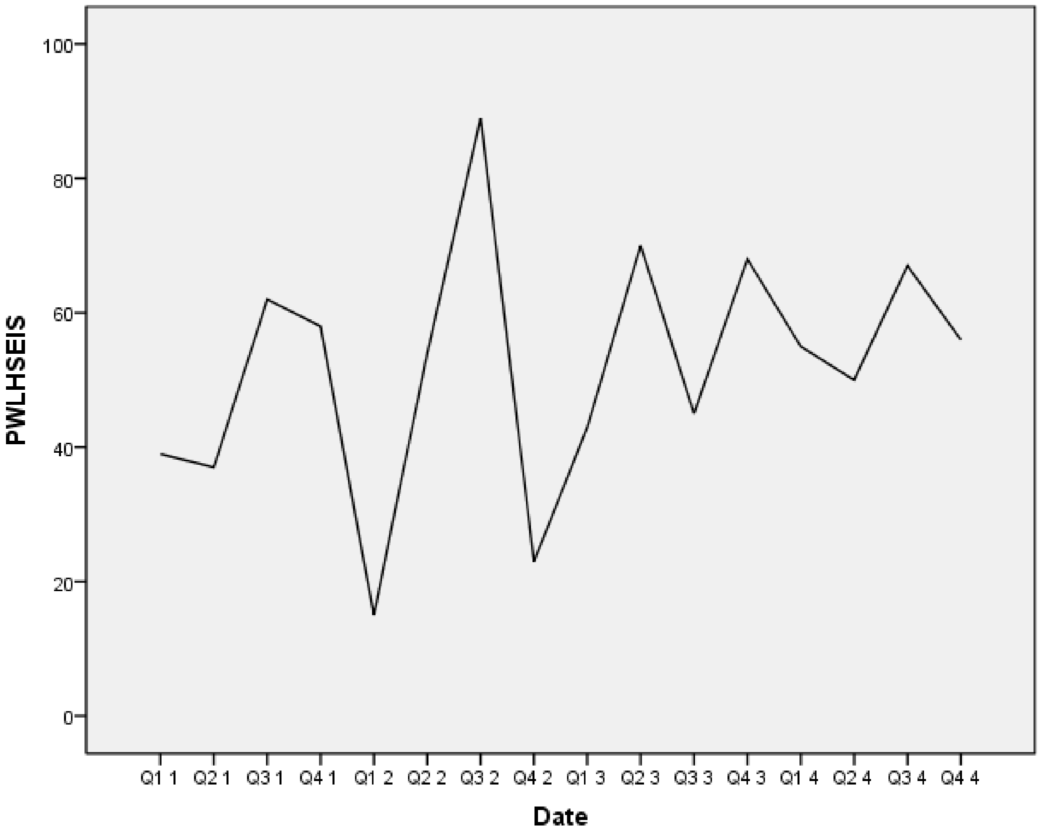
ΚΥΛΙΟΜΕΝΟΙ ΜΕΣΟΙ 1



Τώρα ο κυλιόμενος μέσος 3ων περιόδων έχει δώσει μια νέα μεταβλητή πάνω στην οποία θα εξεταστεί η τάση και η εποχικότητα αλλά πρώτα ας παρατηρηθούν οι γραφικές παραστάσεις που δίνουν οι μεταβλητές χωρίς τη χρήση κυλιόμενου μέσου.κοιτώντας το παρακάτω γράφημα φαίνεται ότι οι πωλήσεις αυξομειώνονται και διακρίνεται μια μικρή ανοδική τάση.Είναι σημαντικό να προσθεθεί πως δεν υπάρχει κάποιο σημάδι εποχικότητας καθώς δεν διακρίνονται χαμηλότερες και υψηλότερες τιμές στα ίδια 3μηνα.

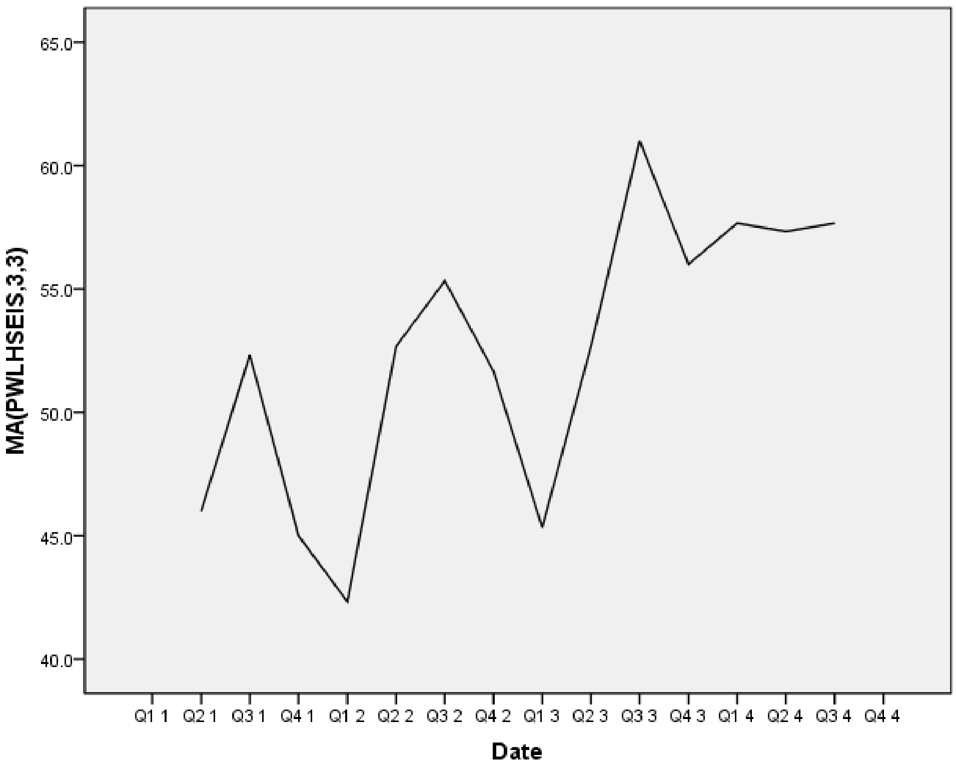
Για να κατασκευασθεί το διάγραμμα μιας χρονοσειράς γίνεται η επιλογή Analyze/Forecasting/Sequence Charts και και στο Variables καταχωρείται η μεταβλητή της οποίας οι παρατηρήσεις χρειάζεται να απεικονιστούν, δηλαδή τη μεταβλητή Pwlhseis, και στο Time Axis Labels τη μεταβλητή YEAR που συμβολίζει τη χρονική περίοδο. Το SPSS επιπλέον παρέχει στον χρήστη τη δυνατότητα να σχεδιάσει τη χρονοσειρά έχοντας εφαρμόσει λογαριθμικό μετασχηματισμό στα δεδομένα ή μετασχηματισμό σε κινητούς μέσους, διαλέγοντας τις αντίστοιχες επιλογές. Επίσης, αν ο χρήστης θελήσει να σχεδιάσει περισσότερες από μια χρονοσειρές μπορεί να επιλέξει αν θα σχεδιαστούν όλες μαζί ή κάθε μία χωριστά. Αν χρειάζεται στο διάγραμμα να περιλαμβάνεται και ο μέσος της χρονοσειράς μπορεί ο χρήστης να πατήσει το κουμπί Format και να διαλέξει Reference line at mean of series και Continue.

ΚΥΛΙΟΜΕΝΟΙ ΜΕΣΟΙ 2



Ας παρατηρηθεί όμως τι γίνεται με την γραφική παράσταση των πωλήσεων έχοντας εφαρμόσει τον κινητό μέσο 3^{ων} περίοδων.

ΚΥΛΙΟΜΕΝΟΙ ΜΕΣΟΙ 3



Όπως γίνεται αντιληπτό το τοπίο έχει πλέον ξεκαθαρίσει και η τάση φαίνεται πεντακάθαρα καθώς είναι ανοδική. Το πιο σημαντικό βέβαια είναι πως τα $1^α$ τρίμηνα του έτους έχουν τις χαμηώτερες τιμές και τα $3^α$ τρίμηνα έχουν τις υψηλότερες και αυτό είναι σημάδι εποχικότητας.Τέλος Το συμπέρασμα είναι πως με την μέθοδο κινητών μέσων γίνονται πιο ορατές η τάση και η εποχικότητα καθώς έχει εξαληφθεί την τυχαία συνιστώσα.Η χρήση των κινητών μέσων είναι από τις πιο διαδεδομένες τεχνικές εξομάλυνσης χρονολογικών σειρών.

Σε δεύτερη φάση θα παρουσιαστεί ένα παράδειγμα εκτίμηση τάσης με την παραπάνω μέθοδο των κυλιόμενων μέσων και θα χρησιμοποιηθεί το υπόδειγμα που περιγράφει την μεταβολή της τάσης.

Τα υποδείγματα μεταβολής της τάσης είναι τρία.Το γραμμικό, το εκθετικό και το δευτέρου βαθμού με τους εξής τύπους:

$Y=b0+b1*t$

$Y=b0*b1^t$

$Y=b0+b1*t*b2^2*t^2$

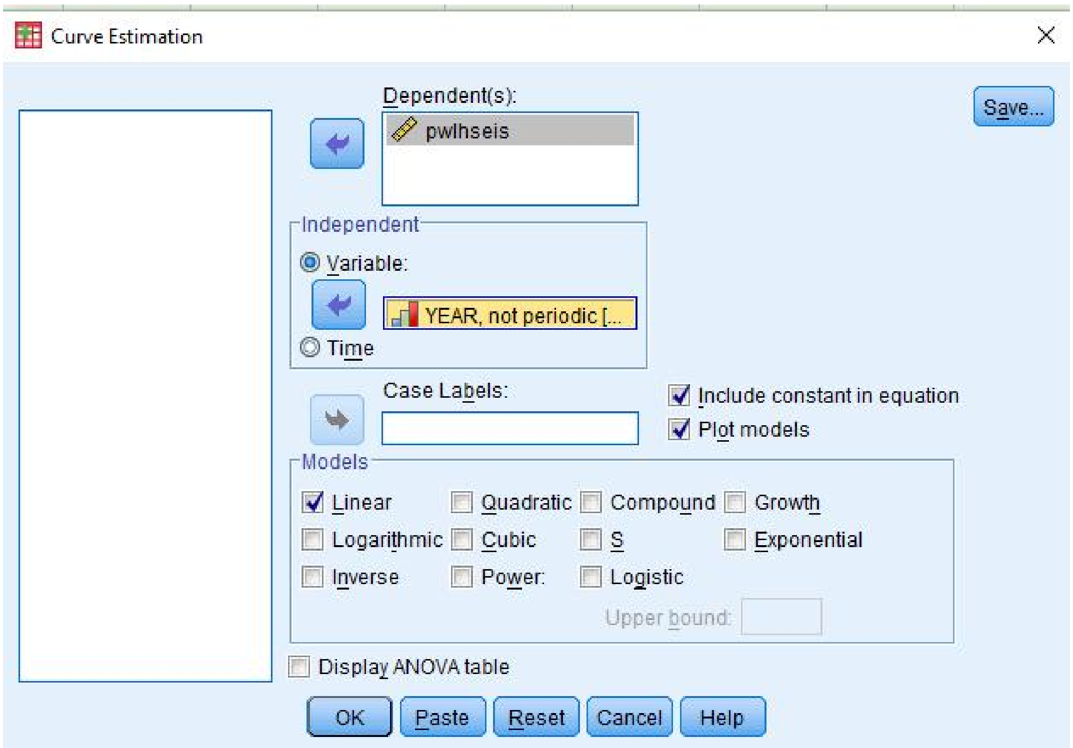
Έστω λοιπόν ότι υπάρχει ένα μαγαζί που πουλάει μπλουζάκια και πρέπει να υπολογιστεί η τάση.Οι ποσότητες που πουλήθηκαν κάθε μήνα φαίνονται παρακάτω:

ΤΑΣΗ 1

	pwlhseis	months
1	87	1
2	88	2
3	89	3
4	92	4
5	94	5
6	96	6
7	99	7
8	100	8
9	102	9
10	106	10
11	110	11
12	115	12
13	118	13
14	119	14

Στη συνέχεια με την βοήθεια της παλινδρόμησης στην επιλογή regression επιλέγεται ποιο υπόδειγμα χρειάζεται για να υπολογιστεί η τάση.Στη συγκεκριμένη περίπτωση θα πραγματοποιηθεί το γραμμικό γιατί έχει τον πιο απλό τύπο όσον αφορά τις πράξεις.Επίσης πρέπει να σημειωθεί πως με την βοήθεια της παλινδρόμησης ο χρήστης είναι σε θέση να εκτελέσει προβλέψεις για τις μελλοντικές πωλήσεις.

ΤΑΣΗ 2



Για το επόμενο βήμα ο χρήστης προχωράει στο output,συγκεκριμένα στο πινακάκι με τα coefficients και κοιτάει τις τιμές στο constant και στην πάνω μεταβλητή. Οι τιμές αυτές είναι τα νούμερα τα οποία θα αντικαταστήσουν τα b0 και b1 στην στο γραμμικό υπόδειγμα.

ΤΑΣΗ 3

	Unstandardized Coefficients		Standardized Coefficients	t	Sig.
	B	Std. Error	Beta		
YEAR, not periodic	2.604	.124	.987	21.041	.000
(Constant)	81.538	1.054		77.368	.000

Οπότε παίρνει την y=b0+b1*t και βάζει όπου B0 το 81,538 και όπου το B1 την τιμή 2,604 και στην θέση του t την χρονική περίοδο στην οποία θέλει να υπολογίσει την τιμή της Y.

Είναι σημαντικό να σημειωθεί πως στο παραπάνω πινακάκι υπάρχει ένα Sig. Το Sig(significance) είναι η σημαντικότητα των δεδομένων.Τώρα οσον αφορά το sig πριν γίνει η συνέχεια στην πρόβλεψη των τιμών για τους επόμενους μήνες πρέπει να να εξακριβωθεί πως έχει τιμή μικρότερη του 0.05 όπου δείχνει ότι η μεταβλητή είναι στατιστικά σημαντική.

Στο output έχει εμφανιστεί και ένα άλλο πινακάκι το οποίο φαίνεται παρακάτω:

ΤΑΣΗ 4

Model Summary

R	R Square	Adjusted R Square	Std. Error of the Estimate
.987	.974	.971	1.867

The independent variable is YEAR, not periodic.

Στο πινακάκι αυτό καλό είναι να να παρατηρηθούν οι τιμές του R και του R Square. Αυτές οι τιμές παρέχουν στον χρήστη σημαντικές πληροφορίες για την χρονοσειρά.Το R πληροφορεί τον χρήστη για το πόσο ισχυρή είναι η συχέτιση του υποδείγματος και το R^2 μετράει το κατά πόσο η ανεξάρτητη μεταβλητή επηρεάζει την εξαρτημένη, δηλαδή κατά πόσο ο χρόνος επηρεάζει τις πωλήσεις.Στη συγκεκριμένη περίπτωση ο χρόνος επηρεάζει κατά 97.4% την αύξηση των πωλήσεων.Σε όλα τα υποδείγματα:γραμμικό, εκθετικό και δευτέρου βαθμού υπάρχει το πινακάκι με το R και το R^2 και πάντα θα επιλέγεται το υπόδειγμα με το μεγαλύτερο R^2.

Αφού λοιπόν έχει υπολογίστει η τάση παραπάνω ας γίνει μια προσπάθεια πρόβλεψης για τους 2 επόμενους μήνες.Αυτό γίνεται με τον εξής τρόπο: επιλέγεται το γραμμικό μοντέλο με τις τιμές που εμφανίστηκαν πριν και μπάινει όπου t το 15 και το 16. Y=81,538+2,604*15=120 και Y=81,538+2,604*16=123.Με αυτόν τον τρόπο δημιουργήθηκε μια αρχική πρόβλεψη για τους επόμενους 2 μήνες.

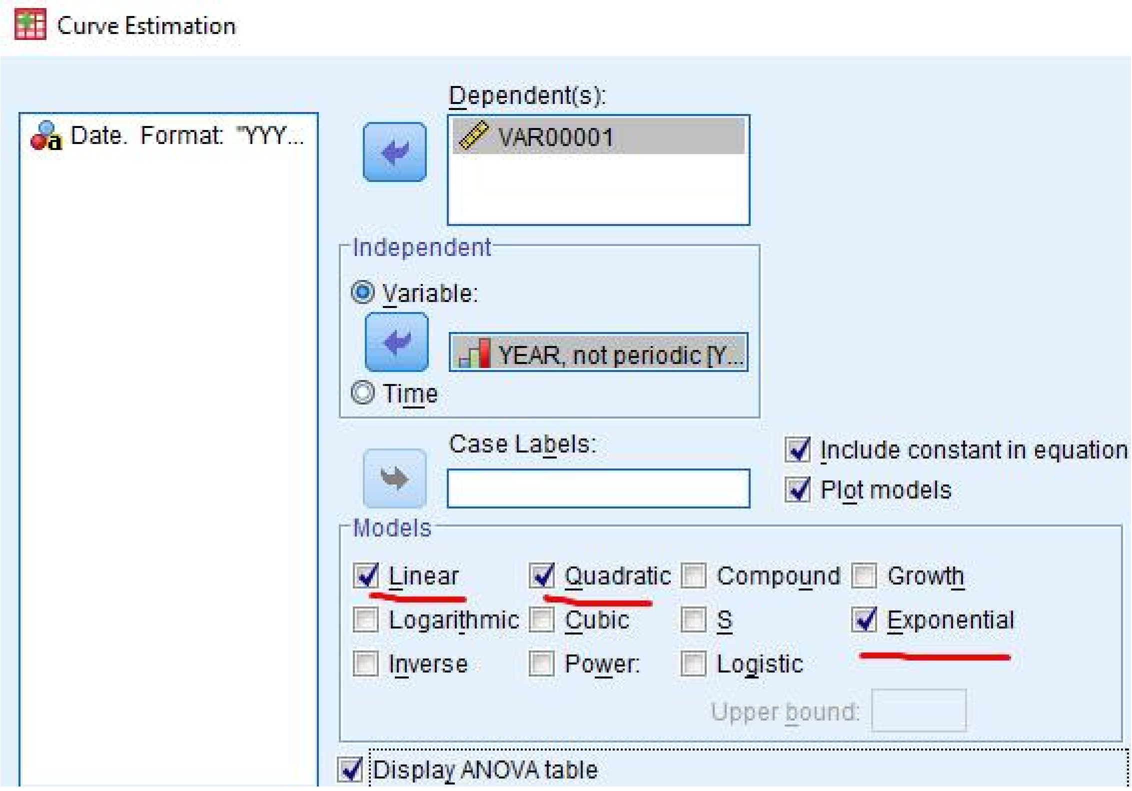
Ας παρατηρηθεί λίγο μια χρονοσειρά στην οποία θα δοκιμάστηκαν και τα 3 υποδείγματα και θα γίνει επιλογή του καλύτερου. Παρακάτω φαίνονται οι τιμές της χρονολογικής σειράς

ΠΑΛΙΝΔΡΟΜΗΣΗ 1

	VAR00001	YEAR_	DATE_
1	.50	1	1
2	.60	2	2
3	1.30	3	3
4	2.70	4	4
5	4.10	5	5
6	6.90	6	6
7	10.80	7	7
8	19.20	8	8

Μετά επιλέγεται η μέθοδος παλινδρόμησης και με τα 3 υποδείγματα

ΠΑΛΙΝΔΡΟΜΗΣΗ 2



και Στη συνέχει φαίνεται το output. Όπως αναγράφεται στα coefficients του γραμμικου το sig της σταθεράς constant είναι >0.05 άρα δεν είναι στατιστικά σημαντικός και δεν μπορεί να χρησιμοποιηθεί, επίσης οι συσχετίσεις δεν είναι αρκετά ισχυρές.

ΠΑΛΙΝΔΡΟΜΗΣΗ 3

Linear

Model Summary

R	R Square	Adjusted R Square	Std. Error of the Estimate
.901	.812	.780	3.036

The independent variable is YEAR, not periodic.

ANOVA

	Sum of Squares	df	Mean Square	F	Sig.
Regression	238.333	1	238.333	25.856	.002
Residual	55.305	6	9.218		
Total	293.639	7			

The independent variable is YEAR, not periodic.

Coefficients

	Unstandardized Coefficients		Standardized Coefficients	t	Sig.
	B	Std. Error	Beta		
YEAR, not periodic	2.382	.468	.901	5.085	.002
(Constant)	-4.957	2.366		-2.095	.081

Στη συνέχεια βλέποντας και τα πινακάκια στο υπόδειγμα 2^{ου} βαθμού (quadratic) πάλι τα sig τους δεν είναι στατιστικά σημαντικά ενώ οι συσχετίσεις τους είναι δυνατές, παρ όλα αυτά όσο οι δείκτες β0 ή β1 δεν είναι στατιστικά σημαντικοί κι επομένως δεν μπορεί να χρησιμοποιηθεί το υπόδειγμα.

ΠΑΛΙΝΔΡΟΜΗΣΗ 4

Quadratic

Model Summary

R	R Square	Adjusted R Square	Std. Error of the Estimate
.989	.978	.970	1.126

The independent variable is YEAR, not periodic.

ANOVA

	Sum of Squares	df	Mean Square	F	Sig.
Regression	287.301	2	143.650	113.322	.000
Residual	6.338	5	1.268		
Total	293.639	7			

The independent variable is YEAR, not periodic.

Coefficients

	Unstandardized Coefficients		Standardized Coefficients	t	Sig.
	B	Std. Error	Beta		
YEAR, not periodic	-2.477	.801	-.937	-3.093	.027
YEAR, not periodic ** 2	.540	.087	1.882	6.215	.002
(Constant)	3.141	1.571		2.000	.102

Τέλος υπάρχει και το εκθετικό υπόδειγμα (exponential) στο οποίο υπάρχουν πάρα πολύ ισχυρές συσχετίσεις και τα sig τους είναι στατιστικά σημαντικά οπότε χρησιμοποιείται αυτό για μελλοντικές προβλέψεις με τον τύπο που δώθηκε παραπάνω.

ΠΑΛΙΝΔΡΟΜΗΣΗ 5

Exponential

Model Summary

R	R Square	Adjusted R Square	Std. Error of the Estimate
.995	.991	.989	.137

The independent variable is YEAR, not periodic.

ANOVA

	Sum of Squares	df	Mean Square	F	Sig.
Regression	12.276	1	12.276	654.670	.000
Residual	.113	6	.019		
Total	12.389	7			

The independent variable is YEAR, not periodic.

Coefficients

	Unstandardized Coefficients		Standardized Coefficients	t	Sig.
	B	Std. Error	Beta		
YEAR, not periodic	.541	.021	.995	25.587	.000
(Constant)	.261	.028		9.372	.000

Τελειώνοντας με το συγκεκριμένο παράδειγμα ένας ακόμα τρόπος σύγκρισης αυτών των υποδειγμάτων είναι με τους τύπους της μέσης απόλυτης απόκλισης MAD και του αθροίσματος τετραγώνων σφαλμάτων πρόβλεψης καθώς το υπόδειγμα με τις τιμές που είναι κοντά στο 0 είναι το καλύτερο.

Στη συνέχεια ακολουθεί ένα παράδειγμα εποχικότητας καθώς η εποχικότητα είναι ένα φαινόμενο το οποίο απασχολεί αρκετά επιχειρήσεις που πουλάνε προϊόντα τα οποία σημειώνουν υψηλές πωλήσεις κατά την διάρκεια κάποιων μηνών μέσα στο έτος ενώ πολύ χαμηλές σε άλλες περιόδους.

Μία από τις πολλές επιχειρήσεις που συναντούν εποχικότητα στα έσοδα τους είναι ένα ξενοδοχείο στην Ελλάδα όπου ο τουρισμός αυξάνεται κατά τη διάρκεια του καλοκαιριού. Πιο κάτω πρατηρούνται οι 3μηνιαίες αφίξεις τουριστών σε χιλλιάδες στη χώρα από το 2013 μέχρι το 2016. Σκοπός είναι η εξάλειψη της εποχικότητας και να παρατηρηθεί αν αυξάνονται ή μειώνονται οι αφίξεις κατά την πάροδο του χρόνου.

ΕΠΟΧΙΚΟΤΗΤΑ 1

	afikseis	YEAR_	QUARTER_	DATE_
1	1029	2013	1	Q1 2013
2	4255	2013	2	Q2 2013
3	6821	2013	3	Q3 2013
4	1864	2013	4	Q4 2013
5	1145	2014	1	Q1 2014
6	4173	2014	2	Q2 2014
7	6324	2014	3	Q3 2014
8	1671	2014	4	Q4 2014
9	1384	2015	1	Q1 2015
10	4377	2015	2	Q2 2015
11	7104	2015	3	Q3 2015
12	1900	2015	4	Q4 2015
13	1448	2016	1	Q1 2016
14	4823	2016	2	Q2 2016
15	7660	2016	3	Q3 2016
16	2108	2016	4	Q4 2016

Αφού λοιπόν προσθέθηκαν οι αφίξεις στη συνέχει υπάρχει η εντολή seasonal decomposition του Pasw statistics, επιλέγεται το προσθετικό μοντέλο Y=T+C+S+I για αρχή καθώς θα χρησιμοποιηθεί και το πολλαπλασιαστικό.

ΕΠΟΧΙΚΟΤΗΤΑ 2

Seasonal Decomposition

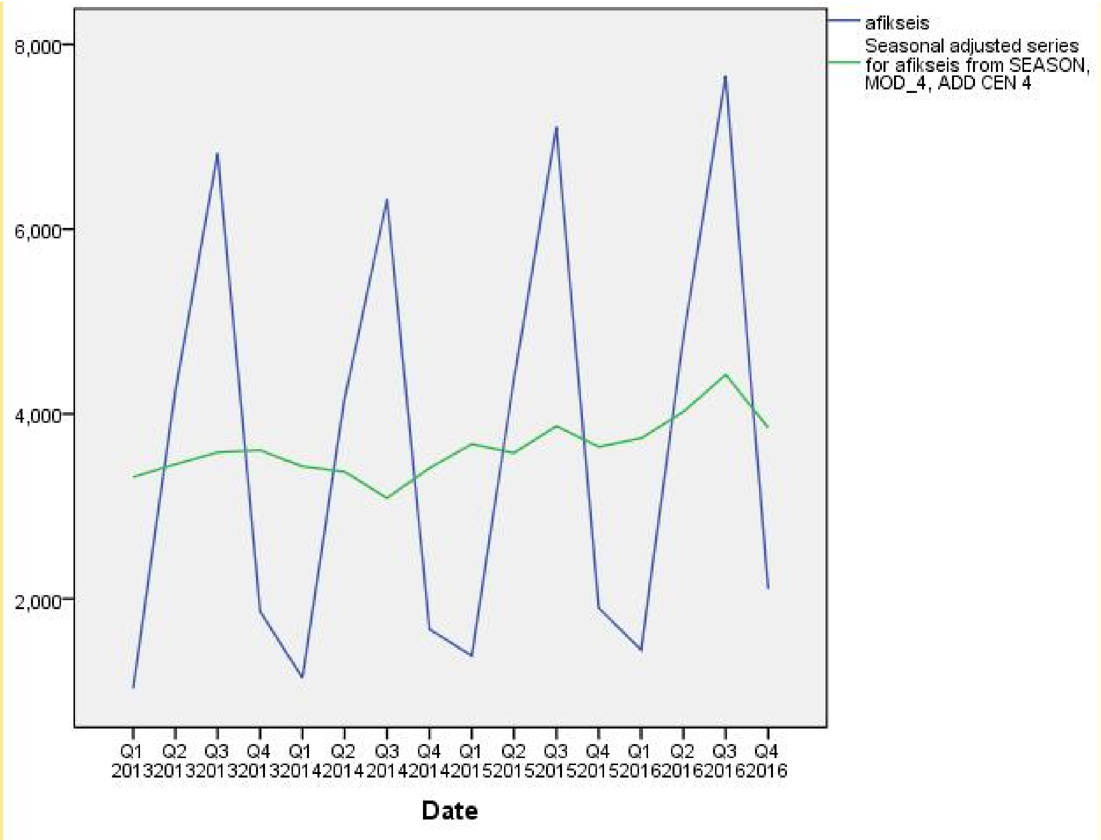
Series Name:afikseis

DATE_	Original Series	Moving Average Series	Difference of Original Series from Moving Average Series	Seasonal Factor	Seasonally Adjusted Series	Smoothed Trend-Cycle Series	Irregular (Error) Component
Q1 2013	1029.000	.	.	-2288.813	3317.813	3422.625	-104.813
Q2 2013	4255.000	.	.	798.063	3456.938	3453.604	3.333
Q3 2013	6821.000	3506.75	3314.250	3234.938	3586.063	3515.563	70.500
Q4 2013	1864.000	3511.00	-1647.000	-1744.188	3608.188	3521.799	86.389
Q1 2014	1145.000	3438.63	-2293.625	-2288.813	3433.813	3438.090	-4.278
Q2 2014	4173.000	3352.38	820.625	798.063	3374.938	3354.882	20.056
Q3 2014	6324.000	3358.13	2965.875	3234.938	3089.063	3328.229	-239.167
Q4 2014	1671.000	3413.50	-1742.500	-1744.188	3415.188	3413.688	1.500
Q1 2015	1384.000	3536.50	-2152.500	-2288.813	3672.813	3551.646	121.167
Q2 2015	4377.000	3662.63	714.375	798.063	3578.938	3653.326	-74.389
Q3 2015	7104.000	3699.25	3404.750	3234.938	3869.063	3718.118	150.944
Q4 2015	1900.000	3763.00	-1863.000	-1744.188	3644.188	3749.799	-105.611
Q1 2016	1448.000	3888.25	-2440.250	-2288.813	3736.813	3871.424	-134.611
Q2 2016	4823.000	3983.75	839.250	798.063	4024.938	3988.326	36.611
Q3 2016	7660.000	.	.	3234.938	4425.063	4100.729	324.333
Q4 2016	2108.000	.	.	-1744.188	3852.188	4156.931	-304.743

Παραπάνω αναγράφονται κάποια πινακάκια τα οποία εμφανίστηκαν μετά την εντολή. Εδώ φαίνονται οι αρχικές τιμές τις χρονοσειράς, οι κυλιόμενοι μέσοι (οι οποίοι προσαρμόστηκαν μόνοι τους), η διαφορά της αρχικής σειράς με τους κυλιόμενους μέσους, τους εποχικούς δείκτες ,την εποχιακά προσαρμοσμένη χρονολογική σειρά, την σειρά εξομάλυνσης της τάσης και τέλος τα σφάλματα.

Όπως φαίνεται στην στήλη του seasonal factor παρατηρείται πως οι τιμές σε κάθε Q1 Q2 Q3 Q4 είναι ίδιες. Αυτό σημαίνει ότι κάθε τρίμηνο έχει τον δικό του δείκτη. Οι δείκτες της χρονολογικής σειράς παίζουν πολύ σημαντικό ρόλο καθώς με την βοήθειά τους κατασκευάζεται η εποχικά προσαρμοσμένη χρονολογική σειρά (Seasonal Adjusted Series). Αυτό γίνεται με τον εξής τρόπο: επιλέγεται κάθε εποχικός δείκτη για παράδειγμα ο πρώτος ο οποίος είναι ίσος με -2288.813 και θεωρείται πως σε κάθε πρώτο τρίμηνο του έτους οι αφίξεις είναι κατά 2288 λιγότερες από το μέσο τριμηνιαίο επίπεδο του έτους, και ακολουθεί η πράξη (original series-seasonal factor) οπότε 1029-(-2288)=3317 (seasonal adjusted series). Έχοντας λοιπόν αφαιρέσει την εποχικότητα φαίνεται πως οι αφίξεις στο πρώτο τρίμηνο του 2013 είναι κατά 116 λιγότερες από το πρώτο τρίμηνο του 2014 όπου αυτό είναι σημάδι σταθερής ανόδου.

ΕΠΟΧΙΚΟΤΗΤΑ 3



Παραπάνω στο διάγραμμα φαίνεται πεντακάθαρα ότι υπάρχει μια μικρή αλλά σταθερή άνοδος μετά το 2014 καθώς η πράσσινη γραμμή η οποία είναι η εποχιακά προσαρμοσμένη σειρά δείχνει ανοδική πορεία.

Ας πραγματοποιηθεί όμως και η εξάλειψη εποχικότητας και με το πολλαπλασιαστικό υπόδειγμα.Παρακάτω μπορεί να παρατηρηθεί πως έχουν εμφανιστεί πάλι οι ίδιες στείλες με τη διαφορά πως οι εποχικοί δείκτες είναι πλέον ποσοστιαίοι.

Seasonal Decomposition

Series Name:afikseis

DATE_	Original Series	Moving Average Series	Ratio of Original Series to Moving Average Series (%)	Seasonal Factor (%)	Seasonally Adjusted Series	Smoothed Trend-Cycle Series	Irregular (Error) Component
Q1 2013	1029.000	.	.	37.2	2768.915	3204.499	.864
Q2 2013	4255.000	.	.	120.8	3521.949	3283.396	1.073
Q3 2013	6821.000	3506.75	194.5	191.6	3559.323	3441.190	1.034
Q4 2013	1864.000	3511.00	53.1	50.4	3699.433	3483.898	1.062
Q1 2014	1145.000	3438.63	33.3	37.2	3081.057	3378.833	.912
Q2 2014	4173.000	3352.38	124.5	120.8	3454.076	3348.903	1.031
Q3 2014	6324.000	3358.13	188.3	191.6	3299.980	3360.679	.982
Q4 2014	1671.000	3413.50	49.0	50.4	3316.391	3452.722	.961
Q1 2015	1384.000	3536.50	39.1	37.2	3724.178	3562.017	1.046
Q2 2015	4377.000	3662.63	119.5	120.8	3622.931	3646.491	.994
Q3 2015	7104.000	3699.25	192.0	191.6	3706.998	3725.466	.995
Q4 2015	1900.000	3763.00	50.5	50.4	3770.882	3792.717	.994
Q1 2016	1448.000	3888.25	37.2	37.2	3896.394	3879.918	1.004
Q2 2016	4823.000	3983.75	121.1	120.8	3992.094	3968.656	1.006
Q3 2016	7660.000	.	.	191.6	3997.129	4057.639	.985
Q4 2016	2108.000	.	.	50.4	4183.694	4102.130	1.020

Επειδή όμως οι δείκτες είναι ποσοστιαίοι η εποχιακά προσαρμοσμένη χρονολογική σειρά υπολογίζεται με διαφορετικό τρόπο. Έχονταις για παράδειγμα τον εποχικό δείκτη του πρώτου τριμήνου. Πολλαπλασιάζεται λοιπόν το (37.2-1) με το 100 όπου ισούται με 3620 και αφαιρείται η τιμή της αρχικής χρονοσειράς το πρώτο τρίμηνο δηλαδή το 1029, οπότε ισχύει 3720-1029=2591 όπου είναι η τελική τιμή. Επίσης πρέπει να προσθεθεί πως με το πολλαπλασιαστικό υπόδειγμα ο εποχικό δείκτης εκφράζεται αλλίως, δηλαδή για το 37,2 σημαινει πως οι αφίξεις είναι κατά 37,2 λιγότερες από το μέσο τριμηνιαίο επίπεδο.

Στο σημείο που βρίσκεται το παράδειγμα θα συνεχιστούν η στασιμότητα και οι αυτοσυσχετίσεις. Σε πρακτικό επίπεδο ο έλεγχος της στασιμότητας γίνεται με την χρήση γραφικών παραστάσεων. Την πρώτη ιδέα για την στασιμότητα την παίρνει ο χρήστης από το sequence chart, επίσης η ύπαρξη τάσης κι εποχικότητας είναι σημάδι μη στασιμότητας. Τώρα ο βασικότερος έλεγχος όπως έχει προαναφερθεί, γίνεται στο διάγραμμα αυτοσυσχετίσεων. Καλό είναι όμως όλα αυτά να παρουσιαστούν σε ένα παράδειγμα παρακάτω.

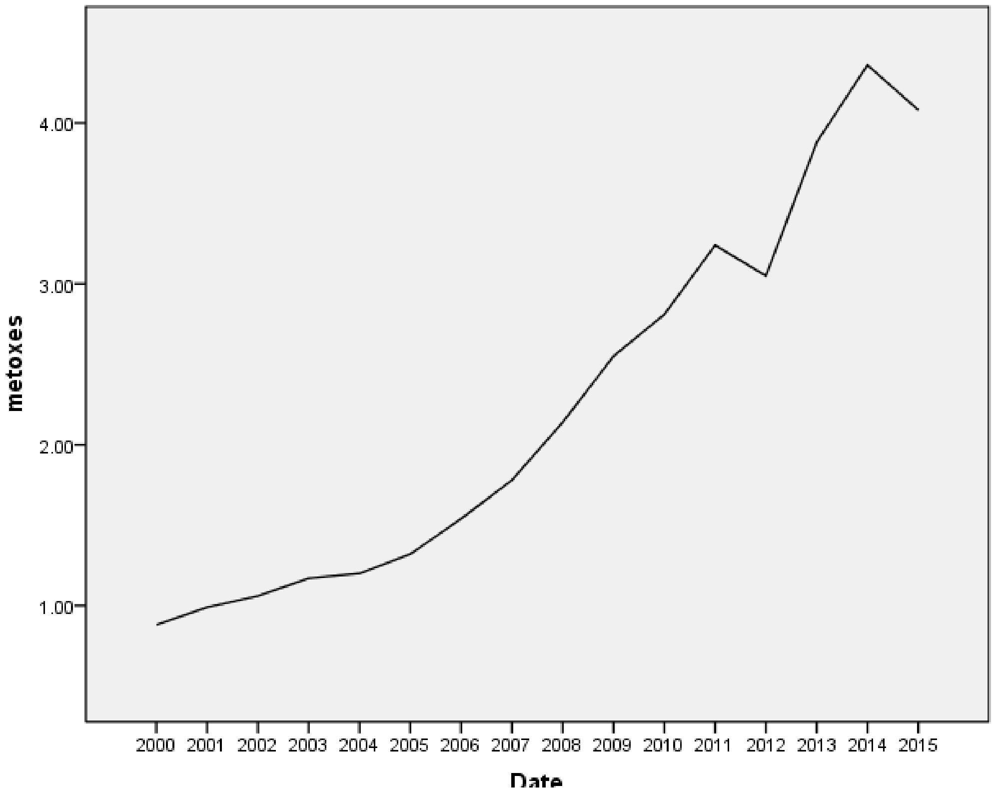
Έστω λοιπόν ότι υπάρχουν οι μετοχές μιας επιχείρησης από το 2000 και πρέπει να πραγματοποιηθεί έλεγχος στασιμότητας πάνω στην χρονοσειρά.

ΣΤΑΣΙΜΟΤΗΤΑ 3

metoxes	YEAR_	DATE_
.88	2000	2000
.99	2001	2001
1.06	2002	2002
1.17	2003	2003
1.20	2004	2004
1.32	2005	2005
1.54	2006	2006
1.78	2007	2007
2.14	2008	2008
2.55	2009	2009
2.81	2010	2010
3.24	2011	2011
3.05	2012	2012
3.88	2013	2013
4.36	2014	2014
4.08	2015	2015

Όπως θα εμφανιστεί στο γράφημα η χρονοσειρά έχει έντονη ανοδική τάση και δεν διατηρεί σταθερό μέσο, αυτά τα δύο αρκούνε για να εξακριβωθεί πως η χρονοσειρά δεν είναι σε καμία περίπτωση στάσιμη.

ΣΤΑΣΙΜΟΤΗΤΑ 4



Επιπλέον τα sig της χρονοσειράς στο διάγραμμα αυτοσυσχετίσεων είναι μικρότερα του 0.05 (να σημειωθελί ότι αυτά τα sig δεν έχουν να κάνουν με τα προηγούμενα που έχουν εμφανιστεί καθώς δεν πραγματοποιείται έλεγχος σημαντικότητας αλλά στασιμότητας).

Για να υπολογιστούν τις αυτοσυσχετίσεις και μερικές αυτοσυσχετίσεις της χρονοσειράς γίνεται η επιλογή Analyze/Forecasting/Autocorrelations και μετακινείται στο Variables τη μεταβλητή metoxes που περιέχει τις τιμές της χρονοσειράς. Στο Display αφήνονται ενεργοποιημένα τα Autocorrelations και Partial autocorrelations. Αν χρειαστεί γίνεται να υπολογιστούν οι αυτοσυσχετίσεις και μερικές αυτοσυσχετίσεις της χρονοσειράς έχοντας εφαρμόσει λογαριθμικό μετασχηματισμό στις παρατηρήσεις ή έχοντας εφαρμόσει διαφορές. Πατώντας το κουμπί Options μπορεί να καταχωρηθεί ο αριθμός των αυτοσυσχετίσεων και μερικών αυτοσυσχετίσεων που χρειάζεται να υπολογιστούν. Γίνεται επιστροφή στην προηγουμένη οθόνη πατώντας Continue

ΣΤΑΣΙΜΟΤΗΤΑ 5

Autocorrelations

Series:metoxes

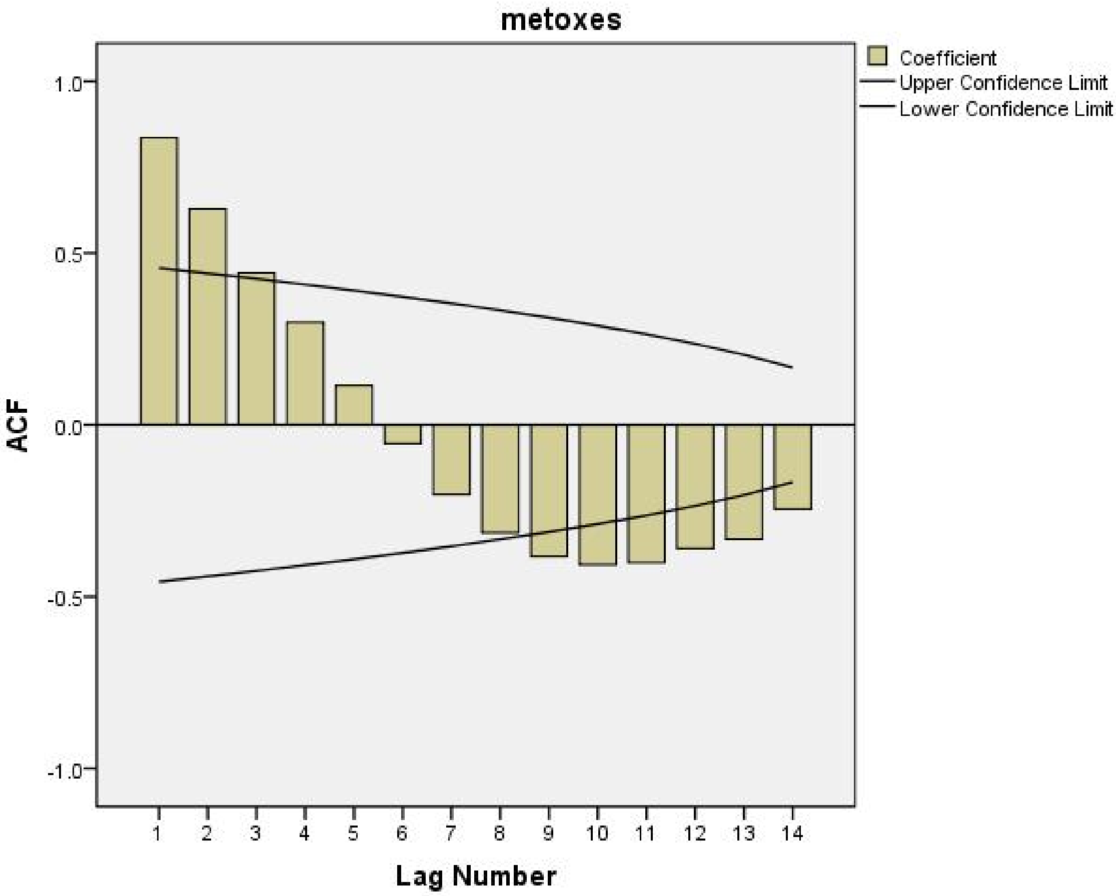
Lag	Autocorrelation	Std. Error[a]	Box-Ljung Statistic		
			Value	df	Sig.[b]
1	.836	.228	13.424	1	.000
2	.628	.220	21.540	2	.000
3	.443	.212	25.880	3	.000
4	.298	.204	28.016	4	.000
5	.114	.195	28.355	5	.000
6	-.054	.186	28.440	6	.000
7	-.203	.177	29.759	7	.000
8	-.314	.167	33.319	8	.000
9	-.383	.156	39.356	9	.000
10	-.407	.144	47.318	10	.000
11	-.401	.132	56.568	11	.000
12	-.360	.118	65.894	12	.000
13	-.333	.102	76.552	13	.000
14	-.245	.083	85.191	14	.000

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

b. Based on the asymptotic chi-square approximation.

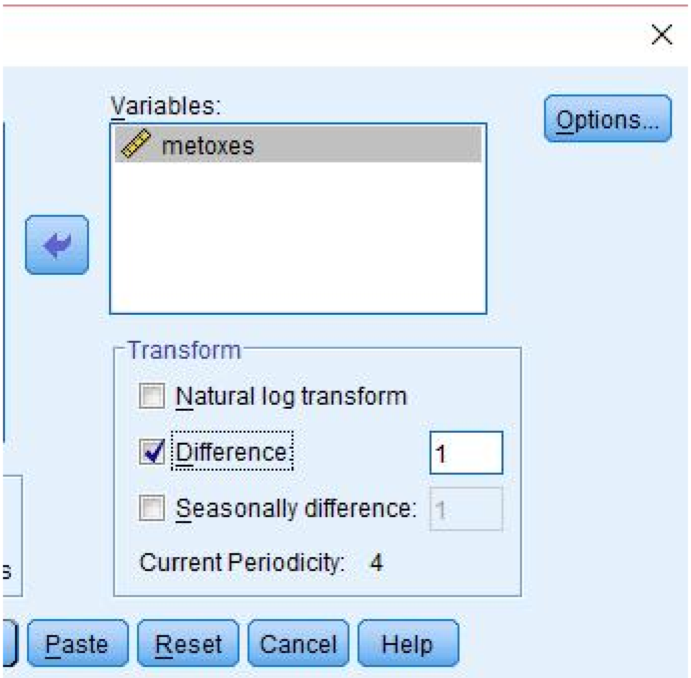
Τέλος για το παράδειγμα υπάρχει και το διάγραμμα των αυτοσυσχετίσεων ACF

ΣΤΑΣΙΜΟΤΗΤΑ 6



Επιλέγοντας ΟΚ στο SPSS Statistics Viewer προκύπτουν τα διάγραμματα αυτοσυσχετίσεων (ACF) και μερικών αυτοσυσχετίσεων (PACF) της χρονοσειράς στα οποία εκτός από τους συντελεστές αυτοσυσχέτισης απεικονίζεται και ένα 95% διάστημα εμπιστοσύνης που βοηθάει στην επιλογή των σημαντικών αυτοσυσχετίσεων. Όσες αυτοσυσχετίσεις βρίσκονται εκτός των ορίων αυτού του διαστήματος θεωρούνται σημαντικές. Επιπρόσθετα, εμφανίζονται οι τιμές των δειγματικών αυτοσυσχετίσεων και μερικών αυτοσυσχετίσεων της χρονοσειράς. Στον πίνακα των αυτοσυσχετίσεων το SPSS δίνει τις τιμές της στατιστικής Q των Box και Ljung με τα αντίστοιχα P-Values για τον έλεγχο της μηδενικής υπόθεση ότι η χρονοσειρά είναι λευκός θόρυβος. Παρατηρείται ότι οι αυτοσυσχετίσεις της χρονοσειράς φθίνουν με αργό ρυθμό και δεν συγκλίνουν στο μηδέν οπότε μπορεί να θεωρηθεί μη στάσιμη. Από τις τιμές της στατιστικής Q συμπεραίνεται ότι οι τιμές της χρονοσειράς δεν αποτελούν λευκό θόρυβο, δηλαδή συσχετίζονται μεταξύ τους.

Έχοντας λοιπόν δει μη στάσιμες χρονολογικές σειρές σειρά έχουν οι ολοκληρωμένες σειρές d τάξης. Όπως έχει προαναφερθεί και στην θεωρία κάποιες μη στάσιμες χρονολογικές σειρές μπορούν να μετατραπούν σε στάσιμες με την χρήση 1^{ων} διαφορών. Έχοντας λοιπόν τη μη στάσιμη σειρά παραπάνω με τις μετοχές θα γίνει προσπάθεια μετατροπής σε στάσιμη χρησιμοποιώντας τις 1^{ες} διαφορές. Πως εφαρμόζονται οι πρώτες διαφορες;



Στη συνέχεια ανοίγει το analyze-forecasting-autocorrelations. Στο πλαίσιο που ανοίγει το πρόγραμμα μεταφέρεται η μεταβλητή μετοχές μέσα και επιλέγεται το differnce με τον αριθμό 1 για να δώσει τις $1^{ες}$ διαφορές.

metoxes	YEAR_	DATE_	metoxes_1.1
.88	2000	2000	.
.99	2001	2001	.11
1.06	2002	2002	.07
1.17	2003	2003	.11
1.20	2004	2004	.03
1.32	2005	2005	.12
1.54	2006	2006	.22
1.78	2007	2007	.24
2.14	2008	2008	.36
2.55	2009	2009	.41
2.81	2010	2010	.26
3.24	2011	2011	.43
3.05	2012	2012	-.19
3.88	2013	2013	.83
4.36	2014	2014	.48
4.08	2015	2015	-.28

Αφού εφαρμόστηκαν οι πρώτες διαφορές εμφανίζεται μια νέα σειρά.

Στο γίνεται έλεγχος αυτοσυσχετίσεων στην καινούριγια σειρά.

ΣΤΑΣΙΜΟΤΗΤΑ 8

DIFF(metoxes,1)

Autocorrelations

Series:DIFF(metoxes,1)

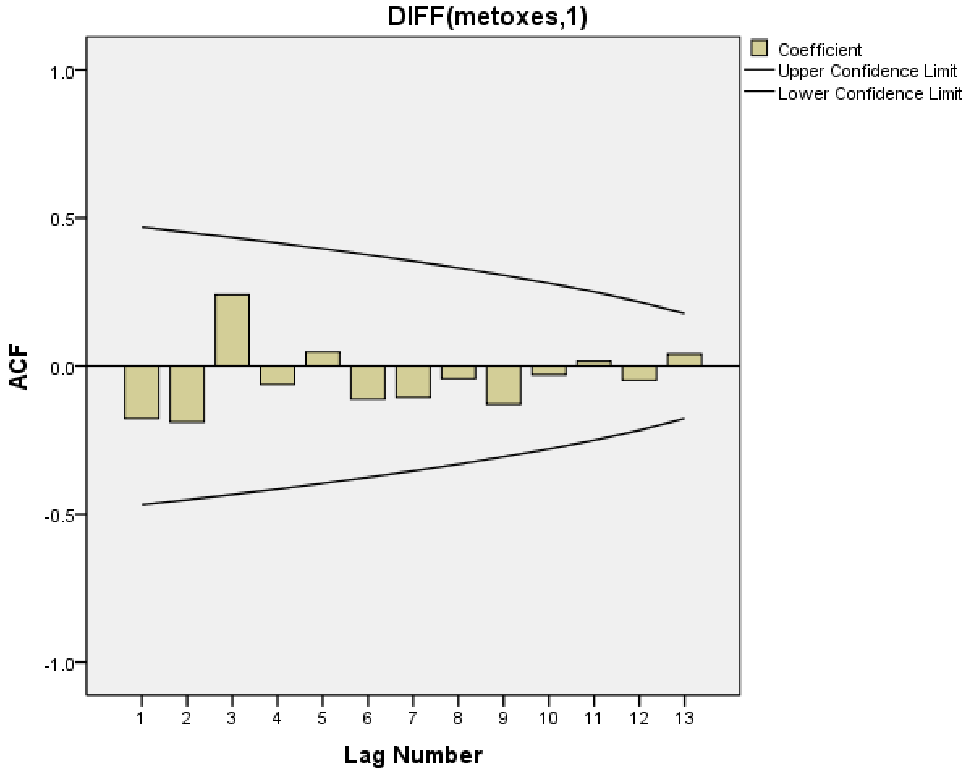
Lag	Autocorrelation	Std. Error[a]	Box-Ljung Statistic		
			Value	df	Sig.[b]
1	-.177	.234	.569	1	.451
2	-.189	.226	1.268	2	.531
3	.241	.217	2.503	3	.475
4	-.062	.208	2.593	4	.628
5	.048	.198	2.652	5	.753
6	-.111	.188	3.003	6	.809
7	-.106	.177	3.360	7	.850
8	-.042	.166	3.425	8	.905
9	-.129	.153	4.130	9	.903
10	-.030	.140	4.174	10	.939
11	.015	.125	4.190	11	.964
12	-.048	.108	4.389	12	.975
13	.041	.089	4.604	13	.983

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

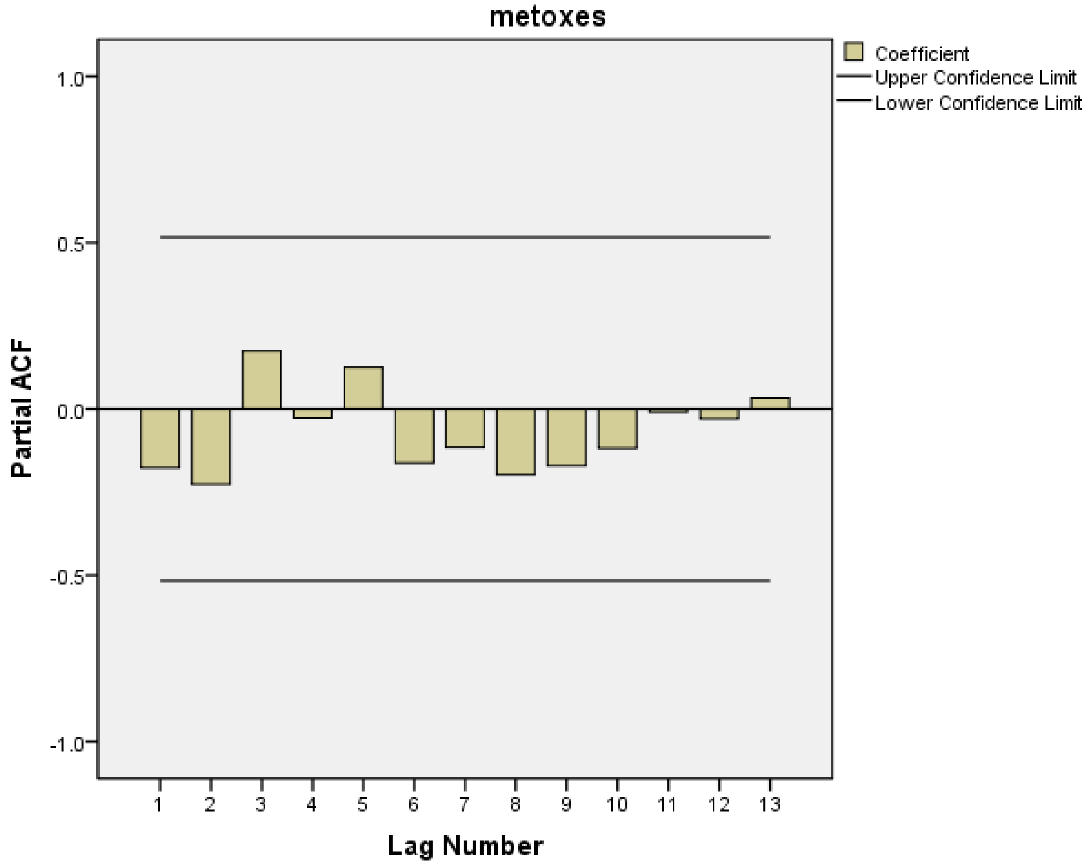
b. Based on the asymptotic chi-square approximation.

Στο πινακάκι των αυτοσυσχετίσεων το πρώτο που πρέπει να κοιτάει ο χρήστης είναι οι τιμές των sig στην μέθοδο box-ljung όπου πρέπει να είναι μεγαλύτερες του 0,05 για να υπάρχει στασιμότητα.

ΣΤΑΣΙΜΟΤΗΤΑ 9



Παραπάνω φαίνεται πως οι αυτοσυσχετίσεις πλέον φθήνουν πάρα πολύ γρήγορα προς το 0 και εδώ λοιπόν ο χρήστης καταλήγει στο συμπέρασμα όπου η στασιμότητα είναι σίγουρη.Στη συνέχεια παρατηρείται και το διάγραμμα μερικών αυτοσυσχετίσεων.

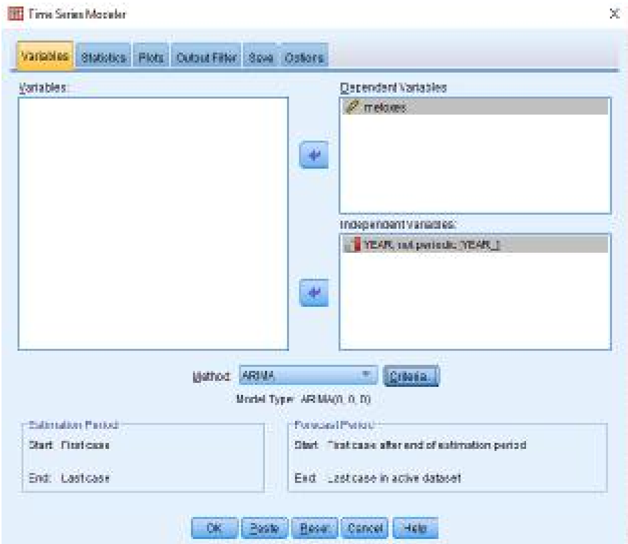


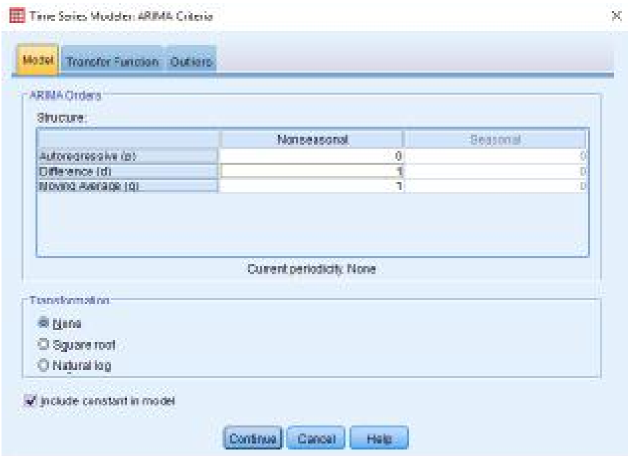
Όπως γίνεται αντιληπτό κι εδώ καμία τιμή δεν περνάει τα όρια οπότε το μοντέλο που έχει δημιουργηθεί είναι ένα μοντέλο ARIMA(0,1,0) αυτό επειδή καμία τιμή των ACF δεν περνάει τα όρια ομοίως και των PACF κι έχουν χρησιμοποιηθεί 1^{ες} διαφορές.q=0 i=1 p=0

Παρακάτω αναλύεται ένα παράδειγμα για το πώς γίνεται σωστά η εκτίμηση των ARIMA(p,d,q) υποδειγμάτων στο SPSS χωρίς να πραγματοποιούνται συνέχεια διαφορές και να παρουσιάζονται συνεχώς άλλα δεδομένα. Για να εκτιμηθεί ένα ARIMA(p, d, q) γίνεται επιλογή Analyze/Forecasting/Create Models και ανοίγει το παράθυρο του Time Series Modeler. Έστω ότι πρέπει να εκτιμηθεί ένα υπόδειγμα ARIMA(0, 1, 1) για τη χρονοσειρά του παραδείγματός. Επιλέγεται η μεταβλητή metoxes που περιέχει τις τιμές της χρονοσειράς και μετακινείται στο πλαίσιο dependent Variable. Στην κυλιόμενη μπάρα Method επιλέγεται ARIMA και πατιέται το κουμπί Criteria για να ανοίξει το παράθυρο Time Series Modeler: ARIMA Criteria. Στο πλαίσιο του ARIMA Orders Nonseasonal πληκτρολογείται p = 0, d = 1 και q = 1. Στις επιλογές Transformation υπάρχει η δυνατότητα εκτίμησης του υποδείγματος μετασχηματίζοντας τις τιμές της χρονοσειράς με λογαρίθμους ή με τετραγωνικές ρίζες. Στην περίπτωση που εξετάζεται μένει η επιλογή None. Η επιλογή include constant in model αναφέρεται στο μέσο όρο της χρονοσειράς. Ειδικότερα, εάν η χρονοσειρά έχει μη μηδενικό μέσο όρο κατά τη διαδικασία της εκτίμησης, η εκτίμηση του υποδείγματος γίνεται σε αποκλίσεις από τον μέσο όρο οπότε θα πρέπει να επιλέγεται το include constant in model. Στην χρονοσειρά που εκτιμάται εφαρμόζεται μετασχηματισμός σε πρώτες διαφορές, δηλαδή ο μέσος όρος της χρονοσειράς θα είναι μηδέν όποτε δεν πρέπει να επιλεγεί το include constant in model. Αφού εισαχθούν οι προηγούμενες επιλογές πατώντας Continue γίνεται επιστροφή στο παράθυρο του Time Series Modeler.

Σε αυτό το σημείο γίνεται αντιληπτό πως όλη αυτή η διαδικασία που ακολουθέι παρακάτω είναι η μεθοδολογία box-jenkins στην οποία όπως έχει γίνει αναφορά, πρώτα γίνεται ταυτοποίηση του υποδείγματος και φαίνονται οι τιμές των p,d,q μετά σε δεύτερη φάση πραγματοποιείται εκτίμηση των παραμέτρων α και θ και τέλος διεξάγεται διαγνωστικός έλεγχος για 4 κριτήρια (1. Τη σημαντικότητα των συντελεστών του υποδείγματος. 2. Τη σταθερότητα των συντελεστών του υποδείγματος. 3. Τις ιδιότητες των καταλοίπων. 4. Την προβλεπτική ικανότητα του υποδείγματος.)

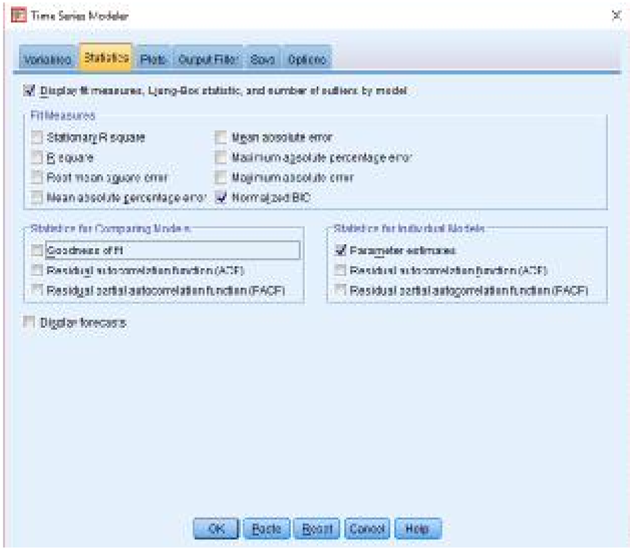
ΕΚΤΙΜΗΣΗ ARIMA 1





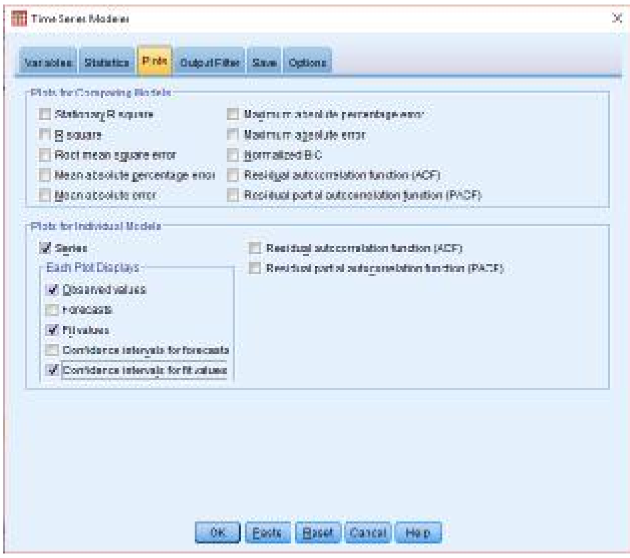
Στη συνέχεια αφού πατηθεί το κουμπί Statistics και γίνει μεταφορά στο παράθυρο Time Series Modeler:Statistics. Στις επιλογές Fit Measures υπάρχουν διαθέσιμα κάποια στατιστικά μέτρα που χρησιμεύουν στην αξιολόγηση της ερμηνευτικής ικανότητας του υποδείγματος που θα εκτιμηθεί και στην επιλογή του καλύτερου υποδείγματος όταν εκτιμώνται διαφορετικά υποδείγματα. Τα πιο αξιόπιστα από αυτά, για την ανάλυση χρονοσειρών, είναι τα μέτρα που ανήκουν στην κατηγορία των πληροφοριακών κριτηρίων όπως το AIC και το BIC τα οποία προσφέρονται σε παλιότερες εκδόσεις του SPSS. Μεταξύ πολλών εκτιμηθέντων υποδειγμάτων καλύτερο θεωρείται αυτό που έχει τη μικρότερη τιμή στα δύο προηγούμενα κριτήρια. Στο SPSS 18 υπάρχει μια τροποποιημένη εκδοχή του κριτηρίου BIC που ονομάζεται Normalized BIC η οποία επιλέγεται. Στις επιλογές Statistics for Comparing Models δεν επιλέγεται τίποτα ενώ από τις επιλογές Statistics for Individuals Models χρησιμποιούνται Parameter estimates. Οι επιλογές Residual autocorrelation_function και Residual partial autocorrelation_function υπολογίζουν τις αυτοσυχετίσεις και μερικές αυτοσυσχετίσεις των σφαλμάτων από το εκτιμηθέν υπόδειγμα. Οι αυτοσυσχετίσεις αυτές που μπορούν να υπολογιστούν και με άλλο τρόπο θα εξεταστούν αναλυτικά σε επόμενη ενότητα και για το λόγο αυτό προς το παρόν δεν θα επιλεχθεί ο υπολογισμός τους

ΕΚΤΙΜΗΣΗ ARIMA 2



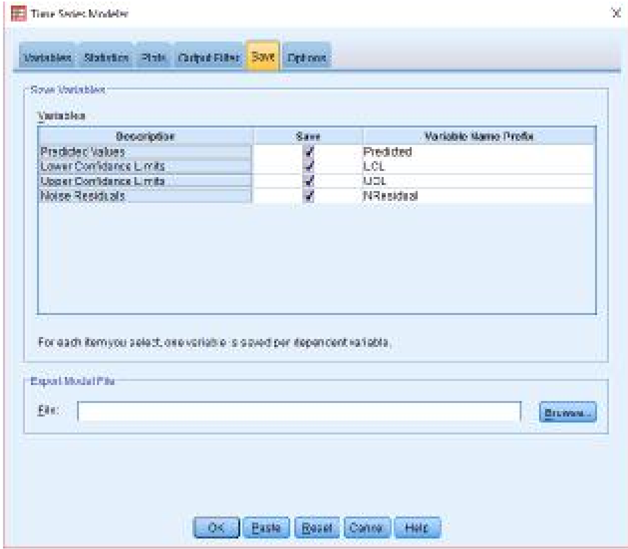
Πατώντας το κουμπί plots μπορεί να επιλεχθεί η κατασκευή κάποιων διαγραμμάτων Αυτό που ενδιαφέρει εδώ είναι η κατασκευή ενός διαγράμματος που να απεικονίζει ταυτόχρονα τις πραγματικές τιμές της χρονοσειράς μαζί με τις εκτιμημένες τιμές της και τα κάτω και άνω όρια ενός διαστήματος εμπιστοσύνης εντός του οποίου βρίσκονται οι εκτιμήσεις. Το διάγραμμα αυτό προκύπτει διαλέγοντας από τις επιλογές του Plots for Individuals Models το Series και από τις επιλογές του Each Plot Displays τα Observed values, Fit values και Confidence intervals for fit values. Εάν επιλεχθούν και οι επιλογές Residual autocorrelation_function και Residual partial autocorrelation_function (δεν θα γίνει προς το παρόν) θα κατασκευαστούν και τα διαγράμματα των αυτοσυσχετίσεων και μερικών αυτοσυσχετίσεων των σφαλμάτων από το εκτιμηθέν υπόδειγμα.

ΕΚΤΙΜΗΣΗ ARIMA 3



Αφού πατηθεί το κουμπί Save στο πλαίσιο Variables του Save Variables επιλέγονται τα Predicted Values, Lower Confidence Limits, Upper Confidence Limits και Noise Residuals για να δημιουργηθούν στο data file τέσσερεις νέες μεταβλητές που θα περιλαμβάνουν τις εκτιμημένες τιμές, το κάτω όριο του διαστήματος εμπιστοσύνης των εκτιμήσεων, το άνω όριο του διαστήματος εμπιστοσύνης των εκτιμήσεων και τα σφάλματα που θα προκύψουν από την εκτίμηση του υποδείγματος

ΕΚΤΙΜΗΣΗ ARIMA 4



Πατώντας OK στο SPSS Statistics Viewer προκύπτουν τα αποτελέσματα από την εκτίμηση του υποδείγματος. Πιο συγκεκριμένα:

ΕΚΤΙΜΗΣΗ ARIMA 5

Model Description

			Model Type
Model ID	metoxes	Model_1	ARIMA(0,1,1)

Στο Model Description περιγράφεται ο τύπος του υποδείγματος που εκτιμήθηκε, δηλαδή ARIMA(0, 1, 1).

ΕΚΤΙΜΗΣΗ ARIMA 6

Model Statistics

Model	Number of Predictors	Model Fit statistics	Ljung-Box Q(18)			Number of Outliers
		Normalized BIC	Statistics	DF	Sig.	
metoxes-Model_1	1	-2.489	.	0	.	0

Στο Model Statistics περιλαμβάνονται η τιμή του πληροφοριακού κριτηρίου Normalized BIC και η τιμή της στατιστικής Q των Box και Ljung για 18 χρονικές υστερήσεις με το P-Value για τον έλεγχο της μηδενικής υπόθεση ότι η χρονοσειρά των σφαλμάτων είναι λευκός θόρυβος.

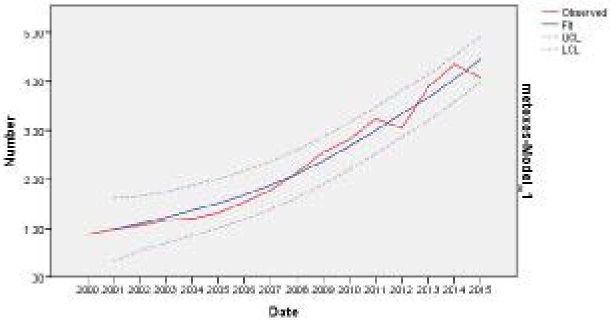
ΕΚΤΙΜΗΣΗ ARIMA 7

[illegible]

					[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]		[illegible]	[illegible]	[illegible]	[illegible]
			[illegible]		[illegible]			
			[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]

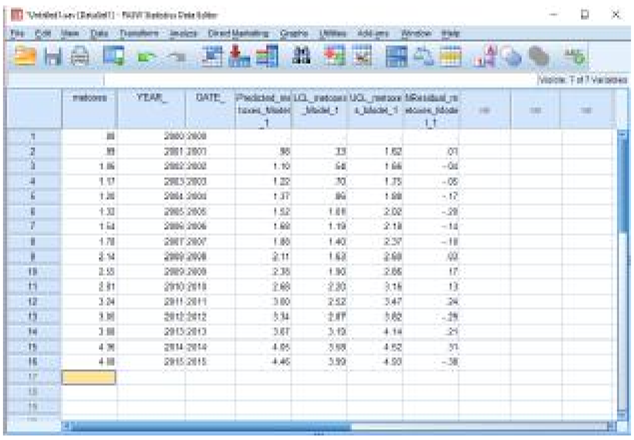
Στο ARIMA Model Parameters περιλαμβάνονται οι τιμές των παραμέτρων του εκτιμηθέντος υποδείγματος, δηλαδή στην περίπτωση που εξετάζεται η τιμή του όρου του κινητού μέσου πρώτης τάξης μαζί με το τυπικό σφάλμα της εκτίμησης, την τιμή της t στατιστικής για τον έλεγχο της στατιστικής του σημαντικότητας και το P-Value του ελέγχου

ΕΚΤΙΜΗΣΗ ARIMA 8



Στο προηγούμενο διάγραμμα απεικονίζονται οι τιμές της χρονοσειράς μαζί με τις εκτιμήσεις τους και τα κάτω και άνω όρια του διαστήματος εμπιστοσύνης των εκτιμήσεων. Επιπρόσθετα, στο data file έχουν δημιουργηθεί τέσσερεις νέες μεταβλητές που περιλαμβάνουν τις εκτιμημένες τιμές, το κάτω όριο και το άνω όριο του διαστήματος εμπιστοσύνης των εκτιμήσεων και τα σφάλματα από την εκτίμηση του υποδείγματος.

ΕΚΤΙΜΗΣΗ ARIMA 9



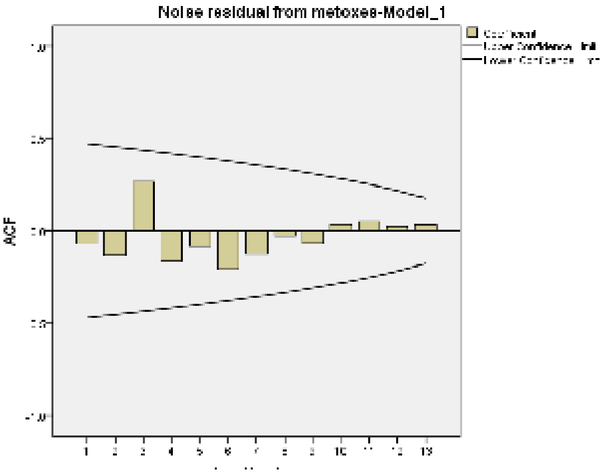
	metoxes	YEAR_	DATE_	Predicted_metoxes_Model_1	LCL_metoxes_Model_1	UCL_metoxes_Model_1	NResidual_metoxes_Model_1
1	.88	2000	2000				
2	.89	2001	2001	.98	.33	1.62	.01
3	1.06	2002	2002	1.10	.58	1.66	-.04
4	1.17	2003	2003	1.22	.70	1.75	-.05
5	1.20	2004	2004	1.37	.86	1.88	-.17
6	1.32	2005	2005	1.52	1.01	2.02	-.20
7	1.54	2006	2006	1.68	1.19	2.18	-.14
8	1.78	2007	2007	1.88	1.40	2.37	-.10
9	2.14	2008	2008	2.11	1.62	2.60	.03
10	2.53	2009	2009	2.36	1.90	2.86	.17
11	2.81	2010	2010	2.68	2.20	3.16	.13
12	3.24	2011	2011	3.00	2.52	3.47	.24
13	3.05	2012	2012	3.34	2.87	3.82	-.29
14	3.88	2013	2013	3.67	3.19	4.14	.21
15	4.36	2014	2014	4.05	3.58	4.52	.31
16	4.08	2015	2015	4.46	3.99	4.93	-.38

Αφού έχει κατασκευαστεί το συγκεκριμένο υπόδειγμα και υπάρχουν οι πρβλέψεις και τα σφάλματα καλό είναι να πραγματοποιηθούν κάποιοι έλεγχοι.

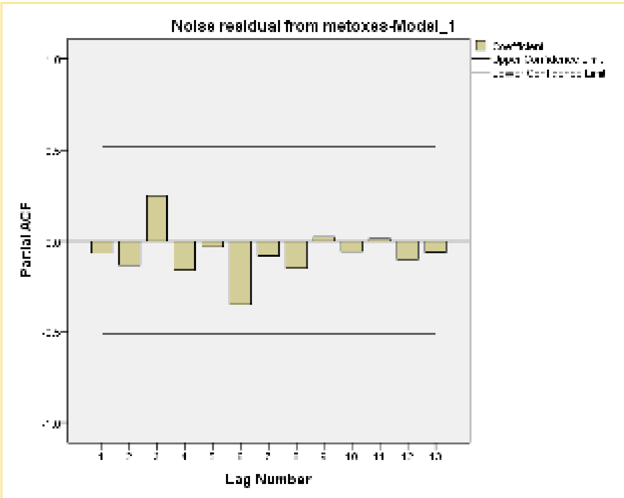
Για να είναι το υπόδειγμα ικανοποιητικό πρέπει: Οι αυτοσυσχετίσεις και μερικές αυτοσυσχετίσεις της χρονοσειράς των σφαλμάτων να μην διαφέρουν σημαντικά από το 0. Μία ή δύο συσχετίσεις υψηλής τάξης μπορεί να υπερβαίνουν το 95% διάστημα εμπιστοσύνης αλλά αν η πρώτη ή η δεύτερη συσχέτιση είναι πολύ μεγάλη τότε η χρονοσειρά δεν είναι ταυτοποιημένη σωστά. Τα σφάλματα πρέπει να είναι λευκός θόρυβος. Ένα τέστ για αυτόν τον έλεγχο γίνεται με τη στατιστική Q των Box και Ljung. Εξετάζεται η τιμή της Q για αριθμό χρονικών υστερήσεων περίπου ίσο με το ένα τέταρτο του δείγματος (αλλά όχι περισσότερο από 50). Ο υπολογισμός των αυτοσυσχετίσεων και μερικών αυτοσυσχετίσεων των σφαλμάτων, καθώς και των τιμών της στατιστικής Q γίνεται όπως δείχτηκε στην ενότητα 5, δηλαδή επιλέγεται Analyze/Forecasting/Autocorrelations και στο Variables μετακινείται η μεταβλητή NResidual_metoxes_Model_1που περιέχει τις τιμές των σφαλμάτων. Από τα διαγράμματα των αυτοσυσχετίσεων και μερικών αυτοσυσχετίσεων διαπιστώνεται ότι δεν υπάρχουν συσχετίσεις που να υπερβαίνουν τα όρια των διαστημάτων εμπιστοσύνης.

Επίσης, η υψηλή τιμή του p-value για τον έλεγχο με τη στατιστική Q των Box και Ljung οδηγεί στο συμπέρασμα ότι τα σφάλματα έχουν συμπεριφορά λευκού θορύβου.

ΕΚΤΙΜΗΣΗ ARIMA 10



ΕΚΤΙΜΗΣΗ ARIMA 11



ΕΚΤΙΜΗΣΗ ARIMA 12

Autocorrelations

Series:Noise residual from metoxes-Model_1

Lag	Autocorrelation	Std. Error[a]	Box-Ljung Statistic Value	df	Sig.[b]
1	-.071	.234	.092	1	.762
2	-.133	.226	.440	2	.803
3	.263	.217	1.913	3	.591
4	-.165	.208	2.545	4	.637
5	-.084	.198	2.726	5	.742
6	-.206	.188	3.924	6	.687
7	-.126	.177	4.433	7	.729
8	-.029	.166	4.463	8	.813
9	-.069	.153	4.664	9	.863
10	.034	.140	4.724	10	.909
11	.050	.125	4.883	11	.937
12	.023	.108	4.928	12	.960
13	.032	.089	5.059	13	.974

Επιπρόσθετα, μια άλλη τεχνική που χρησιμοποιείται για τον έλεγχο της σωστής ταυτοποίησης της χρονοσειράς είναι η εκτίμηση ενός διευρυμένου υποδείγματος είτε από την πλευρά του αυτοπαλίνδρομου μέρους είτε από την πλευρά του μέρους των κινητών μέσων όρων. Για να είναι το εκτιμηθέν υπόδειγμα ικανοποιητικό θα πρέπει στα διευρυμένα υποδείγματα οι συντελεστές να είναι στατιστικά μη σημαντικοί. Στην περίπτωση της χρονοσειράς που εξετάζεται εκτιμήθηκε ένα υπόδειγμα ARIMA(0, 1, 1) οπότε ως διευρυμένα υποδείγματα θα πρέπει να εκτιμηθούν το ARIMA(1, 1, 1) και το ARIMA(0, 1, 2). Εκτιμώντας τα υποδείγματα αυτό με τη διαδικασία που περιγράφηκε προηγουμένως προκύπτουν οι παρακάτω πίνακες εκτιμήσεων:

ΕΚΤΙΜΗΣΗ ARIMA 13

[illegible]

					[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]		[illegible]	[illegible]	[illegible]	[illegible]
			[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
			[illegible]		1			
			[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]

ΕΚΤΙΜΗΣΗ ARIMA 14

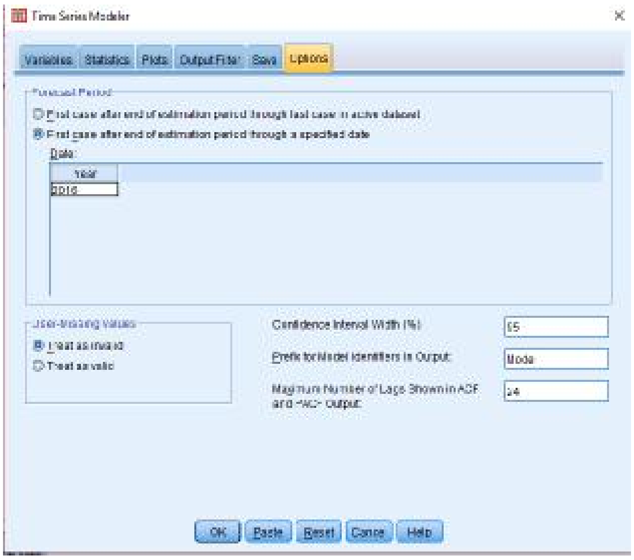
ARIMA Model Parameters

					Estimate	SE	t	Sig.
metoxes-Model_1	metoxes	No Transformation	Constant		-25.900	24.907	-1.040	.321
			Difference		1			
			MA	Lag 1	.561	11.991	.047	.964
				Lag 2	.431	5.402	.080	.938
	YEAR, not periodic	No Transformation	Numerator	Lag 0	.013	.012	1.049	.316

Από τα προηγούμενα αποτελέσματα διαπιστώνεται ότι στα διευρυμένα υποδείγματα οι πρόσθετοι όροι δεν είναι στατιστικά σημαντικοί. Κατά συνέπεια, η χρονοσειρά metoxes που αναλύθηκε έχει ταυτοποιηθεί σωστά ως ARIMA(0, 1, 1) διαδικασία.

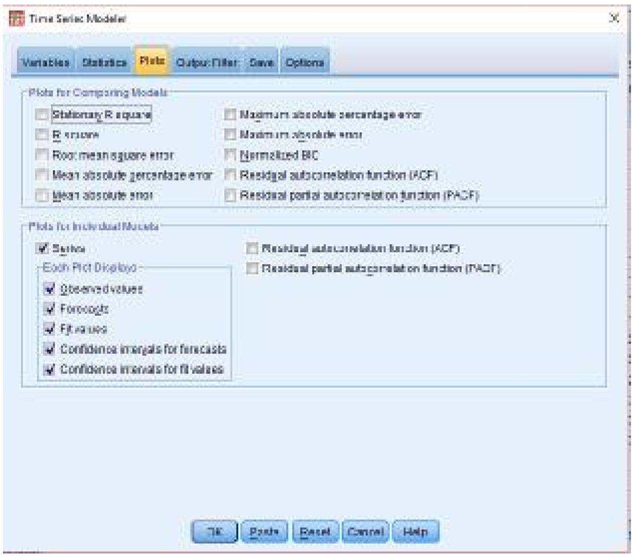
Για να διενεργηθούν προβλέψεις για τις μελλοντικές τιμές της χρονοσειράς εκτιμάται κανονικά το υπόδειγμα όπως δείχτηκε στην ενότητα , αφού όμως προηγουμένως πατηθεί το κουμπί Options και στο πλαίσιο Forecast Period επιλεχθεί First case after end of estimation period through a specified date και εισαχθεί μέχρι πια χρονική περίοδο το SPSS θα υπολογίσει τις προβλέψεις. Ας υποτεθεί ότι χρειάζονται προβλέψεις μέχρι το έτος 2016 οπότε στο πλαίσιο πληκτρολογείται το 2016.

ΕΚΤΙΜΗΣΗ ARIMA 15



Εάν πατηθεί το κουμπί plots μπορεί να επιλεχθεί στο διάγραμμα με τις πραγματικές τιμές της χρονοσειράς, τις εκτιμήσεις και τα διαστήματα εμπιστοσύνης που έχει αναφερθεί να περιληφθούν και οι προβλέψεις.

ΕΚΤΙΜΗΣΗ ARIMA 16



Πατώντας OK προκύπτουν στο SPSS Statistics Viewer τα γνωστά αποτελέσματα με τη διαφορά ότι στο διάγραμμα περιλαμβάνονται και οι προβλέψεις ενώ στο data file θα δημιουργηθούν οι μεταβλητές που όμως θα περιλαμβάνουν και τις προβλέψεις μέχρι το 2016

ΕΚΤΙΜΗΣΗ ARIMA 17

	metoxes	YEAR_	DATE_	Predicted_metoxes_Model_1
1	.88	2000	2000	.
2	.99	2001	2001	.98
3	1.06	2002	2002	1.10
4	1.17	2003	2003	1.22
5	1.20	2004	2004	1.37
6	1.32	2005	2005	1.52
7	1.54	2006	2006	1.68
8	1.78	2007	2007	1.88
9	2.14	2008	2008	2.11
10	2.55	2009	2009	2.38
11	2.81	2010	2010	2.68
12	3.24	2011	2011	3.00
13	3.05	2012	2012	3.34
14	3.88	2013	2013	3.67
15	4.36	2014	2014	4.05
16	4.08	2015	2015	4.46
17	.	2016	2016	4.87

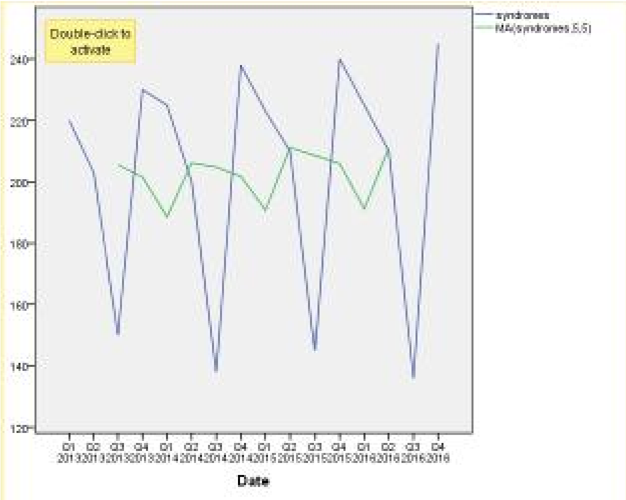
Τέλος όμως πέρα από τις ασκήσεις θα παρουσιαστεί ολοκληρωμένα η πρόβλεψη που γίνεται σε έναν πραγματικό οργανισμό.Έχουν αντληθεί τα δεδομένα από ένα γυμναστήριο κοντά στο τεί της πάτρας με βασική μεταβλητή τις συνδρομές που τρέχουν ανά μήνα.Σκοπός είναι να πραγματοποιηθεί πρόβλεψη των μελλοντικών συνδρομών στα επόμενα εξάμηνα. Παρακάτω παρουσιάζεται το πινακάκι στο SPSS με τους αριθμούς των συνδρομών. Για διευκόλυνση έγινε διαχωρισμός στις τιμές ανά τρίμηνο.

ΠΡΟΒΛΕΨΗ ΠΕΛΑΤΕΙΑΣ 1

	syndromes	YEAR_	QUARTER_	DATE_
1	220	2013	1	Q1 2013
2	203	2013	2	Q2 2013
3	150	2013	3	Q3 2013
4	230	2013	4	Q4 2013
5	225	2014	1	Q1 2014
6	200	2014	2	Q2 2014
7	138	2014	3	Q3 2014
8	238	2014	4	Q4 2014
9	223	2015	1	Q1 2015
10	210	2015	2	Q2 2015
11	145	2015	3	Q3 2015
12	240	2015	4	Q4 2015
13	225	2016	1	Q1 2016
14	210	2016	2	Q2 2016
15	136	2016	3	Q3 2016
16	245	2016	4	Q4 2016

Στη συνέχεια θα πραγματοποιηθεί σύγκριση του διαγράμματος των κυλιόμενων μέσων με των αρχικών τιμών για να γίνει έλεγχος υπάρξης σταδιακής αύξησης ή μείωσης μέσα στα έτη.

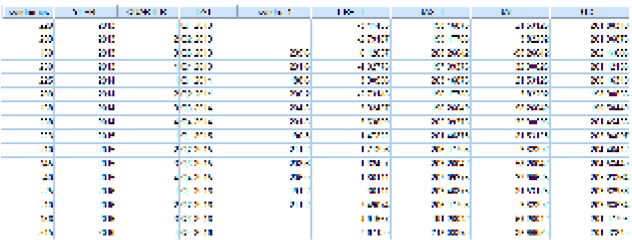
ΠΡΟΒΛΕΨΗ ΠΕΛΑΤΕΙΑΣ 2



Από το παραπάνω διάγραμμα είίναι αντιληπτό πως υπάρχει σίγουρα εποχικότητα στις συνδρομές κι επίσης μετά το 2015 φαίνεται μια μικρή αύξηση των τιμών. Η εποχικότητα υπάρχει καθώς οι υψηλότερες τιμές εμφανίζονται στο 4^{o} τρίμηνο του έτους και οι χαμηλότερες στο 3^{o} τρίμηνο. Αυτό συμβαίνει επειδή οι πελάτες του γυμναστηρίου είναι κυρίως φοιτητές οι οποίοι αγοράζουν μεγαλύτερα πακέτα το φθινόπορο και τον χειμώνα ενώ τους καλοκαιρινούς μήνες φεύγουν για διακοπές και δεν υπάρχει αρκετός κόσμος λόγο καύσωνα.

Στο επόμενο βήμα πάμε προχωράμε στην εξάλειψη εποχικότητας με την βοήθεια του seasonal decomposition.

ΠΡΟΒΛΕΨΗ ΠΕΛΑΤΕΙΑΣ 3



Στη στήλη Seasonal Adjusted Series είναι η χρονοσειρά της πελατείας χωρίς εποχικότητα επομένως μπορεί να πραγματοποιηθεί πρόβλεψη.

Παρακάτω αναγράφονται οι αυτοσυσχετίσεις.

ΠΡΟΒΛΕΨΗ ΠΕΛΑΤΕΙΑΣ 4

Autocorrelations

Series:Seasonal adjusted series for syndromes from SEASON, MOD_17, ADD CEN 4

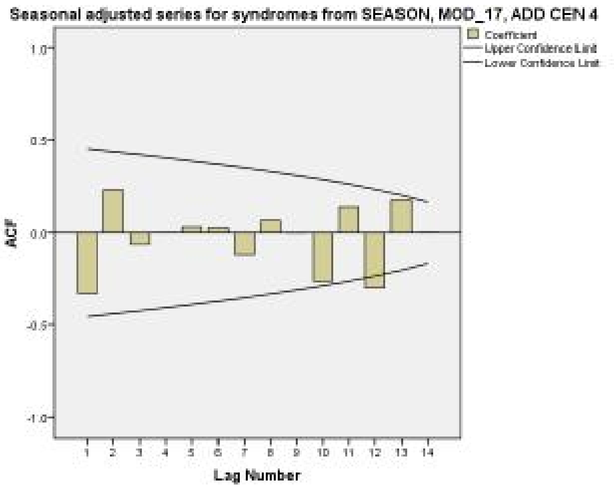
Lag	Autocorrelation	Std. Error[a]	Box-Ljung Statistic		
			Value	df	Sig.[b]
1	-.330	.228	2.086	1	.149
2	.232	.220	3.189	2	.203
3	-.061	.212	3.272	3	.352
4	.001	.204	3.272	4	.513
5	.030	.195	3.296	5	.654
6	.024	.186	3.313	6	.769
7	-.118	.177	3.761	7	.807
8	.066	.167	3.916	8	.865
9	-.002	.156	3.916	9	.917
10	-.262	.144	7.210	10	.705
11	.139	.132	8.331	11	.683
12	-.299	.118	14.756	12	.255
13	.177	.102	17.751	13	.167
14	.002	.083	17.751	14	.218

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

b. Based on the asymptotic chi-square approximation.

Από τη στιγμή που τα sig είναι μεγαλύτερα του 0,05 χωρίς να έχουν χρησιμοποιηθεί διαφορές δεν χρειάζεται να γίνει κάτι άλλο σε αυτό το σημείο έχει παρουσιαστεί ένα μοντέλο ARIMA(1,0,0)

ΠΡΟΒΛΕΨΗ ΠΕΛΑΤΕΙΑΣ 5



Τρέχοντας το forecasting στο Create models για ARIMA(1,0,0) εμφανίζονται αυτές οι τιμές για την χρονοσειρά χωρίς εποχικότητα

ΠΡΟΒΛΕΨΗ ΠΕΛΑΤΕΙΑΣ 6

200.73266
206.38763
204.85624
206.76210

Οι πραγματικές τιμές όμως με εποχικότητα θα είναι οι εξής:221

210

146

238

Οι τιμές αυτές υπολογίστηκαν με την βοήθεια των εποχικών δεικτών.

Επίσης να προσθεθεί πως έγινε έλεγχος και στα μοντέλα ARIMA(2,0,0) και ARIMA(1,0,1) αλλά το ARIMA(1,0,0) ήταν το πιο αξιόπιστο λόγο των p-value(sig).

Σε αυτό το σημείο τελείωσε η πρόβλεψή.

ΣΥΜΠΕΡΑΣΜΑΤΑ

Έχοντας δει ανάλυση χρονολογικών σειρών τόσο σε θεωρητικό επίπεδο όσο και σε πρακτικό, είναι πλέον αντιληπτό πως η διαδικασία ανάλυσης είναι μια πολύ αξιόπιστη διαδικασία καθώς βασίζεται σε μαθηματικούς τύπους οι οποίοι πολύ σωστά εφαρμόζονται πάνω στα δεδομένα που παρουσιάστηκαν και κατάφεραν να εξάγουν πληροφορίες οι οποίες χρησιμοποιήθηκαν για την πρόβλεψη μελλοντικών τιμών.Το κομμάτι της πρόβλεψης όμως είναι αυτό το οποίο σημειώνει κάποιους προβληματισμούς.Είναι άξιο σημείωσης πως όλες οι διαδικασίες και οι πράξεις τρέχουν με τη βοήθεια του PASW STATISTICS το οποίο χειρίζεται ο άνθρωπος προς διυκόλυνσή του, γι αυτό το λόγο είναι αρκετά εύκολο για κάποιον να κάνει κάποιο λάθος είτε κατά τη διάρκεια της ανάλυσης αλλά και της πρόβλεψης, κάτι το οποίο μπορεί να αποτελέσει πρόβλημα καθώς θα καταλήξει σε λανθασμένη πρόβλεψη.Ο παραπάνω λόγος είναι αυτός που καθιστά την διαδικασία ανάλυσης και πρόβλεψης ως χρονοβόρα.

Τώρα όσον αφορά καθαρά το κομμάτι της πρόβλεψης είναι ευδιάκριτο πως δεν γίνεται να προβλεφθούν τα πάντα με ακρίβεια καθώς υπάρχει η τυχαία συνιστώσα.Για παράδειγμα αν μιλάμε για μια επιχείρηση η οποία έχει τις εγκαταστάσεις της σε μια χώρα όπου η οικονομία και η κυβέρνηση δεν είναι σταθερές με αποτέλεσμα να αλλάζουν συνεχώς νόμοι και να αυξομειώνεται με μεγάλους ρυθμούς η αγορά, τότε οι προβλεπόμενες τιμές που έχουν εκπονηθεί από το πρόγραμμα «μπορεί» να μην είναι ιδιαίτερα αξιόπιστες, ενώ αν η επιχείρηση βρίσκεται σε σταθερό περιβάλλον τότε οι προβλεπόμενες τιμές έχουν μεγάλη βαρύτητα και πρέπει να ληφθούν σοβαρά υπόψη από τον λήπτη αποφάσεων.Το κυριότερο συμπέρασμα είναι το εξής: Οι τεχνικές προβλέψεων και ελέγχου που χρησιμοποιούνται σήμερα είναι αρκετά αξιόπιστες και σχεδόν πάντα πρέπει να λαμβάνονται υπόψη τα τελικά αποτελέσματα,υπάρχει όμως η τυχαία συνιστώσα που κάνει αυτές τις τεχνικές να μην έιναι ιδιαίτερα ακριβείς γι αυτό θα πρέπει οι λήπτες αποφάσεων στις επιχειρήσεις να συμβουλέυονται τα αποτελέσματα και τα σχόλια του τμήματος που πραγματοποιέι τις προβλέψεις αλλά σε καμία περίπτωση δεν πρέπει να κινηθούν εξ’ ολοκλήρου με βάση αυτά.

Τέλος είναι αξιοσημείωτο πως το pasw statistics σαν πρόγραμμα είναι ιδιαίτερα εύχρηστο καθώς δεν χρειάζεται να πραγματοποιούνται συνεχώς πράξεις από τον άνθρωπο και με αυτόν τον τρόπο αποφεύγονται πολλά λάθη.Επιπλέον το πρόγραμμα αυτό δεν απαιτεί απο τον χρήστη να είναι ιδιαίτερα καλός στις μαθηματικές συναρτήσεις και στους τύπους που παρουσιάστηκαν στο θεωρητικό κομμάτι, αρκεί να μάθει να το διαχειρίζεται αποτελεσματικά και σωστά.

ΒΙΒΛΙΟΓΡΑΦΙΑ

1.(ΒΙΒΛΙΟ)ΣΤΑΤΙΣΤΙΚΗ ΣΚΕΨΗ ΣΤΟΝ ΚΟΣΜΟ ΤΩΝ ΕΠΙΧΕΙΡΗΣΕΩΝ(ACZEL D. AMIR).

2.(ΒΙΒΛΙΟ)ΑΝΑΛΥΣΗ ΧΡΟΝΟΣΕΙΡΩΝ(ΣΤΡΙΝΤΖΗΣ ΜΙΧΑΛΗΣ-ΓΕΡΑΣΙΜΟΣ).

3.(ΣΗΜΕΙΩΣΕΙΣ)ΧΡΟΝΟΛΟΓΙΚΕΣ ΣΕΙΡΕΣ(Φ. ΚΟΥΝΤΟΥΡΗ).

4.(ΣΗΜΕΙΩΣΕΙΣ)ΣΗΜΕΙΩΣΕΙΣ ΟΙΚΟΝΟΜΕΤΡΙΑΣ(e-class ΑΤΕΙ ΠΑΤΡΩΝ).

5.(ΣΗΜΕΙΩΣΕΙΣ)ΑΝΑΛΥΣΗ ΧΡΟΝΟΛΟΙΚΩΝ ΣΕΙΡΩΝ(Χρυσανθη Παπαθανασοπουλου e-class).

6.(ΣΗΜΕΙΩΣΕΙΣ)ΑΝΑΛΥΣΗ ΧΡΟΝΟΛΟΓΙΚΩΝ ΣΕΙΡΩΝ(ΚΟΥΝΕΤΑΣ e-class).

7.(ΒΙΒΛΙΟ)ΤΕΧΝΙΚΕΣ ΟΙΚΟΝΟΜΕΤΡΙΑΣ ΚΑΙ ΑΝΑΛΥΣΗΣ ΧΡΟΝΟΛΟΓΙΚΩΝ ΣΕΙΡΩΝ(ΘΕΟΔΩΡΟΣ ΣΙΜΟΣ)

8.(ΒΙΒΛΙΟ) Theil, H., (1967) Applied Economic Forecasting, Rand McNally.

9.(ΒΙΒΛΙΟ) Γεωργίου Κ. Χρήστου, (2007) Εισαγωγή στην Οικονομετρία, Β΄ Τόμος, Γ΄ Έκδοση, Αθηνά, Gutenberg.

10.(ΒΙΒΛΙΟ) Nelson, C. R., (1973) Applied Time Series Analysis for Managerial Forecasting, Holden – Day.

11(ΒΙΒΛΙΟ) Box, G. E. P. and Pierce, D.A (1970) Distribution of residual auto-correlation in autoregressive-integrated moving average time-series models. J. Amer. Statist. Ass. 65, 1509-26