ΤΕΧΝΟΛΟΓΙΚΟ ΕΚΠΑΙΔΕΥΤΙΚΟ ΙΔΡΥΜΑ ΔΥΤΙΚΗΣ ΕΛΛΑΔΑΣ

ΣΧΟΛΗ ΔΙΟΙΚΗΣΗΣ ΚΑΙ ΟΙΚΟΝΟΜΙΑΣ

ΤΜΗΜΑ ΔΙΟΙΚΗΣΗΣ ΕΠΙΧΕΙΡΗΣΕΩΝ

Τίτλος Εργασίας:

Η χρήση της μεθοδολογίας Box – Jenkins στην ανάλυση χρονοσειρών

Φοιτητής:

Μαρκόπουλος Άγγελος – Κωνσταντίνος

Ντεντης Ιωάννης

Παρασκευόπουλος Νικόλαος

Επιβλέπων καθηγητής:

Δρ. Δημήτριος Παπαδόπουλος

Πάτρα, Φεβρουάριος 2015

Περίληψη

Αντικείμενο της εργασίας αυτής είναι η μελέτη και παρουσίαση των τεχνικών πρόβλεψης που χρησιμοποιούνται για την ανάλυση χρονοσειρών. Συγκεκριμένα, το ενδιαφέρον της εργασίας εστιάζεται στην ανάλυση και εφαρμογή της μεθοδολογίας Box-Jenkins και των μοντέλων ARIMA, αφού βεβαίως νωρίτερα έχει προηγηθεί η παρουσίαση των αυτοπαλίνδρομων διαδικασιών (AR) και των διαδικασιών κινητού μέσου (MA).

Abstract

The subject of this paper is the study and presentation of forecasting techniques used in time series analysis. Specifically, interesting work focuses on the analysis and application of the Box-Jenkins methodology and ARIMA models, after earlier certainly preceded the presentation of the autoregressive process (AR) and moving average processes (MA).

ΠΕΡΙΕΧΟΜΕΝΑ

ΠΙΝΑΚΑΣ ΕΙΚΟΝΩΝ

ΕΙΣΑΓΩΓΗ

Στην ανάλυση χρονοσειρών, η μέθοδος Box-Jenkins (το όνομά της προήλθε από τους στατιστικολόγους Jeorge Box και Jenkins Gwilym), εφαρμόζει αυτοπαλίνδρομα μοντέλα κινητού μέσου όρου ΑΡΜΑ ή ARIMA, για να βρει την καλύτερη προσαρμογή του μοντέλου στις παρελθοντικές τιμές μιας χρονολογικής σειράς.

Η εκτίμηση των παραμέτρων για τα μοντέλα Box-Jenkins είναι ένα αρκετά πολύπλοκο μη γραμμικό πρόβλημα εκτίμησης. Για το λόγο αυτό, η εκτίμηση των παραμέτρων θα πρέπει να αφεθεί σε ένα υψηλής ποιότητας πρόγραμμα λογισμικού που προσομοιώνει τα μοντέλα Box-Jenkins. Ευτυχώς, πλέον πολλά προγράμματα στατιστικού λογισμικού δίνουν τη δυνατότητα κατασκευής μοντέλων Box-Jenkins.

Στην ανάλυση που ακολουθεί, αρχικά παρουσιάζονται κάποιες βασικές έννοιες των χρονολογικών σειρών, καθώς και οι διάφορες τεχνικές μέτρησης της μακροπρόθεσμης τάσης. Επίσης, αποσαφηνίζονται βασικές έννοιες του αντικειμένου, όπως είναι η μελέτη στασιμότητας, η συνάρτηση αυτοσυσχέτισης και ο τελεστής υστέρησης.

Στα επόμενα κεφάλαια μελετάται η εφαρμογή των αυτοπαλίνδρομων διαδικασιών (AR) και των διαδικασιών κινητού μέσου, καθώς και πως προκύπτουν από το συνδυασμό αυτών οι μικτές διαδικασίες ARMA.

Αφού προηγηθούν τα ανωτέρω, γίνεται μια λεπτομερής παρουσίαση της μεθοδολογίας Box-Jenkins και εν συνεχεία πραγματοποιείται η εύρεση του καταλληλότερου υποδείγματος για τη πρόβλεψη των τιμών της μετοχής εταιρειών εισηγμένων στο χρηματιστήριο Αθηνών.

ΚΕΦΑΛΑΙΟ 1: ΧΡΟΝΟΛΟΓΙΚΕΣ ΣΕΙΡΕΣ

1.1. Έννοια Χρονολογικής Σειράς

Με τον όρο χρονολογική σειρά (time series) εννοούμε μια ακολουθία τιμών μιας μεταβλητής ή μιας σύνθεσης μεταβλητών, οι οποίες έχουν ληφθεί σε διαδοχικές χρονικές περιόδους. Με άλλα λόγια, μια χρονολογική σειρά είναι ένα σύνολο παρατηρήσεων y_1, y_2,…y_T, όπου ο δείκτης T υποδηλώνει ισαπέχοντα (ή μη) χρονικά σημεία ή διαστήματα. Οι προαναφερθείσες παρατηρήσεις y_1, y_2,…y_T, είναι συγκεκριμένες τιμές των τυχαίων μεταβλητών y_1, y_2,…y_T, και αποτελούν ένα τμήμα μόνο από μια άπειρη σειρά τυχαίων μεταβλητών.

Η άπειρη αυτή ακολουθία των τυχαίων μεταβλητών, ονομάζεται στοχαστική (στοχαστική ανέλιξη) ή τυχαία διαδικασία. Η έννοια της στοχαστικής διαδικασίας είναι ανάλογη της έννοιας του πληθυσμού, ενώ η έννοια των παρατηρήσεων y_1, y_2,…y_T, αναφέρεται στην έννοια του δείγματος [4].

1.2. Ανάλυση Χρονολογικών Σειρών

Στην ανάλυση χρονολογικών σειρών συναντώνται δύο μοντέλα, το προσθετικό και το πολλαπλασιαστικό μοντέλο.

Το προσθετικό μοντέλο δίνεται από τον τύπο:

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

ενώ το πολλαπλασιαστικό μοντέλο από τον τύπο:

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

όπου,

Y: η τιμή της μεταβλητής,

T: η μακροχρόνια τάση,

S: η εποχιακή συνιστώσα,

C: η κυκλική συνιστώσα,

R: η τυχαία συνιστώσα.

1.3. Βασικές Συνιστώσες Χρονολογικών Σειρών

Κατά την ανάλυση μιας χρονολογικής σειράς μελετώνται οι ακόλουθες συνιστώσες [5]:

Μακροχρόνια Τάση (Secular Trend):

Η μακροπρόθεσμη τάση είναι η γενική συμπεριφορά μιας δεδομένης μεταβλητής σε μια μακρά χρονική περίοδο. Η τάση αναφέρεται στην επιρροή που έχουν μακροχρόνιοι παράγοντες στην χρονοσειρά απαλλαγμένη από εποχιακές, κυκλικές και τυχαίες επιρροές.

Κυκλική Συνιστώσα (Cyclical Component):

Είναι η ανοδική ή καθοδική τάση που αναφέρεται σε μεγάλες περιόδους χρόνου. Οι διακυμάνσεις αυτές όταν αφορούν μεταβλητές που αντιπροσωπεύουν πωλήσεις, λέγονται εμπορικοί κύκλοι.

Η κυκλική συνιστώσα μετράει τις αποκλίσεις των τιμών γύρω από τη μακροχρόνια τάση μια χρονολογικής σειράς και έχει διάρκεια μεγαλύτερη από ένα έτος.

Εποχική Συνιστώσα (Seasonal Component):

Η βασική διαφορά που παρουσιάζει η εποχική συνιστώσα έναντι της κυκλικής είναι η χρονική διάρκεια, καθώς μπορεί να εμφανιστεί και να εκτιμηθεί σε διαστήματα μικρότερα του έτους (πχ μήνες ή τρίμηνα).

Η μέτρηση της εποχιακής συνιστώσας επιτυγχάνεται με τους εποχιακές δείκτες.

Τυχαία Συνιστώσα (Random Component):

Η τυχαία συνιστώσα αποτελεί στην ουσία την μη συστηματική διακύμανση. Αυτό πρακτικά σημαίνει ότι δεν επηρεάζεται από τις υπόλοιπες συνιστώσες και δεν επαναλαμβάνεται σε τακτά χρονικά διαστήματα.

1.4. Μέτρηση της Μακροπρόθεσμης Τάσης

Αφού σχολιάστηκαν οι συνιστώσες που συναντώνται σε μια χρονολογική σειρά, σε αυτό το σημείο παρατίθενται οι τρόποι ανάλυσης μιας χρονολογικής σειράς και εκτίμησης της τάσης. Συγκεκριμένα αναφέρονται οι ακόλουθες μεθοδολογίες [5]:

- Προσαρμογή τάσης με το ελεύθερο χέρι (freehand fitting).
- Προσαρμογή της τάσης με την αρχή των ελαχίστων τετραγώνων (principles of least squares).
- Προσαρμογή της τάσης με τους κινητούς μέσους (moving averages method).
- Προσαρμογή της τάσης με τους εκθετικά ομαλοποιημένους μέσους (method of exponentially smoothed averages).
- Προσδιορισμός της τάσης όταν η καμπύλη τάσης λογίζεται ως η εκθετική καμπύλη.
- Προσδιορισμός της τάσης όταν η καμπύλη τάσης λογίζεται ως η λογιστική καμπύλη.

1.4.1. Freehand fitting

Η συγκεκριμένη μέθοδος δεν θεωρείται αξιόπιστη και στην ουσία πρόκειται μια πρόχειρη μεθοδολογία προκειμένου κάποιος να εξάγει γρήγορα συμπεράσματα σχετικά με τη τάση της χρονοσειράς. Τα στάδια εφαρμογής της μεθόδου είναι τα εξής:

Α) Κατασκευή του στικτού διαγράμματος για τη τυχαία μεταβλητή y, με την εύρεση των σημείων (x_i,y_i), $i=1(1)n$.

Β) Σχεδιασμός της ευθείας γραμμής που προκύπτει από την προσαρμογή της στα σημεία (x_i,y_i), $i=1(1)n$.

Γ) Εύρεση των τιμών α και β της γραμμικής εξίσωσης $y = α + βx$, με τη χρήση δύο σημείων της ευθείας.

1.4.2. Προσαρμογή της τάσης με την αρχή των ελαχίστων τετραγώνων

Με βάση την υπόθεση ότι η τάση εκφράζεται με μια ευθεία γραμμή $Y_c = a + bt$ (όπου Y_c είναι η τιμή της τάσης για το χρονικό σημείο t), τότε με τη χρήση της αρχής των ελαχίστων τετραγώνων θα γίνει η εύρεση των τιμών a και b.

Το γραμμικό μοντέλο δίνεται από τη σχέση:

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

Όπου $R(\varepsilon)=0$ και $\Delta(\varepsilon)=\sigma^2$.

Σύμφωνα με την αρχή των ελαχίστων τετραγώνων, οι αμερόληπτοι εκτιμητές των α και β, δίνονται από τις σχέσεις:

$$b = \frac{n\sum t_i y_i - (\sum t_i)(\sum y_i)}{n\sum (t_i)^2 - (\sum t_i)^2}$$

$$a = \bar{y} - b\bar{t}$$

1.4.3. Προσαρμογή της τάσης με τη χρήση των κινητών μέσων

Για να γίνει καλύτερα κατανοητή η συγκεκριμένη μεθοδολογία, είναι απαραίτητο να οριστεί πριν ο κινητός μέσος, όπου στην ουσία είναι μια βοηθητική χρονολογική σειρά στην οποία η τιμή για μια ορισμένη χρονική περίοδο αντικαθίσταται από τον μέσο αυτής τις τιμής και τις τιμές ενός αριθμού προηγούμενων και επόμενων χρονικών περιόδων.y_{1980}

Για παράδειγμα έστω ότι ο ακόλουθος πίνακας περιλαμβάνει μια χρονολογική σειρά

1980	1981	…	…	2009	2010
Y_{1980}	Y_{1981}	…	…	Y_{2009}	Y_{2010}

Τότε η σειρά που προκύπτει και της οποίας οι όροι είναι οι τριετείς κινητοί μέσοι, δίνεται στον ακόλουθο πίνακα:

1980	
1981	$Y'_{1980} = (Y_{1980} + Y_{1981} + Y_{1982})/3$
…	…
…	…
2009	$Y'_{2009} = (Y_{2008} + Y_{2009} + Y_{2010})/3$
2010	

Το σημαντικό πλεονέκτημα της σειράς των κινητών μέσων, είναι ότι το στικτό διάγραμμα της σειράς αυτής φανερώνει τη τάση της αρχικής χρονολογικής σειράς. Από την άλλη, σημαντικό μειονέκτημα αποτελεί το γεγονός ότι παραλείπονται οι τιμές στην αρχή και στο τέλος της σειρά. Σημαντικότερο όμως είναι το μειονέκτημα ότι η μεθοδολογία αυτή δεν καταλήγει σε μοντελοποίηση του προβλήματος και εξαγωγή της εξίσωσης, ώστε εν συνεχεία να διατυπωθούν προβλέψεις (Με εξαίρεση τον τελευταίο κινητό μέσο που και πάλι μπορεί να χρησιμοποιηθεί σε ορισμένες περιπτώσεις).

1.4.4. Exponential smoothing method

Η προσαρμογή της τάσης με τη μέθοδο της εκθετικής εξομάλυνσης, προβλέπει την κατασκευή μιας καινούριας χρονολογικής σειράς με βάση τα δεδομένα της αρχικής χρονολογικής σειράς. Στη νέα αυτή σειρά που προκύπτει, κάθε παρατήρηση είναι γραμμικός συνδυασμός των προηγούμενων παρατηρήσεων (όπου κάθε παρατήρηση έχει μεγαλύτερη τιμή από τη προηγούμενη).

Οι παρατηρήσεις της εκθετικά ομαλοποιημένης σειράς προέρχονται από τις παρατηρήσεις της αρχικής σειράς και δίνονται με βάση το τύπο:

$$\bar{Y}_t = \bar{Y}_{t-1} + a(Y_t - \bar{Y}_{t-1})$$

Όπου

Y_t : η παρατήρηση το χρονικό σημείο t,

$\bar{Y}_t$: η εκθετικά ομαλοποιημένη παρατήρηση τη χρονική στιγμή t,

α : ο συντελεστής που ονομάζεται παράγοντας ομαλοποίησης και ο οποίος δέχεται τιμές στο κλειστό διάστημα [0,1]. Οι συνήθεις τιμές για τον συντελεστή α είναι το 0.1 και το 0.3.

Η ανωτέρω εξίσωση συνήθως γράφεται στη μορφή:

$$\bar{Y}_t = \alpha Y_t + (1 - \alpha)\bar{Y}_{t-1}$$

1.4.5. Προσαρμογής της τάσης με τη χρήση δύο μέσων σημείων

Σύμφωνα με τη τεχνική αυτή, η τάση προσδιορίζεται με βάση τα ακόλουθα στάδια:

Α) Έστω ότι έχουμε τις παρατηρήσεις (t_1,y_1) , (t_2,y_2) , … , (t_n,y_n). Τότε χωρίζουμε τις παρατηρήσεις σε δύο ομάδες με βάση το n/2, δηλαδή:

- (t_1,y_1) , (t_2,y_2) , … , $(t_{n/2},y_{n/2})$ και
- $(t_{(n/2)+1},y_{(n/2)+1})$, $(t_{(n/2)+2},y_{(n/2)+2})$, … , (t_n,y_n).

Β) Στο δεύτερο στάδιο είναι εύρεση των σημείων (x'_1,y'_1) και (x'_2,y'_2),

όπου,

x'_1, x'_2 : οι αριθμητικοί μέσοι των χρονικών περιόδων $(t_1, t_2,\dots, t_{n/2})$ και $(t_{(n/2)+1},\dots, t_n)$ αντίστοιχα,

y_1, y_2 : οι αριθμητικοί μέσοι των παρατηρήσεων $(y_1, y_2,\dots, y_{n/2})$ και $(y_{(n/2)+1},\dots, y_n)$ αντίστοιχα.

Γ) Στο τελευταίο στάδιο περιγράφεται η εξίσωση της ευθείας που εκφράζει τη τάση της χρονολογικής σειράς και η οποία προσδιορίζεται από το τύπο:

$$\frac{x - x'_1}{x'_2 - x'_1} = \frac{\bar{y} - y'_1}{y'_2 - y'_1}$$

1.4.6. Προσδιορισμός της τάσης όταν ως καμπύλη τάσης λογίζεται η εκθετική καμπύλη

Υποθέτοντας ότι οι μεταβλητές t (χρόνος) και y (παρατήρηση τη χρονική στιγμή t), συνδέονται με την ακόλουθη εξίσωση μακροπρόθεσμης τάσης:

$$y = ae^{\beta t}$$

Προκειμένου να γίνει εκτίμηση των παραμέτρων, αρχικά λογαριθμίζουμε την ανωτέρω εξίσωση, λαμβάνοντας την έκφραση:

$$lny = lna + \beta t$$

Εν συνεχεία, θέτοντας Y = ln y και A = ln α, η ανωτέρω εξίσωση παίρνει τη μορφή:

$$Y = A + \beta t$$

Με βάση την αρχή των ελαχίστων τετραγώνων γίνεται η εκτίμηση των παραμέτρων A και β, η οποία προκύπτει από τη σχέσεις $\bar{A}$ = ln α και b = β.

Με αντικατάσταση στην αρχική εξίσωση καταλήγουμε στην εξίσωση της τάσης:

$$\bar{y} = e^{\bar{A}} \cdot e^{bt}$$

1.4.7. Προσδιορισμός της τάσης όταν ως καμπύλη τάσης λογίζεται η λογιστική καμπύλη

Υποθέτοντας ότι οι μεταβλητές t (χρόνος) και y (παρατήρηση τη χρονική στιγμή t), συνδέονται με την ακόλουθη εξίσωση μακροπρόθεσμης τάσης:

$$y = \frac{1}{1 + \alpha e^{\beta t}}$$

Προκειμένου να γίνει εκτίμηση των παραμέτρων, μετασχηματίζουμε την ανωτέρω σχέση:

$$1 - y = 1 - \frac{1}{1 + \alpha e^{\beta t}} = y \cdot \alpha \cdot e^{\beta t}$$

Οπότε καταλήγουμε στη μαθηματική έκφραση:

$$Y = A + \beta t$$

όπου, Α = ln α.

1.5. *Μέτρηση της Εποχικής Διακύμανσης*

Για να γίνει καλύτερα κατανοητός ο ρόλος της εποχικής διακύμανσης, πρέπει αρχικά να δοθεί ένας εννοιολογικός ορισμός για τον εποχικό κύκλο, ο οποίος είναι ένα σύνολο χρονικών περιόδων για τις οποίες η τιμή της χρονολογικής σειράς για μια συγκεκριμένη χρονική περίοδο, έχει σχέση με τη θέση της στο σύνολο των χρονικών περιόδων (πχ όταν ο εποχικός κύκλος είναι ο χρόνος και οι εποχές είναι οι μήνες, τότες έχουμε 12 περιόδους).

Για τη μέτρηση της εποχικής διακύμανσης, χρησιμοποιείται η μέθοδος *ratio to moving average*, η οποία έχει τα ακόλουθα στάδια.

Α) Υπολογισμός του 12μηνου κινητού μέσου για κάθε χρονική περίοδο.

Β) Υπολογισμός του πηλίκου προς τον κινητό μέσο που προκύπτει από τη διαίρεση του κινητού μέσου για κάθε χρονική περίοδο με την αρχική τιμή.

Γ) Με την ανωτέρω διαδικασία προκύπτουν οι «εποχικοί δείκτες» (seasonal indexes), οι οποίοι χρησιμοποιούνται για την απαλλαγή της χρονοσειράς από την εποχικότητα.

1.6. *Μέτρηση της Κυκλικής Διακύμανσης*

Στη περίπτωση όπου μια χρονολογική σειρά Y αποτελείται από ετήσια δεδομένα, υποθέτουμε ότι απουσιάζει η εποχική και η τυχαία διακύμανση και έτσι το πολλαπλασιαστικό μοντέλο παίρνει τη μορφή:

$$Y = T \cdot C$$

όπου,

T : η τάση και

C : η κυκλική διακύμανση

Από τον μετασχηματισμό της ανωτέρω εξίσωσης προκύπτει η εξίσωση ως προς το C, δηλαδή:

$$C = \frac{Y}{T}$$

Η σχέση αυτή υποδηλώνει ότι η κυκλική διακύμανση προκύπτει από τη διαίρεση των αρχικών τιμών της σειράς Y με τις αντίστοιχες τιμές της τάσης (οι οποίες βρίσκονται με βάση κάποια μέθοδο υπολογισμού, πχ αρχή ελαχίστων τετραγώνων).

Η τιμή που επιστρέφεται από τον υπολογισμό του πηλίκου (πολλαπλασιασμένη με το 100) είναι η κυκλική σχετική διακύμανση (cyclical relative).

ΚΕΦΑΛΑΙΟ 2: ΑΝΑΛΥΣΗ ΧΡΟΝΟΛΟΓΙΚΩΝ ΣΕΙΡΩΝ

2.1. Υποδείγματα

Με τον όρο χρονολογική σειρά (time series) εννοούμε μια ακολουθία τιμών μιας μεταβλητής ή μιας σύνθεσης μεταβλητών, οι οποίες έχουν ληφθεί σε διαδοχικές χρονικές περιόδους. Με άλλα λόγια, μια χρονολογική σειρά είναι ένα σύνολο παρατηρήσεων $y_1, y_2, \ldots y_T$ [1].

2.2. Στασιμότητα (Stationary)

Μια από τις σημαντικότερες έννοιες στην ανάλυση χρονολογικών σειρών, είναι αυτή της στασιμότητας. Νωρίτερα, κατά τον ορισμό της χρονολογικής σειράς, επισημάνθηκε ότι αυτή είναι ένα σύνολο παρατηρήσεων y_t μιας τυχαίας μεταβλητής Y, όπου κάθε μια λαμβάνεται μια συγκεκριμένη χρονική στιγμή t. Με άλλα λόγια η χρονολογική σειρά, μπορεί να θεωρηθεί ως η στοχαστική διαδικασία πεπερασμένου πλήθους παρατηρήσεων, δηλαδή η πραγματοποίηση μιας διαδικασίας $Y_1, Y_2, \ldots, Y_t$.

Για τον ορισμό ενός μοντέλου χρονολογικών σειρών για ένα σύνολο παρατηρήσεων y_t απαιτείται ο καθορισμός όλων των κοινών κατανομών μιας ακολουθίας Y_t, πραγματοποίηση των οποίων πρέπει να είναι οι παρατηρήσεις $\{ y_t \}$. Ο προσδιορισμός λοιπόν του μοντέλου, συνεπάγεται τον ορισμό όλων των κοινών κατανομών των $(Y_1, Y_2, \ldots, Y_n)'$, n = 1, 2, …, δηλαδή των πιθανοτήτων $P(Y_1 \leq y_1, Y_2 \leq y_2, \ldots, Y_n \leq y_n,)$, όπου $-\infty \leq y_1 \leq y_2 \leq \ldots \leq y_n \leq +\infty$, με n=1,2,..

Λόγω του ότι η διαδικασία που περιγράφεται ανωτέρω είναι εξαιρετικά δύσκολη, προτείνεται αντ' αυτής η χρήση ροπών μικρότερης τάξης και μια από τις σημαντικότερες υποθέσεις στη μελέτη ροπών είναι αυτή της στασιμότητας. Στο σημείο αυτό παρατίθενται κάποιοι χρήσιμοι ορισμοί σχετικά με την έννοια της στασιμότητας.

Μια χρονολογική σειρά καλείται στάσιμη, όταν δεν υπάρχει συστηματική αλλαγή του μέσου όρου και της διασποράς της στο χρόνο. Πρακτικά αυτό σημαίνει ότι στη περίπτωση που μια χρονολογική σειρά εμφανίζει τάση, τότε δεν μπορεί να είναι στάσιμη. Από την άλλη, μια στοχαστική διαδικασία είναι αυστηρώς ή πλήρως στάσιμη, όταν οι ιδιότητες της δεν επηρεάζονται από μια μεταβολή στην αρχή μέτρησης του χρόνου. Οι διαδικασίες αυτές, συνήθως καλούνται στάσιμες n-τάξης. Πρακτικά αυτό σημαίνει ότι η από κοινού συνάρτηση κατανομής με αρχή το χρονικό σημείο t είναι η ίδια με την από κοινού συνάρτηση κατανομής με αρχή το χρονικό σημείο t+s. Η ανωτέρω φράση διατυπώνεται καταλλήλως από τη σχέση:

$$F(y_t, y_{t+1}, \dots, y_{t+T}) = F(y_{t+s}, y_{t+1+s}, \dots, y_{t+T+s})$$

Όπως είναι εμφανές, το s αντιπροσωπεύει μια αυθαίρετη μετακίνηση κατά μήκος του άξονα του χρόνου (είτε αύξηση είτε μείωση).

Λόγω της δυσκολίας εφαρμογής της στασιμότητας που προκύπτει από τις συνθήκες που πρέπει να ικανοποιούνται, ο ορισμός της στασιμότητας έγινε πιο ευέλικτος προσδιορίζοντας αυτό το οποίο ονομάζουμε ασθενής στάσιμη διαδικασία n-τάξης, με την προϋπόθεση ότι υπάρχουν οι ροπές n-τάξεως και οι οποίες είναι ανεξάρτητες του t.

Διατυπώνοντας την ανωτέρω πρόταση σε μαθηματική μορφή, για να είναι μια διαδικασία n-τάξης ασθενής στάσιμη θα πρέπει να ισχύουν:

$$E(Y_t) = \mu$$

$$Var(Y_t) = \sigma^2$$

$$Cov(Y_t, Y_{t+s}) = Cov(Y_{t+m}, Y_{t+s+m}) = \gamma_s$$

Στη περίπτωση όπου οι ανωτέρω συνθήκες προσδιορίζουν μόνο τις πρώτες και δεύτερες ροπές, τότε προκύπτουν οι ασθενείς στάσιμες διαδικασίες 2ης τάξης (ή κατά συνδιακύμανση στάσιμες).

2.3. ***Συνάρτηση Αυτοσυσχέτισης***

Ο συντελεστής συσχέτισης ρ, όπως είναι γνωστό, αποτελεί ένα μέτρο για το βαθμό της σχέσης μεταξύ δύο μεταβλητών. Οι τιμές οι οποίες δέχεται ανήκουν στο κλειστό διάστημα [-1,1] και ανάλογα με το ποια είναι η ακριβής τιμή του, προσδιορίζεται και η συσχέτιση. Συγκεκριμένα ισχύουν τα εξής:

- Αν ρ > 0, τότε υπάρχει θετική συσχέτιση (όσο πιο κοντά είναι η τιμή του ρ στο 1, τόσο πιο ισχυρή είναι η συσχέτιση).
- Αν ρ < 0, τότε υπάρχει αρνητική συσχέτιση (όσο πιο κοντά είναι η τιμή του ρ στο -1, τόσο πιο ισχυρή είναι η συσχέτιση).
- Αν ρ = 1 ή ρ = -1, τότε υπάρχει η μέγιστη δυνατή συσχέτιση (θετική στη περίπτωση που ρ = 1 και αρνητική στη περίπτωση που ρ = -1).
- Αν ρ = 0, τότε δεν υπάρχει συσχέτιση.

Κατά την ανάλυση χρονολογικών σειρών, ο συντελεστής συσχέτισης μεταξύ δύο μεταβλητών ονομάζεται συντελεστής αυτοσυσχέτισης (ACF) και εκφράζεται από τη συνάρτηση:

$$\rho_s(Y) = \frac{\gamma_s(Y)}{\gamma_\rho(Y)} = \frac{\gamma_s}{\gamma_\rho} = \rho_s\,, \qquad \acute{o}\pi o\upsilon$$

$$\rho_s = (\sum_{i=1}^{N-s}(Y(t_i) - \mu_Y)\,(Y(t_{i+s}) - \mu_Y))/\sum_{i=1}^{N}(Y(t_i) - \mu_Y)^2)$$

2.4. *Τελεστής Υστέρησης*

Πολλές φορές στην ανάλυση χρονοσειρών ανακύπτει πλήθος αλγεβρικών πράξεων που πρέπει να υπολογιστούν. Σε πολλές περιπτώσεις η χρήση του τελεστή υστέρησης L (Lag operator), είναι σημαντικά βοηθητική ως προς αυτή την κατεύθυνση.

Ο τελεστής υστέρησης εκφράζεται από τον ακόλουθο τύπο:

$$Y_{t-s} = L^s Y_t$$

όπου ό εκθέτης s του L, υποδηλώνει το πλήθος των υστερήσεων της μεταβλητής Y.

Για παράδειγμα :

$1^η$ υστέρηση: Y_{t-1} **à** LYt

2η υστέρηση: Y_{t-2} **à** L^2Yt

3η υστέρηση: Y_{t-3} **à** L^3Yt

κτλ

ΚΕΦΑΛΑΙΟ 3: ΑΥΤΟΠΑΛΙΝΔΡΟΜΗ ΔΙΑΔΙΚΑΣΙΑ

Στη γενική του μορφή, ένα αυτοπαλίνδρομο υπόδειγμα p τάξης συμβολίζεται με AR(p) και εκφράζεται από τη σχέση [3]:

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

όπου,

α_i, i = 0 (1) p : οι σταθεροί παράμετροι και

ε_t : ο λευκός θόρυβος (white noise), ο οποίος μετρά τυχαία σφάλματα.

Οι όροι του ανωτέρω υποδείγματος είναι ανεξάρτητες τυχαίες μεταβλητές με σταθερή διακύμανση και μέσο ίσο με το μηδέν. Στο αυτοπαλίνδρομο υπόδειγμα, η εξαρτημένη μεταβλητή Y_t παλινδρομείται στις προηγούμενες τιμές της. Η τάξη του αυτοπολίνδρομου υποδείγματος συμβολίζεται με p και προσδιορίζει το μήκος της υστέρησης, ενώ τα Y_{t-1}, Y_{t-2}, …, Y_{t-p} είναι οι τιμές της χρονοσειράς με υστέρηση.

3.1. Αυτοπαλίνδρομα Υποδείγματα Πρώτης Τάξης – AR(1)

Από τη γενική μορφή των αυτοπαλίνδρομων υποδειγμάτων που αναφέρθηκε ανωτέρω, προκύπτει εύκολα η μορφή του αυτοπαλίνδρομου υποδείγματος πρώτης τάξης, η οποία είναι:

$$Y_t = a_0 + a_1 Y_{t-1} + \varepsilon_t$$

Με βάση την υπόθεση ότι η χρονοσειρά είναι στάσιμη, τότε για κάθε χρονικό σημείο t θα ισχύει μ = $E(Y_t)$ και $E(\varepsilon_t) = 0$, οπότε το Y_t μπορεί να εκφραστεί σε αποκλίσεις από το μέσο του ως εξής:

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

Αντικαθιστώντας στην ανωτέρω εξίσωση τα μ = $E(Y_t)$ και $E(\varepsilon_t) = 0$ προκύπτει:

$$\mu = \alpha_0 + \alpha_1 \mu \Leftrightarrow \mu = \frac{\alpha_0}{1 - \alpha_1}$$

Αφαιρώντας από τη γενική μορφή του αυτοπαλίνδρομου υποδείγματος τον μέσο μ (και από τα δύο μέλη) και αντικαθιστώντας σύμφωνα με την ανωτέρω εξίσωση (εκφρασμένη ως προς το $α_0$), προκύπτει:

$$Y_t - \mu = a_0 + a_1 Y_{t-1} + \varepsilon_t - \mu \Leftrightarrow Y_t - \mu = \mu - \alpha_1 \mu + a_1 Y_{t-1} + \varepsilon_t - \mu \Leftrightarrow Y_t - \mu = a_1 (Y_{t-1} - \mu) + \varepsilon_t$$

Για $y_t = Y_t - \mu$, $t=1(1)n$, η ανωτέρω εξίσωση γίνεται:

$$y_t = a_1 y_{t-1} + \varepsilon_t$$

Στη περίπτωση του υποδείγματος AR(1), για να είναι στάσιμη η χρονοσειρά θα πρέπει να ισχύει ότι $|α_1| < 1$.

Για $|α_1| < 1$, η διασπορά της διαδικασίας $\{ y_t \}$, δίνεται από τον τύπο:

$$\gamma_0 = V(y_t) = \frac{\sigma^2}{1 - \alpha_1^2}$$

Αντίστοιχα η συνδιακύμανση (ή αυτοδιακύμανση) δίνεται από τον τύπο:

$$\gamma_s = Cov(y_t, y_{t-s})$$

Τέλος ο συντελεστής αυτοσυσχέτισης δίνεται από την έκφραση:

$$\rho_s = \frac{\gamma_s}{\gamma_0} = \alpha_1^s$$

Λόγω του ότι $γ_0 \neq 0$ και επίσης δεν μπορεί να λάβει αρνητικές τιμές, για να είναι η χρονοσειρά στάσιμη θα πρέπει να ισχύει $|α_1| < 1$.

Επίσης, για $α_1 > 0$, η συνάρτηση αυτοσυσχέτισης φθίνει γεωμετρικά (λαμβάνοντας ως αρχική τιμή τη μονάδα) και τείνει προς το 0 καθώς το s αυξάνει. Αντίστοιχη είναι και η συμπεριφορά της συνάρτησης αυτοσυσχέτισης για $α_1 < 0$, με τη διαφορά ότι έχει αρνητικό πρόσημο.

3.2. *Αυτοπαλίνδρομα Υποδείγματα Δεύτερης Τάξης* $-AR(2)$

Από τη γενική μορφή των αυτοπαλίνδρομων υποδειγμάτων και σύμφωνα με όσα αναφέρθηκαν νωρίτερα, προκύπτει εύκολα η μορφή του αυτοπαλίνδρομου υποδείγματος δεύτερης τάξης, η οποία είναι:

$$Y_t = a_0 + a_1 Y_{t-1} + a_2 Y_{t-2} + \varepsilon_t$$

Με βάση την υπόθεση ότι η χρονοσειρά είναι στάσιμη, τότε για κάθε χρονικό σημείο t θα ισχύει $\mu = E(Y_t)$ και $E(\varepsilon_t) = 0$, οπότε το Y_t μπορεί να εκφραστεί σε αποκλίσεις από το μέσο του. Οπότε μια χρονολογική σειρά Y_t με μέσο $\mu = 0$ είναι:

$$y_t = a_1 y_{t-1} + a_2 y_{t-2} + \varepsilon_t$$

Ο μέσο μ για το αυτοπαλίνδρομο υπόδειγμα AR(2), δίνεται από την εξίσωση:

$$\mu = \frac{\alpha_0}{1 - \alpha_1 - \alpha_2}$$

Η διασπορά της διαδικασίας { y_t }, δίνεται από τον τύπο:

$$\gamma_0 = Var(Y_t) = a_1 \gamma_1 + \alpha_2 \gamma_2 + \sigma^2$$

Αντίστοιχα η συνδιακύμανση (ή αυτοδιακύμανση) δίνεται από τον τύπο:

$$\gamma_s = Cov(y_t, y_{t-s}) = a_1 \gamma_{s-1} + \alpha_2 \gamma_{s-2}$$

Τέλος ο συντελεστής αυτοσυσχέτισης δίνεται από την έκφραση:

$$\rho_s = a_1 \rho_{s-1} + \alpha_2 \rho_{s-2}, \qquad s > 0$$

3.3. *Αυτοπαλίνδρομα Υποδείγματα Τάξης* $p - AR(p)$

Στη γενική του μορφή, ένα αυτοπαλίνδρομο υπόδειγμα p τάξης συμβολίζεται με AR(p) και εκφράζεται από τη σχέση:

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

Για μια αυτοπαλίνδρομη διαδικασία AR(p) ισχύουν οι ακόλουθες συνθήκες:

$$\gamma_0 = Var(Y_t) = a_1\gamma_1 + \alpha_2\gamma_2 + \cdots + \alpha_p\gamma_p + \sigma^2$$

$$\gamma_s = Cov(y_t, y_{t-s}) = a_1\gamma_{s-1} + \alpha_2\gamma_{s-2} + \cdots + \alpha_p\gamma_{s-p}, \qquad s > 0$$

$$\rho_s = a_1\rho_{s-1} + \alpha_2\rho_{s-2} + \cdots + \alpha_p\rho_{s-p}, \qquad s > 0$$

Από τη τελευταία ισότητα προκύπτουν p εξισώσεις (Yule-Walker) για s = 1, 2, …, p, οι οποίες δίνονται ακολούθως:

$$\text{Για } s = 1: \qquad \rho_1 = a_1 + \alpha_2\rho_1 + \alpha_3\rho_2 + \cdots + \alpha_p\rho_{p-1}$$

$$\text{Για } s = 2: \qquad \rho_2 = a_1\rho_1 + \alpha_2 + \alpha_3\rho_1 + \cdots + \alpha_p\rho_{p-2}$$

…. …. …. …. …. … …

$$\text{Για } s = p: \qquad \rho_p = a_1\rho_{p-1} + \alpha_2\rho_{p-2} + \alpha_3\rho_{p-3} + \cdots + \alpha_p$$

Σύμφωνα με τα παραπάνω, είναι εμφανές ότι η συνάρτηση αυτοσυσχέτισης ενός αυτοπαλίνδρομου υποδείγματος AR(p), καθορίζεται από τις τιμές των παραμέτρων α_i, για i=1(1)p.

Οι παραπάνω εξισώσεις μπορούν να αναπαραστούν με τη μορφή πινάκων όπως ακολούθως:

$$\begin{bmatrix} \rho_1 \\ \rho_2 \\ \vdots \\ \rho_p \end{bmatrix} = \begin{bmatrix} 1 & \rho_1 & \cdots & a_1 \\ \rho_1 & 1 & \cdots & a_2 \\ \vdots & \vdots & \vdots & \vdots \\ \rho_{p-1} & \rho_{p-2} & \cdots & 1 \end{bmatrix}$$

3.4. *Συνάρτηση Μερικής Αυτοσυσχέτισης για Αυτοπαλίνδρομα Υποδείγματα AR(p)*

Πολλές φορές κατά την ανάλυση αυτοπαλίνδρομων υποδειγμάτων, είναι αρκετά δύσκολο να προσδιοριστεί με ακρίβεια η τάξη του υποδείγματος που περιγράφει τη σειρά με βάση τη συνάρτηση αυτοσυσχέτισης. Αυτό οφείλεται στο γεγονός ότι οι αυτοπαλίνδρομες διαδικασίες έχουν συναρτήσεις αυτοσυσχέτισης οι οποίες βαίνουν φθίνουσες, καθώς αυξάνει το μήκος της υστέρησης. Γι' αυτόν ακριβώς το λόγο, χρησιμοποιείται ένα επιπλέον μέτρο που είναι η συνάρτηση μερικής αυτοσυσχέτισης (Partial Autocorellation Function – PACF).

Ως μερική αυτοσυσχέτιση μεταξύ δύο μεταβλητών Y_t και Y_{t-s}, αναφέρεται η συσχέτιση μεταξύ των Y_t και Y_{t-s}, απαλλαγμένη από τις γραμμικές επιδράσεις των ενδιάμεσων μεταβλητών Y_{t-1}, Y_{t-2}, …, $Y_{t-(s-1)}$. Ο συντελεστής αυτοσυσχέτισης τάξεως s μεταξύ των Y_t και Y_{t-s} (για s=1(1)p) συμβολίζεται με ρ_{ss} και είναι ο συντελεστής μερικής παλινδρόμησης της μεταβλητής y_{t-s} στο υπόδειγμα

$$y_t = \rho_{1s} y_{t-1} + \rho_{2s} y_{t-2} + \cdots + \rho_{ss} y_{t-s} + \varepsilon_t$$

Όπως παρατηρούμε, στον μερικό συντελεστή αυτοσυσχέτισης υπάρχουν δύο δείκτες. Ο πρώτος (αριστερά) δηλώνει τη χρονική υστέρηση της μεταβλητής, ενώ ο δεύτερος (δεξιά) τη μέγιστη τάξη της παλινδρόμησης. Όπως είναι λογικό, για τα αυτοπαλίνδρομα υποδείγματα AR(p) ισχύει ότι για s > p, $\rho_{ss} = 0$. Οι σχέσεις που δημιουργούνται με βάση αυτή τη συνθήκη δίνονται ακολούθως:

<u>Υπόδειγμα AR(1)</u>

$\rho_{11} = \rho_1 = \alpha_1$

$\rho_{ss} = 0, \quad s > 1$

<u>Υπόδειγμα AR(2)</u>

$\rho_{11} = \rho_1$

$\rho_{22} = \dfrac{\rho_2 - \rho_1^2}{1 - \rho_1^2}$

$\rho_{ss} = 0, \quad s > 2$

<u>Υπόδειγμα AR(p)</u>

$\rho_{11} = \rho_1$

$\rho_{22} \neq 0, \dots, \rho_{pp} \neq 0$

$\rho_{ss} = 0, \quad s > 2$

3.5. *Έλεγχος Στατιστικής Σημαντικότητας*

Για τον έλεγχο της στατιστικής σημαντικότητας χρησιμοποιούνται οι τιμές που επιστρέφουν οι συναρτήσεις της αυτοσυσχέτισης και της μερικής αυτοσυσχέτισης. Πρακτικά επειδή οι πραγματικές τιμές αυτών των συναρτήσεων δεν είναι γνώστες, χρησιμοποιούνται οι αντίστοιχες τιμές που έχουν προκύψει ως εκτιμήσεις τους από το δείγμα [1,3].

Για δείγματα με μεγάλο πλήθος τιμών, οι εκτιμήσεις των αυτοσυσχετίσων έχουν κανονική κατανομή, με μέση τιμή μηδέν και διακύμανση 1/T, όπου T είναι το μέγεθος του δείγματος. Ανάλογη είναι και η συμπεριφορά των εκτιμήσεων των μερικών αυτοσυσχετίσεων για υστερήσεις μεγαλύτερης από τη τάξης p της διαδικασίας AR.

Για να πραγματοποιηθεί ο έλεγχος στατιστικής σημαντικότητας του συντελεστή αυτοσυσχέτισης ρ_s, ελέγχουμε τις υποθέσεις H_0 και H_1:

$$H_0 : \hat{\rho}_s = 0$$
$$H_1 : \hat{\rho}_s \neq 0$$

όπου $\hat{\rho}_s$, ο εκτιμητής του συντελεστή αυτοσυσχέτισης.

Η πραγμάτωση του ελέγχου γίνεται με τη χρήση της στατιστικής

$$t_s = \frac{\hat{\rho}_s}{\sqrt{\frac{1}{T}}} = \hat{\rho}_s\sqrt{T}$$

Με βάση αυτά, για δεδομένο επίπεδο σημαντικότητας α = 5%, αν ισχύει $|t_s| > 2$ η μηδενική υπόθεση (H_0) απορρίπτεται και γίνεται αποδεκτή η H_1, ενώ σε διαφορετική περίπτωση αποδεκτή γίνεται η μηδενική υπόθεση. Το 95% διάστημα εμπιστοσύνης για αυτόν τον έλεγχο είναι:

$$\hat{\rho}_s - \frac{2}{\sqrt{T}} \leq \rho_s \leq \hat{\rho}_s$$

Ανάλογος είναι ο τρόπος ελέγχου σημαντικότητας και για τους συντελεστές μερικής αυτοσυσχέτισης. Σε αυτή τη περίπτωση ο μερικός συντελεστής αυτοσυσχέτισης είναι σημαντικός εφόσον ισχύει $\left|\hat{\rho}_{ss}\sqrt{T}\right| < 2$.

Ο έλεγχος σημαντικότητας των συντελεστών της μερικής αυτοσυσχέτισης μπορεί να προσδιορίσει την αλγεβρική τάξη μιας αυτοπαλίνδομης διαδικασίας (AR). Με βάση τη τελευταία σημαντική τιμή του t_s επιλέγεται η τάξη της σειράς, πχ αν η τελευταία σημαντική τιμή του t_s προκύπτει για s = 3, δηλαδή ο τελευταίος στατιστικά συντελεστής είναι ο ρ_{33}, τότε εξάγεται το συμπέρασμα ότι του αυτοπαλίνδρομο υπόδειγμα είναι $3^{ης}$ τάξης ή αλλιώς AR(3).

ΚΕΦΑΛΑΙΟ 4: ΔΙΑΔΙΚΑΣΙΕΣ ΚΙΝΗΤΟΥ ΜΕΣΟΥ – MA(q)

Για την περιγραφή φαινομένων, όπου τα γεγονότα παράγουν ένα άμεσο αποτέλεσμα που η επίδρασή του συνεχίζεται και μετά τη παύση του γεγονότος, γίνεται συχνά η χρήση υποδειγμάτων κινητού μέσου [1].

Στη γενική του μορφή, ένα υπόδειγμα κινητού μέσου q τάξης συμβολίζεται με MA(q) και εκφράζεται από τη σχέση:

$$Y_t = \mu + \varepsilon_t - \theta_1\varepsilon_{t-1} - \theta_2\varepsilon_{t-2} - \cdots - \theta_q\varepsilon_{t-q}$$

όπου,

θ_i, i = 0 (1) q : οι σταθεροί παράμετροι και

ε_t : ο λευκός θόρυβος (white noise), ο οποίος μετρά τυχαία σφάλματα.

Στις διαδικασίες κινητού μέσου – MA(q), η χρονολογική σειρά Y_t θεωρείται ότι δημιουργείται ως ένας σταθμικός μέσος τυχαίων σφαλμάτων των q προηγούμενων περιόδων.

4.1. Διαδικασίες Κινητού Μέσου Πρώτης Τάξης – MA(1)

Από τη γενική μορφή των υποδειγμάτων κινητού μέσου που αναφέρθηκε ανωτέρω, προκύπτει εύκολα η μορφή του υποδείγματος κινητού μέσου πρώτης τάξης, η οποία είναι:

$$Y_t = \varepsilon_t - \theta_1\varepsilon_{t-1}$$

Στο συγκεκριμένο υπόδειγμα ισχύουν οι ακόλουθες συνθήκες:

Αρχικά, ισχύει ότι $E(Y_t) = 0$.

Επίσης η διασπορά της διαδικασίας $\{ y_t \}$, δίνεται από τον τύπο:

$$\gamma_0 = Var(Y_t) = \sigma_\varepsilon^2(1 + \theta_1^2)$$

Αντίστοιχα η συνδιακύμανση (ή αυτοδιακύμανση) δίνεται από τον τύπο:

$$\gamma_k = Cov(y_t, y_{t-k}) = \begin{cases} -\theta_1 \sigma_\varepsilon^2, & k = 1 \\ 0, & k > 1 \end{cases}$$

Τέλος ο συντελεστής αυτοσυσχέτισης δίνεται από την έκφραση:

$$\rho_k = \frac{\gamma_k}{\gamma_0} = \begin{cases} -\dfrac{\theta_1}{1 + \theta_1^2}, & k = 1 \\ 0, & k > 1 \end{cases}$$

Από τις ανωτέρω εκφράσεις και συγκεκριμένα από τις συναρτήσεις που αφορούν τις αυτοσυνδιακυμάνσεις και κατ' επέκταση και τις αυτοσυσχετίσεις, γίνεται αντιληπτό ότι αυτές θα έχουν τιμή 0 για όλα τα k, εκτός από k = 1. Πρακτικά αυτό συνεπάγεται ότι μια παρατήρηση της χρονοσειράς Y_t, σχετίζεται αποκλειστικά είτε με την επόμενη είτε με τη προηγούμενη παρατήρηση και με καμία άλλη.

Σε αντίθεση με τα αυτοπαλινδρομα υποδείγματα – AR(p), όπου η συνάρτηση αυτοσυσχέτισης ρ_s, φθίνει γεωμετρικά προς το μηδέν (αλλά δεν μηδενίζεται), στις διαδικασίες κινητών μέσων – MA(q) η συνάρτηση αυτοσυσχέτισης ρ_k μηδενίζεται μετά από q χρονικές υστερήσεις, δηλαδή:

$$\rho_k = 0, \qquad \gamma\iota\alpha\ k > q$$

4.2. *Διαδικασίες Κινητού Μέσου Δεύτερης Τάξης – MA(2)*

Από τη γενική μορφή των υποδειγμάτων κινητού μέσου που αναφέρθηκε ανωτέρω, προκύπτει εύκολα η μορφή του υποδείγματος κινητού μέσου πρώτης τάξης, η οποία είναι:

$$Y_t = \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2}$$

Στο συγκεκριμένο υπόδειγμα ισχύουν οι ακόλουθες συνθήκες:

Αρχικά, ισχύει ότι $E(Y_t) = 0$.

Επίσης η διασπορά της διαδικασίας $\{ y_t \}$, δίνεται από τον τύπο:

$$\gamma_0 = Var(Y_t) = \sigma_\varepsilon^2 (1 + \theta_1^2 + \theta_2^2)$$

Αντίστοιχα η συνδιακύμανση (ή αυτοδιακύμανση) δίνεται από τον τύπο:

$$\gamma_k = Cov(y_t, y_{t-k}) = \begin{cases} -\theta_1(1-\theta_2)\sigma_\varepsilon^2, & k = 1 \\ -\theta_2\sigma_\varepsilon^2, & k = 2 \\ 0, & k > 2 \end{cases}$$

Τέλος ο συντελεστής αυτοσυσχέτισης δίνεται από την έκφραση:

$$\rho_k = \frac{\gamma_k}{\gamma_0} = \begin{cases} -\dfrac{\theta_1(1-\theta_2)}{1+\theta_1^2+\theta_2^2}, & k = 1 \\ -\dfrac{\theta_2}{1+\theta_1^2+\theta_2^2}, & k = 2 \\ 0, & k > 2 \end{cases}$$

Αντίστοιχα με τη περίπτωση της 1ης τάξης, έτσι και στο υπόδειγμα κινητού μέσου 2ης τάξης – MA(q), οι συναρτήσεις που αφορούν τις αυτοσυνδιακυμάνσεις και κατ' επέκταση και τις αυτοσυσχετίσεις, έχουν τιμή 0 για όλα τα k, εκτός από k = 1, 2. Πρακτικά αυτό συνεπάγεται ότι η Y_t επηρεάζεται μόνο από τις Y_{t-1} και Y_{t-2}.

4.3. *Διαδικασίες Κινητού Μέσου q Τάξης – MA(q)*

Στη γενική του μορφή, ένα υπόδειγμα κινητού μέσου q τάξης συμβολίζεται με MA(q) και εκφράζεται από τη σχέση:

$$Y_t = \varepsilon_t - \theta_1\varepsilon_{t-1} - \theta_2\varepsilon_{t-2} - \cdots - \theta_q\varepsilon_{t-q}$$

Για μια διαδικασία κινητού μέσου MA(q) ισχύουν οι ακόλουθες συνθήκες:

$$\mu = E(Y_t) = 0$$

$$\gamma_0 = Var(Y_t) = \sigma_\varepsilon^2(1+\theta_1^2+\theta_2^2+\ldots+\theta_q^2)$$

$$\gamma_k = Cov(y_t, y_{t-k}) = (-\theta_k + \theta_1\theta_{k+1} + \theta_2\theta_{k+2} + \cdots + \theta_q\theta_{k+q})\sigma_\varepsilon^2 = 0$$

$$\rho_k = \frac{\gamma_k}{\gamma_0} = \frac{-\theta_k + \theta_1\theta_{k+1} + \theta_2\theta_{k+2} + \cdots + \theta_q\theta_{k+q}}{1+\theta_1^2+\theta_2^2+\ldots+\theta_q^2}$$

Επίσης ισχύουν και οι ακόλουθες συνθήκες:

$$\varphi_{11} = \rho_1$$

$$\varphi_{22} = \frac{\begin{bmatrix} 1 & \rho_1 \\ \rho_1 & \rho_2 \end{bmatrix}}{\begin{bmatrix} 1 & \rho_1 \\ \rho_1 & 1 \end{bmatrix}} = \frac{\rho_2 - \rho_1^2}{1 - \rho_1^2}$$

$$\varphi_{33} = \frac{\rho_1^3}{1 - 2\rho_1^2}$$

4.4. Αντιστρεψιμότητα

Η έννοια της αντιστρεψιμότητας στις διαδικασίες κινητού μέσου, συνδέεται με τη μετατροπή ενός υποδείγματος κινητού μέσου – MA(q) σε αυτοπαλίνδρομο υπόδειγμα άπειρης τάξης – AR(∞). Αντίστοιχα, αντιστρέψιμο είναι ένα αυτοπαλίνδρομο υπόδειγμα – AR(p) αν μπορεί μετατραπεί σε υπόδειγμα κινητού μέσου άπειρης τάξης – MA(∞).

Για να ισχύει η αντιστρεψιμότητα σε υποδείγματα κινητού μέσου – MA(q), θα πρέπει να πληρούνται οι ακόλουθες συνθήκες:

<u>Υποδείγματα MA(1):</u>

$$|\theta_1| < 1$$

<u>Υποδείγματα MA(2):</u>

$$\theta_1 + \theta_2 < 1$$

$$\theta_2 - \theta_1 < 1$$

$$|\theta_2| < 1$$

<u>Υποδείγματα MA(q):</u>

Οι ρίζες του πολυωνύμου Θ(L) = 0, πρέπει να βρίσκονται έξω από το μοναδιαίο κύκλο.

ΚΕΦΑΛΑΙΟ 5: **ΜΕΙΚΤΕΣ ΔΙΑΔΙΚΑΣΙΕΣ** – ARMA(p,q)

Στο κεφάλαιο αυτό μελετώνται οι μικτές διαδικασίες – ARMA(p,q), οι οποίες προκύπτουν από τον συνδυασμό των αυτοπαλίνδρομων διαδικασιών AR(p) και των διαδικασιών κινητού μέσου – MA(q). Τέτοιου είδους υποδείγματα μελετώνται όταν κατά τον έλεγχο στασιμότητας μιας χρονοσειράς παρατηρείται ότι τα δεδομένα της έχουν συναρτήσεις αυτοσυσχέτισης ή μερικής αυτοσυσχέτισης που φθίνουν με αργό ρυθμό αλλά δεν φαίνεται σαφώς να μηδενίζονται μετά από κάποιο σημείο.

Στη γενική του μορφή, ένα υπόδειγμα ARMA(p,q) προσδιορίζεται από το συνδυασμό των υποδειγματών AR(p) και MA(q). Πρακτικά αυτό σημαίνει ότι το μοντέλο θα εμπεριέχει τόσο τους όρους ενός αυτοπαλίνδρομου υποδείγματος όσο και ενός υποδείγματος κινητού μέσου. Κάπως έτσι προκύπτει η γενική μορφή του μικτού υποδείγματος – ARMA(p,q) που δίνεται από την ακόλουθη έκφραση:

$$Y_t = a_0 + a_1 Y_{t-1} + a_2 Y_{t-2} + \cdots + a_p Y_{t-p} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \cdots - \theta_q \varepsilon_{t-q}$$

όπου,

α_i, i = 0 (1) p : σταθεροί παράμετροι,

θ_i, i = 0 (1) q : σταθεροί παράμετροι και

ε_t : ο λευκός θόρυβος (white noise), ο οποίος μετρά τυχαία σφάλματα.

Από τη γενική μορφή του μικτού υποδείγματος, είναι προφανές ότι τόσο ένα αυτοπαλίνδρομο υπόδειγμα όσο και ένα υπόδειγμα κινητού μέσου, μπορούν να θεωρηθούν ως υποδείγματα ARMA όπου θα ισχύει AR(p) = ARMA(p,0) και MA(q) = ARMA(0,q) αντίστοιχα.

*5.1. **Μικτές Διαδικασίες Πρώτης Τάξης** – ARMA(1,1)*

Από τη γενική μορφή των μικτών διαδικασιών – ARMA(p,q) προκύπτει εύκολα η μικτή διαδικασία ARMA(1,1) που είναι και η πιο απλή περίπτωση των συγκεκριμένων υποδειγμάτων.

Έτσι λοιπόν για p = 1 και q = 1, προκύπτει το μικτό υπόδειγμα ARMA(1,1) που εκφράζεται από τον τύπο:

$$Y_t = a_0 + a_1 Y_{t-1} + \varepsilon_t - \theta_1 \varepsilon_{t-1}$$

Το μικτό υπόδειγμα ARMA(1,1) μπορεί να διατυπωθεί τόσο ως μια διαδικασία κινητού μέσου όσο και ως μια αυτοπαλίνδρομη διαδικασία άπειρης τάξης.

Για τη μετατροπή της μικτής διαδικασίας σε διαδικασία κινητού μέσου, αρχικά χρησιμοποιούμε τη διαδοχική υστέρηση της ανωτέρω εξίσωσης (που αντιπροσωπεύει το μοντέλο ARMA(1,1)) και αντικαθιστούμε τους όρους Y_{t-i}, καταλήγοντας στην σχέση:

$$Y_t = a_0 \sum_{i=0}^{\infty} a_1^i + \varepsilon_t + \sum_{i=0}^{\infty} a_1^i (a_1 + \theta_1) \varepsilon_{t-1-i}$$

Για να ισχύει η στασιμότητα στην ανωτέρω σειρά, θα πρέπει το άθροισμα που περιγράφεται ακολούθως να συγκλίνει:

$$\sum_{i=0}^{\infty} a_1^i (a_1 + \theta_1)$$

Για να συγκλίνει το ανωτέρω άθροισμα θα πρέπει να ισχύει ότι $|\alpha_1| < 1$. Στη περίπτωση αυτή η σειρά είναι στάσιμη και το υπόδειγμα εκφράζεται από τον τύπο:

$$Y_t = \frac{\alpha_0}{1 - \alpha_1} + \varepsilon_t + \sum_{i=0}^{\infty} a_1^i (a_1 + \theta_1) \varepsilon_{t-1-i}$$

όπου,

$$\sum_{i=0}^{\infty} a_1^i = \frac{\alpha_0}{1 - \alpha_1}$$

Ο τύπος της σειράς που προκύπτει, φανερώνει μια διαδικασία κινητού μέσου άπειρης τάξης (είναι εμφανές από τους άπειρους όρους που περιέχει η συνάρτηση), η οποία θα μπορούσε να προσεγγιστεί με έναν πεπερασμένο αριθμό περιορισμένων όρων αφού η σημαντικότητα των συντελεστών είναι φθίνουσα. Το γεγονός αυτό, πρακτικά σημαίνει ότι από ένα σημείο και πέρα οι όροι μπορούν να παραλειφθούν.

Τα παραπάνω, δίνουν μια σαφή απόδειξη για τη σπουδαιότητα των μικτών υποδειγμάτων, καθώς θα απαιτούνταν μια πολύ υψηλής τάξης διαδικασία κινητού μέσου για να προσδιοριστεί το μοντέλο ARMA(1,1), το οποίο περιέχει μόνο δύο συντελεστές.

Αντίστοιχη είναι και η διαδικασία που χρησιμοποιείται προκειμένου το μικτό υπόδειγμα ARMA(1,1), να διατυπωθεί ως αυτοπαλίνδρομο υπόδειγμα άπειρης τάξης.

Σύμφωνα λοιπόν και με τη προηγούμενη προσέγγιση, αντικαθιστώνας τα ε_{t-i}, προκύπτει η έκφραση:

$$Y_t = \frac{\alpha_0}{1-\theta_1} + \varepsilon_t + \sum_{i=0}^{\infty} \theta_1^i (a_1 + \theta_1) Y_{t-1-i}$$

Για να είναι δυνατή η αντιστρεψιμότητα της ανωτέρω σειράς θα πρέπει να ισχύει η ανισότητα $|\theta_1| < 1$.

Στο σημείο αυτό παρατίθενται οι συνθήκες που πρέπει να ικανοποιούνται για το μικτό υπόδειγμα – ARMA(1,1):

$$\mu = E(Y_t) = \frac{a_0}{1-a_1}$$

$$\gamma_0 = \frac{1+\theta_1^2 + 2\alpha_1\theta_1\sigma^2}{1-\alpha_1^2}$$

$$\gamma_1 = \alpha_1\gamma_0 + \theta_1\sigma^2$$

$$\gamma_s = \alpha_1\gamma_{s-1}, \qquad s > 1$$

$$\rho_1 = \alpha_1 \frac{\theta_1}{\gamma_0}\sigma^2 = \frac{(1+\alpha_1\theta_1)(\alpha_1+\theta_1)}{1+\theta_1^2+2\alpha_1\theta_1}$$

$$\rho_s = a_1\rho_{s-1}, \qquad s > 1$$

Τα συμπεράσματα που εξάγονται από τις ανωτέρω εκφράσεις μας δίνουν κάποια σημαντικά στοιχεία για τη συμπεριφορά του μικτού υποδείγματος. Αρχικά, από τη συνάρτηση αυτοσυσχέτισης είναι εμφανές ότι υπεισέρχεται ο συντελεστής από τη διαδικασία κινητού μέσου πρώτης τάξης – MA(1), μόνο όμως για την αυτοσυσχέτιση πρώτης τάξης (ρ_1). Οι υπόλοιπες αυτοσυσχετίσεις εξαρτώνται αποκλειστικά από την αυτοπαλίνδρομη διαδικασία που ενυπάρχει στο μικτό υπόδειγμα.

Επίσης, η συνάρτηση αυτοσυσχέτισης του μικτού υποδείγματος ARMA(1,1) φθίνει γεωμετρικώς καθώς το s αυξάνει. Αντίστοιχα, η συνάρτηση αυτοσυσχέτισης του υποδείγματος ARMA(1,1) φθίνει γεωμετρικώς (όπως και στη περίπτωση του υποδείγματος κινητού μέσου πρώτης τάξης).

*5.2. **Μικτές Διαδικασίες** – ARMA(p,q)*

Όπως τονίστηκε και νωρίτερα, στη γενική του μορφή, ένα υπόδειγμα ARMA(p,q) προσδιορίζεται από το συνδυασμό των υποδειγματών AR(p) και MA(q). Πρακτικά αυτό σημαίνει ότι το μοντέλο θα εμπεριέχει τόσο τους όρους ενός αυτοπαλίνδρομου υποδείγματος όσο και ενός υποδείγματος κινητού μέσου. Κάπως έτσι προκύπτει η γενική μορφή του μικτού υποδείγματος – ARMA(p,q) που δίνεται από την ακόλουθη έκφραση:

$$Y_t = a_0 + a_1 Y_{t-1} + a_2 Y_{t-2} + \cdots + a_p Y_{t-p} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \cdots - \theta_q \varepsilon_{t-q}$$

όπου,

α_i, i = 0 (1) p : σταθεροί παράμετροι,

θ_i, i = 0 (1) q : σταθεροί παράμετροι και

ε_t : ο λευκός θόρυβος (white noise), ο οποίος μετρά τυχαία σφάλματα.

Σε ένα μικτό υπόδειγμα ARMA(p,q), όταν ισχύει $s \leq q$, οι αυτοσυσχετίσεις εξαρτώνται από το σύνολο των συντελεστών αi και θi του αυτοπαλίνδρομου υποδείγματος – AR(p) και του υποδείγματος κινητού μέσου – MA(q) αντίστοιχα. Όταν ισχύει $s > q$, οι αυτοσυνδιακυμάνσεις και οι αυτοσυσυσχετίσεις είναι παρόμοιες με εκείνες που επιστρέφονται από μια αυτοπαλίνδρομη διαδικασία AR(p) και οι οποίες περιγράφονται από τις ακόλουθες εξισώσεις:

$$\gamma_s = Cov(y_t, y_{t-s}) = a_1\gamma_{s-1} + \alpha_2\gamma_{s-2} + \cdots + \alpha_p\gamma_{s-p}, \qquad s > q$$

$$\rho_s = a_1\rho_{s-1} + \alpha_2\rho_{s-2} + \cdots + \alpha_p\rho_{s-p}, \qquad s > q$$

Γενικά σε μια διαδικασία μικτού υποδείγματος ARMA(p,q) η συνάρτηση αυτοσυσχέτισης της συμπεριφέρεται όπως εκείνη ενός AR(p) υποδείγματος, ενώ η συνάρτηση μερικής αυτοσυσχέτισης συμπεριφέρεται όπως εκείνη ενός MA(q) υποδείγματος για s > q – p.

Στον πίνακα που ακολουθεί βλέπουμε συνοπτικά τα χαρακτηριστικά των μικτών υποδειγμάτων ARMA(p,q)

Υπόδειγμα	**Αυτοσυσχετίσεις ACF**	**Μερικές αυτοσυσχετίσεις PACF**
Λευκός θόρυβος	Όλες μηδέν: $\rho_k = 0$	Όλες μηδέν: $\varphi_{kk} = 0$
$AR(1)$	Φθίνουν προς το μηδέν από ρ_1 Ευθέως αν $\alpha > 0$ Με πριονωτή μορφή αν $\alpha < 0$	Μηδέν μετά από φ_{11}
$AR(2)$	Φθίνουν Ευθέως από ρ_2 για πραγματικές ρίζες Με ημιτονοειδή τρόπο για μιγαδικές ρίζες	Μηδέν μετά από φ_{22}
$AR(p)$	Φθίνουν προς το μηδέν από ρ_q	Μηδέν μετά το φ_{pp}
$MA(1)$	Μηδέν μετά το ρ_1	Φθίνει σχεδόν γεωμετρικά από το φ_{11}
$MA(2)$	Μηδέν μετά το ρ_2	Φθίνει σχεδόν γεωμετρικά από το φ_{22}

$MA(q)$	Μηδέν μετά το ρ_q	Φθίνει σχεδόν γεωμετρικά από το φ_{qq}
$ARMA(1,1)$	Φθίνει γεωμετρικά από το ρ_1	Φθίνει γεωμετρικά ή κυματιστά από το φ_{11}
$ARMA(p,q)$	Φθίνει γεωμετρικά από το ρ_q	Φθίνει γεωμετρικά ή κυματιστά από το φ_{pp}

Πίνακας 1: Χαρακτηριστικά των Μικτών Υποδειγμάτων

Εν συνεχεία στον πίνακα που ακολουθεί παρουσιάζονται συγκεντρωτικά οι συνθήκες στασιμότητας και αντιστρεψιμότητας των μικτών υποδειγμάτων – ARMA(p,q)

Υπόδειγμα	Εξίσωση	Στασιμότητα	Αντιστρεψιμότητα
$AR(1)$	$y_t = \alpha_1 y_{t-1} + \varepsilon_t$	$\lvert a_1 \rvert < 1$	Αντιστρέψιμο εφόσον στάσιμο
$AR(2)$	$y_t = \alpha_1 y_{t-1} + \alpha_2 y_{t-2} + \varepsilon_t$	$\alpha_1 + \alpha_2 < 1$ $\alpha_2 - \alpha_1 < 1$ $\lvert \alpha_2 \rvert < 1$	Αντιστρέψιμο εφόσον στάσιμο
$AR(p)$	$\mathrm{A}(L) y_t = \varepsilon_t$	Οι ρίζες του πολυωνύμου $\mathrm{A}(L)$ να κείνται εκτός του μοναδιαίου κύκλου	Αντιστρέψιμο εφόσον στάσιμο

$MA(1)$	$y_t = \varepsilon_t - \theta_1 \varepsilon_{t-1}$	Πάντα στάσιμο	$\lvert\theta_1\rvert < 1$
$MA(2)$	$y_t = \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2}$	Πάντα στάσιμο	$\theta_1 + \theta_2 < 1$ $\theta_2 - \theta_1 < 1$ $\lvert\theta_2\rvert < 1$
$MA(q)$	$y_t = \Theta(L)\varepsilon_t$	Πάντα στάσιμο	Οι ρίζες του πολυωνύμου $\Theta(L)$ να κείνται εκτός του μοναδιαίου κύκλου
$ARMA(1,1)$	$y_t = \alpha_1 y_{t-1} + \varepsilon_t - \theta_1 \varepsilon_{t-1}$	$\lvert a_1\rvert < 1$	$\lvert\theta_1\rvert < 1$
$ARMA(p,q)$	$\mathrm{A}(L) y_t = \Theta(L)\varepsilon_t$	Οι ρίζες του πολυωνύμου $\mathrm{A}(L)$ να κείνται εκτός του μοναδιαίου κύκλου	Οι ρίζες του πολυωνύμου $\Theta(L)$ να κείνται εκτός του μοναδιαίου κύκλου

Πίνακας 2: Συνθήκες Στασιμότητας και Αντιστρεψιμότητας των Μικτών Υποδειγμάτων

ΚΕΦΑΛΑΙΟ 6: **ΜΕΘΟΔΟΛΟΓΙΑ** BOX – JENKINS

Όλα τα υποδείγματα τα οποία μελετήθηκαν μέχρι στιγμής στη παρούσα εργασία αναφέρονται σε στάσιμες διαδικασίες. Αυτό πρακτικά σημαίνει ότι ο μέσος, η διακύμανση και οι αυτοδιακυμάνσεις δεν εξαρτώνται από τον χρόνο (t), γι' αυτό κιόλας ο μέσος και η διακύμανση παραμένουν σταθεροί, ενώ οι αυτοδιακυμάνσεις εξαρτώνται μόνο από τη χρονική υστέρηση s. Στη συνέχεια της εργασίας θα εξεταστούν οι μη στάσιμες διαδικασίες μέσω της μεθοδολογίας Box – Jenkins που θα αναλυθεί [6].

Συνήθως οι σειρές οι οποίες δεν παρουσιάζουν στασιμότητα, είναι εκείνες όπου εμφανίζουν τάση ή εποχικές διακυμάνσεις (πχ επενδύσεις), καθώς και οι σειρές τυχαίας διαδρομής (random walk). Αν σε μια σειρά υπάρχει διαχρονική μετατόπιση του μέσου τότε η σειρά είναι μη στάσιμη ως προς το μέσο, ενώ αν μια σειρά δεν έχει σταθερή διακύμανση τότε η σειρά είναι μη στάσιμη ως προς τη διακύμανση [7].

Στη περίπτωση που η σειρά που εξετάζεται δεν είναι στάσιμη, τότε αυτή μπορεί να μετατραπεί σε στάσιμη μέσω της μεθοδολογίας Box – Jenkins. Η μεθοδολογία αυτή προτείνει τη μετατροπή μιας χρονοσειράς σε στάσιμη με τη χρήση διαφορών. Άρα λοιπόν το πρώτο βήμα της μεθοδολογίας Box – Jenkins είναι ο έλεγχος στασιμότητας της σειράς και η μετατροπής της σε στάσιμη (εφόσον δεν είναι). Για μια σειρά Yt οι διαφορές δίνονται από την έκφραση:

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

ή με τη χρήση τελεστών υστέρησης:

$$\Delta Y_t = (1 - L)Y_t$$

Αφού προσδιοριστεί ο αριθμός των d διαφορών που χρησιμοποιήθηκαν για τη μετατροπή της σειράς σε στάσιμη, ακολουθείτε η ανάλυση προσαρμογής του καταλληλότερου υποδείγματος ARMA(p,q) που προσδιορίζει τη νέα στάσιμη σειρά που έχει προκύψει.

Στη περίπτωση που η νέα σειρά που έχει προκύψει είναι στάσιμη με τη χρήση των πρώτων διαφορών (d=1), τότε το υπόδειγμα που προσαρμόζεται καλείται ολοκληρωμένο υπόδειγμα 1ης τάξης ARMA(p,q) ή αυτοπαλίνδρομο ολοκληρωμένο υπόδειγμα κινητών μέσων – ARIMA (p,1,q).

Γενικά ένα ολοκληρωμένο υπόδειγμα χρησιμοποιεί d αριθμό διαφορών προκειμένου να μετατραπεί σε στάσιμο, σύμφωνα πάντα με το τύπο:

$$\Delta^d Y_t = (1 - L)^d Y_t$$

Μια σειρά χωρίς στασιμότητα ονομάζεται ολοκληρωμένης τάξης και συμβολίζεται με I(d), όταν χρησιμοποιείται d αριθμός διαφορών προκειμένου να μετατραπεί σε στάσιμη. Μια στάσιμη σειρά, όπως είναι για παράδειγμα ο λευκός θόρυβος, λογίζεται ως ολοκληρωμένη σειρά μηδενικής τάξης I(0).

6.1. Υπόδειγμα ARIMA(p,d,q)

Όπως αναφέρθηκε και νωρίτερα, αυτοπαλίνδρομο ολοκληρωμένο υπόδειγμα κινητού μέσου τάξεως (p,d,q) – ARIMA(p,d,q), ονομάζεται το μικτό υπόδειγμα ARMA(p,q) που εφαρμόζεται σε μια ολοκληρωμένη σειρά d τάξεως.

Στη γενική του μορφή το υπόδειγμα ARIMA(p,d,q) δίνεται από την έκφραση:

$$A(L)(1 - L)^d Y_t = \delta + \Theta(L)\varepsilon_t, \qquad \acute{o}\pi o\upsilon$$

$$A(L) = 1 - a_1 L - a_2 L^2 - \cdots - a_p L^p \quad \kappa\alpha\iota$$

$$\Theta(L) = 1 - \theta_1 L - \theta_2 L^2 - \cdots - \theta_p L^p$$

Όπου:

d: ο αριθμός των διαφορών που απαιτούνται προκειμένου να μετατραπεί η σειρά σε στάσιμη,

p: οι παράμετροι της αυτοπαλίνδρομης διαδικασίας,

q: οι παράμετροι της διαδικασίας κινητού μέσου.

Γενικά ένα υπόδειγμα ARIMA(p,d,q) μπορεί να διατυπωθεί ως:

- Εξίσωση διαφοράς (Difference equation), δηλαδή ως συνάρτηση των παρελθοντικών τιμών της και της τρέχουσας και παρελθοντικών τιμών του διαταρακτικού όρου.
- Αντίστροφης μορφής (Inverted form), δηλαδή ως συνάρτηση των παρελθοντικών τιμών της και της τρέχουσας τιμής του διαταρακτικού όρου.
- Τυχαία διαταραχή (Random shock), δηλαδή ως συνάρτηση της τρέχουσας τιμής και των παρελθοντικών τιμών του διαταρακτικού όρου.

Στην ανάλυση χρονολογικών σειρών η μεθοδολογία Box – Jenkins χρησιμεύει στην εύρεση ενός στατιστικού υποδείγματος ARIMA(p,d,q) το οποίο πρέπει να προσδιορίζει ικανοποιητικά τη στοχαστική διαδικασία από την οποία προήλθαν τα δεδομένα (δείγμα).

6.2. Εποχικό υπόδειγμα ARIMA (SARIMA)

Μια ειδική κατηγορία των υποδειγμάτων ARIMA είναι τα εποχικά υποδείγματα SARIMA. Η δομή του εποχικού μέρους ενός υποδείγματος ARIMA είναι ίδια με εκείνη του μη εποχικού υποδείγματος, αυτό σημαίνει ότι το εποχικό υπόδειγμα μπορεί να διαθέτει ένα παράγοντα MA, έναν παράγοντα AR και ενδεχομένως μια τάξη διαφορών. Στο εποχικό μέρος του υποδείγματος, οι προαναφερθέντες παράγοντες, διεξάγουν πολλαπλασιασμού της χρονικής υστέρησης s (όπου s το πλήθος των περιόδων σε μια εποχή)[1].

Το εποχικό υπόδειγμα ARIMA (δηλαδή το υπόδειγμα SARIMA) διατυπώνεται ως:

$$ARIMA(p,d,q)x(P,D,Q),$$

όπου:

d: ο αριθμός των διαφορών που απαιτούνται προκειμένου να μετατραπεί η σειρά σε στάσιμη,

p: οι παράμετροι της αυτοπαλίνδρομης διαδικασίας,

---

[1] http://nemertes.lis.upatras.gr/jspui/bitstream/10889/899/1/%CE%94%CE%B9%CF%80%CE%BB%CF%89%CE%BC%CE%B1%CF%84%CE%B9%CE%BA%CE%AE%CE%B5%CF%81%CE%B3%CE%B1%CF%83%CE%AF%CE%B1.pdf.pdf

q: οι παράμετροι της διαδικασίας κινητού μέσου.

P: το πλήθος των εποχικών αυτοπαλίνδρομων όρων (SAR)

D: το πλήθος των εποχικών διαφορών και

Q: το πλήθος των εποχικών όρων κινητού μέσου (SMA).

Όπως και στη περίπτωση των υποδειγμάτων ARIMA όπου το πρώτο βήμα ήταν ο καθορισμός των διαφορών (εφόσον χρειάζεται), έτσι και στα εποχικά υποδείγματα το πρώτο βήμα είναι ο καθορισμός της αναγκαιότητας της εποχικής διαφοράς (είτε μαζί με μια μη εποχική διαφορά είτε αυτόνομα). Θα πρέπει ωστόσο να σημειωθεί ότι δεν πρέπει να χρησιμοποιούνται περισσότερες από μια εποχικές διαφορές, καθώς ούτε περισσότερες από δυο διαφορές συνολικά (εποχικές και μη). Στη περίπτωση που ένα εποχικό υπόδειγμα είναι σταθερό συναρτήσει του χρόνου (υπό την έννοια ότι διατηρείται σχετικά σταθερό στις διάφορες εποχές), τότε είναι πολύ πιθανό να πρέπει να χρησιμοποιηθεί μια εποχική διαφορά (ανεξάρτητα από το αν θα πρέπει να χρησιμοποιηθεί ή όχι μια μη εποχική διαφορά). Ενδεικτικό παράδειγμα είναι οι εποχικές διακυμάνσεις που παρατηρούνται σε σειρές μεταξύ καλοκαιρινών μηνών και χειμερινών.

Σε ένα εποχικό υπόδειγμα, ο έλεγχος αυτοσυσχετίσεων προσδιορίζει τους όρους SAR και SMA που θα προστεθούν στο υπόδειγμα. Συγκεκριμένα, αν κατά την εποχική περίοδο η αυτοσυσχέτιση είναι θετική, προστίθεται ένα όρος SAR στο υπόδειγμα, ενώ αν κατά την εποχική περίοδο η αυτοσυσχέτιση είναι αρνητική, προστίθεται ένα όρος SMA στο υπόδειγμα. Στις περισσότερες περιπτώσεις, ένας όρος SAR και SMA (SAR(1) και SMA(1)) είναι επαρκείς και πολύ σπάνια συναντώνται διαδικασίες με περισσότερους όρους.

Το πλέον χρησιμοποιούμενο υπόδειγμα SARIMA είναι το ARIMA(0,1,1)x(0,1,1) το οποίο είναι στην ουσία το υπόδειγμα MA(1) x SMA(1) και το οποίο διαθέτει μια εποχική και μια μη εποχική διαφορά.

6.3. *Στάδια μεθοδολογίας Box – Jenkins*

Η μεθοδολογία Box – Jenkins περιγράφεται από τρία βασικά στάδια, τα οποία είναι:

- Η ταυτοποίηση (identification)
- Η εκτίμηση (estimation)
- Ο διαγνωστικός έλεγχος (diagnostic checking)

6.3.1. Ταυτοποίηση

Στο στάδιο της ταυτοποίησης γίνεται ο καθορισμός των τιμών p, d και q του υποδείγματος ARIMA (p,d,q) με βάση τις πληροφορίες που προέρχονται από το δείγμα.

Συγκεκριμένα, αρχικά προσδιορίζεται ο αριθμός d των διαφορών που απαιτούνται για τη μετατροπή της σειράς σε στάσιμη (εφόσον δεν είναι).

Για τον έλεγχο της στασιμότητας, εξετάζεται η δειγματική συνάρτηση αυτοσυσχέτισης. Αν οι αυτοσυσχετίσεις συγκλίνουν με υψηλό ρυθμό προς το μηδέν, τότε είναι πολύ πιθανό η σειρά να είναι στάσιμη. Από την άλλη, όταν οι αυτοσυσχετίσεις φθίνουν με χαμηλό ρυθμό προς το μηδέν, τότε είναι πολύ πιθανό η σειρά να μην είναι στάσιμη και να πρέπει να μετατραπεί. Σε αυτή τη τελευταία περίπτωση γίνεται χρήση των διαφορών ώστε να μετατραπεί σε στάσιμη η σειρά.

Εν συνεχεία γίνεται ο καθορισμός των τάξεων p και q της αυτοπαλίνδρομης διαδικασίας και της διαδικασίας κινητού μέσου αντίστοιχα, προσδιορίζοντας κατά αυτό τον τρόπο τη τάξη του υποδείγματος ARIMA με βάση τις απλές και μερικές δειγματικές αυτοσυσχετίσεις.

6.3.2. Εκτίμηση

Μετά το στάδιο της ταυτοποίησης, έπεται το στάδιο της εκτίμησης. Σε αυτό το σημείο η μεθοδολογία Box – Jenkins προτείνει την εκτίμηση των συντελεστών του υποδείγματος ARIMA(p,d,q). Πιο συγκεκριμένα στο στάδιο αυτό γίνεται η εκτίμηση των p παραμέτρων (α_i, i=1(1)p) του αυτοπαλίνδρομου υποδείγματος και των q παραμέτρων (θ_i, i=1(1)q) του υποδείγματος κινητού μέσου.

Γενικά για τη μέθοδο εκτίμησης των παραμέτρων ισχύει ότι αυτή προσδιορίζεται από τη μαθηματική μορφή του υποδείγματος. Για παράδειγμα, αν η σειρά που μελετάται είναι μια αυτοπαλίνδρομη διαδικασία, οι παράμετροί της μπορούν να εκτιμηθούν χρησιμοποιώντας τη μέθοδο των ελαχίστων τετραγώνων. Από την άλλη όμως, αν η σειρά περιλαμβάνει και όρους της διαδικασίας κινητού μέσου, η εκτίμηση των παραμέτρων του κινητού μέσου θα γίνει με τη χρήση μη γραμμικών μεθόδων εκτίμησης. Άρα όπως συμπεραίνεται, η πλέον κατάλληλη μέθοδος εκτίμησης που πρέπει να ακολουθηθεί εξαρτάται από τη μορφή της στοχαστικής διαδικασίας (που διαμορφώνει τη χρονολογική σειρά) και από τον αριθμό των διαφορών (εφόσον απαιτούνται για τη μετατροπή της χρονοσειράς σε στάσιμη).

6.3.3. **Διαγνωστικός Έλεγχος**

Αφού ολοκληρωθούν τα στάδιο της ταυτοποίησης και της εκτίμησης, πλέον έχει προσδιοριστεί μια μορφή του υποδείγματος, της οποίας ο σκοπός είναι να χρησιμοποιηθεί σε μελλοντικές προβλέψεις. Για να χρησιμοποιηθεί όμως το εξαχθέν υπόδειγμα ως μοντέλο πρόβλεψης θα πρέπει αυτό να είναι απαραιτήτως αρκούντος ικανοποιητικό.

Σύμφωνα λοιπόν με την ανωτέρω πρόταση θα πρέπει το υπόδειγμα να ελεγχθεί για τη καταλληλότητα του με βάση τέσσερεις άξονες, οι οποίοι είναι:

- Η σημαντικότητα των συντελεστών του υποδείγματος
- Η σταθερότητα των συντελεστών του υποδείγματος
- Οι ιδιότητες των καταλοίπων
- Η προβλεπτική ικανότητα του υποδείγματος

Λαμβάνοντας υπόψη τα όσα αναφέρθηκαν, στο στάδιο αυτό πρέπει να γίνει έλεγχος για το πόσο καλά προσαρμόζεται το υπόδειγμα στα δεδομένα της σειράς (έλεγχος καταλληλότητας). Γενικά μπορεί να ειπωθεί ότι ο βέλτιστος τρόπος για τον έλεγχο της προσαρμοστικότητας του υποδείγματος είναι μέσω της εξέτασης της προβλεπτικής του ικανότητας πέραν της δειγματοληπτικής περιόδου, δηλαδή της ικανότητας του υποδείγματος να προβλέπει με τη μέγιστη δυνατή ακρίβεια μελλοντικές τιμές της σειράς. Επειδή όμως, είναι σύνηθες το φαινόμενο να μην είναι επαρκές το πλήθος των δεδομένων για αυτή τη διαδικασία, γι’ αυτό και στα τρία στάδια υλοποιούνται με τη χρήση του ίδιου δείγματος παρατηρήσεων.

Στο σημείο όπου το υπόδειγμα το οποίο εξετάζεται θεωρηθεί κατάλληλο, δηλαδή εκφράζει ικανοποιητικά τη διαδικασία από την οποία προέρχονται τα δεδομένα, τότε τα κατάλοιπα που προκύπτουν θα πρέπει να συμπεριφέρονται ως μια διαδικασία λευκού θορύβου. Πρακτικά αυτό σημαίνει ότι αν γίνει έλεγχος αυτοσυσχετίσεων μεταξύ των καταλοίπων, θα πρέπει να διαπιστωθεί ότι δεν υπάρχει αυτοσυσχέτιση μεταξύ τους. Για τον προαναφερθέν έλεγχο των καταλοίπων ακολουθείται η στατιστική Q των Box – Pierce, με βάση την οποία γίνεται ο έλεγχος της σημαντικότητας n πλήθους συντελεστών αυτοσυσχέτισης.

Η στατιστική Q των Box – Pierce, δίνεται από την έκφραση:

$$Q_{BP} = T\sum_{s=1}^{n} \hat{\rho}_s^2$$

Όπου:

T: το πλήθος των παρατηρήσεων,

n: το πλήθος των συντελεστών αυτοσυσχέτισης και

$\hat{\rho}_s$: οι δειγματικές αυτοσυσχετίσεις των καταλοίπων.

Επιπροσθέτως, ισχύει ότι η τετραγωνική ρίζα του πλήθους των παρατηρήσεων ισούται με το πλήθος των αυτοσυσχετίσεων, το οποίο δηλώνεται ως εξής:

$$n = \sqrt{T}$$

Επιπλέον, η στατιστική Q ακολουθεί τη κατανομή X^2 με n – p – q βαθμούς ελευθερία. Οι υποθέσεις που ελέγχονται σε σχέση με τη στατιστική Q είναι:

$$H_0: Q_{BP} > X_a^2$$

$$H_1: Q_{BP} < X_a^2$$

Η μηδενική υπόθεση (H_0) θα απορρίπτεται εφόσον η τιμή της στατιστικής Q είναι μεγαλύτερη από τη κρίσιμη τιμή της κατανομής X_α^2, για δεδομένο επίπεδο σημαντικότητας α (συνήθως χρησιμοποιείται το α = 0,05). Αντίστοιχα θα απορρίπτεται η υπόθεση H_1 αν η τιμή της στατιστικής Q είναι μικρότερη από τη κρίσιμη τιμή της κατανομής X_α^2, για δεδομένο επίπεδο σημαντικότητας α.

Στο σημείο αυτό πρέπει να σημειωθεί ότι το κριτήριο το οποίο χρησιμοποιείται σήμερα ευρέως στη μεθοδολογία Box – Jenkins, είναι η στατιστική Q των Ljung και Box, η οποία δίνεται από την έκφραση:

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

Όπως και στη περίπτωση της στατιστικής Q κατά Box – Pierce, έτσι και στη περίπτωση κατά Ljung και Box η στατιστική ακολουθεί προσεγγιστικά τη κατανομή X_α^2, με n – p – q βαθμούς ελευθερίας. Επιπλέον, ανάλογα με τη στατιστική Q κατά Box – Pierce, έτσι και στο Q_{LB} η μηδενική υπόθεση (H_0) θα απορρίπτεται εφόσον η τιμή της στατιστικής Q είναι μεγαλύτερη από τη κρίσιμη τιμή της κατανομής X_α^2, για δεδομένο επίπεδο σημαντικότητας α (συνήθως χρησιμοποιείται το α = 0,05). Αντίστοιχα θα απορρίπτεται η υπόθεση H_1 αν η τιμή της στατιστικής Q είναι μικρότερη από τη κρίσιμη τιμή της κατανομής X_α^2, για δεδομένο επίπεδο σημαντικότητας α.

6.4. *Κριτήρια Επιλογής*

Στην υπό-ενότητα αυτή θα σχολιαστεί ο έλεγχος της καταλληλότητας του εκτιμημένου υποδείγματος μέσω της υπερπροσαρμογής (overfitting). Σύμφωνα με τη διαδικασία αυτή, ο έλεγχος πραγματοποιείται μέσω της σύγκρισης του εκτιμημένου υποδείγματος με ένα υπόδειγμα υψηλότερης τάξης. Για παράδειγμα, αν το εκτιμώμενο μοντέλο είναι ένα μικτό υπόδειγμα ARMA(1,2), αυτό θα πρέπει να συγκριθεί με τα μικτά υποδείγματα ARMA(2,2) και ARMA(1,3).

Στη περίπτωση που οι πρόσθετοι συντελεστές που προκύπτουν από τα υποδείγματα υψηλότερης τάξης δεν είναι στατιστικά σημαντικοί (δεν είναι στατιστικά διαφορετικοί από το μηδέν), τότε εξάγεται το συμπέρασμα ότι το εκτιμώμενο υπόδειγμα είναι το καταλληλότερο, δηλαδή περιγράφει καλύτερα τη διαδικασία. Σε αντίθετη περίπτωση, δηλαδή στη περίπτωση που οι πρόσθετοι συντελεστές που προκύπτουν από τα υποδείγματα υψηλότερης τάξης είναι στατιστικά σημαντικοί, τότε εξάγεται το συμπέρασμα ότι το εκτιμώμενο υπόδειγμα δεν είναι το καταλληλότερο και άρα υπάρχει κάποιο άλλο υπόδειγμα που προσδιορίζει καλύτερα τη διαδικασία.

Όπως είναι λογικό, προσθέτοντας υστερήσεις είτε στο τμήμα που αφορά την αυτοπαλίδρομη διαδικασία είτε στο τμήμα που αφορά τη διαδικασία κινητού μέσου, επιτυγχάνεται η αύξηση της τάξης του υποδείγματος και ταυτόχρονα μειώνεται το άθροισμα των τετραγώνων των καταλοίπων. Επιπροσθέτως, από τη στιγμή που θα απαιτείται η εκτίμηση περισσότερων παραμέτρων, αυτό συνεπάγεται ότι θα μειώνονται και οι βαθμοί ελευθερίας.

Τα όσα αναφέρθηκαν στην παραπάνω παράγραφο αποτελούν εισαγωγικά προκειμένου να παρουσιαστούν εκείνα τα κριτήρια τα οποία υποβοηθούν στη διαδικασία επιλογής του καταλληλότερου υποδείγματος. Δύο δημοφιλή κριτήρια που χρησιμοποιούνται στην ανάλυση χρονολογικών σειρών είναι τα εξής:

- Το κριτήριο Schwartz (Schwartz Bayesian Criterion – SBC)
- Το κριτήριο πληροφοριών Akaike (Akaike Information Criterion – AIC)

Το κριτήριο Schwartz, δίνεται από το τύπο:

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

Αντίστοιχα, το κριτήριο πληροφοριών Akaike, δίνεται από το τύπο:

$$SBC = \ln(s^2) + nln(T)$$

όπου:

s^2 : η εκτίμηση της διακύμανσης των καταλοίπων,

n : το πλήθος των εκτιμούμενων παραμέτρων του υποδείγματος (p + q + c) όπου το c είναι σταθερά (συνήθως ως σταθερά λογίζεται η μονάδα) και

T : το πλήθος των παρατηρήσεων που χρησιμοποιούνται στη παλινδρόμηση.

Σύμφωνα με όσα αναφέρθηκαν και νωρίτερα, η προσθήκη μιας επιπλέον μεταβλητής στο υπόδειγμα οδηγεί σε μείωση του αθροίσματος των τετραγώνων των καταλοίπων και συνεπώς σε μείωση της διακύμανσης ($σ^2$) και αύξηση του πλήθους n.

Για την επιλογή του κατάλληλου υποδείγματος, ελέγχεται η τιμή των κριτηρίων και συγκεκριμένα επιλέγεται το υπόδειγμα εκείνο που επιστρέφει τη μικρότερη τιμή στα κριτήρια του Akaike (AIC) και του Schwartz (SBC). Μεταξύ των δύο κριτηρίων, το SBC θεωρείται ασυμπτωτικά καλύτερο. Λόγο της έκφρασης $\ln(T) > 2$, το κριτήριο SBC επιβάλει μεγαλύτερη "ποινή" από το κριτήριο AIC στις επιπλέον εκτιμούμενες παραμέτρους. Σύμφωνα με τη παραπάνω πρόταση, το κριτήριο SBC θα οδηγεί πάντα στην επιλογή ενός υποδείγματος του οποίου το πλήθος των παραμέτρων θα είναι μικρότερο ή ίσο (στην έσχατη περίπτωση) με το πλήθος των παραμέτρων του υποδείγματος που επιλέχθηκε με βάση το κριτήριο AIC.

ΚΕΦΑΛΑΙΟ 7: ΕΦΑΡΜΟΓΗ ΜΕΘΟΔΟΛΟΓΙΑΣ BOX – JENKINS ΜΕ ΤΗ ΧΡΗΣΗ ΤΟΥ SPSS

Όλα τα υποδείγματα τα οποία μελετήθηκαν μέχρι στιγμής στη παρούσα εργασία αναφέρονται σε στάσιμες διαδικασίες. Αυτό πρακτικά σημαίνει ότι ο μέσος, η διακύμανση και οι αυτοδιακυμάνσεις δεν εξαρτώνται από τον χρόνο (t), γι’ αυτό κιόλας ο μέσος και η διακύμανση παραμένουν σταθεροί, ενώ οι αυτοδιακυμάνσεις εξαρτώνται μόνο από τη χρονική υστέρηση s. Στη συνέχεια της εργασίας θα εξεταστούν οι μη στάσιμες διαδικασίες μέσω της μεθοδολογίας Box – Jenkins που θα αναλυθεί.

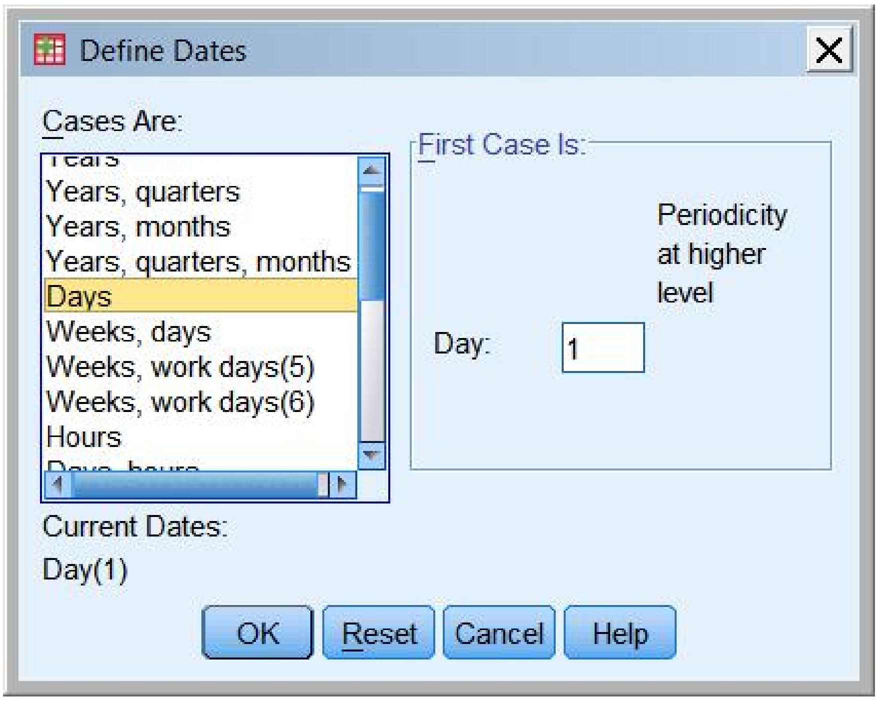
Συνήθως οι σειρές οι οποίες δεν παρουσιάζουν στασιμότητα, είναι εκείνες όπου εμφανίζουν τάση ή εποχικές διακυμάνσεις (πχ επενδύσεις), καθώς και οι σειρές τυχαίας διαδρομής (random walk). Αν σε μια σειρά υπάρχει διαχρονική μετατόπιση του μέσου τότε η σειρά είναι μη στάσιμη ως προς το μέσο [6].

7.1. Τράπεζα της Ελλάδας

Στη μελέτη που θα ακολουθήσει ελέγχουμε τις τιμές κλεισίματος της μετοχής της Τράπεζας της Ελλάδας για τη περίοδο 21/11/2013 – 19/02/2015, όπως αυτές παρέχονται από την ιστοσελίδα της Ναυτεμπορικής[2].

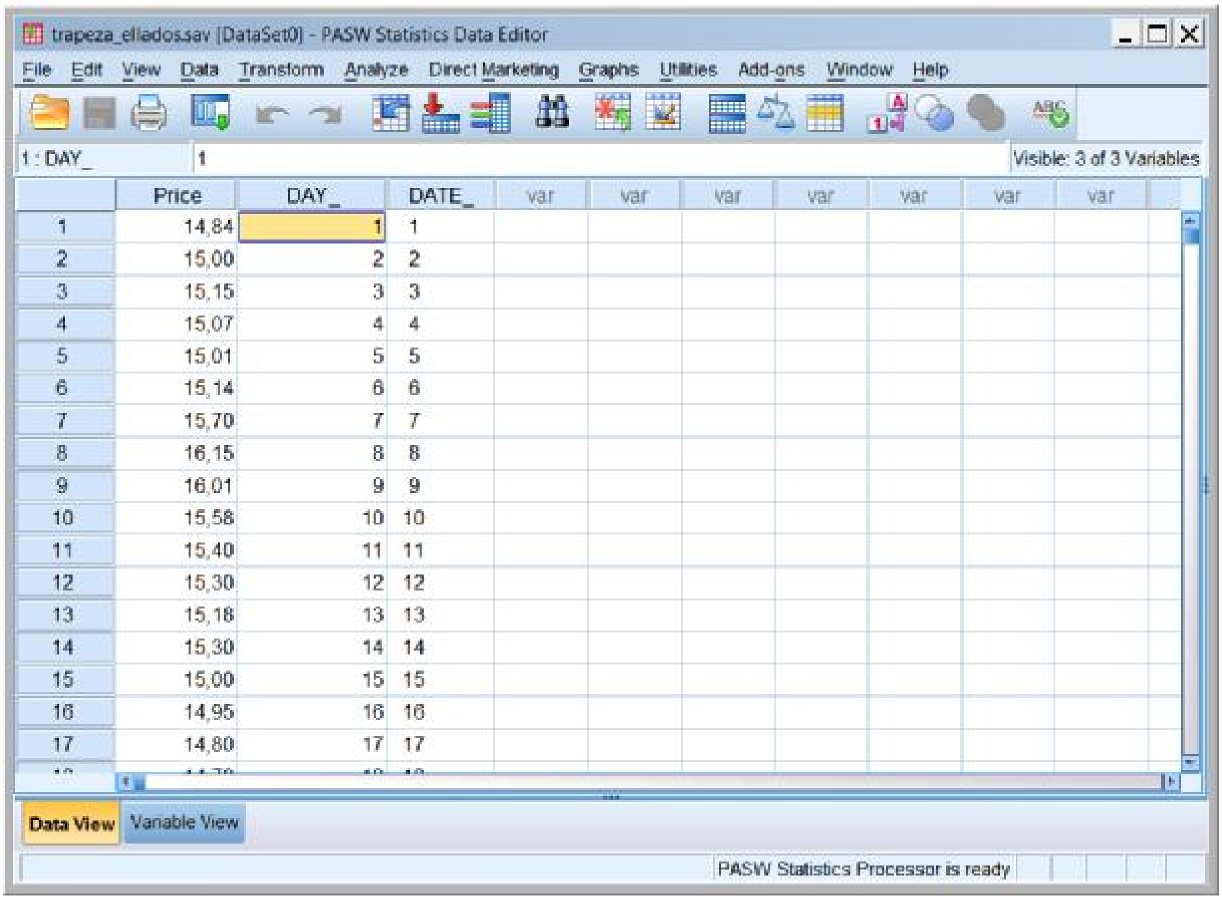
Αφού αντιγραφούν τις τιμές κλεισίματος της μετοχής στο SPSS, ορίζεται η περιοδικότητα των δεδομένων. Συγκεκριμένα, από τη διαδρομή Data **à** Define Dates **à** Days, ορίζεται Day = 1, γεγονός που δηλώνει ότι η περιοδικότητα των δεδομένων είναι ανά ημέρα.

[2] http://www.naftemporiki.gr/finance/quote/tell.ath/ellados-trapeza-ko?tab=4&mode=3

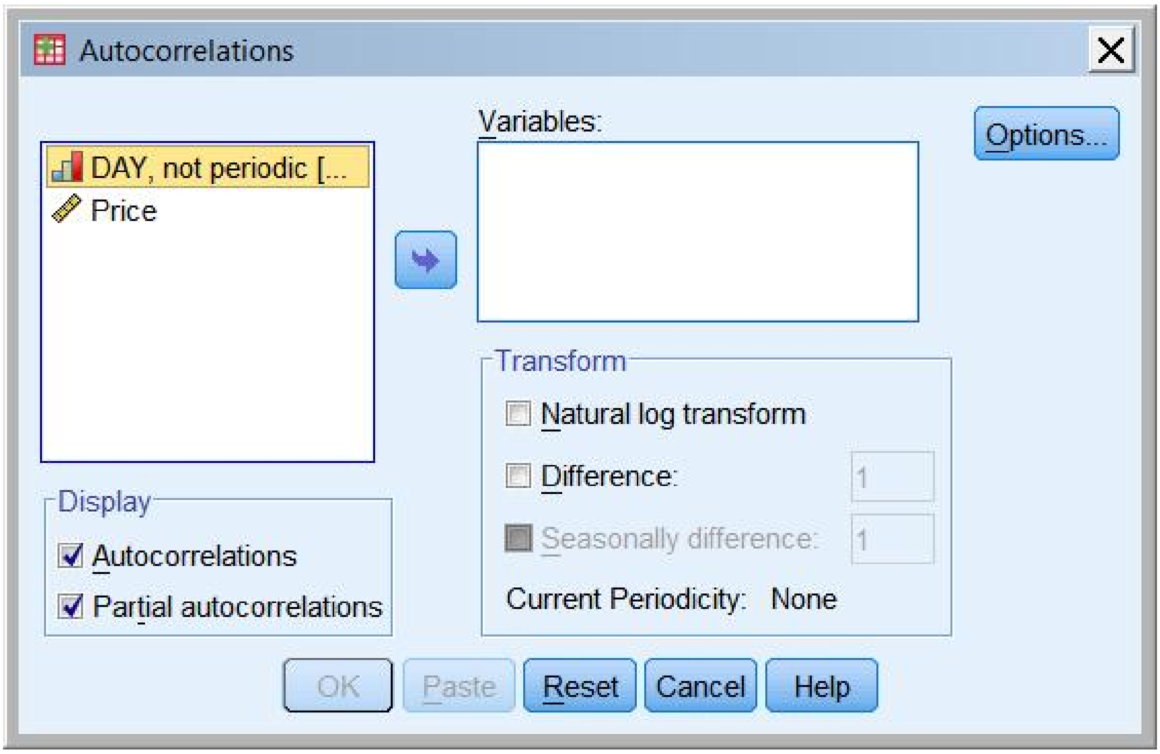


Εικόνα 1: Ορισμός Περιοδικότητας των δεδομένων της Τ.Ε.

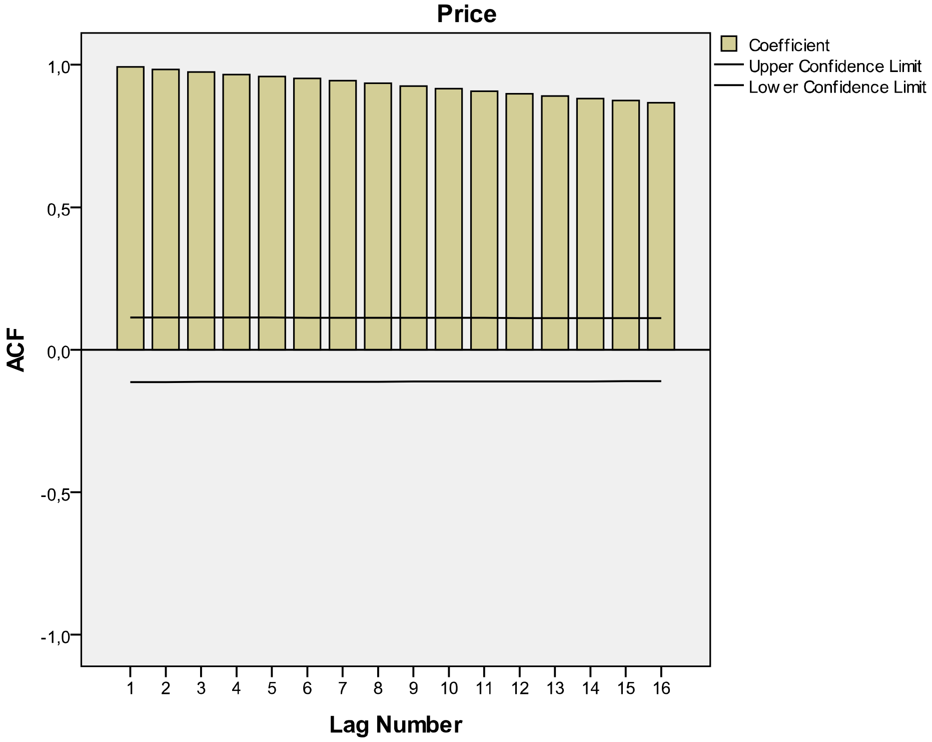
Αυτό έχει ως συνέπεια την εισαγωγή των μεταβλητών "DAY_ και DATE_" στο SPSS.



Σύμφωνα με το πρώτο στάδιο της μεθοδολογίας Box – Jenkins, θα πρέπει αρχικά να γίνει έλεγχος στασιμότητας της χρονοσειράς, προκειμένου να προσδιοριστεί το κατάλληλο μοντέλο. Ακλουθώντας τη διαδρομή Analyze **à** Forecasting **à** Autocorrelations προκύπτει το ακόλουθο παράθυρο διαλόγου.



Στο πεδίο των μεταβλητών (Variables) εισέρχεται η μεταβλητή «Price» που υποδηλώνει τη τιμή της μετοχής σε ένα δεδομένο χρονικό σημείο. Στην αρχική δοκιμή, απαιτείται να εμφανιστούν τα αποτελέσματα των αυτοσυσχετίσεων. Επιπλέον, θεωρείται αρχικά ότι δεν χρειάζεται να οριστούν διαφορές (Difference). Τα αποτελέσματα που λαμβάνονται, εμφανίζονται στον ακόλουθα πίνακα αυτοσυσχετίσεων.



Εικόνα 2: Διάγραμμα αυτοσυσχετίσεων της Τ.Ε.

Autocorrelations

Series:Price

Lag	Autocorrelation	Std. Error[a]	Box-Ljung Statistic		
			Value	df	Sig.[b]
1	,992	,057	306,112	1	,000
2	,984	,057	608,075	2	,000
3	,974	,057	905,360	3	,000
4	,966	,056	1198,437	4	,000
5	,959	,056	1488,076	5	,000
6	,952	,056	1774,650	6	,000
7	,944	,056	2057,140	7	,000
8	,935	,056	2335,377	8	,000
9	,925	,056	2608,763	9	,000
10	,916	,056	2877,560	10	,000
11	,907	,056	3141,867	11	,000
12	,898	,056	3402,240	12	,000
13	,890	,056	3658,623	13	,000
14	,882	,055	3911,136	14	,000

15	,874	,055	4160,241	15	,000
16	,867	,055	4406,095	16	,000

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

b. Based on the asymptotic chi-square approximation.

Στον πίνακα των αυτοσυσχετίσεων που επιστρέφεται από τον πρόγραμμα, εμφανίζονται πληροφορίες που σχετίζονται με:

Lag: Χρονολογικές υστερήσεις.

Autocorrelation: Οι τιμές του συντελεστή αυτοσυσχέτισης για κάθε χρονολογική υστέρηση.

Std. Error: Τυπικό σφάλμα.

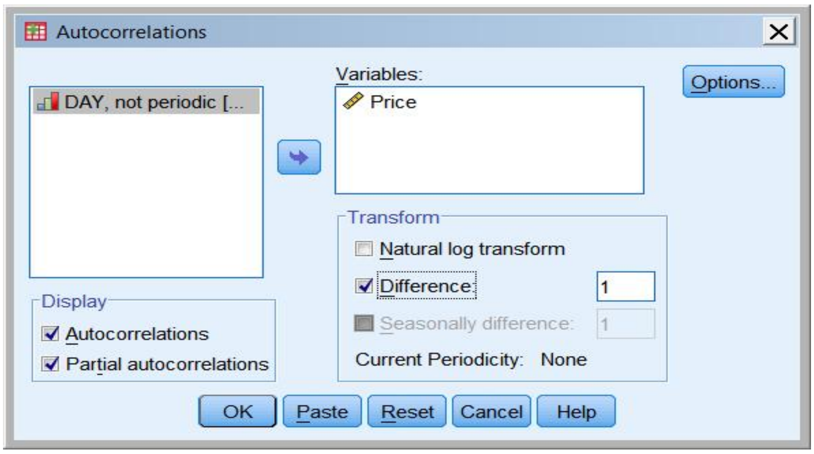
Value: Οι τιμές της στατιστικής συνάρτησης Box – Ljung

df: Οι βαθμοί ελευθερίας.

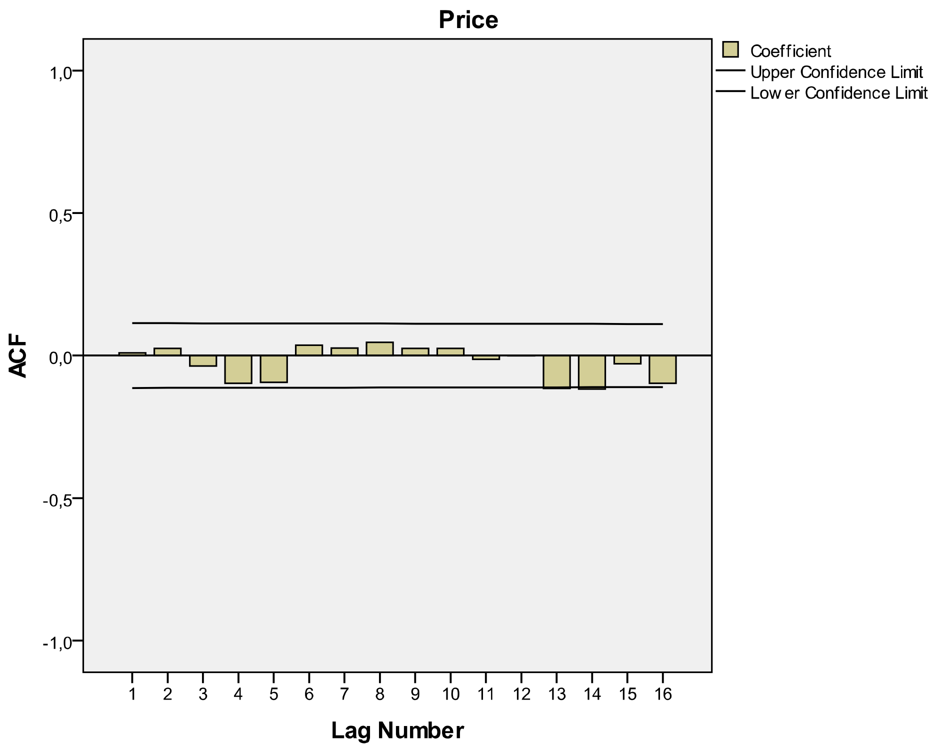
Sig: Οι τιμές πιθανότητας της στατιστικής συνάρτησης.

Σύμφωνα με τα αποτελέσματα που εξήχθησαν, οι τιμές του συντελεστή αυτοσυσχέτισης ρ είναι διάφορες του μηδενός και πολύ κοντά στη τιμή 1, γεγονός που σημαίνει ότι υπάρχει θετική αυτοσυσχέτιση και μάλιστα ισχυρή.

Άρα λοιπόν κατά τον πρώτο έλεγχο που εφαρμόστηκε διαπιστώθηκε ότι η σειρά δεν είναι στάσιμη και άρα θα πρέπει να επαναληφθεί ο έλεγχος στασιμότητας λαμβάνοντας υπόψη τις πρώτες διαφορές.



Τα αποτελέσματα που προκύπτουν δίνονται ακολούθως.



Εικόνα 3: Διάγραμμα αυτοσυχετίσεων της Τ.Ε. (πρώτες διαφορές)

Autocorrelations

Series:Price

Lag	Autocorrelation	Std. Error[a]	Box-Ljung Statistic		
			Value	df	Sig.[b]

1	,010	,057	,028	1	,867
2	,025	,057	,227	2	,893
3	-,037	,057	,647	3	,886
4	-,097	,057	3,589	4	,464
5	-,094	,056	6,377	5	,271
6	,036	,056	6,782	6	,341
7	,026	,056	6,995	7	,429
8	,046	,056	7,677	8	,466
9	,025	,056	7,872	9	,547
10	,025	,056	8,071	10	,622
11	-,013	,056	8,125	11	,702
12	-,001	,056	8,125	12	,775
13	-,115	,056	12,389	13	,496
14	-,117	,056	16,838	14	,265
15	-,028	,055	17,100	15	,313
16	-,097	,055	20,152	16	,213

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

b. Based on the asymptotic chi-square approximation.

Πλέον, παρατηρείται ότι οι τιμές του συντελεστή για τις διάφορες φάσεις υστέρησης έχει μεταβληθεί. Ενδεικτικά ακολουθεί ο αναλυτικός έλεγχος στασιμότητας για όλες τις χρονολογικές υστερήσεις.

Για τον έλεγχο στασιμότητας θα χρησιμοποιήσουμε δυο υποθέσεις:

- Την μηδενική υπόθεση H_0: Sig = 0, που δηλώνει ότι δεν υπάρχει αυτοσυσχέτιση και άρα η χρονοσειρά είναι στάσιμη και
- Την υπόθεση H_1: Sig ≠ 0 που δηλώνει ότι υπάρχει αυτοσυσχέτιση και άρα η χρονοσειρά δεν είναι στάσιμη.

Σύμφωνα με το παραπάνω σενάριο ακολουθεί ο έλεγχος στασιμότητας και για τις 16 χρονικές υστερήσεις.

Lag = 1: $Sig_1 = 0.867 > α = 0.05$, άρα αποδοχή της υπόθεσης H_0.

Lag = 2: $Sig_2 = 0.893 > α = 0.05$, άρα αποδοχή της υπόθεσης H_0.

Lag = 3: Sig_3 = 0.886 > α = 0.05, άρα αποδοχή της υπόθεσης H_0.

Lag = 4: Sig_4 = 0.464 > α = 0.05, άρα αποδοχή της υπόθεσης H_0.

Lag = 5: Sig_5 = 0.271 > α = 0.05, άρα αποδοχή της υπόθεσης H_0.

Lag = 6: Sig_6 = 0.341 > α = 0.05, άρα αποδοχή της υπόθεσης H_0.

Lag = 7: Sig_7 = 0.429 > α = 0.05, άρα αποδοχή της υπόθεσης H_0.

Lag = 8: Sig_8 = 0.466 > α = 0.05, άρα αποδοχή της υπόθεσης H_0.

Lag = 9: Sig_9 = 0.547 > α = 0.05, άρα αποδοχή της υπόθεσης H_0.

Lag = 10: Sig_{10} = 0.622 > α = 0.05, άρα αποδοχή της υπόθεσης H_0.

Lag = 11: Sig_{11} = 0.702 > α = 0.05, άρα αποδοχή της υπόθεσης H_0.

Lag = 12: Sig_{12} = 0.775 > α = 0.05, άρα αποδοχή της υπόθεσης H_0.

Lag = 13: Sig_{13} = 0.496 > α = 0.05, άρα αποδοχή της υπόθεσης H_0.

Lag = 14: Sig_{14} = 0.265 > α = 0.05, άρα αποδοχή της υπόθεσης H_0.

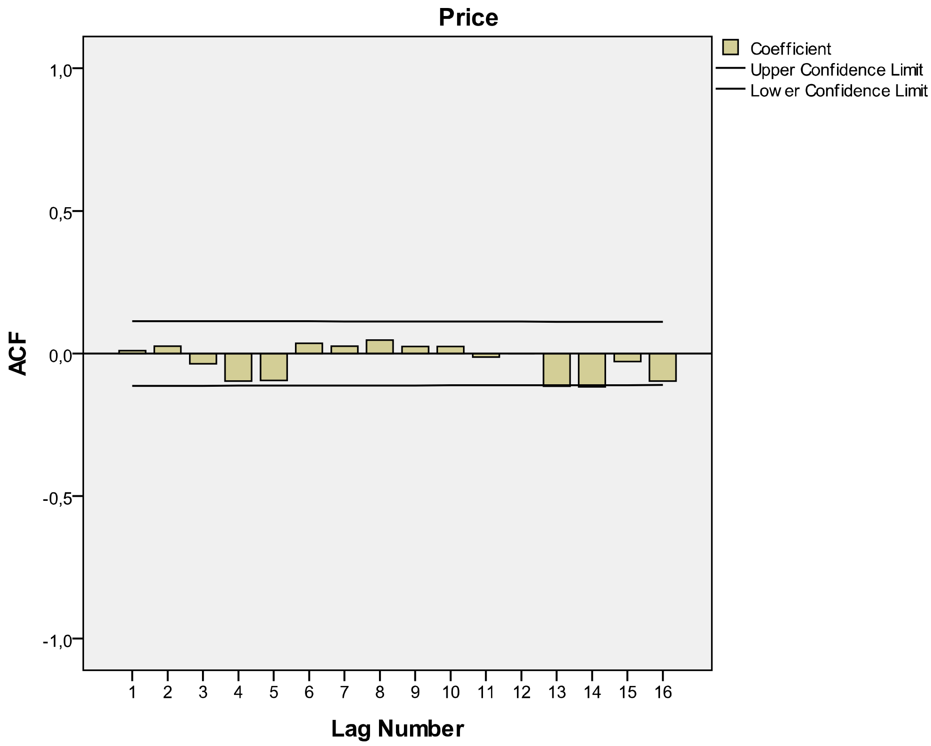
Lag = 15: Sig_{15} = 0.313 > α = 0.05, άρα αποδοχή της υπόθεσης H_0.

Lag = 16: Sig_{16} = 0.213 > α = 0.05, άρα αποδοχή της υπόθεσης H_0.

Με βάση την ανωτέρω ανάλυση, διαπιστώθηκε ότι η χρονοσειρά παραμένει στάσιμη για όλες τις χρονικές υστερήσεις, άρα θεωρείται στάσιμη λαμβάνοντας υπόψη τις πρώτες διαφορές.

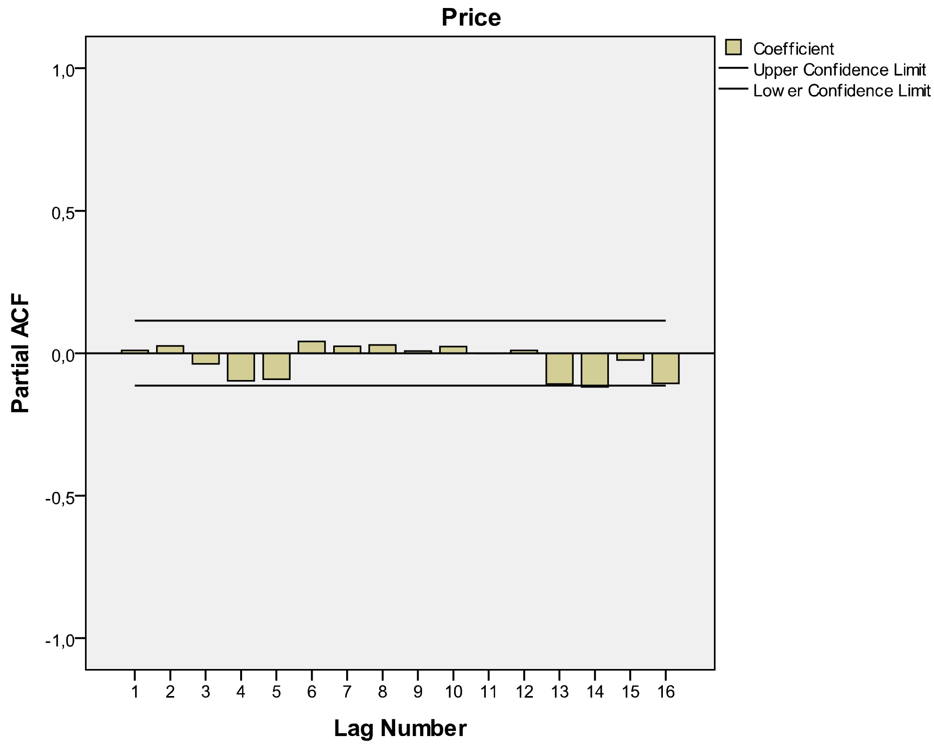
Στο επόμενο στάδιο γίνεται ο προσδιορισμός της τάξεως q και p με τη χρήση των γραφημάτων αυτοσυσχετίσεων (ACF) και των μερικών αυτοσυσχετίσεων (Partial ACF) αντίστοιχα.

Παρατηρώντας το γράφημα ACF διαπιστώνεται ότι υπάρχει δύο τιμές που είναι στο όριο ή εκτός ορίου του διαστήματος εμπιστοσύνης και συγκεκριμένα στο κάτω όριο για τις χρονικές υστέρηση lag=13,14. Αυτό σημαίνει ότι θα θεωρηθεί ότι το q = 2.



Εικόνα 4: Μελέτη αυτοσυσχετίσεων για τον προσδιορισμό για το προσδιορισμό της τάξης q

Ακολούθως από το γράφημα Partial ACF διαπιστώνεται ότι υπάρχει δύο τιμές που είναι στο όριο ή εκτός ορίου του διαστήματος εμπιστοσύνης και συγκεκριμένα στο κάτω όριο για τη χρονική υστέρηση lag=13,14. Αυτό σημαίνει ότι θα θεωρηθεί ότι το p = 2.



Εικόνα 5: Μελέτη των μερικών αυτοσυσχετίσεων για τον προσδιορισμό για το προσδιορισμό της τάξης q

Άρα λοιπόν από τη μελέτη των διαγραμμάτων ACF και Partial ACF προκύπτει ότι q=2 και p=2. Σύμφωνα με τα δεδομένα αυτά, το γενικό υπόδειγμα ARIMA(p,d,q) το οποίο προσδιορίζει καλύτερα τη χρονοσειρά παίρνει τη μορφή:

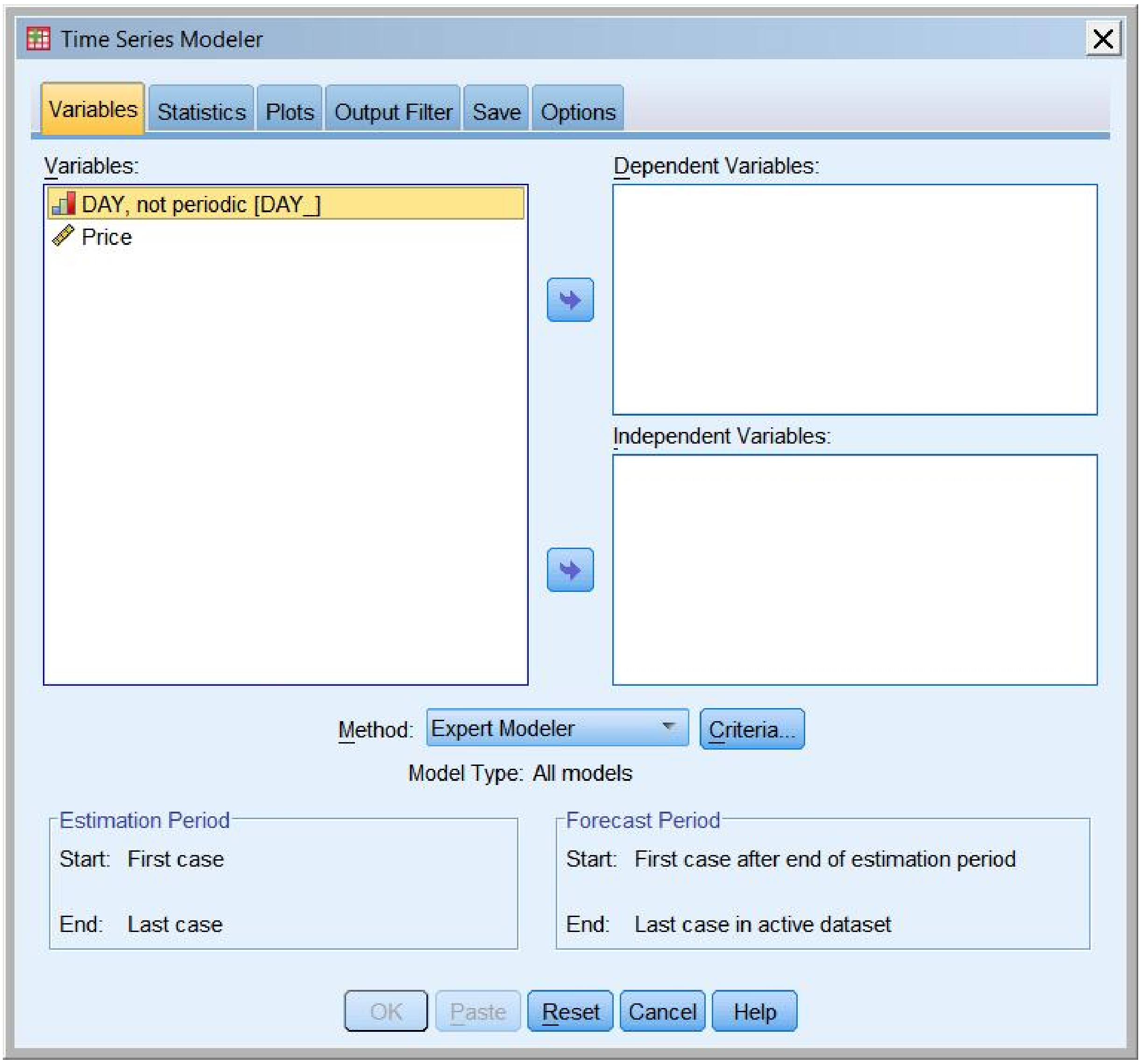
ARIMA(2,1,2)

Η θεωρητική μορφή του υποδείγματος προσδιορίζεται από την έκφραση:

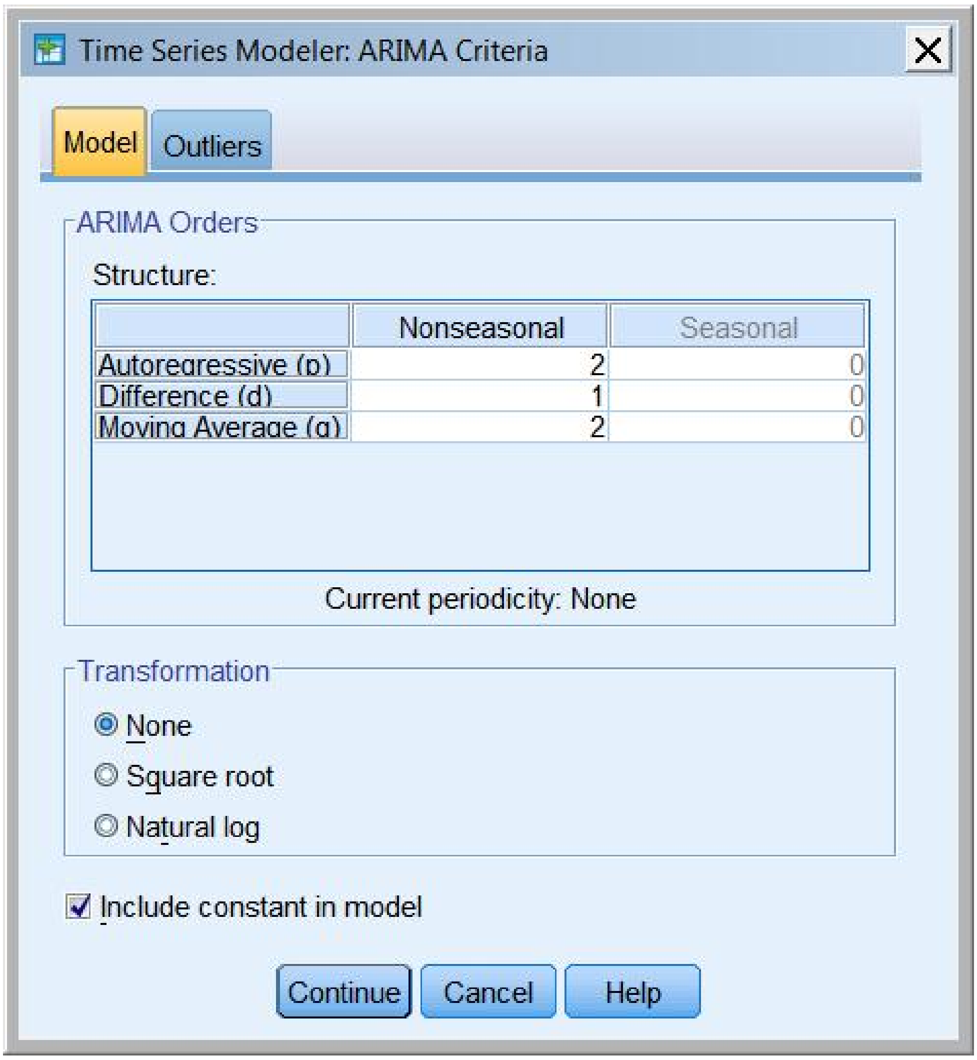
$$Y_t = a_0 + a_1 y_{t-1} + a_2 y_{t-2} + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \varepsilon_t$$

Στο επόμενο στάδιο προσδιορίζονται οι παράμετροι α_0, α_1, α_2, θ_1 και θ_2 και στη συνέχεια πραγματοποιείται ο έλεγχος στατιστικής σημαντικότητας αυτών των παραμέτρων.

Για τον προσδιορισμό των παραμέτρων, αρχικά δηλώνουμε το μοντέλο μας στο SPSS ακολουθώντας τη διαδρομή Analyze **à** Forecasting **à** Create Models…, η οποία μας οδηγεί στο ακόλουθο παράθυρο.

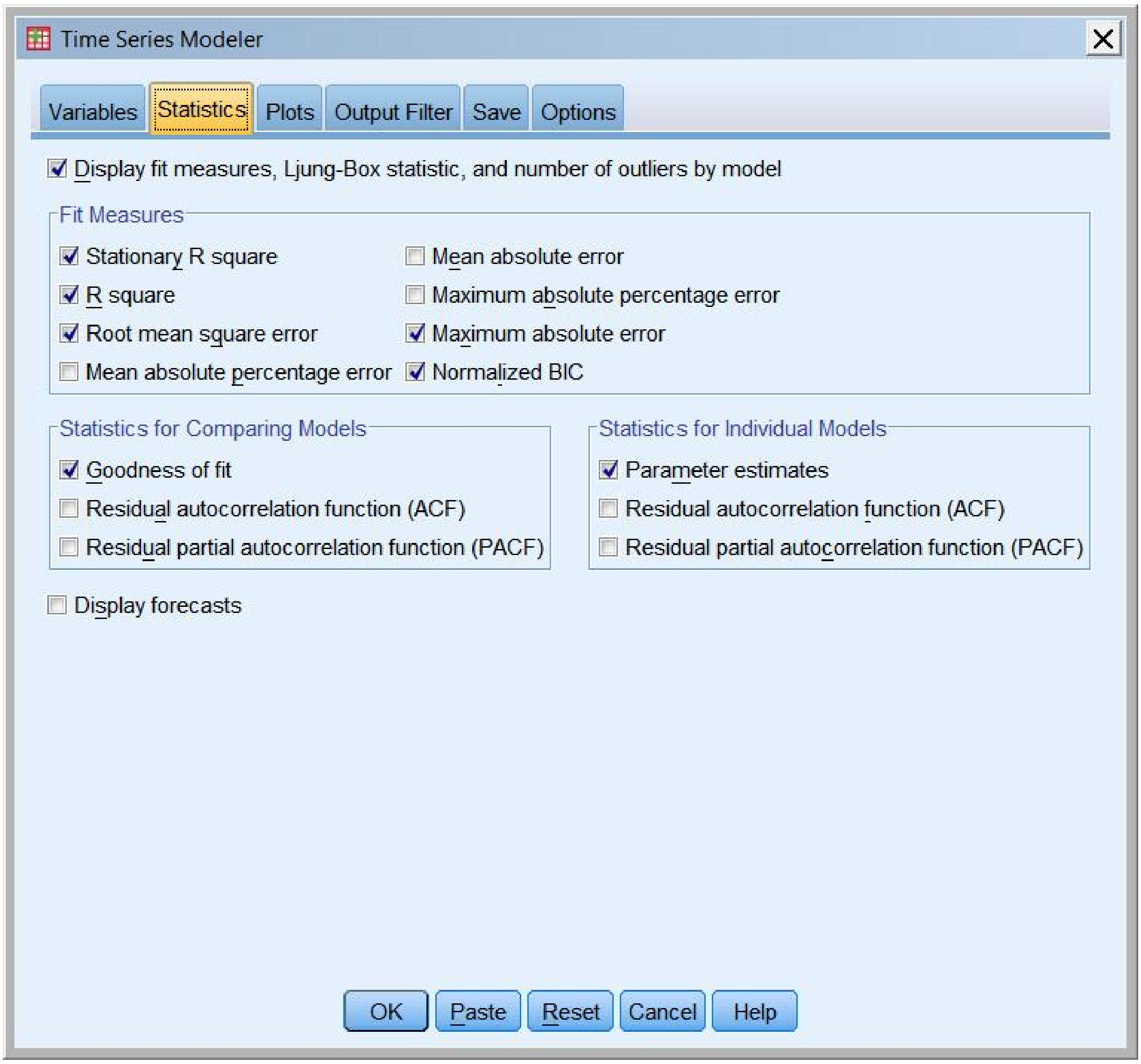


Από τα πεδία που εμφανίζονται στο παράθυρο, στο πεδίο των εξαρτημένων μεταβλητών (Dependent Variables) καταχωρείται η μεταβλητή "Price" που δηλώνει τη τιμή της μετοχής. Εν συνεχεία ως επιλεγμένη μέθοδο (Method) επιλέγεται το υπόδειγμα ARIMA. Έπειτα, στα κριτήρια (Criteria…) καταχωρούνται οι τιμές των p,d,q που ορίστηκαν νωρίτερα και συγκεκριμένα στη στήλη «Nonseasonal».

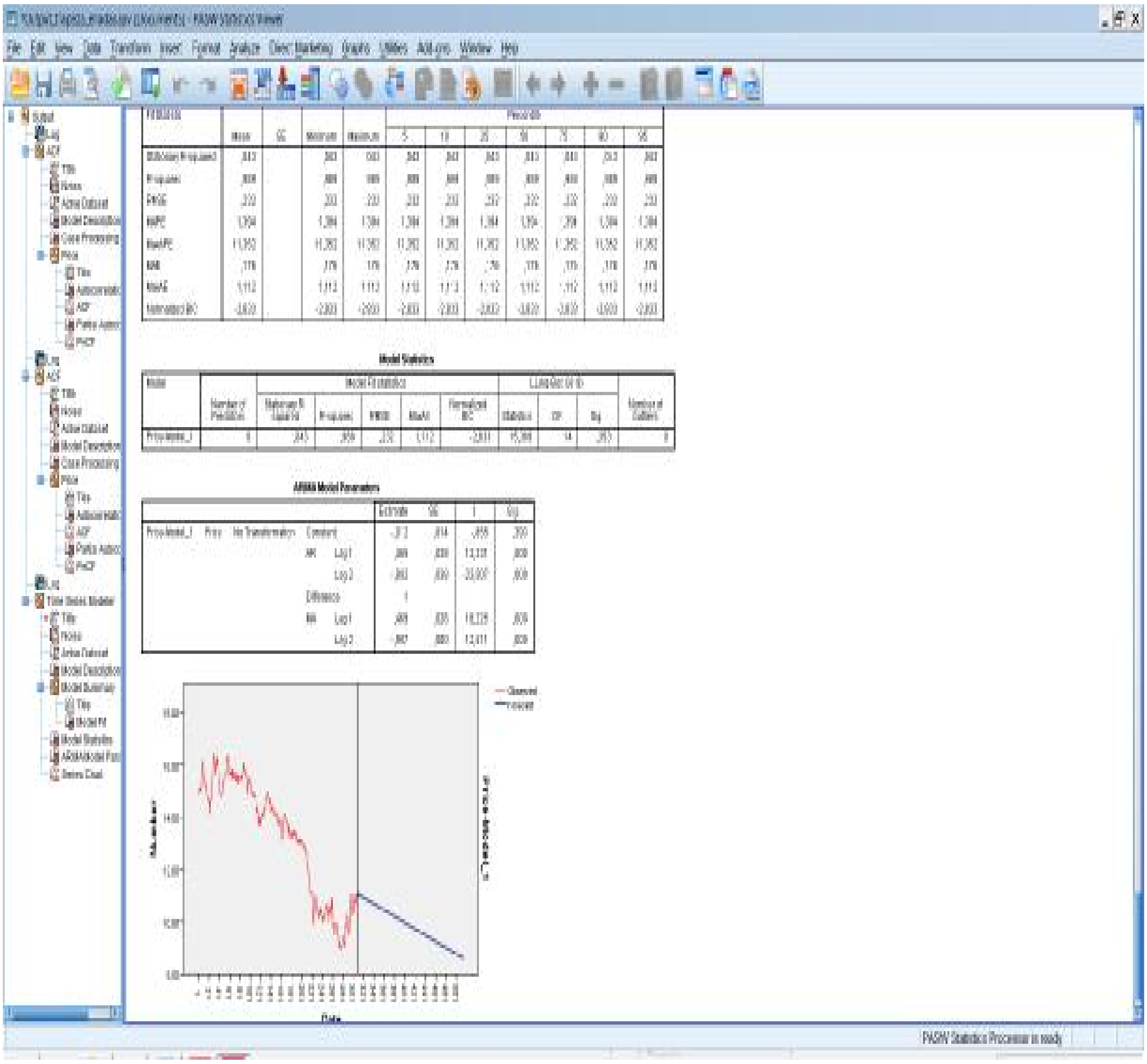


Εικόνα 6: Προσδιορισμός του μοντέλου ARIMA(2,1,2)

Εν συνεχεία, επιλέγοντας την καρτέλα Statistics ορίζονται οι μετρικές που θέλει ο χρήστης να χρησιμοποιήσει:



Αφού ολοκληρωθεί η διαδικασία εισαγωγής της μεθόδου, το πρόγραμμα επιστρέφει στην έξοδό του (output) τα αποτελέσματα μεταξύ των οποίων και τον πίνακα “ARIMA Model Parameters”, στον οποίο εμφανίζονται οι υπολογιζόμενες τιμές των παραμέτρων α_0, α_1, α_2, θ_1 και θ_2.



ARIMA Model Parameters

					Estimate	SE	t	Sig.
Price-Model_1	Price	No Transformation	Constant		-,012	,014	-,855	,393
			AR	Lag 1	,465	,038	12,231	,000
				Lag 2	-,892	,038	-23,507	,000
			Difference		1			
			MA	Lag 1	,469	,026	18,225	,000
				Lag 2	-,997	,080	-12,511	,000

Εικόνα 7: Υπολογισμός των παραμέτρων του υποδείγματος

Όπως εμφανίζεται λοιπόν και στα αποτελέσματα του πίνακα “ARIMA Model Parameters”, οι τιμές των παραμέτρων είναι:

- α_0 = -0.012
- α_1 = 0.485
- α_2 = -0.892
- θ_1 = 0.469
- θ_2 = -0.997

Με την ολοκλήρωση της διαδικασίας εύρεσης των παραμέτρων, επόμενο βήμα είναι ο έλεγχος στατιστικής σημαντικότητας τους υποδείγματος ARIMA(2,1,2) προκειμένου να βρεθούν οι παράμετροι που στατιστικά σημαντικοί.

Για τον έλεγχο λαμβάνονται υπόψη οι ακόλουθες υποθέσεις:

- Η μηδενική υπόθεση H_0: Sig = 0, που δηλώνει ότι ο εκτιμητής είναι στατιστικά μη σημαντικός και
- Η υπόθεση H_1: Sig ≠ 0 που δηλώνει ότι ο εκτιμητής είναι στατιστικά σημαντικός.

Σύμφωνα με τα παραπάνω προκύπτει ότι:

Για την παράμετρο α_0: $Sig_1 = 0.393 > \alpha = 0.05$, άρα αποδοχή της υπόθεσης H_0.

Για την παράμετρο α_1: $Sig_1 = 0 < \alpha = 0.05$, άρα αποδοχή της υπόθεσης H_1.

Για την παράμετρο α_2: $Sig_1 = 0 < \alpha = 0.05$, άρα αποδοχή της υπόθεσης H_1.

Για την παράμετρο θ_1: $Sig_1 = 0 < \alpha = 0.05$, άρα αποδοχή της υπόθεσης H_1.

Για την παράμετρο θ_2: $Sig_1 = 0 < \alpha = 0.05$, άρα αποδοχή της υπόθεσης H_1.

Από τον έλεγχο που προηγήθηκε διαπιστώνεται ότι όλοι οι εκτιμητές είναι στατιστικά σημαντικοί (πλην του σταθερού όρου ο οποίος δεν επηρεάζει), άρα θεωρείται ότι το υπόδειγμα ARIMA(2,1,2) ως το καταλληλότερο.

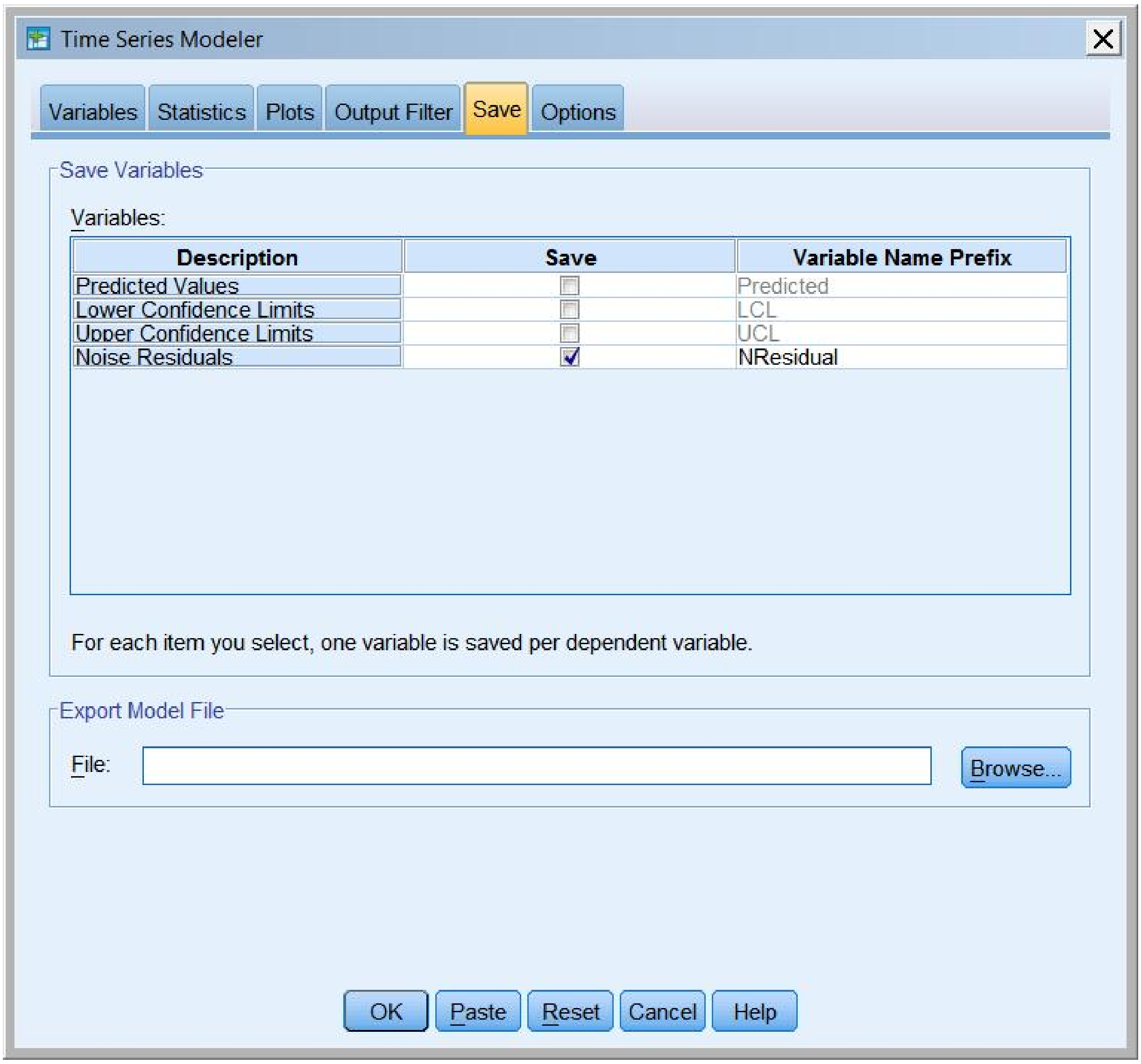
Η μαθηματική διατύπωση του υποδείγματος δίνεται πλέον από την έκφραση:

$$Y_t = -0.012 + 0.465y_{t-1} - 0.892y_{t-2} + 0.469\varepsilon_{t-1} - 0.997\varepsilon_{t-2} + \varepsilon_t$$

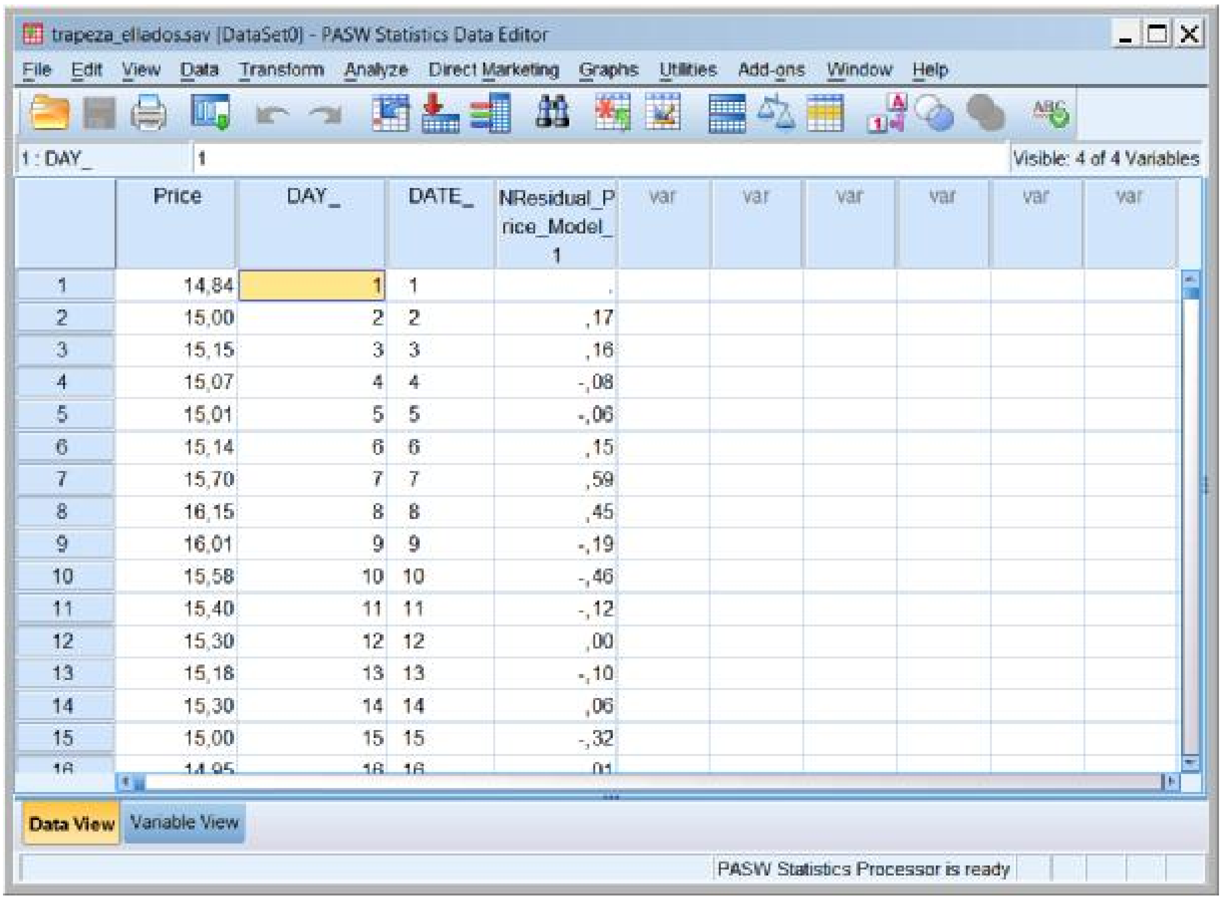
Προκειμένου το υπόδειγμα που εξήχθηκε να θεωρηθεί αποδεκτό, θα πρέπει να περιγράφει τη διαδικασία που παρήγαγε τα δεδομένα. Αυτό σημαίνει ότι τα κατάλοιπα (residuals) θα πρέπει να είναι λευκός θόρυβος ή με άλλα λόγια θα πρέπει τα κατάλοιπα να μην αυτοσυσχετίζονται.

Συνεπώς, και πάλι θα γίνει έλεγχος στασιμότητας, αυτή τη φορά όμως για τα κατάλοιπα επαληθεύοντας με αυτό τον τρόπο ότι όντως τα αποτελέσματα είναι αληθή για το υπόδειγμα ARIMA(2,1,2).

Από τη διαδρομή Analyze **à** Forecasting **à** Create Models ανοίγει το παράθυρο “Time Series Modeler” όπου στη καρτέλα “Save” επιλέγεται η ενέργεια “Noise Residuals”.

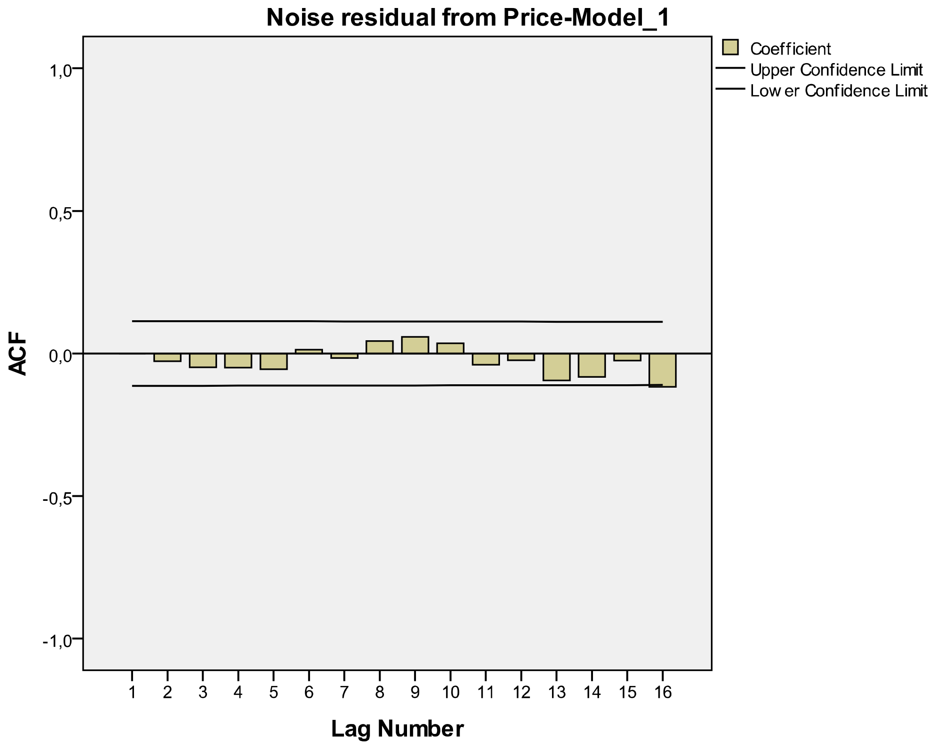


Με αυτό τον τρόπο εμφανίζεται η μεταβλητή "NResidual_Price_Model_1" στη περιοχή εισαγωγής δεδομένων (data entry) στο SPSS, η οποία φανερώνει τα κατάλοιπα.



	Price	DAY_	DATE_	NResidual_Price_Model_1	var	var	var	var	var	var
1	14,84	1	1	.						
2	15,00	2	2	,17						
3	15,15	3	3	,16						
4	15,07	4	4	-,08						
5	15,01	5	5	-,06						
6	15,14	6	6	,15						
7	15,70	7	7	,59						
8	16,15	8	8	,45						
9	16,01	9	9	-,19						
10	15,58	10	10	-,46						
11	15,40	11	11	-,12						
12	15,30	12	12	,00						
13	15,18	13	13	-,10						
14	15,30	14	14	,06						
15	15,00	15	15	-,32						
16	14,95	16	16	,01						

Εν συνεχεία κατά τα γνωστά, γίνεται ο έλεγχος στασιμότητας για τη μεταβλητή "NResidual_Price_Model_1" μέσω της μελέτης του πίνακα αυτοσυσχετίσεων.



Εικόνα 8: Έλεγχος Αυτοσυσχετίσεων των καταλοίπων

Autocorrelations

Series:Noise residual from Price-Model_1

Lag	Autocorrelation	Std. Error[a]	Box-Ljung Statistic		
			Value	df	Sig.[b]
1	,000	,057	,000	1	,999
2	-,028	,057	,239	2	,888
3	-,048	,057	,961	3	,811
4	-,049	,057	1,723	4	,786
5	-,055	,056	2,690	5	,748
6	,013	,056	2,746	6	,840
7	-,016	,056	2,827	7	,901
8	,044	,056	3,427	8	,905
9	,058	,056	4,491	9	,876
10	,036	,056	4,904	10	,898
11	-,040	,056	5,409	11	,910
12	-,023	,056	5,584	12	,936
13	-,095	,056	8,500	13	,810
14	-,082	,056	10,677	14	,711

15	-,025	,055	10,878	15	,761
16	-,117	,055	15,332	16	,500

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

b. Based on the asymptotic chi-square approximation.

Όπως διαφαίνεται και από το διάγραμμα των αυτοσυσχετίσων για τα κατάλοιπα, δεν υπάρχει αυτοσυσχέτιση για καμία χρονική υστέρηση και άρα η σειρά είναι στάσιμη.

Ως επαλήθευση χρησιμοποιούμε τα στοιχεία του ανωτέρω πίνακα και ακολουθούμε τη διαδικασία που περιγράφτηκε νωρίτερα για τον έλεγχο στασιμότητας.

Χρησιμοποιήσουμε λοιπόν τις εξής δυο υποθέσεις:

- Την μηδενική υπόθεση H_0: Sig = 0, που δηλώνει ότι δεν υπάρχει αυτοσυσχέτιση και άρα η χρονοσειρά είναι στάσιμη και
- Την υπόθεση H_1: Sig ≠ 0 που δηλώνει ότι υπάρχει αυτοσυσχέτιση και άρα η χρονοσειρά δεν είναι στάσιμη.

Σύμφωνα με το παραπάνω σενάριο ακολουθεί ο έλεγχος στασιμότητας και για τις 16 χρονικές υστερήσεις.

Lag = 1: $Sig_1 = 0.999 > \alpha = 0.05$, άρα αποδοχή της υπόθεσης H_0.

Lag = 2: $Sig_2 = 0.888 > \alpha = 0.05$, άρα αποδοχή της υπόθεσης H_0.

Lag = 3: $Sig_3 = 0.811 > \alpha = 0.05$, άρα αποδοχή της υπόθεσης H_0.

Lag = 4: $Sig_4 = 0.786 > \alpha = 0.05$, άρα αποδοχή της υπόθεσης H_0.

Lag = 5: $Sig_5 = 0.748 > \alpha = 0.05$, άρα αποδοχή της υπόθεσης H_0.

Lag = 6: $Sig_6 = 0.840 > \alpha = 0.05$, άρα αποδοχή της υπόθεσης H_0.

Lag = 7: $Sig_7 = 0.901 > \alpha = 0.05$, άρα αποδοχή της υπόθεσης H_0.

Lag = 8: $Sig_8 = 0.905 > \alpha = 0.05$, άρα αποδοχή της υπόθεσης H_0.

Lag = 9: $Sig_9 = 0.876 > α = 0.05$, άρα αποδοχή της υπόθεσης H_0.

Lag = 10: $Sig_{10} = 0.898 > α = 0.05$, άρα αποδοχή της υπόθεσης H_0.

Lag = 11: $Sig_{11} = 0.910 > α = 0.05$, άρα αποδοχή της υπόθεσης H_0.

Lag = 12: $Sig_{12} = 0.936 > α = 0.05$, άρα αποδοχή της υπόθεσης H_0.

Lag = 13: $Sig_{13} = 0.810 > α = 0.05$, άρα αποδοχή της υπόθεσης H_0.

Lag = 14: $Sig_{14} = 0.711 > α = 0.05$, άρα αποδοχή της υπόθεσης H_0.

Lag = 15: $Sig_{15} = 0.761 > α = 0.05$, άρα αποδοχή της υπόθεσης H_0.

Lag = 16: $Sig_{16} = 0.500 > α = 0.05$, άρα αποδοχή της υπόθεσης H_0.

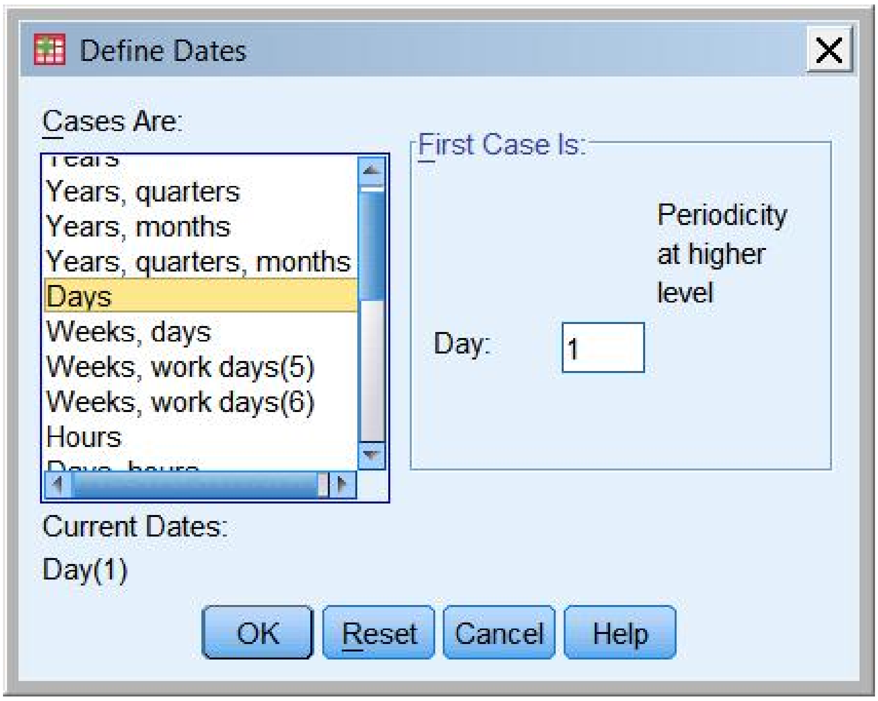
Όπως λοιπόν αποδείχτηκε και από τον έλεγχο των καταλοίπων, η διαδικασία είναι στάσιμη για όλες τις χρονολογικές υστερήσεις. Άρα λοιπόν το υπόδειγμα ARIMA(2,1,2) επαληθεύεται ότι είναι το πλέον κατάλληλο για να περιγράψει τη διαδικασία.

7.2. Motoroil

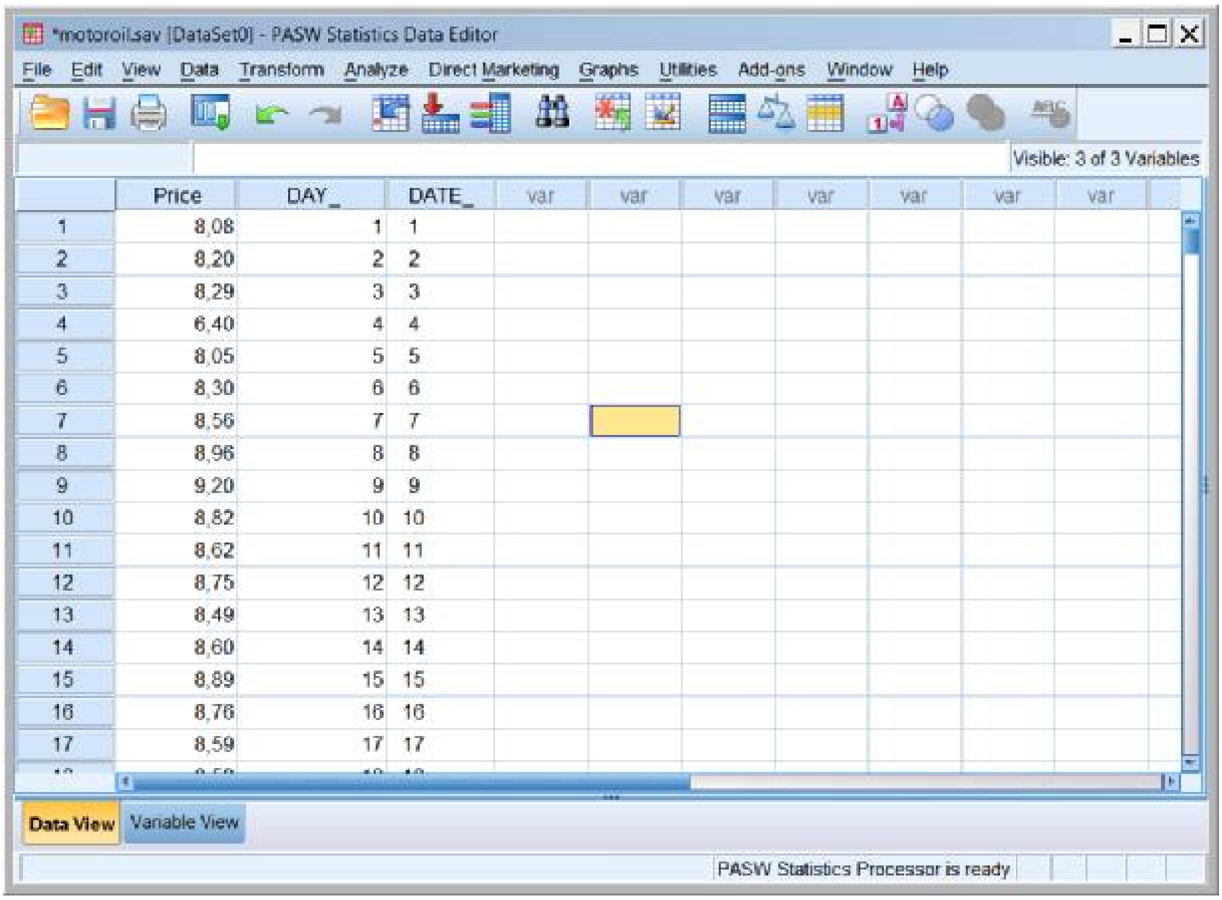
Στη συνέχεια θα ακολουθήσει η ίδια μεθοδολογία για τις τιμές κλεισίματος της μετοχής της Motoroil για τη περίοδο 21/11/2013 – 19/02/2015, όπως αυτές παρέχονται από την ιστοσελίδα της Ναυτεμπορικής[3].

Αφού αντιγραφούν τις τιμές κλεισίματος της μετοχής στο SPSS, ορίζεται η περιοδικότητα των δεδομένων. Συγκεκριμένα, από τη διαδρομή Data **à** Define Dates **à** Days, ορίζεται Day = 1, γεγονός που δηλώνει ότι η περιοδικότητα των δεδομένων είναι ανά ημέρα.

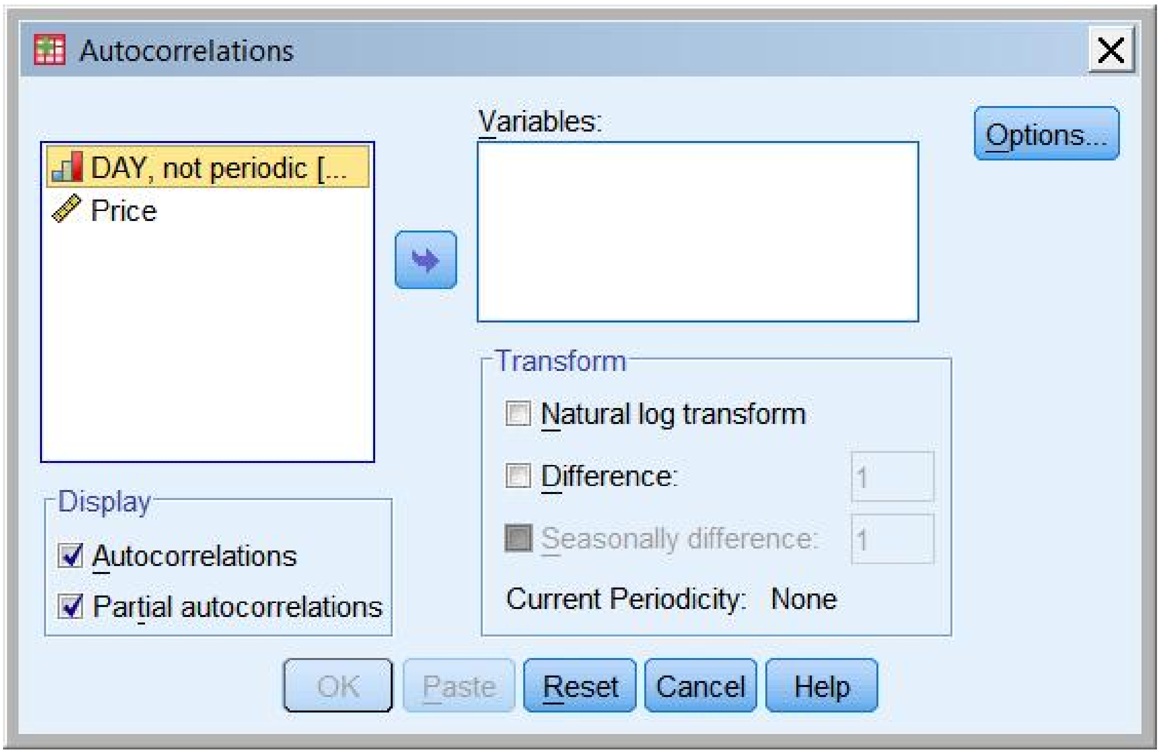
[3] http://www.naftemporiki.gr/finance/quote/tell.ath/ellados-trapeza-ko?tab=4&mode=3



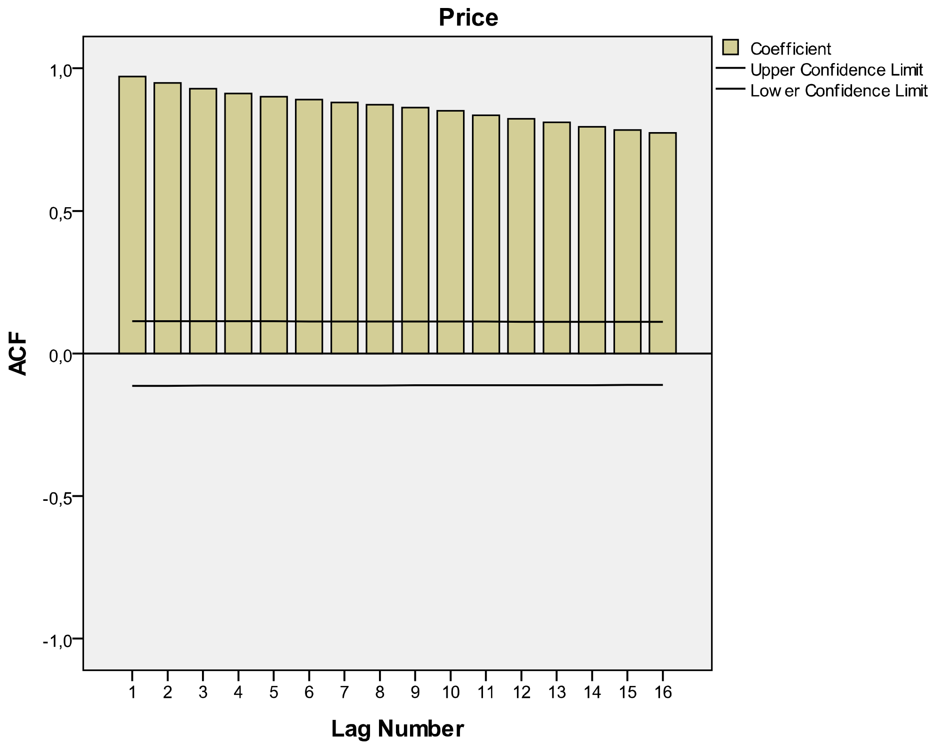
Αυτό έχει ως συνέπεια την εισαγωγή των μεταβλητών “DAY_ και DATE_” στο SPSS.



Σύμφωνα με το πρώτο στάδιο της μεθοδολογίας Box – Jenkins, θα πρέπει αρχικά να γίνει έλεγχος στασιμότητας της χρονοσειράς, προκειμένου να προσδιοριστεί το κατάλληλο μοντέλο. Ακλουθώντας τη διαδρομή Analyze **à** Forecasting **à** Autocorrelations προκύπτει το ακόλουθο παράθυρο διαλόγου.



Στο πεδίο των μεταβλητών (Variables) εισέρχεται η μεταβλητή «Price» που υποδηλώνει τη τιμή της μετοχής σε ένα δεδομένο χρονικό σημείο. Στην αρχική δοκιμή, απαιτείται να εμφανιστούν τα αποτελέσματα των αυτοσυσχετίσεων. Επιπλέον, θεωρείται αρχικά ότι δεν χρειάζεται να οριστούν διαφορές (Difference). Τα αποτελέσματα που λαμβάνονται, εμφανίζονται στον ακόλουθα πίνακα αυτοσυσχετίσεων.



Εικόνα 9: Έλεγχος στασιμότητας βάση των αυτοσυσχετίσεων της μετοχής της Motoroil

Autocorrelations

Series:Price

Lag	Autocorrelation	Std. Error[a]	Box-Ljung Statistic		
			Value	df	Sig.[b]
1	,971	,057	293,281	1	,000
2	,949	,057	574,209	2	,000
3	,929	,057	844,280	3	,000
4	,912	,056	1105,389	4	,000
5	,900	,056	1360,559	5	,000
6	,890	,056	1610,974	6	,000
7	,881	,056	1856,908	7	,000
8	,872	,056	2098,810	8	,000
9	,862	,056	2336,019	9	,000
10	,851	,056	2567,934	10	,000
11	,835	,056	2792,308	11	,000
12	,822	,056	3010,510	12	,000
13	,810	,056	3223,112	13	,000

14	,795	,055	3428,377	14	,000
15	,784	,055	3628,638	15	,000
16	,773	,055	3824,025	16	,000

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

b. Based on the asymptotic chi-square approximation.

Στον πίνακα των αυτοσυσχετίσεων που επιστρέφεται από τον πρόγραμμα, εμφανίζονται πληροφορίες που σχετίζονται με:

Lag: Χρονολογικές υστερήσεις.

Autocorrelation: Οι τιμές του συντελεστή αυτοσυσχέτισης για κάθε χρονολογική υστέρηση.

Std. Error: Τυπικό σφάλμα.

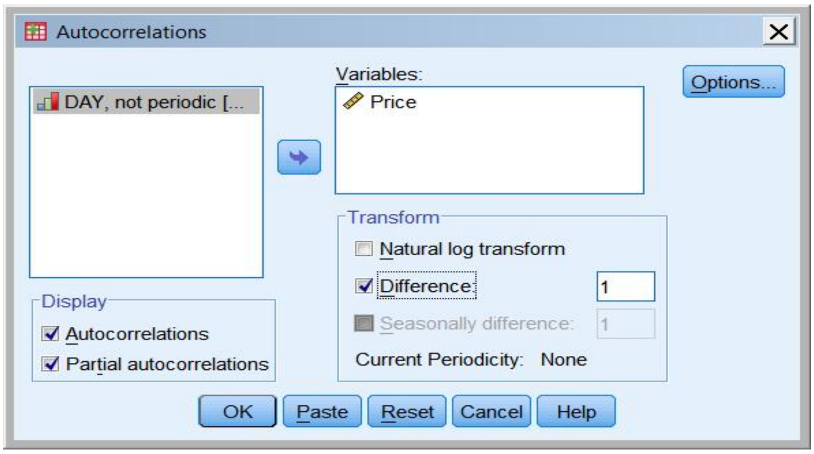
Value: Οι τιμές της στατιστικής συνάρτησης Box – Ljung

df: Οι βαθμοί ελευθερίας.

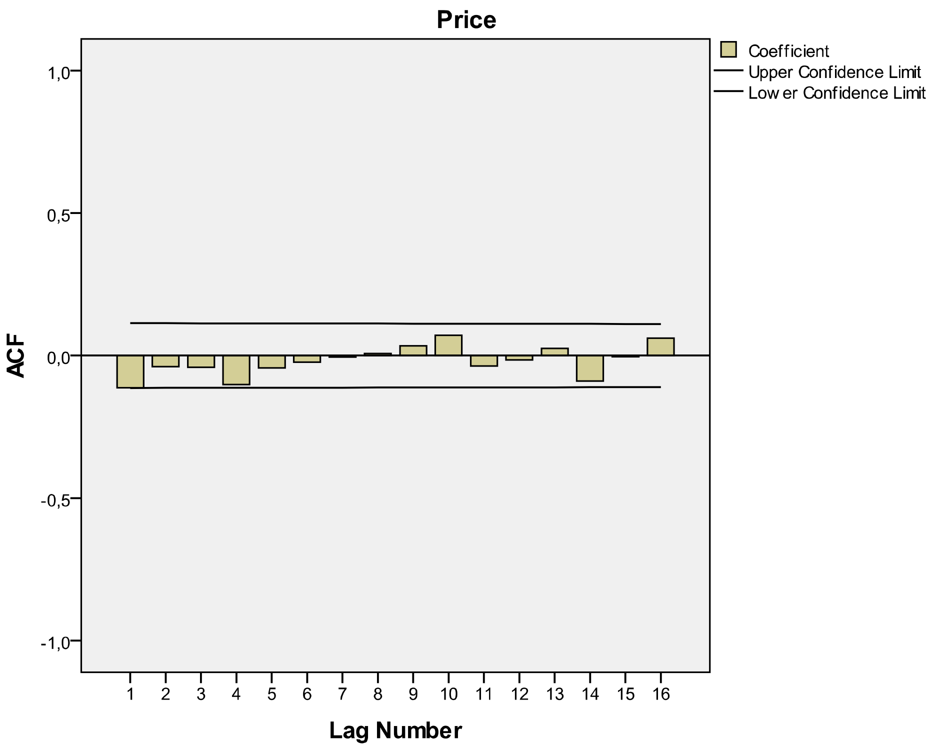
Sig: Οι τιμές πιθανότητας της στατιστικής συνάρτησης.

Σύμφωνα με τα αποτελέσματα που εξήχθησαν, οι τιμές του συντελεστή αυτοσυσχέτισης ρ είναι διάφορες του μηδενός και πολύ κοντά στη τιμή 1, γεγονός που σημαίνει ότι υπάρχει θετική αυτοσυσχέτιση και μάλιστα ισχυρή.

Άρα λοιπόν κατά τον πρώτο έλεγχο που εφαρμόστηκε διαπιστώθηκε ότι η σειρά δεν είναι στάσιμη και άρα θα πρέπει να επαναληφθεί ο έλεγχος στασιμότητας λαμβάνοντας υπόψη τις πρώτες διαφορές.



Τα αποτελέσματα που προκύπτουν δίνονται ακολούθως.



Εικόνα 10: Αυτοσυσχετίσεις μετά από την εφαρμογή των πρώτων διαφορών

Autocorrelations

Series:Price

Lag	Autocorrelation	Std. Error[a]	Box-Ljung Statistic

			Value	df	Sig.[b]
1	-,114	,057	3,996	1	,046
2	-,038	,057	4,453	2	,108
3	-,041	,057	4,985	3	,173
4	-,102	,057	8,223	4	,084
5	-,044	,056	8,824	5	,116
6	-,023	,056	8,994	6	,174
7	-,005	,056	9,001	7	,253
8	,007	,056	9,015	8	,341
9	,034	,056	9,391	9	,402
10	,071	,056	10,983	10	,359
11	-,037	,056	11,414	11	,409
12	-,015	,056	11,484	12	,488
13	,025	,056	11,683	13	,554
14	-,090	,056	14,297	14	,428
15	-,004	,055	14,302	15	,503
16	,061	,055	15,498	16	,489

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

b. Based on the asymptotic chi-square approximation.

Πλέον, παρατηρείται ότι οι τιμές του συντελεστή για τις διάφορες φάσεις υστέρησης έχει μεταβληθεί. Ενδεικτικά ακολουθεί ο αναλυτικός έλεγχος στασιμότητας για όλες τις χρονολογικές υστερήσεις.

Για τον έλεγχο στασιμότητας θα χρησιμοποιήσουμε δυο υποθέσεις:

- Την μηδενική υπόθεση H_0: Sig = 0, που δηλώνει ότι δεν υπάρχει αυτοσυσχέτιση και άρα η χρονοσειρά είναι στάσιμη και
- Την υπόθεση H_1: Sig ≠ 0 που δηλώνει ότι υπάρχει αυτοσυσχέτιση και άρα η χρονοσειρά δεν είναι στάσιμη.

Σύμφωνα με το παραπάνω σενάριο ακολουθεί ο έλεγχος στασιμότητας και για τις 16 χρονικές υστερήσεις.

Lag = 1: $Sig_1 = 0.046 < α = 0.05$, άρα αποδοχή της υπόθεσης H_1.

Lag = 2: $Sig_2 = 0.108 > \alpha = 0.05$, άρα αποδοχή της υπόθεσης H_0.

Lag = 3: $Sig_3 = 0.173 > \alpha = 0.05$, άρα αποδοχή της υπόθεσης H_0.

Lag = 4: $Sig_4 = 0.084 > \alpha = 0.05$, άρα αποδοχή της υπόθεσης H_0.

Lag = 5: $Sig_5 = 0.116 > \alpha = 0.05$, άρα αποδοχή της υπόθεσης H_0.

Lag = 6: $Sig_6 = 0.174 > \alpha = 0.05$, άρα αποδοχή της υπόθεσης H_0.

Lag = 7: $Sig_7 = 0.253 > \alpha = 0.05$, άρα αποδοχή της υπόθεσης H_0.

Lag = 8: $Sig_8 = 0.341 > \alpha = 0.05$, άρα αποδοχή της υπόθεσης H_0.

Lag = 9: $Sig_9 = 0.402 > \alpha = 0.05$, άρα αποδοχή της υπόθεσης H_0.

Lag = 10: $Sig_{10} = 0.359 > \alpha = 0.05$, άρα αποδοχή της υπόθεσης H_0.

Lag = 11: $Sig_{11} = 0.409 > \alpha = 0.05$, άρα αποδοχή της υπόθεσης H_0.

Lag = 12: $Sig_{12} = 0.488 > \alpha = 0.05$, άρα αποδοχή της υπόθεσης H_0.

Lag = 13: $Sig_{13} = 0.554 > \alpha = 0.05$, άρα αποδοχή της υπόθεσης H_0.

Lag = 14: $Sig_{14} = 0.428 > \alpha = 0.05$, άρα αποδοχή της υπόθεσης H_0.

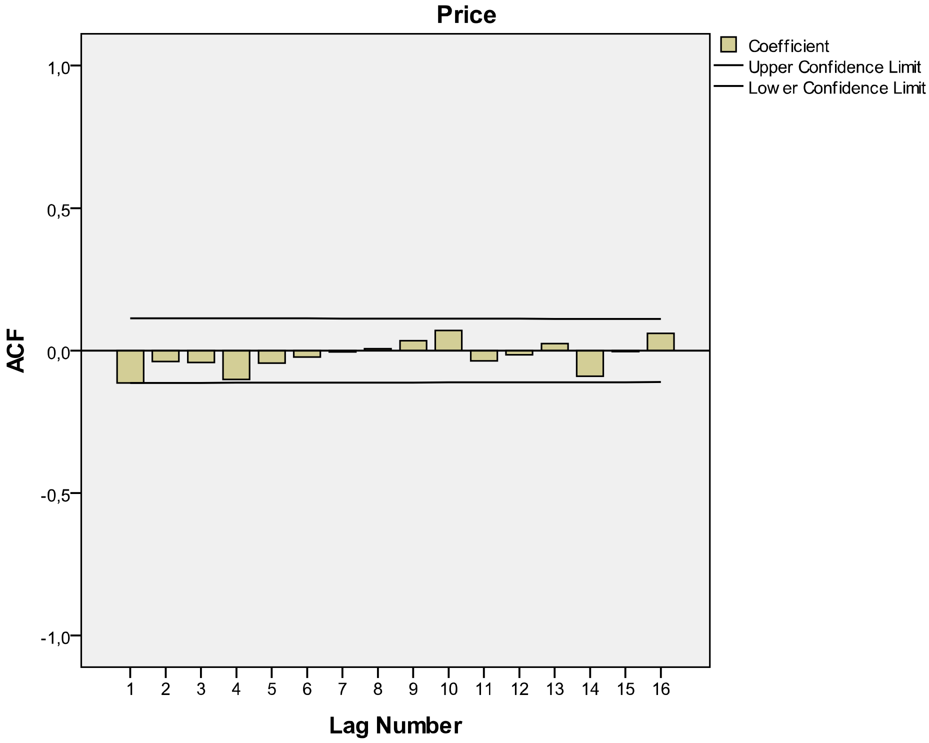
Lag = 15: $Sig_{15} = 0.503 > \alpha = 0.05$, άρα αποδοχή της υπόθεσης H_0.

Lag = 16: $Sig_{16} = 0.489 > \alpha = 0.05$, άρα αποδοχή της υπόθεσης H_0.

Με βάση την ανωτέρω ανάλυση, διαπιστώθηκε ότι η χρονοσειρά παραμένει στάσιμη για όλες τις χρονικές υστερήσεις (πλην της πρώτης για 0.002), άρα θεωρείται στάσιμη λαμβάνοντας υπόψη τις πρώτες διαφορές.

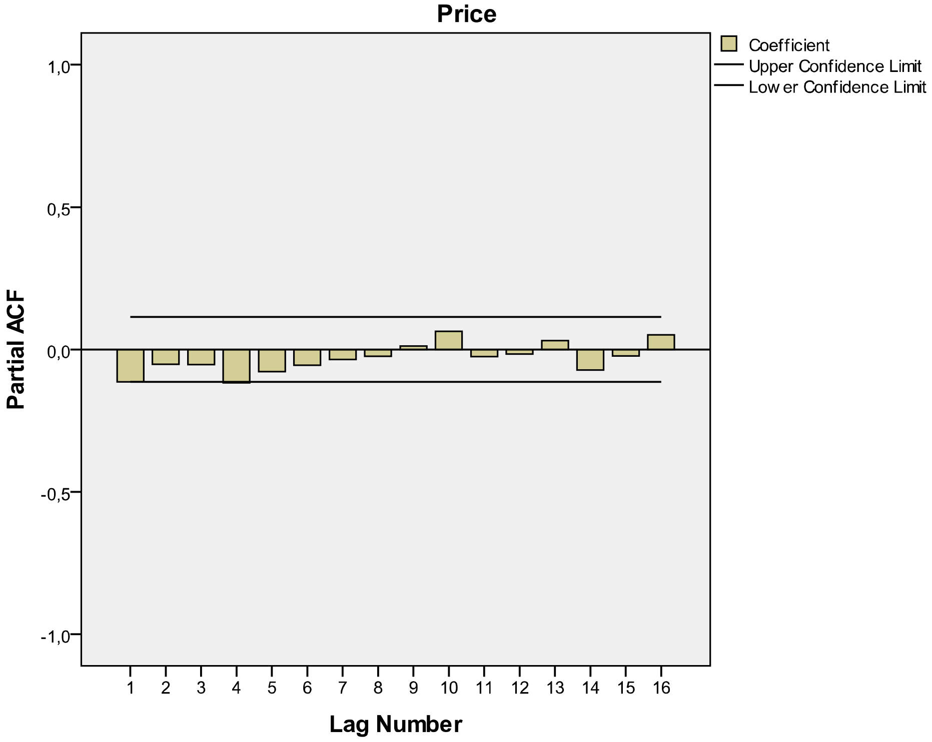
Στο επόμενο στάδιο γίνεται ο προσδιορισμός της τάξεως q και p με τη χρήση των γραφημάτων αυτοσυσχετίσεων (ACF) και των μερικών αυτοσυσχετίσεων (Partial ACF) αντίστοιχα.

Παρατηρώντας το γράφημα ACF διαπιστώνεται ότι υπάρχει μόνο μια τιμή που είναι στο όριο ή εκτός ορίου του διαστήματος εμπιστοσύνης και συγκεκριμένα στο κάτω όριο για τη χρονική υστέρηση lag=1. Αυτό σημαίνει ότι θα θεωρηθεί ότι το q = 1.



Εικόνα 11: Προσδιορισμός της τάξεως q

Ακολούθως από το γράφημα Partial ACF διαπιστώνεται ότι υπάρχει μια τιμή που είναι στο όριο ή εκτός ορίου του διαστήματος εμπιστοσύνης και συγκεκριμένα στο κάτω όριο για τη χρονική υστέρηση lag=4. Αυτό σημαίνει ότι θα θεωρηθεί ότι το p = 1.



Εικόνα 12: Προσδιορισμός της τάξεως p

Άρα λοιπόν από τη μελέτη των διαγραμμάτων ACF και Partial ACF προκύπτει ότι q=1 και p=1. Σύμφωνα με τα δεδομένα αυτά, το γενικό υπόδειγμα ARIMA(p,d,q) το οποίο προσδιορίζει καλύτερα τη χρονοσειρά παίρνει τη μορφή:

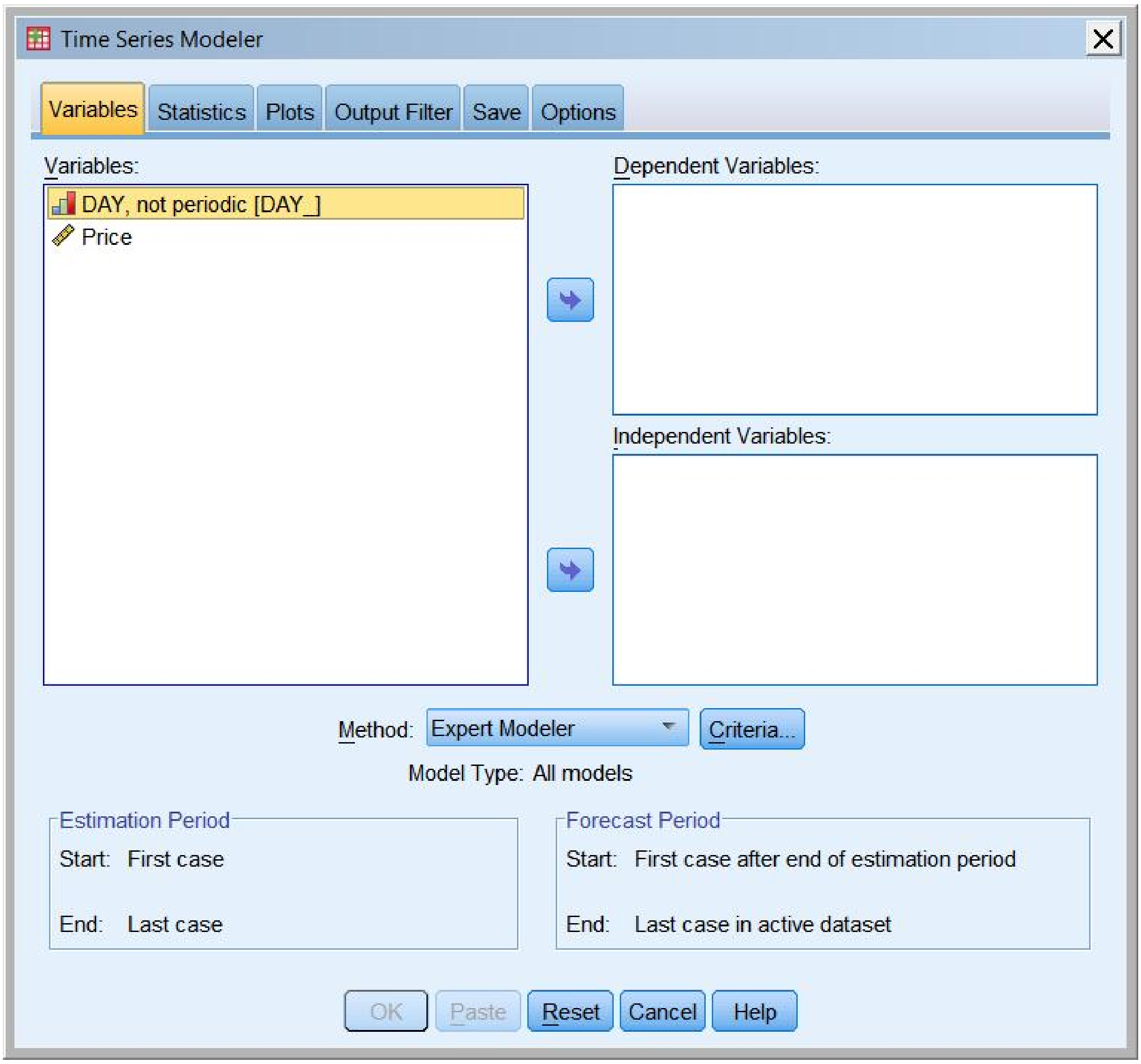
ARIMA(1,1,1)

Η θεωρητική μορφή του υποδείγματος προσδιορίζεται από την έκφραση:

$$Y_t = a_0 + a_1 y_{t-1} + \theta_1 \varepsilon_{t-1} + \varepsilon_t$$

Στο επόμενο στάδιο προσδιορίζονται οι παράμετροι α_0, α_1, και θ_1 και στη συνέχεια πραγματοποιείται ο έλεγχος στατιστικής σημαντικότητας αυτών των παραμέτρων.

Για τον προσδιορισμό των παραμέτρων, αρχικά δηλώνουμε το μοντέλο μας στο SPSS ακολουθώντας τη διαδρομή Analyze **à** Forecasting **à** Create Models…, η οποία μας οδηγεί στο ακόλουθο παράθυρο.



Από τα πεδία που εμφανίζονται στο παράθυρο, στο πεδίο των εξαρτημένων μεταβλητών (Dependent Variables) καταχωρείται η μεταβλητή “Price” που δηλώνει τη τιμή της μετοχής. Εν συνεχεία ως επιλεγμένη μέθοδο (Method) επιλέγεται το υπόδειγμα ARIMA. Έπειτα, στα κριτήρια (Criteria…) καταχωρούνται οι τιμές των p,d,q που ορίστηκαν νωρίτερα και συγκεκριμένα στη στήλη «Nonseasonal».

Αφού ολοκληρωθεί η διαδικασία εισαγωγής της μεθόδου, το πρόγραμμα επιστρέφει στην έξοδό του (output) τα αποτελέσματα μεταξύ των οποίων και τον πίνακα “ARIMA Model Parameters”, στον οποίο εμφανίζονται οι υπολογιζόμενες τιμές των παραμέτρων α_0, α_1 και θ_1.

ARIMA Model Parameters

					Estimate	SE	t	Sig.
Price-Model_1	Price	No Transformation	Constant		-,004	,009	-,428	,669
			AR	Lag 1	,683	,137	4,985	,000
			Difference		1			
			MA	Lag 1	,819	,108	7,573	,000

Εικόνα 13: **Προσδιορισμός συντελεστών του υποδείγματος** ARIMA(1,1,1)

Όπως εμφανίζεται λοιπόν και στα αποτελέσματα του πίνακα “ARIMA Model Parameters”, οι τιμές των παραμέτρων είναι:

- $\alpha_0 = -0.004$
- $\alpha_1 = 0.683$
- $\theta_1 = 0.819$

Με την ολοκλήρωση της διαδικασίας εύρεσης των παραμέτρων, επόμενο βήμα είναι ο έλεγχος στατιστικής σημαντικότητας τους υποδείγματος ARIMA(1,1,1) προκειμένου να βρεθούν οι παράμετροι που στατιστικά σημαντικοί.

Για τον έλεγχο λαμβάνονται υπόψη οι ακόλουθες υποθέσεις:

- Η μηδενική υπόθεση H_0: Sig = 0, που δηλώνει ότι ο εκτιμητής είναι στατιστικά μη σημαντικός και
- Η υπόθεση H_1: Sig $\neq$ 0 που δηλώνει ότι ο εκτιμητής είναι στατιστικά σημαντικός.

Σύμφωνα με τα παραπάνω προκύπτει ότι:

Για την παράμετρο α_0: $Sig_1 = 0.669 > \alpha = 0.05$, άρα αποδοχή της υπόθεσης H_0.

Για την παράμετρο α_1: $Sig_1 = 0 < \alpha = 0.05$, άρα αποδοχή της υπόθεσης H_1.

Για την παράμετρο θ_2: $Sig_1 = 0 < \alpha = 0.05$, άρα αποδοχή της υπόθεσης H_1.

Από τον έλεγχο που προηγήθηκε διαπιστώνεται ότι όλοι οι εκτιμητές είναι στατιστικά σημαντικοί (πλην του σταθερού όρου ο οποίος δεν επηρεάζει), άρα θεωρείται ότι το υπόδειγμα ARIMA(1,1,1) ως το καταλληλότερο.

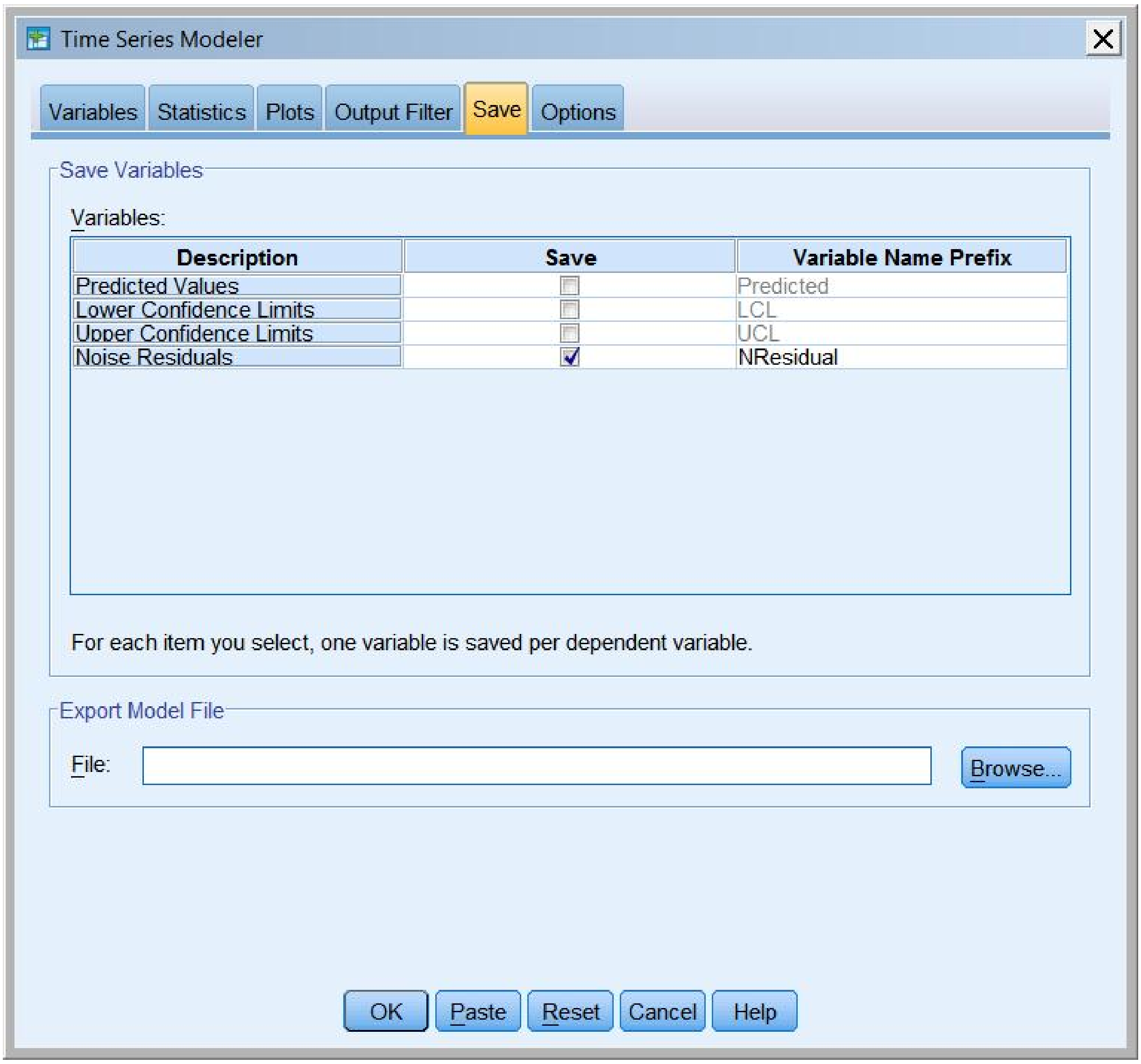
Η μαθηματική διατύπωση του υποδείγματος δίνεται πλέον από την έκφραση:

$$Y_t = -0.004 + 0.683y_{t-1} + 0.819\varepsilon_{t-1} + \varepsilon_t$$

Προκειμένου το υπόδειγμα που εξήχθηκε να θεωρηθεί αποδεκτό, θα πρέπει να περιγράφει τη διαδικασία που παρήγαγε τα δεδομένα. Αυτό σημαίνει ότι τα κατάλοιπα (residuals) θα πρέπει να είναι λευκός θόρυβος ή με άλλα λόγια θα πρέπει τα κατάλοιπα να μην αυτοσυσχετίζονται.

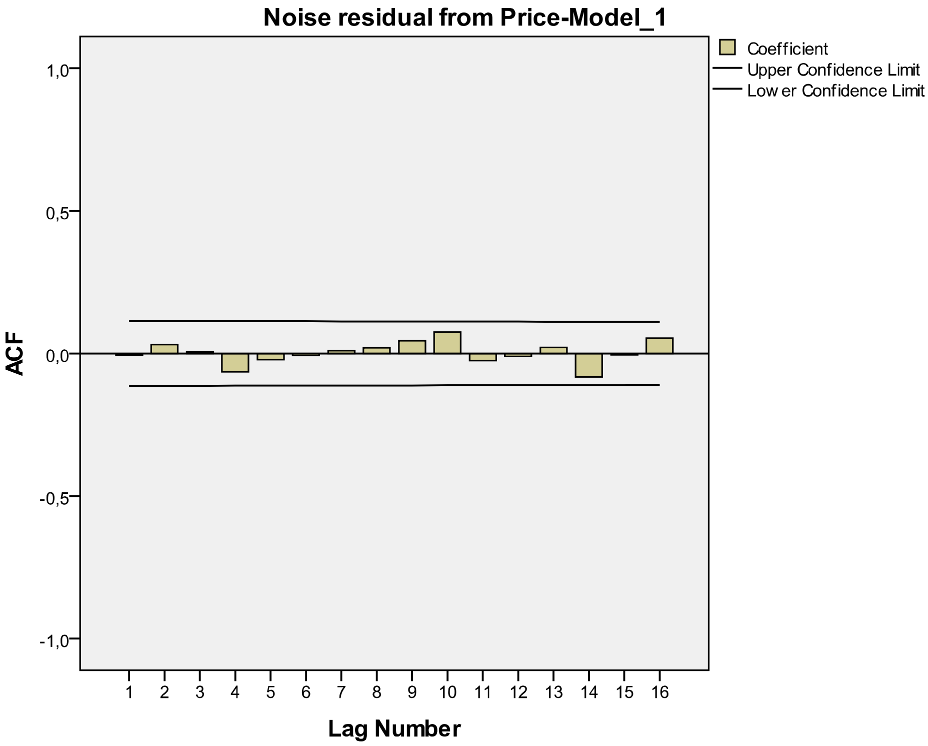
Συνεπώς, και πάλι θα γίνει έλεγχος στασιμότητας, αυτή τη φορά όμως για τα κατάλοιπα επαληθεύοντας με αυτό τον τρόπο ότι όντως τα αποτελέσματα είναι αληθή για το υπόδειγμα ARIMA(1,1,1).

Από τη διαδρομή Analyze **à** Forecasting **à** Create Models ανοίγει το παράθυρο “Time Series Modeler” όπου στη καρτέλα “Save” επιλέγεται η ενέργεια “Noise Residuals”.



Με αυτό τον τρόπο εμφανίζεται η μεταβλητή "NResidual_Price_Model_1" στη περιοχή εισαγωγής δεδομένων (data entry) στο SPSS, η οποία φανερώνει τα κατάλοιπα.

Εν συνεχεία κατά τα γνωστά, γίνεται ο έλεγχος στασιμότητας για τη μεταβλητή "NResidual_Price_Model_1" μέσω της μελέτης του πίνακα αυτοσυσχετίσεων.



Εικόνα 14: Έλεγχος καταλληλότητας του μοντέλου βάση των αυτοσυσχετίσεων των καταλοίπων

Autocorrelations

Series:Noise residual from Price-Model_1

Lag	Autocorrelation	Std. Error[a]	Box-Ljung Statistic		
			Value	df	Sig.[b]
1	-,006	,057	,011	1	,917
2	,031	,057	,304	2	,859
3	,006	,057	,314	3	,957
4	-,064	,057	1,608	4	,807
5	-,022	,056	1,762	5	,881
6	-,007	,056	1,776	6	,939
7	,009	,056	1,804	7	,970
8	,020	,056	1,929	8	,983
9	,045	,056	2,575	9	,979
10	,075	,056	4,360	10	,930
11	-,026	,056	4,569	11	,950
12	-,011	,056	4,605	12	,970
13	,021	,056	4,751	13	,980
14	-,082	,056	6,935	14	,937

15	-,005	,055	6,944	15	,959
16	,054	,055	7,895	16	,952

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

b. Based on the asymptotic chi-square approximation.

Όπως διαφαίνεται και από το διάγραμμα των αυτοσυσχετίσων για τα κατάλοιπα, δεν υπάρχει αυτοσυσχέτιση για καμία χρονική υστέρηση και άρα η σειρά είναι στάσιμη.

Ως επαλήθευση χρησιμοποιούμε τα στοιχεία του ανωτέρω πίνακα και ακολουθούμε τη διαδικασία που περιγράφτηκε νωρίτερα για τον έλεγχο στασιμότητας.

Χρησιμοποιήσουμε λοιπόν τις εξής δυο υποθέσεις:

- Την μηδενική υπόθεση H_0: Sig = 0, που δηλώνει ότι δεν υπάρχει αυτοσυσχέτιση και άρα η χρονοσειρά είναι στάσιμη και
- Την υπόθεση H_1: Sig $\neq$ 0 που δηλώνει ότι υπάρχει αυτοσυσχέτιση και άρα η χρονοσειρά δεν είναι στάσιμη.

Σύμφωνα με το παραπάνω σενάριο ακολουθεί ο έλεγχος στασιμότητας και για τις 16 χρονικές υστερήσεις.

Lag = 1: $Sig_1 = 0.917 > \alpha = 0.05$, άρα αποδοχή της υπόθεσης H_0.

Lag = 2: $Sig_2 = 0.859 > \alpha = 0.05$, άρα αποδοχή της υπόθεσης H_0.

Lag = 3: $Sig_3 = 0.957 > \alpha = 0.05$, άρα αποδοχή της υπόθεσης H_0.

Lag = 4: $Sig_4 = 0.807 > \alpha = 0.05$, άρα αποδοχή της υπόθεσης H_0.

Lag = 5: $Sig_5 = 0.881 > \alpha = 0.05$, άρα αποδοχή της υπόθεσης H_0.

Lag = 6: $Sig_6 = 0.939 > \alpha = 0.05$, άρα αποδοχή της υπόθεσης H_0.

Lag = 7: $Sig_7 = 0.970 > \alpha = 0.05$, άρα αποδοχή της υπόθεσης H_0.

Lag = 8: $Sig_8 = 0.983 > \alpha = 0.05$, άρα αποδοχή της υπόθεσης H_0.

Lag = 9: Sig_9 = 0.979 > α = 0.05, άρα αποδοχή της υπόθεσης H_0.

Lag = 10: Sig_{10} = 0.930 > α = 0.05, άρα αποδοχή της υπόθεσης H_0.

Lag = 11: Sig_{11} = 0.950 > α = 0.05, άρα αποδοχή της υπόθεσης H_0.

Lag = 12: Sig_{12} = 0.970 > α = 0.05, άρα αποδοχή της υπόθεσης H_0.

Lag = 13: Sig_{13} = 0.980 > α = 0.05, άρα αποδοχή της υπόθεσης H_0.

Lag = 14: Sig_{14} = 0.937 > α = 0.05, άρα αποδοχή της υπόθεσης H_0.

Lag = 15: Sig_{15} = 0.959 > α = 0.05, άρα αποδοχή της υπόθεσης H_0.

Lag = 16: Sig_{16} = 0.952 > α = 0.05, άρα αποδοχή της υπόθεσης H_0.

Όπως λοιπόν αποδείχτηκε και από τον έλεγχο των καταλοίπων, η διαδικασία είναι στάσιμη για όλες τις χρονολογικές υστερήσεις. Άρα λοιπόν το υπόδειγμα ARIMA(1,1,1) επαληθεύεται ότι είναι το πλέον κατάλληλο για να περιγράψει τη διαδικασία.

8. ΣΥΜΠΕΡΑΣΜΑΤΑ

Στην ανάλυση που προηγήθηκε, αρχικά παρουσιάστηκαν κάποιες βασικές έννοιες των χρονολογικών σειρών, καθώς και οι διάφορες τεχνικές μέτρησης της μακροπρόθεσμης τάσης. Επίσης, μελετήθηκε η εφαρμογή των αυτοπαλίνδρομων διαδικασιών (AR) και των διαδικασιών κινητού μέσου, καθώς και πως προκύπτουν από το συνδυασμό αυτών οι μικτές διαδικασίες ARMA.

Αφού προηγηθούν τα ανωτέρω, έγινε μια λεπτομερής παρουσίαση της μεθοδολογίας Box-Jenkins και εν συνεχεία πραγματοποιήθηκε η εύρεση του καταλληλότερου υποδείγματος για τη πρόβλεψη των τιμών της μετοχής εταιρειών εισηγμένων στο χρηματιστήριο Αθηνών.

Όπως διαπιστώθηκε δεν υπάρχει ένα συγκεκριμένο υπόδειγμα βάση του οποίου μπορεί κάποιος να οδηγηθεί σε προβλέψεις, ωστόσο έχουν αναπτυχθεί τεχνικές που με βάση τα αποτελέσματα κάποιων μετρικών/δεικτών οδηγούμαστε στο προσδιορισμό των καταλληλότερων μοντέλων από τα οποία επιλέγεται εκείνο που προσδιορίζει καλύτερα τη χρονοσειρά προκειμένου εν συνεχεία να γίνει μια μελλοντική εκτίμηση (πρόβλεψη).

Βιβλιογραφία

1. Δημελή Σ., Σύγχρονες μέθοδοι ανάλυσης χρονολογικών σειρών, 2004.
2. Box, G., E., and Jenkins, G.,M., (1976) Time Series Analysis: Forecasting and Control, 2nd Ed. Holden-Day,Oakland, California.
3. Θαλασσινός Λ., Ανάλυση Χρονολογικών Σειρών, Εκδόσεις Σταμούλης, 1991.
4. Γεωργίου Χρήστου, Εισαγωγή στην Οικονομετρία, Εκδόσεις Gutenberg, 2007.
5. Χρήστος Κ. Φράγκος, Στατιστική Επιχειρήσεων για τις οικονομικές και κοινωνικές επιστήμες, Εκδ. Σταμούλης, Αθήνα, 1998.
6. Κωνσταντίνος Κουνετάς, Σημειώσεις μαθήματος Τεχνικές Προβλέψεων και Ελέγχου, Δεκέμβριος 2012.
7. Theresa Hoang Diem Ngo, Warner Bros. Entertainment Group, Burbank, CA., The Box-Jenkins Methodology for Time Series Models, SAS Global Forum 2013.