



ΤΕΙ ΔΥΤΙΚΗΣ ΕΛΛΑΔΟΣ

ΣΧΟΛΗ ΔΙΟΙΚΗΣΗΣ ΚΑΙ ΟΙΚΟΝΟΜΙΑΣ

ΤΜΗΜΑ ΔΙΟΙΚΗΣΗΣ ΕΠΙΧΕΙΡΗΣΕΩΝ

ΠΤΥΧΙΑΚΗ ΕΡΓΑΣΙΑ

**«ΑΝΑΛΥΣΗ ΧΡΟΝΟΛΟΓΙΚΩΝ ΣΕΙΡΩΝ ΚΑΙ ΟΙ ΕΦΑΡΜΟΓΕΣ ΤΟΥΣ ΣΕ
ΕΠΙΧΕΙΡΗΜΑΤΙΚΑ ΔΕΔΟΜΕΝΑ»**

ΓΙΑΠΙΤΖΑΚΗΣ ΝΙΚΟΛΑΟΣ

ΜΑΝΤΑΣ ΔΗΜΗΤΡΙΟΣ

ΕΠΟΠΤΕΥΟΥΣΑ ΚΑΘΗΓΗΤΡΙΑ: ΠΑΠΑΘΑΝΑΣΟΠΟΥΛΟΥ ΧΡΥΣΑΝΘΗ

ΠΑΤΡΑ, 2018

ΠΕΡΙΛΗΨΗ

Η παρούσα πτυχιακή εργασία πραγματεύεται την ανάλυση και το σχεδιασμό των χρονοσειρών. Στο πρώτο κεφάλαιο παρουσιάζεται ο ορισμός των χρονοσειρών οι ιδιότητες καθώς και η διάκριση σε στάσιμες και μη στάσιμες χρονολογικές σειρές.

Στο δεύτερο κεφάλαιο παρουσιάζονται οι διακρίσεις της ανάλυσης χρονοσειρών συνάρτηση αυτοσυσχέτισης, η μερική αυτοσυσχέτιση και συνάρτηση μερικής αυτοσυσχέτισης και οι διαδικασίες ARIMA και μεθοδολογία δημιουργίας εμπειρικών υποδειγμάτων κατά BOX-JENKINS.

Στο τρίτο κεφάλαιο παρουσιάζονται οι παρεμβάσεις και οι προβλέψεις και η ανάλυση τους μέσα από υποδείγματα.

Στο τέταρτο κεφάλαιο παρουσιάζεται η μελέτη περίπτωσης που πραγματοποιήθηκε και στη συνέχεια ακολουθούν οι βιβλιογραφικές αναφορές που χρησιμοποιήθηκαν.

ABSTRACT

This dissertation deals with the analysis and design of time series. The first chapter presents the definition of time series properties as well as the distinction in stationary and non-stationary time series.

In the second chapter are presented the distinctions of time series analysis of autocorrelation, partial autocorrelation and partial autocorrelation function, and ARIMA procedures and BOX-JENKINS empirical modeling methodology.

In the third chapter are presented the interventions and the forecasts and their analysis through models.

The fourth chapter presents the case study that was carried out, followed by the bibliographic references used.

ΠΕΡΙΕΧΟΜΕΝΑ

ΠΕΡΙΛΗΨΗ	3
ABSTRACT.....	4
ΚΕΦΑΛΑΙΟ 1.....	7
ΑΝΑΛΥΣΗ ΧΡΟΝΟΣΕΙΡΩΝ (TIME SERIES ANALYSIS).....	7
1.1 ΕΙΣΑΓΩΓΗ	7
1.2 ΟΡΙΣΜΟΣ	8
1.3 ΣΤΑΣΙΜΕΣ ΚΑΙ ΜΗ ΣΤΑΣΙΜΕΣ ΧΡΟΝΟΛΟΓΙΚΕΣ ΣΕΙΡΕΣ.	8
1.3.1 ΙΔΙΟΤΗΤΕΣ ΤΩΝ ΣΤΑΣΙΜΩΝ ΔΙΑΔΙΚΑΣΙΩΝ	9
1.3.2 ΟΜΟΓΕΝΩΣ ΜΗ-ΣΤΑΣΙΜΕΣ ΔΙΑΔΙΚΑΣΙΕΣ	10
1.4 ΜΟΝΤΕΛΑ ΧΡΟΝΟΣΕΙΡΩΝ ΚΑΙ ΠΡΟΒΛΕΨΗ.....	11
1.5 ΚΛΑΣΙΚΑ ΥΠΟΔΕΙΓΜΑΤΑ ΑΝΑΛΥΣΗΣ	12
ΚΕΦΑΛΑΙΟ 2.....	13
ΔΙΑΚΡΙΣΕΙΣ ΑΝΑΛΥΣΗΣ ΧΡΟΝΟΣΕΙΡΩΝ	13
2.1 Η ΣΥΝΑΡΤΗΣΗ ΑΥΤΟΣΥΣΧΕΤΙΣΗΣ (AUTOCORRELATION FUNCTION).....	13
2.2 ΜΕΡΙΚΗ ΑΥΤΟΣΥΣΧΕΤΙΣΗ ΚΑΙ ΣΥΝΑΡΤΗΣΗ ΜΕΡΙΚΗΣ ΑΥΤΟΣΥΣΧΕΤΙΣΗΣ.....	15
2.3 ΔΙΑΔΙΚΑΣΙΕΣ ARIMA ΚΑΙ ΜΕΘΟΔΟΛΟΓΙΑ ΔΗΜΙΟΥΡΓΙΑΣ ΕΜΠΕΙΡΙΚΩΝ ΥΠΟΔΕΙΓΜΑΤΩΝ ΚΑΤΑ BOX-JENKINS.....	16
2.3.1 ΔΙΑΔΙΚΑΣΙΕΣ ARIMA	16
2.3.2 ΜΕΘΟΔΟΛΟΓΙΑ ΔΗΜΙΟΥΡΓΙΑΣ ΕΜΠΕΙΡΙΚΩΝ ΥΠΟΔΕΙΓΜΑΤΩΝ ΚΑΤΑ BOX-JENKINS.	18
2.3.2.1 ΤΑΥΤΟΠΟΙΗΣΗ.....	18
2.3.2.2 ΕΚΤΙΜΗΣΗ.....	20
2.3.2.3 ΔΙΑΓΝΩΣΤΙΚΟΣ ΈΛΕΓΧΟΣ.....	20
2.3.2.4 ΚΡΙΤΗΡΙΑ ΕΠΙΛΟΓΗΣ ΥΠΟΔΕΙΓΜΑΤΟΣ.....	21
2.4 ΓΡΑΦΙΚΗ ΑΠΕΙΚΟΝΙΣΗ	22
ΚΕΦΑΛΑΙΟ 3.....	26
ΠΑΡΕΜΒΑΣΕΙΣ (INTERVENTION ANALYSIS) ΚΑΙ ΠΡΟΒΛΕΨΕΙΣ	26
3.1 ΕΙΣΑΓΩΓΗ.....	26
3.2 ΥΠΟΔΕΙΓΜΑΤΑ ΠΑΡΕΜΒΑΣΕΩΝ (INTERVENTION MODELS).....	26
3.3 ΠΡΟΒΛΕΨΕΙΣ	31

3.3.1 ΠΡΟΒΛΕΨΕΙΣ ΜΕ ΥΠΟΔΕΙΓΜΑΤΑ AR(1)	31
3.3.2 ΠΡΟΒΛΕΨΕΙΣ ΜΕ ΥΠΟΔΕΙΓΜΑΤΑ MA (1).....	33
3.3.3 ΠΡΟΒΛΕΨΕΙΣ ΜΕ ΥΠΟΔΕΙΓΜΑΤΑ ARMA.....	35
3.3.4 ΠΡΟΒΛΕΨΕΙΣ ΜΕ ΥΠΟΔΕΙΓΜΑΤΑ ARIMA	35
ΚΕΦΑΛΑΙΟ 4.....	37
ΜΕΛΕΤΗ ΠΕΡΙΠΤΩΣΗΣ - CASE STUDY	37
ΜΟΝΤΕΛΟ ARIMA (1,0,0).....	57
ΜΟΝΤΕΛΟ ARIMA (0,0,1).....	67
ΣΥΜΠΕΡΑΣΜΑΤΑ.....	77
ΒΙΒΛΙΟΓΡΑΦΙΑ.....	79

ΚΕΦΑΛΑΙΟ 1

ΑΝΑΛΥΣΗ ΧΡΟΝΟΣΕΙΡΩΝ (TIME SERIES ANALYSIS)

1.1 ΕΙΣΑΓΩΓΗ

Τα οικονομικά δεδομένα που χρησιμοποιούνται για την εξέταση οικονομικών φαινομένων μπορεί να έχουν τις ακόλουθες μορφές:

- i Διαστρωματικά Δεδομένα (cross-section data): αφορούν μετρήσεις μεταβλητών για διαφορετικές οικονομικές μονάδες άτομα, χώρες, επιχειρήσεις...) σε κάποια χρονική περίοδο.
- i Χρονολογικές Σειρές (time series): αφορούν την εξέλιξη στο χρόνο συγκεκριμένων μεταβλητών.
- i Δεδομένα Panel (panel data): αφορούν την χρονική εξέλιξη μεταβλητών από διαφορετικές οικονομικές μονάδες.

Η ανάλυση χρονοσειρών (time series analysis) είναι το πεδίο έρευνας που μελετά συστήματα, διαδικασίες, σήματα και πρότυπα που εξελίσσονται χρονικά. Η ανάλυση χρονοσειρών έχει δυο βασικούς στόχους:

- a. να μελετήσει και να αναγνωρίσει τη φύση ενός φαινομένου που αναπαρίσταται από μια ακολουθία παρατηρήσεων, και
- b. να προβλέψει τη μελλοντική εξέλιξη του φαινομένου, δηλαδή τις μελλοντικές τιμές της ακολουθίας παρατηρήσεων.

Οι τεχνικές ανάλυσης χρονολογικών σειρών αναπτύχθηκαν, εκτός των άλλων, και για να καλύψουν την ανάγκη της οικονομετρικής ανάλυσης για έγκυρες προβλέψεις των διαφόρων οικονομετρικών φαινομένων. Και οι δύο αυτοί στόχοι απαιτούν ότι το υπόδειγμα των παρατηρούμενων δεδομένων της χρονοσειράς έχει αναγνωριστεί και ουσιαστικά περιγραφηθεί. Από τη στιγμή που το υπόδειγμα εξακριβωθεί μπορούμε να το ερμηνεύσουμε και να το εντάξουμε σε άλλα δεδομένα (δηλαδή μπορούμε να το χρησιμοποιήσουμε στη θεωρία μας για ένα εξερευνούμεο φαινόμενο, για παράδειγμα στις εποχιακές τιμές ενός προϊόντος) (Weatherford, et.al 2003).

1.2 ΟΡΙΣΜΟΣ

Χρονολογική σειρά είναι ένα δείγμα, όπου ο δείκτης T παριστάνει ισαπέχοντα χρονικά σημεία ή χρονικά διαστήματα. Οι παρατηρήσεις είναι συγκεκριμένες τιμές των τυχαίων μεταβλητών και επιπλέον αυτές οι τυχαίες μεταβλητές είναι μέρος μόνο μιας άπειρης ακολουθίας τυχαίων μεταβλητών.

Στα υποδείγματα χρονολογικών σειρών η τρέχουσα τιμή μιας οικονομικής τυχαίας μεταβλητής Y εκφράζεται ως συνάρτηση των προηγούμενων τιμών της, δηλαδή των τιμών της με χρονική υστέρηση, ενώ σε ένα υπόδειγμα παλινδρόμησης η τυχαία μεταβλητή Y είναι συνάρτηση k γενικώς ερμηνευτικών μεταβλητών (Ζάρκος et.al. 1999).

Στόχος της στοχαστικής ανάλυσης των χρονολογικών σειρών είναι η μελέτη της στοχαστικής διαδικασίας διαμόρφωσης των δεδομένων (data generating process) και η διενέργεια προβλέψεων.

1.3 ΣΤΑΣΙΜΕΣ ΚΑΙ ΜΗ ΣΤΑΣΙΜΕΣ ΧΡΟΝΟΛΟΓΙΚΕΣ ΣΕΙΡΕΣ.

Καθώς αναπτύσσουμε υποδείγματα για χρονολογικές σειρές, είναι απαραίτητο να γνωρίζουμε εάν ή όχι η στοχαστική διαδικασία που παρήγαγε τη σειρά, μπορεί να θεωρηθεί ότι ήταν αμετάβλητη, όσον αφορά το χρόνο. Εάν τα χαρακτηριστικά της στοχαστικής διαδικασίας αλλάζουν κατά τη διάρκεια του χρόνου, δηλαδή αν η διαδικασία είναι *μη-στάσιμη*, θα είναι συχνά δύσκολο να αντιπροσωπευθεί η χρονολογική σειρά πάνω σε παρελθόντα και μελλοντικά διαστήματα του χρόνου από ένα απλό αλγεβρικό υπόδειγμα. Αφ' ετέρου, αν η στοχαστική διαδικασία είναι σταθερή στο χρόνο, δηλαδή αν είναι *στάσιμη*, τότε μπορεί να μοντελοποιηθεί μέσω μιας εξίσωσης με σταθερούς συντελεστές, οι οποίοι μπορούν να εκτιμηθούν από παρελθόντα δεδομένα.

Τα υποδείγματα που αναπτύσσονται λεπτομερώς παρακάτω, αντιπροσωπεύουν στοχαστικές διαδικασίες που υποτίθεται ότι είναι σε μια ισορροπία γύρω από ένα σταθερό επίπεδο μέσου. Η πιθανότητα μιας δεδομένης διακύμανσης στη διαδικασία από το επίπεδο του μέσου, υποτίθεται ότι είναι η ίδια σε οποιοδήποτε χρονικό σημείο. Με άλλα λόγια οι στοχαστικές ιδιότητες μιας στάσιμης διαδικασίας υποτίθεται ότι είναι αμετάβλητες όσον αφορά το χρόνο.

Θα μπορούσε κάποιος να ισχυριστεί ότι πολλές από τις χρονολογικές σειρές που συναντιούνται στα οικονομικά όπως για παράδειγμα το GNP, δεν παράγονται από στάσιμες διαδικασίες. Παρόλο που είναι δύσκολο να μοντελοποιήσουμε μη-στάσιμες διαδικασίες, θα δούμε παρακάτω ότι μη-στάσιμες διαδικασίες μπορούν συχνά να μετασχηματιστούν σε στάσιμες ή σε περίπου στάσιμες διαδικασίες.

1.3.1 ΙΔΙΟΤΗΤΕΣ ΤΩΝ ΣΤΑΣΙΜΩΝ ΔΙΑΔΙΚΑΣΙΩΝ

Οποιαδήποτε στοχαστική χρονολογική σειρά y_1, K, y_T μπορεί να θεωρηθεί ότι παράγεται από ένα σύνολο από κοινού κατανεμημένων τυχαίων μεταβλητών. Δηλαδή το σύνολο των σημείων y_1, K, y_T , αντιπροσωπεύουν ένα καθορισμένο αποτέλεσμα της από κοινού συνάρτησης κατανομής πιθανότητας $p(y_1, K, y_T)$. Ομοίως, μια μελλοντική παρατήρηση y_{T+1} μπορεί να θεωρηθεί ότι παράγεται από μια υπό συνθήκη συνάρτηση κατανομής πιθανότητας $p(y_{T+1} | y_1, K, y_T)$ η οποία είναι μια κατανομή πιθανότητας για το y_{T+1} δοθέντος τις παρελθόντες παρατηρήσεις y_1, K, y_T . Ορίζουμε ως στάσιμη διαδικασία, λοιπόν, μια διαδικασία της οποίας η από κοινού κατανομή και η υπό συνθήκη κατανομή είναι και οι δυο αμετάβλητες, όσον αφορά το χρόνο (Δριτσάκη, et.al. 2013). Αν η σειρά είναι στάσιμη, τότε

$$p(y_1, K, y_{t+k}) = p(y_{t+m}, K, y_{t+k+m})$$

και

$$p(y_t) = p(y_{t+m})$$

για κάθε t, k και m .

Σημειώνεται ότι αν η σειρά y_t είναι στάσιμη, ο μέσος της σειράς που ορίζεται ως

$$m_y = E(y_t)$$

πρέπει να είναι επίσης στάσιμος, έτσι ώστε $E(y_t) = E(y_{t+m})$, για κάθε t και m .

Επιπλέον η διακύμανση της σειράς ,

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

πρέπει να είναι στάσιμη, έτσι ώστε $E\left[(y_t - m_y)^2\right] = E\left[(y_{t+m} - m_y)^2\right]$ και τέλος, για κάθε χρονική υστέρηση k , η συνδιακύμανση της σειράς,

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

πρέπει να είναι στάσιμη, έτσι ώστε $Cov(y_t, y_{t+k}) = Cov(y_{t+m}, y_{t+m+k})$.

Εάν μια στοχαστική διαδικασία είναι στάσιμη, η κατανομή πιθανότητας $p(y_t)$ είναι η ίδια για όλο τον χρόνο t και η μορφή της (ή τουλάχιστον μερικές από τις ιδιότητές της) μπορεί να προκύψει εξετάζοντας το ιστόγραμμα των παρατηρήσεων y_1, \dots, y_T που συγκροτούν την παρατηρηθείσα σειρά. Επίσης, μια εκτίμηση του μέσου m_y της διαδικασίας μπορεί να γίνει από τον δειγματικό μέσο της σειράς

$$\bar{y} = \frac{1}{T} \sum_{t=1}^T y_t$$

και μια εκτίμηση της διακύμανσης s_y^2 μπορεί να γίνει από τη δειγματική διακύμανση

$$\hat{s}_y^2 = \frac{1}{T} \sum_{t=1}^T (y_t - \bar{y})^2 .$$

1.3.2 ΟΜΟΓΕΝΩΣ ΜΗ-ΣΤΑΣΙΜΕΣ ΔΙΑΔΙΚΑΣΙΕΣ

Πολύ λίγες από τις χρονολογικές σειρές που συναντάμε στην πράξη είναι στάσιμες. Ευτυχώς, εντούτοις, πολλές από τις μη-στάσιμες χρονολογικές σειρές που

αντιμετωπίζουμε, έχουν την επιθυμητή ιδιότητα, ότι αν αυτές διαφοριστούν¹ μια ή περισσότερες φορές, η σειρά που προκύπτει θα είναι στάσιμη. Μια τέτοια μη-στάσιμη σειρά καλείται *ομογενώς μη-στάσιμη*. Ο αριθμός των φορών, που η αρχική σειρά πρέπει να διαφοριστεί, ώστε να προκύψει μια στάσιμη σειρά, ονομάζεται *τάξη ομοιογένειας*. Έτσι, αν y_t είναι μια πρώτης τάξης ομογενώς μη-στάσιμη σειρά, τότε η σειρά

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

είναι στάσιμη.

Για μια ομογενώς μη-στάσιμη σειρά, οι αυτοσυσχετίσεις θα μειώνονται πολύ αργά καθώς το k στην ACF θα αυξάνει και αυτό είναι ένα κριτήριο για την ύπαρξη ομογενούς μη στασιμότητας στη χρονολογική σειρά. Αντίθετα για μια στάσιμη σειρά, οι αυτοσυσχετίσεις μειώνονται, με πολύ ταχύτερο ρυθμό.

1.4 ΜΟΝΤΕΛΑ ΧΡΟΝΟΣΕΙΡΩΝ ΚΑΙ ΠΡΟΒΛΕΨΗ

Η ανάλυση των χρηματο-οικονομικών δεικτών ως προς την κατανομή και τις συσχετίσεις μας οδηγεί στα παρακάτω συμπεράσματα: 1. Η περιθώρια κατανομή των μεταβολών (αποδόσεων) των δεικτών δεν είναι Γκαουσιανή αλλά έχει παχιές ουρές, που πρακτικά σημαίνει πως μεγάλες μεταβολές των δεικτών εμφανίζονται με μη-μηδενική πιθανότητα. 2. Δεν υπάρχουν σημαντικές συσχετίσεις στις μεταβολές για μικρές υστερήσεις αλλά ενδέχεται οι μεταβολές να έχουν συσχετίσεις μακράς διάρκειας. Το πρώτο πρακτικά σημαίνει πως η βραχυπρόθεσμη πρόβλεψη (με γραμμικά μοντέλα) δε μπορεί να γίνει, τουλάχιστον όχι με σημαντική ακρίβεια (πλήρως αναμενόμενο!). Το δεύτερο αφήνει ανοιχτό το ενδεχόμενο να υπάρχει απόκλιση από τον τυχαίο περίπατο, δηλαδή η αγορά να αποτελεί κάποιο δυναμικό σύστημα ή συσχετισμένη στοχαστική διαδικασία που προκαλεί τις μεταβολές του δείκτη. Στο τελευταίο μέρος θα μελετήσουμε μοντέλα που έχουν χρησιμοποιηθεί στην ανάλυση και πρόβλεψη χρονοσειρών στα χρηματο-οικονομικά. Η μελέτη θα ξεκινήσει με γραμμικά αυτοπαλινδρομούμενα (AR) μοντέλα και μοντέλα τύπου

¹ Διαφορισμένη σειρά, είναι αυτή που προκύπτει, αν από κάθε όρο της αρχικής αφαιρέσουμε τον προηγούμενο

ARIMA, υποθέτοντας πως η χρονοσειρά (με αναφορά στις μεταβολές του χρηματοοικονομικού δείκτη) προέρχεται από μια γραμμική στοχαστική διαδικασία με συσχετίσεις βραχείας κλίμακας. Για τα μοντέλα AR και κατ' επέκταση ARIMA γνωρίζουμε καλά τις στατιστικές τους ιδιότητες και τους περιορισμούς τους. Αν και δεν προσφέρονται για τη μοντελοποίηση χρονοσειρών με μη-Γκαουσιανή κατανομή και συσχετίσεις μακράς κλίμακας χρησιμοποιούνται ως απλά μοντέλα αναφοράς. Στη συνέχεια θα μελετήσουμε μη-γραμμικά μοντέλα και ειδικότερα τοπικά μοντέλα πρόβλεψης που βασίζονται στη θεωρία των δυναμικών συστημάτων και χάους. Έρευνες τα τελευταία χρόνια έδειξαν πως αυτά τα μοντέλα μπορούν να πετύχουν καλύτερες προβλέψεις σε κάποιες περιπτώσεις χρηματο-οικονομικών δεικτών. Τέλος θα μελετήσουμε μοντέλα για τη διασπορά (αυξομείωση) της μεταβολής των δεικτών ως γραμμικές συναρτήσεις προηγούμενων τετραγώνων μεταβολών και διασπορών. Τα μοντέλα αυτά δεν προσφέρονται για προβλέψεις των χρηματοοικονομικών δεικτών αλλά για το μέγεθος της μεταβολής (απόδοσης) των δεικτών.

1.5 ΚΛΑΣΙΚΑ ΥΠΟΔΕΙΓΜΑΤΑ ΑΝΑΛΥΣΗΣ

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

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

- ∅ Το πολλαπλασιαστικό υπόδειγμα
- ∅ Το προσθετικό υπόδειγμα

ΚΕΦΑΛΑΙΟ 2

ΔΙΑΚΡΙΣΕΙΣ ΑΝΑΛΥΣΗΣ ΧΡΟΝΟΣΕΙΡΩΝ

2.1 Η ΣΥΝΑΡΤΗΣΗ ΑΥΤΟΣΥΣΧΕΤΙΣΗΣ (AUTOCORRELATION FUNCTION)

Ενώ είναι συνήθως αδύνατο να ληφθεί μια πλήρης περιγραφή μιας στοχαστικής διαδικασίας, η συνάρτηση αυτοσυσχέτισης είναι εξαιρετικά χρήσιμη επειδή παρέχει μια μερική περιγραφή της διαδικασίας για τη μοντελοποίηση της. Στην ανάλυση χρονολογικών σειρών η σημασία της συνάρτησης αυτοσυσχέτισης είναι πολύ μεγάλη γιατί δείχνει τόσο το βαθμό όσο και το μήκος ή τη χρονική διάρκεια της μνήμης της χρονολογικής σειράς. Μας βοηθά να εξάγουμε συμπεράσματα σχετικά με τον τρόπο αλληλεξάρτησης μεταξύ των όρων της χρονολογικής σειράς.

Όπως είναι γνωστό ο λόγος της συνδιακύμανσης προς το γινόμενο των τετραγωνικών ριζών των διακυμάνσεων δυο μεταβλητών είναι ο συντελεστής συσχέτισής τους. Στην περίπτωση των χρονολογικών σειρών, ο συντελεστής συσχέτισης ανάμεσα στην Y_t και στην Y_{t+k} ονομάζεται συντελεστής αυτοσυσχέτισης (autocorrelation coefficient) και δίνεται από τη σχέση :

$$r_k = \frac{Cov(Y_t, Y_{t+k})}{\sqrt{Var(Y_t)Var(Y_{t+k})}} = \frac{g_k}{g_0}$$

Όπως η αυτοσυνδιακύμανση έτσι και ο συντελεστής αυτοσυσχέτισης δεν εξαρτάται από το t , αλλά από το s . Είναι προφανές ότι $r_k = r_{-k}$, αφού $g_k = g_{-k}$.

Η σχέση που υπάρχει ανάμεσα στο συντελεστή αυτοσυσχέτισης r_k και στο k ονομάζεται *συνάρτηση αυτοσυσχέτισης* (autocorrelation function) και περιγράφεται από τη σχέση :

$$\hat{r}_k = ACF(k) = \frac{\sum_{t=1}^{T-k} (y_t - \bar{y})(y_{t+k} - \bar{y})}{\sum_{t=1}^T (y_t - \bar{y})^2}, \text{ όπου } T \text{ το πλήθος των όρων της σειράς}$$

.Η αυτοσυνδιακύμανση θα δίνεται από τη σχέση : $\hat{g}_k = \frac{\sum_{t=1}^{T-k} (y_t - \bar{y})(y_{t+k} - \bar{y})}{T}$.

Αν για τη δειγματική $ACF(k)$ είναι $k=1, 2, K, q$, $ACF(k) \neq 0$ και για $k > q$ είναι $ACF(k) = 0$, αποδεικνύεται ότι οι δειγματικοί συντελεστές αυτοσυσχέτισης είναι κανονικά κατανεμημένοι με μέση τιμή 0 και διακύμανση $\frac{1}{N} (1 + 2ACF(1)^2 + K + 2ACF(q)^2)$, όπου N το πλήθος των όρων της σειράς.

Έτσι αυξάνοντας διαδοχικά την τιμή του q και αντικαθιστώντας τις ACF με τις αντίστοιχες δειγματικές εκτιμήσεις οι διακυμάνσεις θα είναι :

Για την $ACF(1)$: $Var(ACF(1)) = \frac{1}{N}$.

Για την $ACF(2)$: $Var(ACF(2)) = \frac{1}{N} (1 + 2\hat{A}CF(1)^2)$

.....

Για την $ACF(q)$: $Var(ACF(q)) = \frac{1}{N} (1 + 2\hat{A}CF(1)^2 + K + 2\hat{A}CF(q)^2)$.

Με βάση τα παραπάνω το $(1-a)\%$ διάστημα εμπιστοσύνης για την ACF θα δίνεται από τη σχέση :

$$\hat{A}CF(k) \pm Z_{\alpha/2} \cdot SE(\hat{A}CF(k))$$

όπου $SE(\hat{A}CF(k)) = \sqrt{Var(\hat{A}CF(k))}$.

Ένας έλεγχος για το αν όλοι οι συντελεστές αυτοσυσχέτισης μέχρι τάξης m είναι από κοινού στατιστικά σημαντικοί γίνεται με το *στατιστικό των Ljung και Box* (LBQ).

Η υπόθεση που ελέγχουμε είναι : H_0 : όχι αυτοσυσχέτιση.

Ορίζεται από τη σχέση : $Q = T(T+2) \sum_{k=1}^m (T-k) ACF(k)^2 : X_{m-l}^2$,

όπου m : η τάξη της αυτοσυσχέτισης

l : ο αριθμός των παραμέτρων που υπολογίζουμε.

2.2 ΜΕΡΙΚΗ ΑΥΤΟΣΥΣΧΕΤΙΣΗ ΚΑΙ ΣΥΝΑΡΤΗΣΗ ΜΕΡΙΚΗΣ ΑΥΤΟΣΥΣΧΕΤΙΣΗΣ

Έστω το υπόδειγμα παλινδρόμησης

$$Y_t = b_1 + b_2 X_{2t} + b_3 X_{3t}.$$

Οι συντελεστές b_2, b_3 καλούνται μερικοί συντελεστές παλινδρόμησης.

Ο συντελεστής συσχέτισεως r_{YX_2} ή r_{12} (ο δείκτης 1 αναφέρεται στο Y), αν δε ληφθεί μέριμνα για τη X_3 δεν αντανακλά την πραγματική σχέση μεταξύ Y και X_2 . Είναι φανερό ότι πρέπει να ορίσουμε έναν άλλο συντελεστή μεταξύ Y και X_2 . Αυτό επιτυγχάνεται με τον λεγόμενο *μετρικό συντελεστή συσχέτισεως*. Συμβολίζουμε ως $r_{12.3}$ τον μερικό συντελεστή συσχέτισεως μεταξύ Y και X_2 , κρατώντας τη μεταβλητή X_3 σταθερή. Ο συντελεστής αυτός δίνεται από τη σχέση :

$$r_{12.3} = \frac{r_{12} - r_{13}r_{23}}{\sqrt{(1-r_{13})^2 (1-r_{23})^2}}.$$

Κατ' αντιστοιχία ορίζονται και οι συντελεστές $r_{13.2}$ και $r_{23.1}$.

Σε μια χρονολογική σειρά ορίζεται σαν *μερική αυτοσυσχέτιση* k -τάξης, η μερική συσχέτιση μεταξύ των y_t και y_{t+k} , όταν οι ενδιάμεσοι όροι $y_{t+1}, \dots, y_{t+k-1}$, παραμένουν σταθεροί. Δηλαδή

$$f_{kk} = PACF = Cor(y_t, y_{t+k} | y_{t+1}, y_{t+2}, \dots, y_{t+k-1} = \text{σταθεροί})$$

όπου *PACF* η *συνάρτηση μερικής αυτοσυσχέτισης*.

2.3 ΔΙΑΔΙΚΑΣΤΕΣ ARIMA ΚΑΙ ΜΕΘΟΔΟΛΟΓΙΑ ΔΗΜΙΟΥΡΓΙΑΣ ΕΜΠΕΙΡΙΚΩΝ ΥΠΟΔΕΙΓΜΑΤΩΝ ΚΑΤΑ BOX-JENKINS.

2.3.1 ΔΙΑΔΙΚΑΣΤΕΣ ARIMA

Από το θεώρημα του Wald² μια στοχαστική διαδικασία είναι ένα σταθμισμένο άθροισμα απείρων όρων λευκού θορύβου. Αυτό όμως προϋποθέτει καταρχήν την εύρεση ενός απείρου πλήθους παραμέτρων y_i κάτι που πρακτικά είναι αδύνατο, αφού συνήθως διαθέτουμε ένα πεπερασμένο πλήθος δεδομένων. Εμείς θα μελετήσουμε τα υποδείγματα που προκύπτουν, όταν γίνουν κάποιες πρόσθετες υποθέσεις σχετικά με τη φύση των y_1, y_2, \dots, y_n . Πιο συγκεκριμένα υποθέτουμε ότι το $y(B) = y_0 + y_1 B + y_2 B^2 + \dots + y_n B^n$, όπου B ο τελεστής χρονικής υστέρησης, μπορεί να εκφραστεί ως πηλίκο δυο πολυωνύμων πεπερασμένου βαθμού ως ακολούθως :

$$y(B) = \frac{1 - q_1 B - q_2 B^2 - \dots - q_q B^q}{1 - f_1 B - f_2 B^2 - \dots - f_p B^p} = \frac{\Theta(B)}{\Phi(B)} = \Phi^{-1}(B) \Theta(B).$$

Με βάση το παραπάνω, μια στοχαστική διαδικασία, μπορεί να γραφεί ως εξής :

² Θεώρημα Wald (1934) : Κάθε στάσιμη κατά τη συνδιακύμανση στοχαστική διαδικασία μηδενικού μέσου, μπορεί να αναπαρασταθεί ως ένα σταθμισμένο άθροισμα απείρων όρων λευκού θορύβου ως :

$$Y_t = \sum_{i=1}^{\infty} y_i e_{t-i} + k_t, \text{ όπου } y_0 = 1, \sum_{i=1}^{\infty} y_i^2 < \infty, e_t \text{ λευκός θόρυβος και } k_t \text{ αιτιοκρατική συνιστώσα.}$$

$$Y_t = \Psi(B)e_t \Rightarrow \Phi(B)Y_t = \Theta(B)e_t.$$

Τα $\Phi(B), \Theta(B)$, υπολογίζονται από τα δεδομένα.

Αν θέσουμε $\Theta(B)=1$ το υπόδειγμα γράφεται :

$$\Phi(B)y_t = e_t \Rightarrow (1 - f_1B - f_2B^2 - \dots - f_pB^p)y_t = e_t \Rightarrow y_t = f_1y_{t-1} + f_2y_{t-2} + \dots + f_py_{t-p} + e_t$$

Δηλαδή η y_t εκφράζεται ως ένας γραμμικός συνδυασμός παρελθουσών τιμών μέχρι υστέρηση p συν μια διαδικασία λευκού θορύβου.

Το υπόδειγμα αυτό ονομάζεται *αυτοπαλίνδρομο υπόδειγμα τάξεως p* και συμβολίζεται $AR(p)$ (Autoregressive Model).

Αν θέσουμε $\Phi(B)=1$ το υπόδειγμα γράφεται :

$$y_t = (1 - q_1B - q_2B^2 - \dots - q_qB^q)e_t \Rightarrow y_t = e_t - q_1e_{t-1} - q_2e_{t-2} - \dots - q_qe_{t-q}$$

Δηλαδή η y_t εκφράζεται σαν ένας γραμμικός συνδυασμός από παρελθούσες τιμές μιας διαδικασίας λευκού θορύβου συν την τρέχουσα τιμή του λευκού θορύβου (Βενέτης, 2009).

Το υπόδειγμα αυτό ονομάζεται *υπόδειγμα κινητού μέσου τάξεως q* και συμβολίζεται $MA(q)$ (Moving Average Model).

Στη γενική περίπτωση, η διαδικασία θα είναι ένας συνδυασμός κινητού μέσου και αυτοπαλίνδρομης διαδικασίας. Μια τέτοια διαδικασία ονομάζεται *μικτή διαδικασία τάξεως p, q* και συμβολίζεται $ARMA(p, q)$. Με τις μικτές διαδικασίες μπορούμε να εκφράσουμε καλύτερα μια διαδικασία $AR(p)$ ή $MA(q)$, όταν είτε το p είτε το q λαμβάνουν μεγάλες τιμές.

Όταν μια $ARMA(p, q)$ διαδικασία y_t προέρχεται από μια ομογενώς μη στάσιμη διαδικασία w_t ύστερα από d διαφορίσεις, η w_t ονομάζεται *ολοκληρωμένη μεικτή διαδικασία τάξεως p, d, q* και συμβολίζεται $ARIMA(p, d, q)$ (Autoregressive Integrated Moving Average).

Παραδείγματα

1. Ο τυχαίος περίπατος $y_t = y_{t-1} + e_t$ είναι μια διαδικασία $ARIMA(0,1,0)$.
2. Το υπόδειγμα $y_t = f_1 y_{t-1} + e_t$ (αυτοπαλίνδρομο 1^{ης} τάξης) είναι μια διαδικασία $ARIMA(1,0,0)$.
3. Το υπόδειγμα $y_t = e_t - q_1 e_{t-1} - q_2 e_{t-2}$ (υπόδειγμα κινητού μέσου 2^{ης} τάξης) είναι μια διαδικασία $ARIMA(0,0,2)$.
4. Αν w_t ομογενώς μη στάσιμη διαδικασία και οι πρώτες διαφορές y_t είναι στάσιμες και ακολουθούν το υπόδειγμα $y_t = f_1 y_{t-1} + e_t + q_1 e_{t-1}$ τότε η w_t είναι μια διαδικασία $ARIMA(1,1,1)$.
5. Ο λευκός θόρυβος είναι μια διαδικασία $ARIMA(0,0,0)$.

2.3.2 ΜΕΘΟΔΟΛΟΓΙΑ ΔΗΜΙΟΥΡΓΙΑΣ ΕΜΠΕΙΡΙΚΩΝ ΥΠΟΔΕΙΓΜΑΤΩΝ ΚΑΤΑ BOX-JENKINS.

Η προσέγγιση Box-Jenkins στην ανάλυση χρονολογικών σειρών είναι μια μέθοδος εξευρέσεως ενός στατιστικού υποδείγματος $ARIMA$ που να παριστάνει ικανοποιητικά τη στοχαστική διαδικασία που παρήγαγε τα δεδομένα (δείγμα). Η μέθοδος περιλαμβάνει τα εξής τρία στάδια : α) την *ταυτοποίηση* (identification), β) την *εκτίμηση* (estimation) και γ) το *διαγνωστικό έλεγχο* (diagnostic checking).

2.3.2.1 ΤΑΥΤΟΠΟΙΗΣΗ

Το πρώτο στάδιο είναι η εξειδίκευση ενός $ARIMA$ υποδείγματος με βάση τις πληροφορίες του διαθέσιμου δείγματος. Αυτό σημαίνει καθορισμός των κατάλληλων τιμών των d, p και q . Δηλαδή, καθορισμός του αριθμού (d) των διαφορών που απαιτούνται προκειμένου η σειρά να μετατραπεί σε στάσιμη, αν δεν είναι. Στη

συνέχεια, πρέπει να καθοριστεί η τάξη (p) της αυτοπαλίνδρομης διαδικασίας και η τάξη (q) της διαδικασίας κινητού μέσου.

Το πρώτο βήμα, επομένως, είναι να διαπιστωθεί αν η σειρά είναι στάσιμη ή όχι. Η διαπίστωση αυτή βασίζεται στη συμπεριφορά της δειγματικής συναρτήσεως αυτοσυσχετίσεως. Αν οι αυτοσυσχετίσεις συγκλίνουν ταχύτατα προς το μηδέν, αυτό είναι σοβαρή ένδειξη ότι η σειρά είναι στάσιμη. Αντίθετα, αν οι αυτοσυσχετίσεις φθίνουν με αργό ρυθμό, είναι ένδειξη ότι η σειρά είναι μη στάσιμη.

Αν η σειρά δεν είναι στάσιμη, μετατρέπεται σε στάσιμη παίρνοντας τις πρώτες ή τις δεύτερες κ.ο.κ διαφορές. Με τα δεδομένα της στάσιμης σειράς, στη συνέχεια προσδιορίζεται η τάξη του υποδείγματος ARIMA. Δηλαδή, προσδιορίζονται οι τιμές του p και του q . Ο προσδιορισμός βασίζεται στις δειγματικές, απλές και μερικές αυτοσυσχετίσεις με βάση τον πίνακα (2.1).

Πίνακας 1: Η μορφή της Συνάρτησης Αυτοσυσχετίσεως και Μερικής Αυτοσυσχετίσεως

Διαδικασία	Συνάρτηση υτοσυσχετίσεως ACF	Συνάρτηση Μερικής Αυτοσυσχετίσεως PACF
Λευκός Θόρυβος	Καμιά στατιστικά σημαντική αυτοσυσχέτιση	Καμιά στατιστικά σημαντική αυτοσυσχέτιση
Αυτοπαλίνδρομη Διαδικασία p τάξεως : $AR(p)$	Φθίνει σταδιακά	Μηδενίζεται μετά από p στατιστικά σημαντικές αυτοσυσχετίσεις
Διαδικασία Κινητού Μέσου q τάξεως : $MA(q)$	Μηδενίζεται μετά από q στατιστικά σημαντικές αυτοσυσχετίσεις	Φθίνει σταδιακά

Μικτή $ARMA(p, q)$	Διαδικασία	Φθίνει σταδιακά	Φθίνει σταδιακά
-----------------------	------------	-----------------	-----------------

Πίνακας 2.1

2.3.2.2 ΕΚΤΙΜΗΣΗ

Μετά την εξειδίκευση του υποδείγματος ακολουθεί η εκτίμηση των p παραμέτρων f_1, f_2, \dots, f_p της AR διαδικασίας και των q παραμέτρων q_1, q_2, \dots, q_q της MA διαδικασίας. Αν η σειρά που τακτοποιείται είναι μόνο αυτοπαλίνδρομη (AR), οι συντελεστές μπορούν να εκτιμηθούν με τη μέθοδο ελαχίστων τετραγώνων. Αν, όμως, η σειρά περιέχει και όρους κινητού μέσου (MA), τότε για την εκτίμηση των αντίστοιχων παραμέτρων, χρησιμοποιούνται μη γραμμικές μέθοδοι εκτιμήσεως (Mitchell, 1982).

2.3.2.3 ΔΙΑΓΝΩΣΤΙΚΟΣ ΈΛΕΓΧΟΣ

Στο στάδιο αυτό, γίνεται έλεγχος καλής προσαρμογής του υποδείγματος. Πόσο καλά, δηλαδή, «ταιριάζει» το εκτιμημένο υπόδειγμα με τα δεδομένα, αφού είναι πιθανό κάποιο άλλο υπόδειγμα $ARIMA$ να προσαρμόζεται καλύτερα. Ο διαγνωστικός έλεγχος αφορά, πέραν των στατιστικών ελέγχων για τη σημαντικότητα των συντελεστών, τη συμπεριφορά των καταλοίπων καθώς και την τάξη του υποδείγματος.

Αν το εκτιμημένο $ARMA$ υπόδειγμα ταιριάζει πράγματι με τα δεδομένα, αν δηλαδή, πράγματι εκφράζει ικανοποιητικά τη διαδικασία από την οποία προέρχονται τα δεδομένα, τότε τα κατάλοιπα θα πρέπει να συμπεριφέρονται ως μια διαδικασία λευκού θορύβου. Με άλλα λόγια, τα κατάλοιπα δεν πρέπει να αυτοσυσχετίζονται.

Ο έλεγχος των καταλοίπων γίνεται με τη στατιστική Q των Ljung και Box (LBQ) (βλέπε 2.2.3) με την οποία ελέγχεται η σημαντικότητα από κοινού ενός αριθμού συντελεστών αυτοσυσχέτισης, αλλά και οπτικά, με βάση τα διαγράμματα των ACF και PACF για τις αυτοσυσχετίσεις και τις μερικές αυτοσυσχετίσεις των καταλοίπων.

Η καταλληλότητα του εκτιμώμενου υποδείγματος ελέγχεται, επίσης, συγκρίνοντας το με ένα άλλο υπόδειγμα μεγαλύτερης τάξης. Δηλαδή, το εκτιμημένο υπόδειγμα ARMA(p,q) συγκρίνεται με τα υποδείγματα ARMA(p+1,q) και ARMA(p,q+1). Αν το υπόδειγμα που εκτιμήθηκε περιγράφει τη διαδικασία που παρήγαγε τα δεδομένα, οι επιπλέον συντελεστές στα μεγαλύτερα υποδείγματα δεν θα πρέπει να είναι στατιστικά διαφορετικοί από το μηδέν. Είναι πιθανό, όμως, να προκύψουν και άλλα κατάλληλα υποδείγματα που να περιγράφουν τη διαδικασία που παρήγαγε τα δεδομένα, κάνοντας τον παραπάνω έλεγχο.

2.3.2.4 ΚΡΙΤΗΡΙΑ ΕΠΙΛΟΓΗΣ ΥΠΟΔΕΙΓΜΑΤΟΣ

Αυξάνοντας την τάξη του υποδείγματος, δηλαδή προσθέτοντας υστερήσεις για το αυτοπαλίνδρομο τμήμα ή για το τμήμα κινητού μέσου, θα μειώνεται το άθροισμα των τετραγώνων των καταλοίπων, αλλά θα μειώνονται και οι βαθμοί ελευθερίας, αφού εκτιμώνται περισσότερες παράμετροι. Με άλλα λόγια, δεν υπάρχει μόνο «κέρδος», από την προσθήκη μεταβλητών, αλλά και «κόστος».

Για τη σύγκριση της ερμηνευτικής ικανότητας εναλλακτικών υποδειγμάτων, που διαφέρουν ως προς τον αριθμό των παραμέτρων ή και το μέγεθος του δείγματος, πέραν του συντελεστή προσδιορισμού R^2 χρησιμοποιούνται και άλλα κριτήρια. Στην ανάλυση χρονολογικών σειρών χρησιμοποιούνται ευρέως το μέσο τετράγωνο των καταλοίπων (Residual Mean Square) ή RMS, το κριτήριο πληροφοριών Akaike (Akaike Information Criterion) ή AIC και το Μπεϋσιανό κριτήριο του Schwartz (Schwartz Bayesian Criterion) ή SBC. Τα κριτήρια αυτά ορίζονται ως εξής :

$$RMS = \frac{1}{T} \sqrt{\sum_{i=1}^T \hat{u}_i^2}$$

$$AIC = \ln \frac{\sum_{i=1}^T \hat{u}_i^2}{T} + \frac{2k'}{T}$$

$$SBC = \ln \frac{\sum_{i=1}^T \hat{u}_i^2}{T} + \frac{k'}{T} \ln T$$

όπου : $\sum \hat{u}^2$ το άθροισμα τετραγώνων των καταλοίπων

T ο αριθμός των παρατηρήσεων

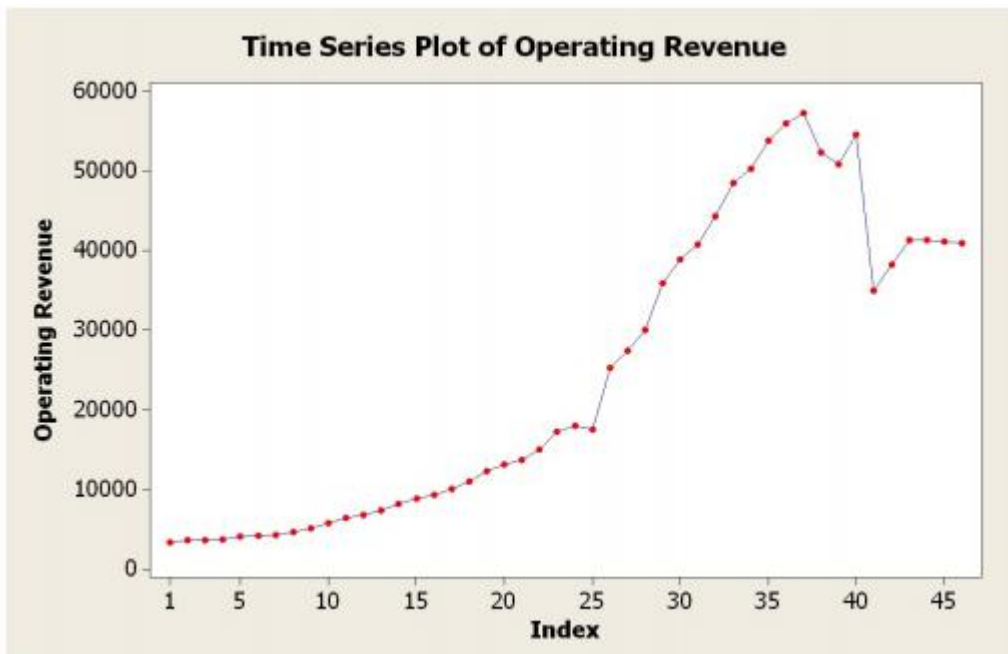
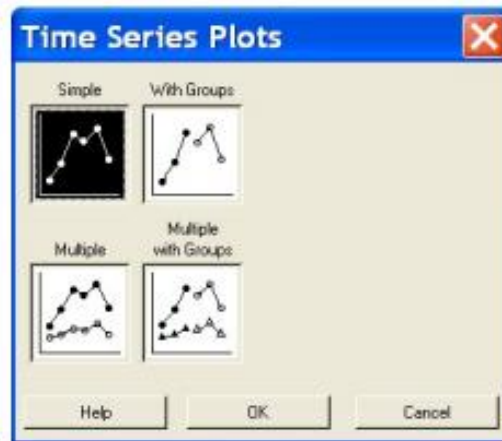
k' ο αριθμός των παραμέτρων που εκτιμώνται $(p+q+1)$.

Με βάση τα παραπάνω κριτήρια, επιλέγεται το υπόδειγμα με τη μικρότερη τιμή. Σημειώνεται ότι τα κριτήρια AIC και SBC μπορούν να πάρουν και αρνητικές τιμές.

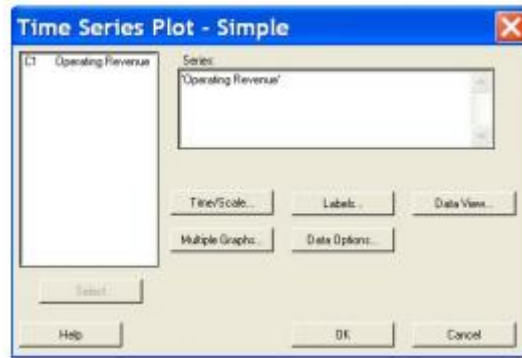
2.4 ΓΡΑΦΙΚΉ ΑΠΕΙΚΟΝΙΣΗ

Ανοίγουμε το αρχείο C:\Forecasting Lab Data\Operating Revenues.MTW. Για να κάνουμε τη γραφική παράσταση της χρονοσειράς

1. Από τη γραμμή μενού επιλέγουμε Stat → Time Series → Time Series Plot.
2. Στο πλαίσιο διαλόγου Time Series Plots που εμφανίζεται επιλέγουμε Simple → OK.

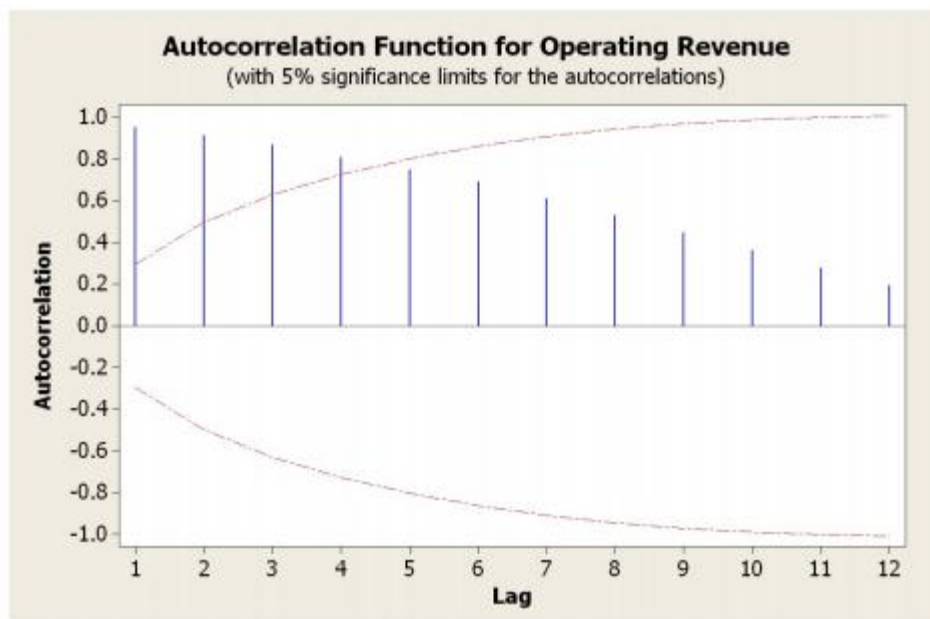


Για να υπολογίσουμε τη συνάρτηση αυτοσυσχέτισης και να πάρουμε τη γραφική της παράσταση 1. Από τη γραμμή μενού επιλέγουμε Stat → Time Series → Autocorrelation. 2. Στο πλαίσιο διαλόγου Autocorrelation Function που εμφανίζεται, διπλοπατάμε C1 Operating Revenue από τον αριστερό κατάλογο.



Στο πλαίσιο διαλόγου Time Series Plots – Simple που εμφανίζεται, διπλοπατάμε C1 Operating Revenue από τον αριστερό κατάλογο.

Πατάμε OK οπότε στο Graph Window παίρνουμε το διάγραμμα.



Πατάμε OK και η γραφική παράσταση εμφανίζεται στο Graph Window με την ονομασία Autocorrelation function for Operating Revenue.

Για να πάρουμε τη συνάρτηση αυτοσυσχέτιση, από τη γραμμή μενού επιλέγουμε Window → Session όπου έχουμε

Autocorrelation Function: Operating Revenue

Lag	ACF	T	LBQ
1	0.957322	6.49	44.97
2	0.915920	3.69	87.07
3	0.870327	2.78	125.96
4	0.810476	2.24	160.49
5	0.750323	1.88	190.81
6	0.692167	1.61	217.26
7	0.611867	1.35	238.45
8	0.530836	1.13	254.83
9	0.449109	0.93	266.86
10	0.363433	0.74	274.97
11	0.277934	0.56	279.84
12	0.194559	0.39	282.30

ΚΕΦΑΛΑΙΟ 3

ΠΑΡΕΜΒΑΣΕΙΣ (INTERVENTION ANALYSIS) ΚΑΙ ΠΡΟΒΛΕΨΕΙΣ

3.1 ΕΙΣΑΓΩΓΗ

Στην ανάλυση των υποδειγμάτων χρονολογικών σειρών μιας μεταβλητής, λίγη προσοχή δίνεται στη συμπεριφορά συγκεκριμένων παρατηρήσεων της σειράς ή στα κατάλοιπα που συσχετίζονται με αυτές. Συγκεκριμένα, η προσοχή αφιερώνεται στο συνολικό υπόδειγμα, στη γραμμική συσχέτιση στις σειρές των καταλοίπων. Είναι ευρέως γνωστό, εντούτοις, ότι χρονολογικές σειρές που περιγράφουν οικονομικά μεγέθη, αλλά και άλλες, συχνά επηρεάζονται από γεγονότα που συμβαίνουν σε συγκεκριμένο χρονικό σημείο. Είναι προφανές, λοιπόν, ότι το να αγνοήσουμε τέτοιους παράγοντες, μπορεί αυτό να οδηγήσει σε προσαρμογή ανεπαρκών υποδειγμάτων και κατά συνέπεια σε ανεπαρκείς προβλέψεις.

Γεγονότα αυτού του τύπου, των οποίων η χρονική στιγμή που συνέβησαν είναι γνωστή, έχουν ονομαστεί ως *παρεμβάσεις* (interventions) από τους Box και Tiao (1975) και μπορούν να ενσωματωθούν σε ένα υπόδειγμα μιας μεταβλητής, ώστε να επεκταθεί για να περιλάβει αιτιοκρατικές μεταβλητές.

Στην παράγραφο αυτή περιγράφεται πως μπορούμε να μοντελοποιήσουμε και να εξετάσουμε τις επιπτώσεις των εξωγενών επιδράσεων πάνω στη χρονολογική σειρά. Αν αυτές οι επιδράσεις είναι παρούσες, πρέπει να μοντελοποιηθούν, έτσι ώστε οι επιπτώσεις τους να μην απεμπολήσουν την ακρίβεια των προβλέψεων. Αφ' ετέρου, δεδομένα χρονολογικών σειρών μπορούν να συλλεχθούν και να αναλυθούν ειδικότερα, ώστε να εξακριβωθούν οι επιπτώσεις από μερικές παρεμβάσεις ή αλλαγές πάνω στις σειρές. Τα υποδείγματα ARIMA είναι στατιστικά κατάλληλα για την αποτίμηση της σημαντικότητας αυτών των επιπτώσεων (Box-Tiao 1975, Glass-Willson-Gottman 1975, McCleary-Hay 1980).

3.2 ΥΠΟΔΕΙΓΜΑΤΑ ΠΑΡΕΜΒΑΣΕΩΝ (INTERVENTION MODELS)

Οι παρεμβάσεις , μπορούν να επηρεάσουν μια χρονολογική σειρά με διάφορους τρόπους. Μπορούν να αλλάξουν το επίπεδο, είτε απότομα, είτε μετά από κάποια καθυστέρηση, να αλλάξουν την τάση ή να οδηγήσουν σε άλλα πιο περίπλοκα υποδείγματα απόκρισης.

Η κατανομή $t - student$, συνηθίζεται να χρησιμοποιείται για τον έλεγχο μιας αλλαγής στο επίπεδο του μέσου, αλλά ένας τέτοιος έλεγχος μπορεί να μην είναι κατάλληλος για χρονολογικές σειρές, διότι διαδοχικές παρατηρήσεις συχνά συσχετίζονται και επίσης, επειδή η επίπτωση μπορεί να μην είναι μια μεταβολή βήματος , όπως προϋποθέτετε στο έλεγχο t .

Χρησιμοποιώντας, λοιπόν, τα υποδείγματα ARIMA μοντελοποιούμε και εξετάζουμε τις επιπτώσεις των εξωγενών επιδράσεων πάνω σε μια χρονολογική σειρά(Song et.al. 2003)..

Ξεκινάμε, υποθέτοντας μια παρέμβαση, η οποία συμβαίνει σε γνωστό χρόνο T . Αν η y_t παράγεται από μια διαδικασία $ARIMA(p, d, q)$, τότε ένα υπόδειγμα παρέμβασης μπορεί να διατυπωθεί ως :

$$\nabla^d y_t = f(I_t) + N_t$$

όπου ∇ είναι το πολυώνυμο διαφορίσης, I_t είναι μια διωνυμική μεταβλητή, η οποία κωδικοποιείται με 1 κατά το διάστημα της παρέμβασης και με 0 αλλού, $f(I_t)$ είναι η συνιστώσα παρέμβασης (Intervention component), που αποσκοπούμε να μοντελοποιήσουμε και να εξετάσουμε και N_t μπορεί να είναι οποιαδήποτε διαδικασία θορύβου.

Οι επιπτώσεις παρέμβασης μπορεί είτε να είναι μόνιμες ή προσωρινές. Συνήθως, η I_t κωδικοποιείται σαν μια συνάρτηση βήματος, η οποία είναι 0 πριν την υποθετική παρέμβαση και 1 μετά. Όταν η I_t είναι συνάρτηση βήματος μόνιμες επιπτώσεις μοντελοποιούνται από αυτή, ενώ προσωρινές «παλμικές» επιπτώσεις μοντελοποιούνται από την $(1 - B)I_t$. Η διωνυμική μορφή της I_t και της $(1 - B)I_t$ είναι :

I_t	0	0	0	1	1	1	1
$(1-B)I_t$	0	0	0	1	0	0	0
Time	$n-3$	$n-2$	$n-1$	n	$n+1$	$n+2$	$n+3$

Η γενική μορφή για τη συνιστώσα παρέμβασης είναι :

$$f(I_t) = \left\{ \begin{array}{l} \frac{U(B)}{S(B)} I_t \quad , \text{για μόνιμες (permanent) επιπτώσεις} \\ \frac{U(B)}{S(B)} (1-B) I_t \quad , \text{για παροδικές (temporary) επιπτώσεις} \end{array} \right\}$$

$$\text{όπου } \frac{U(B)}{S(B)} = \frac{(U_0 + U_1 B + K + U_s B^s)}{(1 - S_1 B - S_2 B^2 - K - S_r B^r)}$$

Η έκφραση $(U_0 + U_1 B + K + U_s B^s)$ καλείται U πολυώνυμο και εκφράζει την αλλαγή στο επίπεδο της σειράς μετά την παρέμβαση.

Η έκφραση $(1 - S_1 B - S_2 B^2 - K - S_r B^r)$ καλείται S πολυώνυμο και εκφράζει το ρυθμό με τον οποίο προσεγγίζεται η ασυμπτωτική επίπτωση. Μικρές τιμές των s_i δείχνουν γρήγορη προσέγγιση της ασυμπτωτικής επίπτωσης, ενώ μεγάλες τιμές των s_i δείχνουν ότι πολλές παρατηρήσεις θα χρειαστούν, ώστε να προσεγγισθεί η ασυμπτωτική επίπτωση.

Οι συνιστώσες παρέμβασης είναι επίσης είτε σταδιακές (gradual) είτε άμεσες (sudden) στην επίτευξη της πλήρους επίπτωσης. Οι διαστάσεις παροδική / μόνιμη και άμεση / σταδιακή είναι ανεξάρτητες η μια από την άλλη και η κάθε μια έχει χαρακτηριστικό τύπο και ορισμό.

Όταν ο τύπος μια παρέμβασης δεν μπορεί να καθοριστεί εκ των προτέρων (*a priori*), βασιζόμενοι στην θεωρία, οι McCleary και Hay προτείνουν μια ανάλυση τριών σταδίων χρησιμοποιώντας τη συνιστώσα παρέμβασης $\frac{U_0}{(1-S_j B_j)}$. Για $j=1$ τα τρία

στάδια είναι :

1. Προσπαθούμε να προσαρμόσουμε μια άμεση, παροδική επίπτωση (sudden, temporary effect) :

$$\frac{U_0}{(1-S_1 B)}(1-B)I_t$$

Αν το S_1 είναι πολύ μεγάλο (μεγαλύτερο από ή πολύ κοντά στο 1.00), αποκλείουμε μια παροδική επίπτωση.

2. Προσπαθούμε να προσαρμόσουμε μια σταδιακή, μόνιμη επίπτωση (gradual, permanent effect) :

$$\frac{U_0}{(1-S_1 B)}I_t$$

Αν το S_1 είναι πολύ μικρό (όχι σημαντικά διαφορετικό του 0), αποκλείουμε μια σταδιακή επίπτωση. Αν $S_1=1$, το αποτέλεσμα της παρέμβασης δε σταθεροποιείται ποτέ, αλλά η σειρά αποκτά μια γραμμική τάση και η παράμετρος U_0 παριστάνει την κλίση.

3. Τέλος, προσπαθούμε να προσαρμόσουμε μια άμεση, μόνιμη επίπτωση (sudden, permanent effect) :

$$U_0 I_t$$

Αν αυτή δεν είναι σημαντική, τότε η μηδενική υπόθεση για μη παρέμβαση δεν μπορεί να απορριφθεί.

Ένα υπόδειγμα είναι αποδεκτό μόνο αν όλες οι παράμετροί του που εκτιμώνται (συνιστώσες παρέμβασης και οι παράμετροι του ARIMA) είναι στατιστικά σημαντικές. Ένα συνηθισμένο λάθος είναι να αποδεχτούμε ένα μοντέλο παρέμβασης στο οποίο το S_1 είναι σημαντικά διαφορετικό του 0, αλλά το U_0 δεν είναι σημαντικό. Αυτό οδηγεί σε μια μη λογική ερμηνεία ότι, το επίπεδο της σειράς μετά την παρέμβαση δεν διαφέρει σημαντικά από το επίπεδο της σειράς πριν την παρέμβαση. Δηλαδή δεν υπάρχει διαφορά στο επίπεδο της σειράς, ενώ υπάρχει σημαντικός ρυθμός προσέγγισης της ασυμπτωτικής επίπτωσης (Βάμβουκας,2007).

3.3 ΠΡΟΒΛΕΨΕΙΣ

Ο κύριος σκοπός της εξειδίκευσης και εκτίμησης ενός υποδείγματος ARIMA είναι η διενέργεια βραχυχρόνιων προβλέψεων. Χρησιμοποιώντας, δηλαδή, το εκτιμώμενο υπόδειγμα και τις υπάρχουσες πληροφορίες μέχρι την χρονική περίοδο T , μπορούμε να κάνουμε πρόβλεψη για την τιμή της T στην περίοδο $T+1$, $T+2$ και γενικά να γίνει πρόβλεψη στην περίοδο $T+h$ (Song et.al. 2003)..

Από την ανάλυση του απλού γραμμικού υποδείγματος, η καλύτερη πρόβλεψη μιας μελλοντικής τιμής Y_f της εξαρτημένης μεταβλητής Y , για δεδομένη τιμή X_f της ερμηνευτικής μεταβλητής X , είναι η :

$$\hat{Y}_f = \hat{b}_0 + \hat{b}_1 X_f$$

η οποία στην ουσία είναι η εκτίμηση της $E(Y_f)$. Η διαφορά εξάλλου: $f = Y_f - \hat{Y}_f$ παριστάνει το σφάλμα πρόβλεψης. Στην περίπτωση των υποδειγμάτων ARIMA έχουμε ανάλογα αποτελέσματα (Σαριαννίδης et.al. 2013)

3.3.1 ΠΡΟΒΛΕΨΕΙΣ ΜΕ ΥΠΟΔΕΙΓΜΑΤΑ AR(1)

Έστω το υπόδειγμα:

$$Y_t = d + a_1 Y_{t-1} + e_t$$

Για $t = T+1$,

$$Y_{T+1} = d + a_1 Y_T + e_{T+1}$$

Αν οι παράμετροι d και a_1 είναι γνωστές, τότε με βάση τις πληροφορίες μέχρι την περίοδο T , μια πρόβλεψη για την περίοδο $T+1$ είναι η υπό συνθήκη προσδοκώμενη τιμή της Y_{T+1} . Δηλαδή, αν \hat{Y}_{T+1} παριστάνει την πρόβλεψη, τότε:

$$\hat{Y}_{T+1} = E(Y_{t+1} | \text{πληροφορίες μέχρι την περίοδο } T)$$

ή

$$\hat{Y}_{T+1} = E(Y_{t+1} | Y_T, Y_{T-1}, \dots, Y_1)$$

Χάρην απλουστεύσεως, παριστάνουμε το δεξιό μέρος της παραπάνω σχέσης με $E_T(Y_{T+1})$, οπότε

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

Γενικά, $\hat{Y}_{T+h} = E_T(Y_{T+h})$ είναι μια πρόβλεψη h περιόδους μπροστά με βάση τις πληροφορίες μέχρι την περίοδο T. Επειδή η παραπάνω πρόβλεψη ελαχιστοποιεί τον μέσο του τετραγώνου του σφάλματος (Mean Square Error ή MSE), δηλαδή ελαχιστοποιεί τη σχέση

$$E(\hat{Y}_{T+h} - Y_{T+h})^2 \text{ θεωρείται άριστη (optimal forecast)}$$

Από τη σχέση $Y_{T+1} = d + a_1 Y_T + e_{T+1}$ έχουμε

$$E_T(Y_{T+1}) = E_T(d + a_1 Y_T + e_{T+1}) \quad \text{ή}$$

$$E_T(Y_{T+1}) = d + a E_T(Y_T) + E_T(e_{T+1})$$

Αν λάβουμε υπόψη ότι στην περίοδο $t = T$ κάθε όρος με δείκτη $t \leq T$ είναι γνωστός, έπεται ότι για $t \leq T$, $E_T(Y_t) = Y_t$ και $E_T(e_t) = e_t$. Για $t > T$ όμως, $E_T(e_t) = E(e_t) = 0$.

Με βάση τα παραπάνω αποτελέσματα η σχέση $E_T(Y_{T+1}) = d + a E_T(Y_T) + E_T(e_{T+1})$ γίνεται:

$E_T(Y_{T+1}) = d + a_1 Y_T$ και η πρόβλεψη για την περίοδο T+1: $\hat{Y}_{T+1} = d + a_1 Y_T$. Για το σφάλμα πρόβλεψης εξάλλου αποδεικνύεται εύκολα ότι $\hat{e}_{T+1} = e_{T+1}$ και επομένως $V(\hat{e}_{T+1}) = V(e_{T+1}) = s^2$.

Για πρόβλεψη στην περίοδο $T+2$, έχουμε :

$$\hat{Y}_{T+2} = d + a_1 \hat{Y}_{T+1}$$

$$\hat{e}_{T+2} = e_{T+2} + a_1 \hat{e}_{T+1} = e_{T+2} + a_1 e_{T+1}$$

$$V(\hat{e}_{T+2}) = s^2 (1 + a_1^2)$$

Γενικά, για προβλέψεις h περιόδους μπροστά, $\hat{Y}_{T+h} = d + a_1 \hat{Y}_{T+h-1}$ και

$V(\hat{e}_{T+h}) = s^2 (1 + a_1^2 + a_1^4 + \dots + a_1^{2(h-1)})$. Από την παραπάνω σχέση λοιπόν, είναι φανερό ότι η διακύμανση του σφάλματος πρόβλεψης αυξάνει μη γραμμικά καθώς αυξάνει η περίοδος πρόβλεψης. Ας τονίσουμε εδώ, ότι στο όριο, καθώς η περίοδος πρόβλεψης μεγαλώνει, η πρόβλεψη συγκλίνει προς τον μέσο. Δηλαδή,

$$\lim_{h \rightarrow \infty} \hat{Y}_{T+h} = \frac{d}{1 - a_1} = m$$

3.3.2 ΠΡΟΒΛΕΨΕΙΣ ΜΕ ΥΠΟΔΕΙΓΜΑΤΑ ΜΑ (1)

Έστω το ΜΑ (1) Υπόδειγμα :

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

Όπως και προηγουμένως, η πρόβλεψη στην περίοδο $T+1$ είναι :

$$\hat{Y}_{T+1} = E_T(m + e_{T+1} + q_1 e_T) \quad \text{ή} \quad \hat{Y}_{T+1} = m + q_1 e_T,$$

ενώ το σφάλμα πρόβλεψης είναι: $\hat{e}_{T+1} = Y_{T+1} - \hat{Y}_{T+1} = e_{T+1}$

Συνεπώς, η διακύμανση του σφάλματος πρόβλεψης είναι : $V(\hat{e}_{T+1}) = V(e_{T+1}) = s^2$

Η πρόβλεψη για δυο περιόδους μπροστά, δηλαδή στην περίοδο $T+2$, είναι:

$$\hat{Y}_{T+2} = E_T(Y_{T+2}) = E_T(\mathbf{m} + \mathbf{e}_{T+2} + \mathbf{q}_1 \mathbf{e}_{T+1}) \quad \text{ή} \quad \hat{Y}_{T+2} = \mathbf{m} \quad \text{αφού}$$

$$E_T(\mathbf{e}_{T+2}) = E_T(\mathbf{e}_{T+1}) = 0$$

Η διακύμανση του σφάλματος πρόβλεψης είναι :

$$V(\hat{e}_{T+1}) = V(\mathbf{e}_{T+2} + \mathbf{q}_1 \mathbf{e}_{T+1}) = s^2(1 + \mathbf{q}_1^2)$$

Γενικά για προβλέψεις $h > 1$ περιόδους μπροστά έχουμε:

$$\hat{Y}_{T+m} = \mathbf{m} \quad \text{και} \quad V(\hat{e}_{T+h}) = s^2(1 + \mathbf{q}_1^2)$$

Από την παραπάνω σχέση είναι φανερό ότι ένα υπόδειγμα κινητού μέσου πρώτης τάξης είναι κατάλληλο για προβλέψεις μόνο μια περίοδο μπροστά, αφού για $h > 1$ η πρόβλεψη είναι πάντα ο μέσος. Από τη σχέση $\hat{Y}_{T+1} = \mathbf{m} + \mathbf{q}_1 \mathbf{e}_T$ βλέπουμε ότι μια πρόβλεψη μια περίοδο μπροστά εξαρτάται από την τιμή του τυχαίου όρου ε στη περίοδο T . Επειδή η τιμή αυτή είναι άγνωστη, στην πράξη αντικαθίσταται από την εκτίμησή της, η οποία προκύπτει από τις προηγούμενες τιμές. Συγκεκριμένα, αρχίζοντας από την περίοδο $t=1$, η τιμή της Y από την σχέση $Y_t = \mathbf{m} + \mathbf{e}_t + \mathbf{q}_1 \mathbf{e}_{t-1}$ είναι : $Y_1 = \mathbf{m} + \mathbf{e}_1 + \mathbf{q}_1 \mathbf{e}_0$. Η πρόβλεψη εξάλλου είναι $\hat{Y}_1 = E_0(Y_1) = \mathbf{m} + \mathbf{q}_1 \mathbf{e}_0$, αφού $E_0(\mathbf{e}_1) = 0$

Η τιμή όμως, \mathbf{e}_0 , που αναφέρεται στην περίοδο, δεν είναι συνήθως γνωστή και αντικαθίσταται από την προσδοκώμενη τιμή της που είναι μηδέν. Επομένως, $\hat{Y}_1 = \mathbf{m}$ και $\hat{e}_1 = Y_1 - \hat{Y}_1 = Y - \mathbf{m}$. Η πρόβλεψη τώρα της Y μια περίοδο μπροστά από την περίοδο $t=1$, η πρόβλεψη δηλαδή της Y_2 , είναι : $\hat{Y}_2 = \mathbf{m} + \mathbf{q}_1 \mathbf{e}_1 = \mathbf{m} + \mathbf{q}_1 \hat{e}_1$ και $\hat{e}_2 = Y_2 - \hat{Y}_2$. Συνεχίζοντας

με αυτόν τον τρόπο εκτιμάμε όλα τα ε , δηλαδή $\hat{e}_1, \hat{e}_2, \dots, \hat{e}_T$. Έτσι έχουμε την εκτίμηση της ε_T για να γίνει πρόβλεψη στην περίοδο $T+1$.

3.3.3 ΠΡΟΒΛΕΨΕΙΣ ΜΕ ΥΠΟΔΕΙΓΜΑΤΑ ARMA

Ας δούμε πρώτα την περίπτωση ενός ARMA (1,1) υποδείγματος, δηλαδή του υποδείγματος:

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

Για $t=T+1$, το παραπάνω υπόδειγμα γίνεται: $Y_{T+1} = d + a_1 Y_T + e_{T+1} + q_1 e_T$

Σύμφωνα με την προηγούμενη ανάλυση, η άριστη πρόβλεψη για μια χρονική περίοδο μπροστά θα είναι:

$$\hat{Y}_{T+1} = E_T(d + a_1 Y_T + e_{T+1} + q_1 e_T) \quad \text{ή} \quad \hat{Y}_{T+1} = d + a_1 Y_T + q_1 e_T$$

Το σφάλμα προβλέψεως και η διακύμανσή του, εξάλλου, δίνονται από τις ακόλουθες σχέσεις:

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

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

Ακολουθώντας την ίδια διαδικασία μπορούμε να κάνουμε προβλέψεις για δύο, τρεις και γενικά h περιόδους μπροστά. Ας σημειωθεί ότι και σ' αυτήν την περίπτωση, στο όριο, καθώς μεγαλώνει η περίοδος προβλέψεως, η πρόβλεψη συγκλίνει προς τον μέσο (Song et.al. 2003).

3.3.4 ΠΡΟΒΛΕΨΕΙΣ ΜΕ ΥΠΟΔΕΙΓΜΑΤΑ ARIMA

Έστω το αυτοπαλίνδρομο ολοκληρωμένο υπόδειγμα πρώτης τάξης :

$$w_t = d + a_1 w_{t-1} + e_t$$

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

Δηλαδή, η μη στάσιμη σειρά $\{Y_t\}$ γίνεται στάσιμη παίρνοντας τις πρώτες διαφορές, οι οποίες είναι AR(1).

Είναι φανερό ότι με το υπόδειγμα $w_t = d + a_1 w_{t-1} + e_t$ μπορεί να γίνει πρόβλεψη της διαφοράς w και στη συνέχεια πρόβλεψη της Y . Κατά τα γνωστά, η άριστη πρόβλεψη της w_{T+1} είναι:

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

Η πρόβλεψη τώρα για την Y_{T+1} είναι:

$$\begin{aligned} \hat{Y}_{T+1} &= Y_t + \hat{w}_{T+1} \\ &= Y_t + d + a_1 w_T \\ &= Y_t + d + a_1 (Y_t - Y_{t-1}) \\ &= d + (1 + a_1) Y_t - a_1 Y_{t-1} \end{aligned}$$

Γενικά, η πρόβλεψη h περιόδους μπροστά, είναι

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

ΚΕΦΑΛΑΙΟ 4

ΜΕΛΕΤΗ ΠΕΡΙΠΤΩΣΗΣ - CASE STUDY

Μια χρονολογική σειρά είναι μια σειρά δεδομένων που βασίζονται σε (ή καταχωρημένα ή με γραφική παράσταση) χρονοδιάγραμμα. Συνήθως, μια χρονολογική σειρά είναι μια ακολουθία που λαμβάνεται σε διαδοχικά ισαπέχουσες χρονικές στιγμές. Έτσι είναι μια ακολουθία δεδομένων με διακριτό χρόνο. Παραδείγματα χρονολογικών σειρών είναι τα ύψη της παλίρροιας των ωκεανών, οι μετρήσεις των ηλιακών κηλίδων και η ημερήσια τιμή κλεισίματος του Dow Jones.

Οι χρονολογικές σειρές παρουσιάζονται πολύ συχνά μέσω γραφημάτων γραμμών. Οι χρονοσειρές χρησιμοποιούνται στη στατιστική, στην επεξεργασία σήματος, στην αναγνώριση προτύπων, στην οικονομετρία, στη μαθηματική χρηματοδότηση, στην πρόγνωση καιρού, στην πρόβλεψη σεισμού, στα ηλεκτροεγκεφαλογράφηματα, στη μηχανική ελέγχου, στην αστρονομία, στη μηχανική επικοινωνιών και σε μεγάλο βαθμό σε οποιοδήποτε τομέα εφαρμοσμένης επιστήμης και μηχανικής που περιλαμβάνει χρονικές μετρήσεις.

Η ανάλυση χρονολογικών σειρών περιλαμβάνει μεθόδους για την ανάλυση δεδομένων με σκοπό την απόσπαση σημαντικών στατιστικών και άλλων χαρακτηριστικών των δεδομένων. Για την πρόβλεψη μέσω χρονολογικών σειρών πραγματοποιείται η χρήση ενός μοντέλου που χρησιμοποιείται για την πρόβλεψη μελλοντικών τιμών βάσει των προηγούμενων τιμών. Ενώ η ανάλυση παλινδρόμησης χρησιμοποιείται συχνά με τέτοιο τρόπο ώστε να ελέγχει τις θεωρίες ότι οι τρέχουσες τιμές μιας ή περισσότερων ανεξάρτητων χρονοσειρών επηρεάζουν την τρέχουσα τιμή μιας άλλης χρονοσειράς, αυτός ο τύπος ανάλυσης των χρονοσειρών δεν ονομάζεται "ανάλυση χρονοσειρών" γιατί αυτή η ανάλυση επικεντρώνεται στη σύγκριση τιμών μιας μόνο χρονοσειράς ή πολλαπλών εξαρτημένων χρονοσειρών σε διαφορετικά χρονικά σημεία. Η ανάλυση διακοπής της χρονικής σειράς είναι η ανάλυση των παρεμβάσεων σε μία μόνο χρονολογική σειρά

Τα δεδομένα χρονοσειρών έχουν μια φυσική χρονική ταξινόμηση. Αυτό καθιστά την ανάλυση των χρονολογικών σειρών διακριτή από τις μελέτες εγκάρσιας τομής, στις οποίες δεν υπάρχει φυσική ταξινόμηση των παρατηρήσεων (π.χ. παρουσίαση των μισθών των ανθρώπων με αναφορά στα αντίστοιχα επίπεδα εκπαίδευσης, όπου τα δεδομένα των ατόμων θα μπορούσαν να εισαχθούν με οποιαδήποτε σειρά). Η ανάλυση χρονοσειρών διακρίνεται επίσης από την ανάλυση χωρικών δεδομένων, όπου οι παρατηρήσεις σχετίζονται τυπικά με γεωγραφικές τοποθεσίες (π.χ. καταγραφή των τιμών των κατοίκων από την τοποθεσία καθώς και των εγγενών χαρακτηριστικών των σπιτιών). Ένα στοχαστικό μοντέλο για μια χρονολογική σειρά θα αντικατοπτρίζει γενικά το γεγονός ότι οι παρατηρήσεις που βρίσκονται κοντά στο χρόνο θα είναι πιο στενά συνδεδεμένες από τις παρατηρήσεις που διαχωρίζονται περαιτέρω. Επιπλέον, τα μοντέλα χρονολογικών σειρών θα χρησιμοποιούν συχνά τη φυσική μονόδρομη ταξινόμηση του χρόνου, έτσι ώστε οι τιμές για μια δεδομένη περίοδο να εκφράζονται ως προκύπτοντες κατά κάποιο τρόπο από προηγούμενες τιμές και όχι από μελλοντικές τιμές.

Η διαδικασία διερεύνησης των χρονολογικών σειρών μέσω του SPSS υπολογίζει την εκθετική εξομάλυνση, τη μονοπύρηνη αυτόματη αντιστροφή του ολοκληρωμένου κινητού μέσου (ARIMA) και τα πολυπαραγοντικά μοντέλα ARIMA (ή μοντέλα λειτουργιών μεταφοράς) για χρονολογικές σειρές και παράγει προβλέψεις. Η διαδικασία περιλαμβάνει τον ορισμό προσαρμοσμένων μοντέλων ARIMA ή εκθετικής εξομάλυνσης προκειμένου να βρεθεί το κατάλληλο μοντέλο για πρόβλεψη.

Στη συνέχεια πρόκειται να πραγματοποιηθεί ανάλυση χρονολογικών σειρών για την περίπτωση των ΚΤΕΛ Ηλείας.

Πίνακας εσόδων του ΚΤΕΛ Ηλείας

2841487	Α τρίμηνο 2007
2790913	Β τρίμηνο 2007
3219054	Γ τρίμηνο 2007
3076181	Δ τρίμηνο 2007
2777332	Α τρίμηνο 2008
3021657	Β τρίμηνο 2008
3160493	Γ τρίμηνο 2008
3160870	Δ τρίμηνο 2008
2731416	Α τρίμηνο 2009
3079454	Β τρίμηνο 2009
3289219	Γ τρίμηνο 2009
3277049	Δ τρίμηνο 2009
2957497	Α τρίμηνο 2010
3207978	Β τρίμηνο 2010
3330909	Γ τρίμηνο 2010
3265513	Δ τρίμηνο 2010
2975918	Α τρίμηνο 2011
3204605	Β τρίμηνο 2011
3306366	Γ τρίμηνο 2011
3320456	Δ τρίμηνο 2011
2902634	Α τρίμηνο 2012
3185954	Β τρίμηνο 2012
3115145	Γ τρίμηνο 2012
3137669	Δ τρίμηνο 2012
2760481	Α τρίμηνο 2013
2934616	Β τρίμηνο 2013
3015369	Γ τρίμηνο 2013
3068402	Δ τρίμηνο 2013
2754552	Α τρίμηνο 2014

2899037	Β τρίμηνο 2014
3075882	Γ τρίμηνο 2014
3019026	Δ τρίμηνο 2014
2706555	Α τρίμηνο 2015
2814064	Β τρίμηνο 2015
2780844	Γ τρίμηνο 2015
2978295	Δ τρίμηνο 2015
2639864	Α τρίμηνο 2016
2730422	Β τρίμηνο 2016
2666406	Γ τρίμηνο 2016
2790134	Δ τρίμηνο 2016

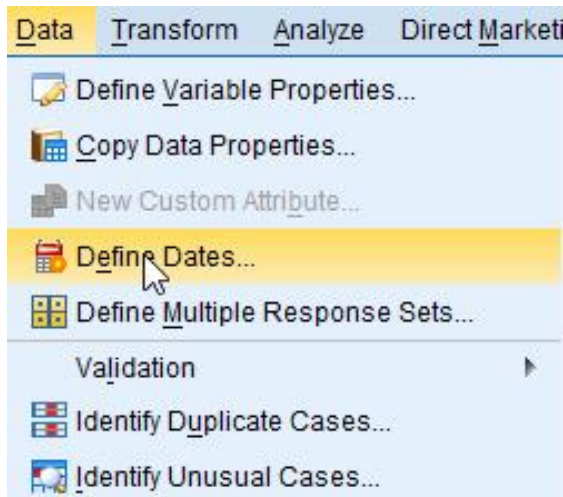
Πίνακας 1: ο πίνακας που προηγείται παρουσιάζει τα έσοδα των ΚΤΕΛ Αχαΐας από το πρώτο τρίμηνο του 2007 έως το τέταρτο του 2016.

Όταν εισαχθούν τα δεδομένα στο SPSS το αρχείο έχει τη μορφή της εικόνας 1

	VAR00001	var	var	var	var	var	var	var	var	var
1	2841487									
2	2790913									
3	3219054									
4	3076181									
5	2777332									
6	3021657									
7	3160493									
8	3160870									
9	2731416									
10	3079454									
11	3289219									
12	3277049									
13	2967497									
14	3207978									
15	3330909									
16	3265513									
17	2976918									
18	3204605									
19	3306366									
20	3320456									
21	2902634									
22	3186964									
23	3116145									
24	3137669									
25	2760481									

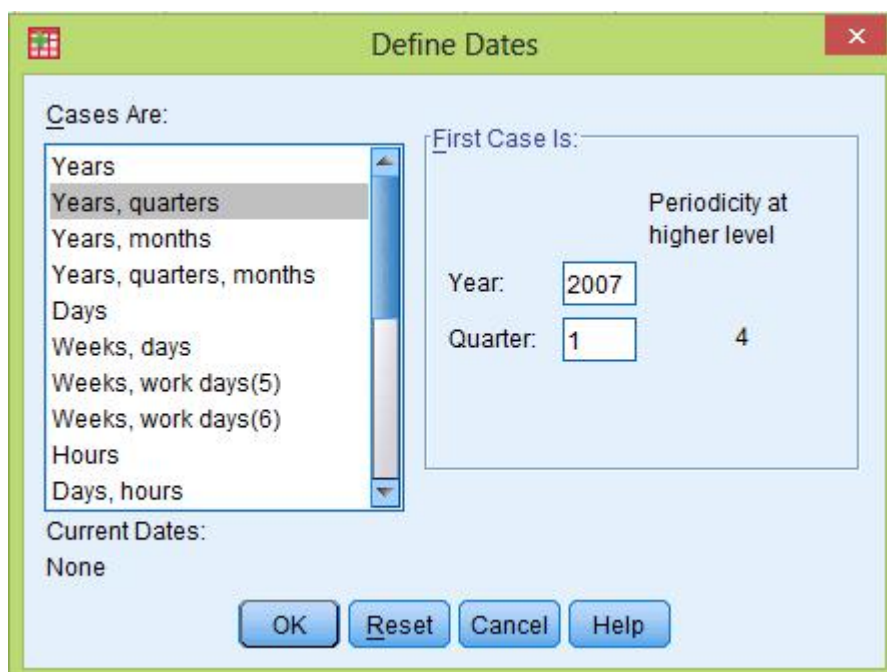
Εικόνα 1 : Εισαγωγή των δεδομένων στο SPSS

Προκειμένου να εισαχθούν οι χρονολογίες στο πρόγραμμα γίνεται η επιλογή του πεδίου Data και έπειτα του Define Dates. (εικόνα 2)



Εικόνα 2 : Εισαγωγή των χρονολογιών

Αφού επιλεγεί το πεδίο Data και έπειτα του Define Dates, εμφανίζεται το παράθυρο Define Dates όπου γίνεται επιλογή του πεδίου Years, quarters και έπειτα δηλώνεται αρχικό έτος το 2007 και τρίμηνο το 1. Έπειτα επιλέγεται το πεδίο OK.



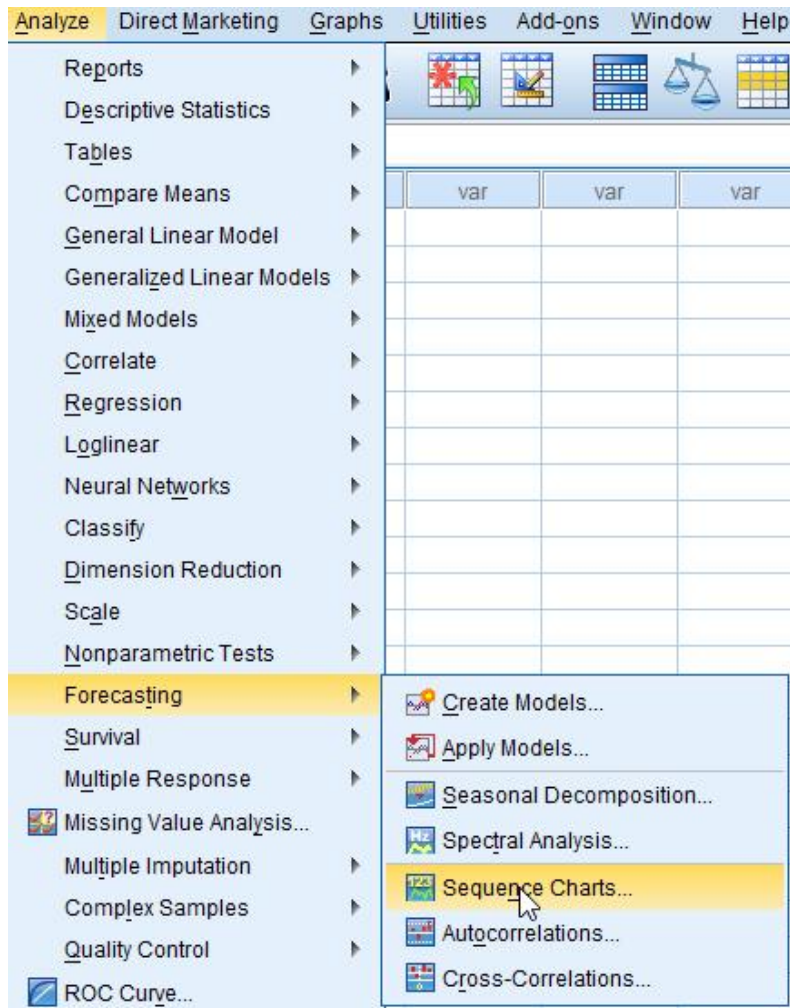
Εικόνα 3 : Εισαγωγή των χρονολογιών #2

Αφού εισαχθούν οι χρονολογίες στο πρόγραμμα πλέον το παράθυρο Data View έχει τη μορφή της εικόνας 4, αφού έχουν εισαχθεί τρεις ακόμη μεταβλητές που ορίζουν, η πρώτη τα έτη, η δεύτερη τα τρίμηνα και η τρίτη και τα δύο μαζί σε αλφαριθμητική μορφή.

	VAR00001	YEAR_	QUARTER_	DATE_
1	2841487	2007	1	Q1 2007
2	2790913	2007	2	Q2 2007
3	3219054	2007	3	Q3 2007
4	3076181	2007	4	Q4 2007
5	2777332	2008	1	Q1 2008
6	3021657	2008	2	Q2 2008
7	3160493	2008	3	Q3 2008
8	3160870	2008	4	Q4 2008
9	2731416	2009	1	Q1 2009
10	3079454	2009	2	Q2 2009
11	3289219	2009	3	Q3 2009
12	3277049	2009	4	Q4 2009
13	2957497	2010	1	Q1 2010
14	3207978	2010	2	Q2 2010
15	3330909	2010	3	Q3 2010
16	3265513	2010	4	Q4 2010
17	2975918	2011	1	Q1 2011

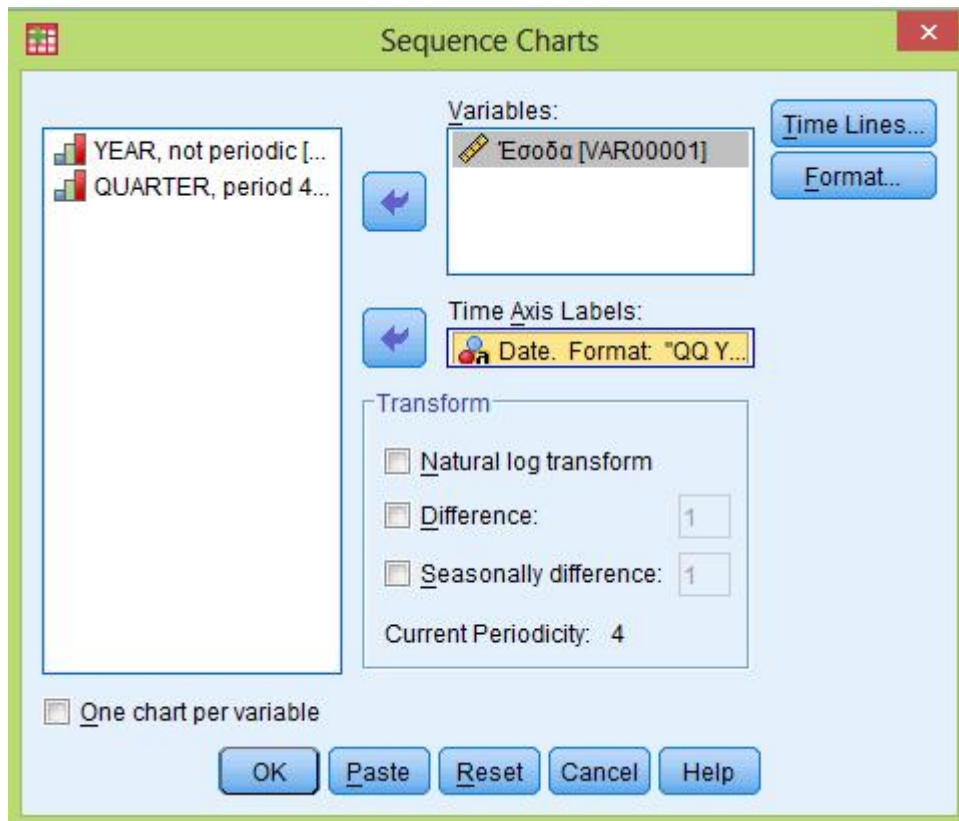
Εικόνα 4 : Στιγμιότυπο έπειτα από την εισαγωγή των χρονολογιών

Αφότου πραγματοποιηθεί η εισαγωγή των δεδομένων και οριστούν οι χρόνοι γίνεται ο ορισμός των χρονολογικών σειρών μέσω της διαδικασίας Analyze à Forecasting και έπειτα Sequence Charts. Αυτό πραγματοποιείται προκειμένου να πραγματοποιηθεί έλεγχος στασιμότητας.



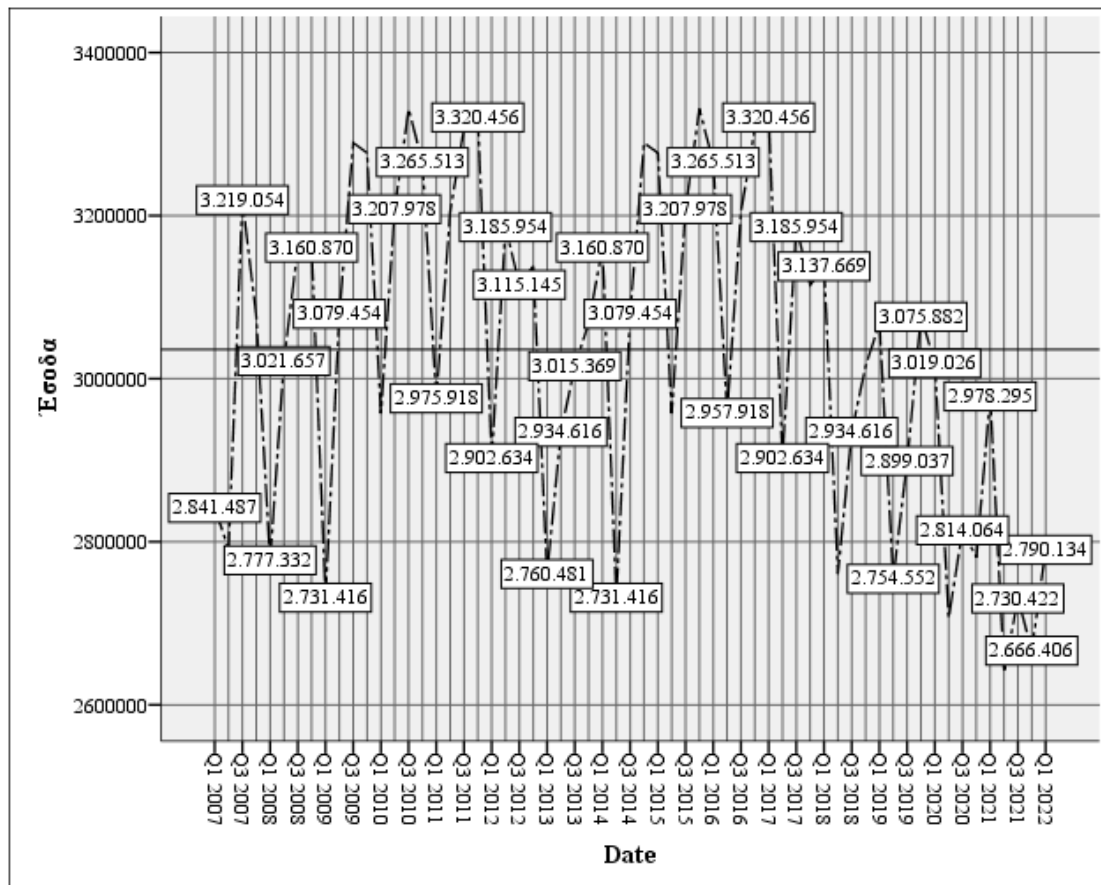
Εικόνα 5 : Ορισμός της χρονοσειράς #1

Στο παράθυρο που εμφανίζεται εισάγονται τα Έσοδα των ΚΤΕΛ και έπειτα εισάγεται και η μεταβλητή που δημιουργήθηκε και πραγματεύεται τα έτη σε συνδυασμό με τα εξάμηνα. Έπειτα επιλέγεται το πεδίο Οκ, προκειμένου να υπολογιστούν οι πίνακες και να γίνει έλεγχος στασιμότητας.

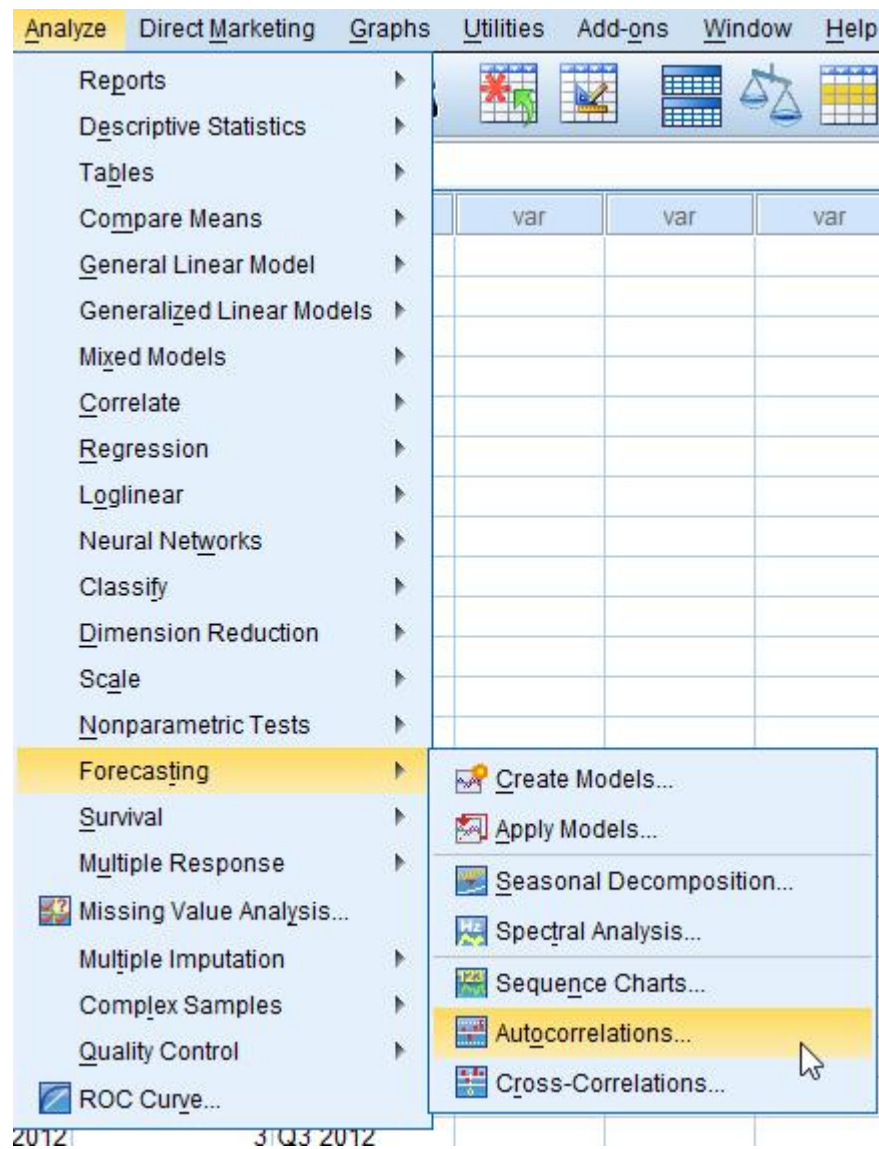


Εικόνα 6 : Ορισμός της χρονοσειράς #2

Όπως είναι εμφανές στο γράφημα 1 η χρονολογική σειρά δείχνει στάσιμη ως προς το μέσο. Συνεπώς θα ακολουθήσει έλεγχος αυτοσυσχετίσεων. Αυτό είναι δυνατό μέσω της διαδικασίας Analyze → Forecasting → Autocorrelations (εικόνα 7)

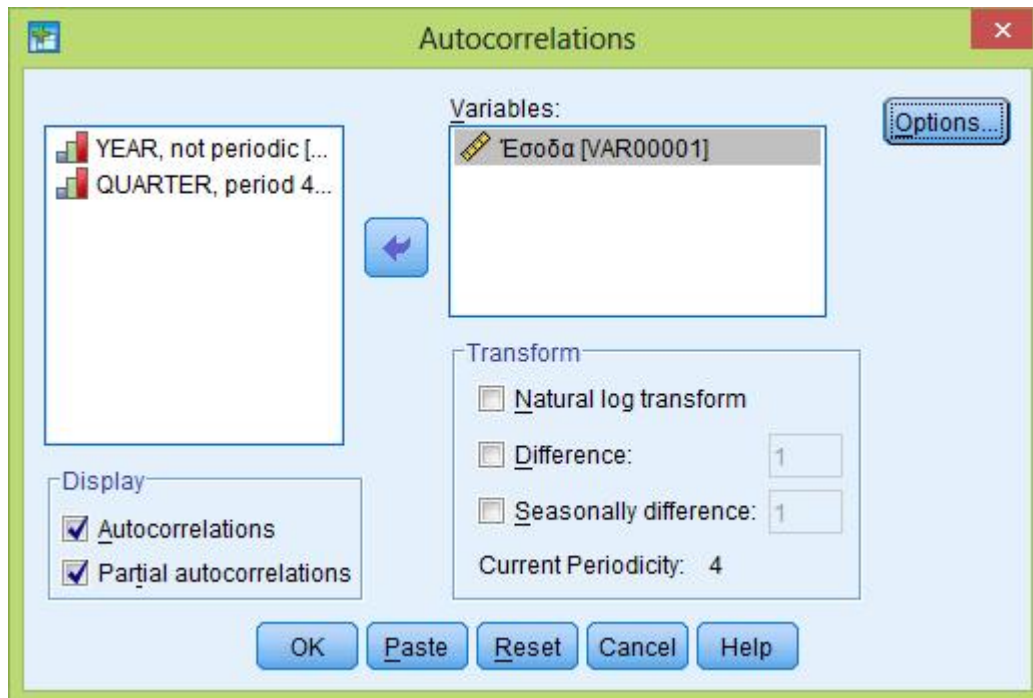


Γράφημα 1: Παρουσίαση της στασιμότητας των εσόδων του ΚΤΕΛ Ηλείας ως προς το χρόνο



Εικόνα 7 : Έλεγχος αυτοσυσχετίσεων

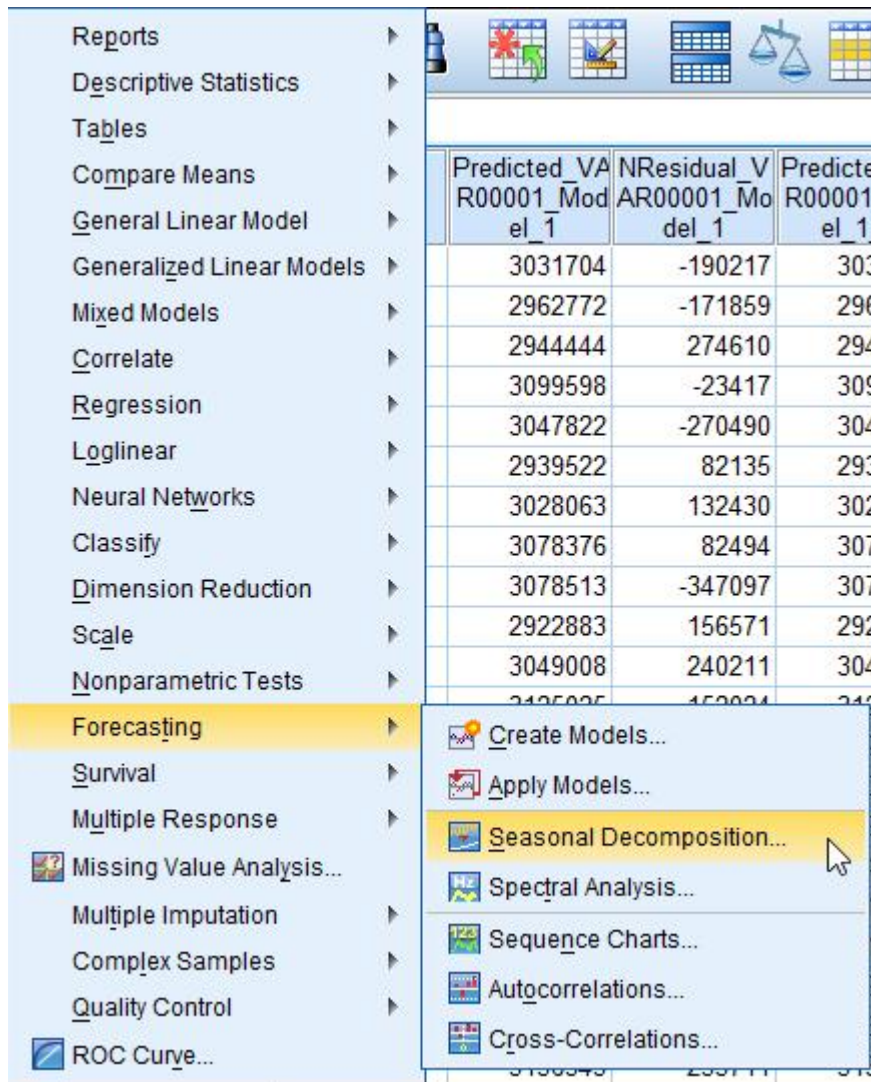
Στο παράθυρο που εμφανίζεται εισάγονται τα δεδομένα των εσόδων και έπειτα το πεδίο OK ώστε να εμφανιστούν τα δεδομένα που απαιτούνται για την ανάλυση της αυτοσυσχέτισης.



Εικόνα 8 : Έλεγχος αυτοσυσχετίσεων #2

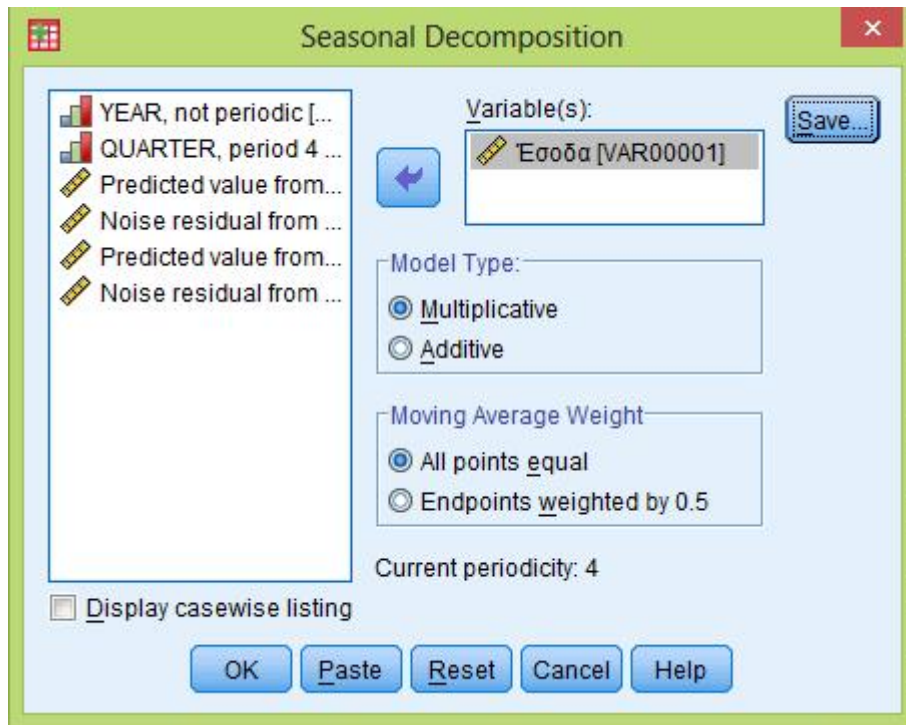
Από τον πίνακα ACF φαίνεται πως οι συντελεστές sig φθίνουν γρήγορα προς το μηδέν όσο περνάει ο χρόνος, όμως υπάρχει μια εμφανής εποχικότητα. Από το p4 και έπειτα φαίνεται να υπάρχει μηδενισμός εξαιρώντας την εποχικότητα.

Στη συνέχεια πραγματοποιείται έλεγχος εποχικότητας μέσω της διαδικασίας Analyze → Forecasting → Seasonal Decomposition.



Εικόνα 9 : Έλεγχος εποχικότητας

Έπειτα επιλέγεται από το παράθυρο που εμφανίζεται στο πεδίο των μεταβλητών η μεταβλητή έσοδα και OK.



Εικόνα 10 : Έλεγχος εποχικότητας #2

Στον πίνακα που ακολουθεί είναι εμφανείς τα ποσοστά που επηρεάζονται τα έσοδα από την εποχικότητα.

Seasonal Factors

Series

Name: Έσοδα

Perio	Seasonal
d	Factor (%)
1	99,1
2	96,1
3	101,8
4	102,9

Πίνακας εποχικότητας.

Autocorrelations

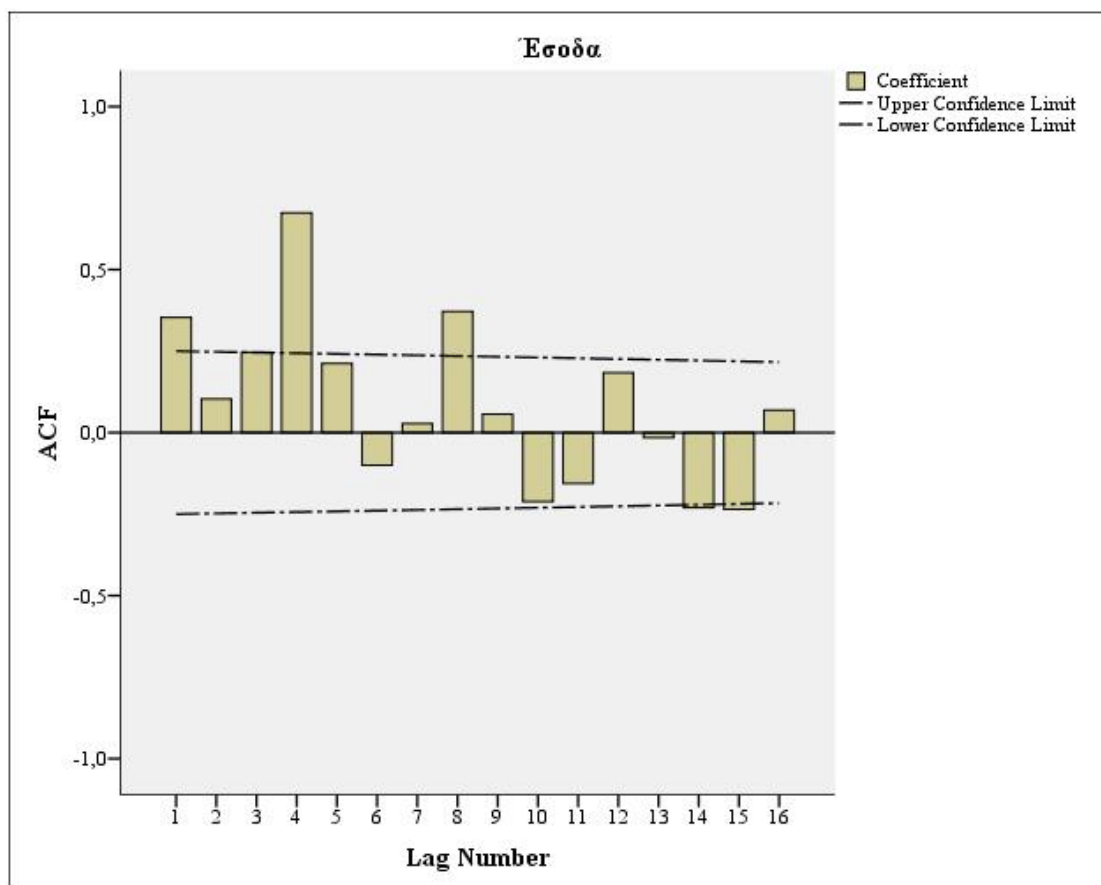
Series: Έσοδα

Lag	Autocorre lation	Std. Error ^a	Box-Ljung Statistic		
			Valu e	df	Sig. ^b
1	,354	,125	8,00	1	,005
			4		
2	,103	,124	8,69	2	,013
			6		
3	,247	,123	12,7	3	,005
			30		
4	,674	,122	43,3	4	,000
			74		
5	,212	,121	46,4	5	,000
			53		
6	-,100	,120	47,1	6	,000
			50		
7	,028	,119	47,2	7	,000
			04		
8	,372	,117	57,2	8	,000
			17		
9	,057	,116	57,4	9	,000
			56		
10	-,211	,115	60,8	10	,000
			20		
11	-,156	,114	62,6	11	,000
			97		
12	,184	,113	65,3	12	,000
			39		

13	-,016	,112	65,3	13	,000
			59		
14	-,231	,111	69,7	14	,000
			09		
15	-,235	,109	74,3	15	,000
			23		
16	,069	,108	74,7	16	,000
			30		

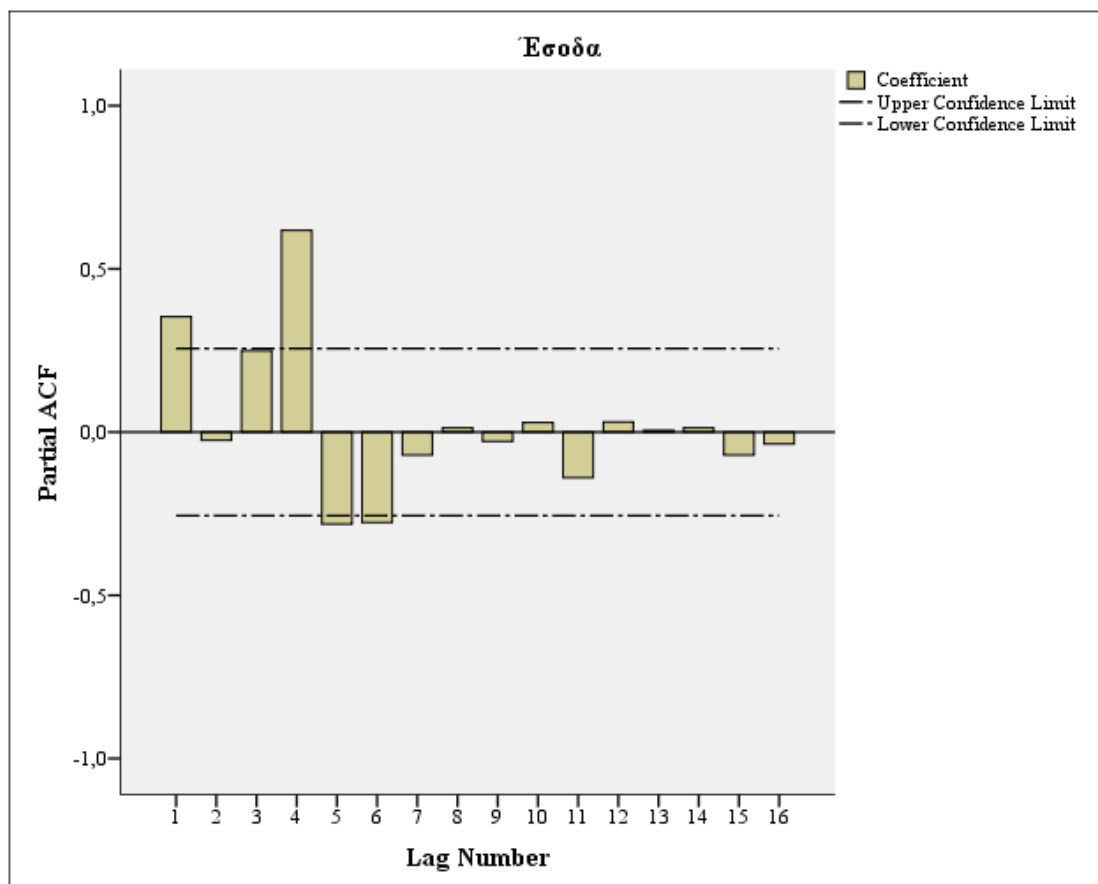
- a. The underlying process assumed is independence (white noise).
- b. Based on the asymptotic chi-square approximation.

Πίνακας 1: πίνακας αυτοσυσχετίσεων των δεδομένων των εσόδων.



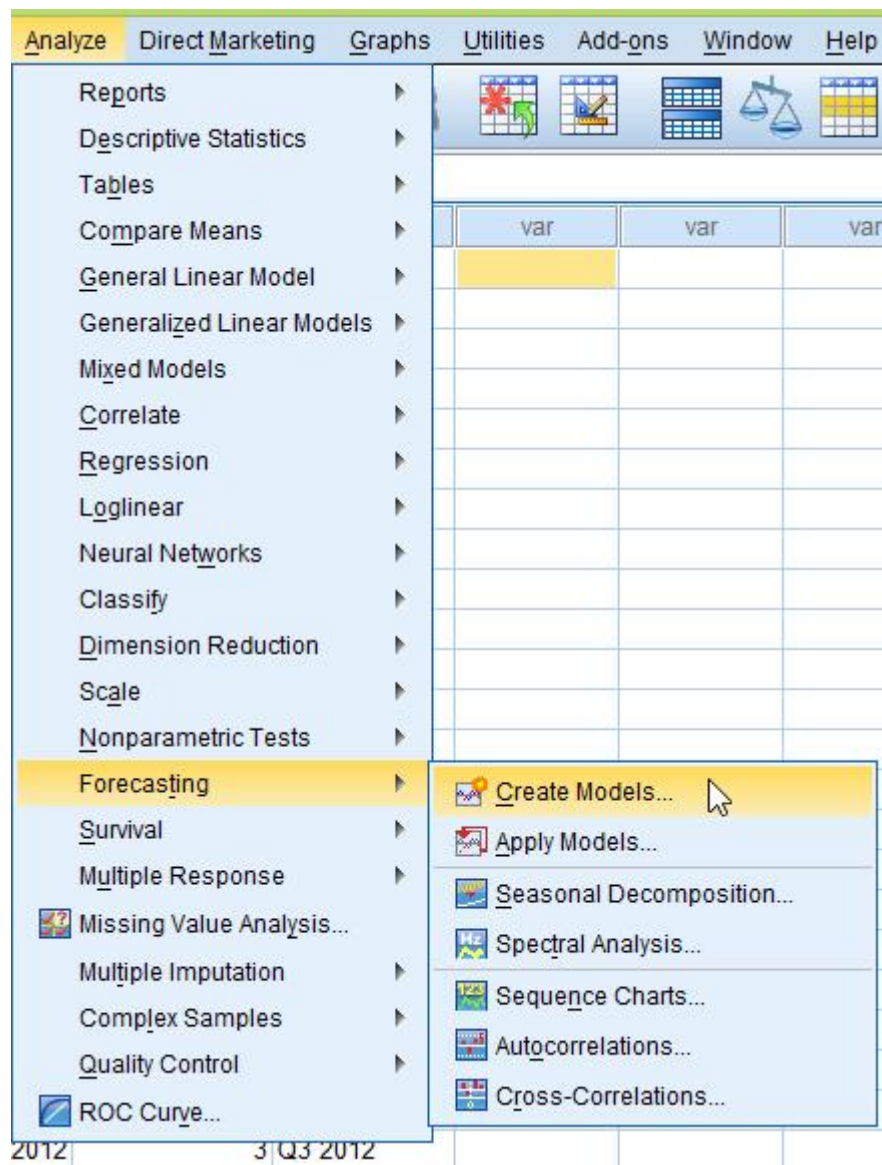
Γράφημα 2: Γράφημα αυτοσυσχέτισης ACF

Από το γράφημα PACF είναι εμφανές πως υπάρχει στις 3 πρώτες τιμές μια μη μηδενική τιμή. Συνεπώς το $p=1$. Από το διάγραμμα ACF υπάρχει και εδώ μια μη μηδενική τιμή άρα το $q=2$ και επειδή υπάρχει στασιμότητα το $d=0$. Άρα το μοντέλο ARIMA που θα διερευνηθεί είναι το ARIMA (1,0,1).



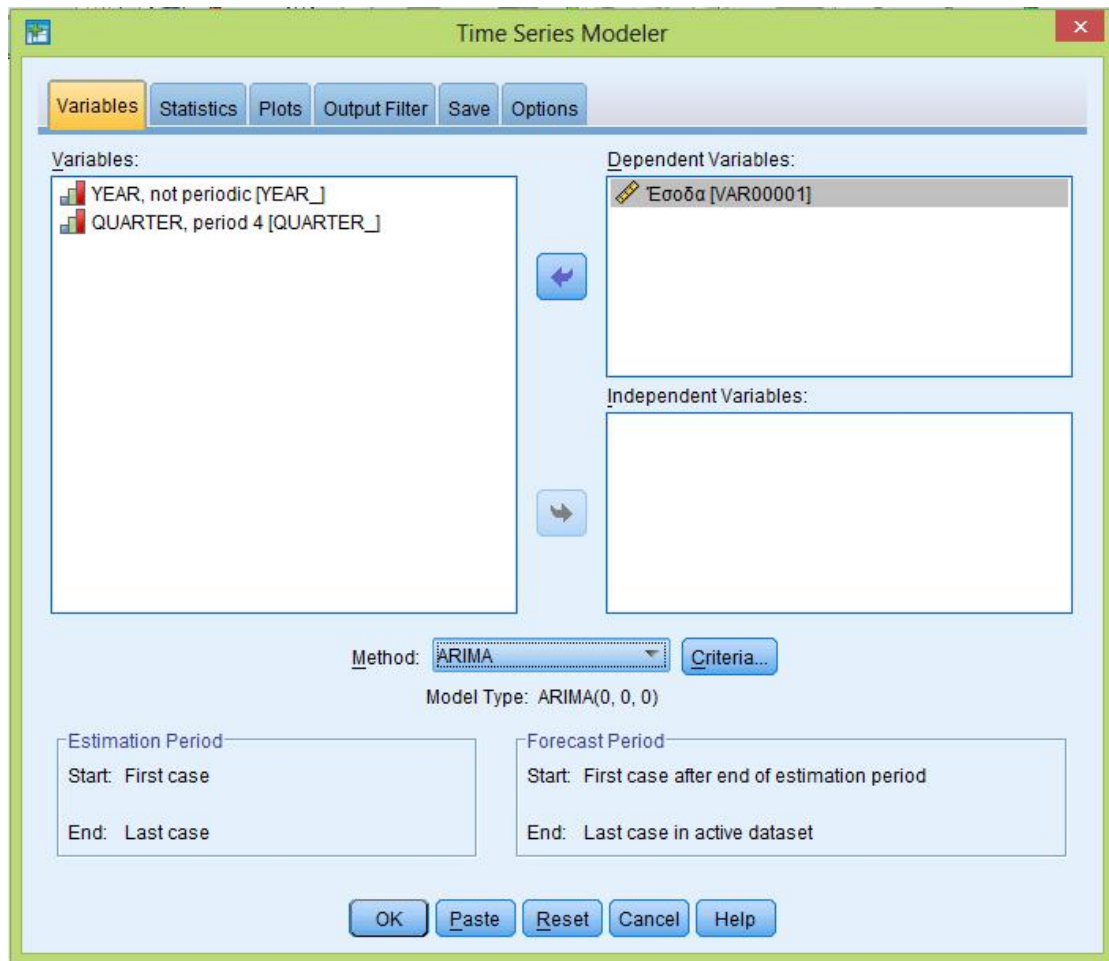
Γράφημα 3: Γράφημα μερικής αυτοσυσχέτισης PACF

Αφού βρέθηκε ο τύπος μοντέλου ARIMA που θα χρησιμοποιηθεί, διενεργείται δημιουργία μοντέλων, μέσω της διεργασίας Analyze à Forecasting à Create Models. (εικόνα 9).



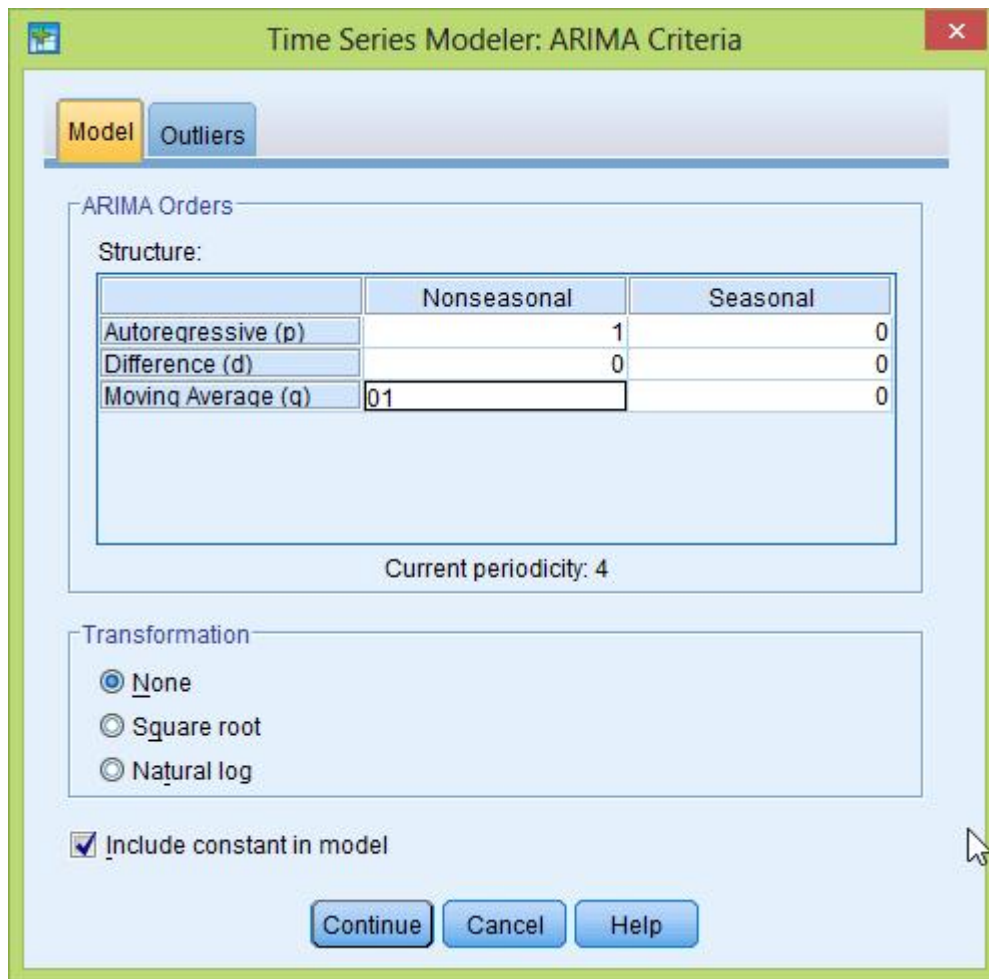
Εικόνα 11 : Δημιουργία μοντέλου

Στο παράθυρο που εμφανίζεται επιλέγονται τα έσοδα ως Εξαρτημένη μεταβλητή, ενώ στη μέθοδο επιλέγεται η επιλογή ARIMA. Εδώ πρέπει να σημειωθεί ότι επιλέγεται το πεδίο criteria στη συνέχεια.



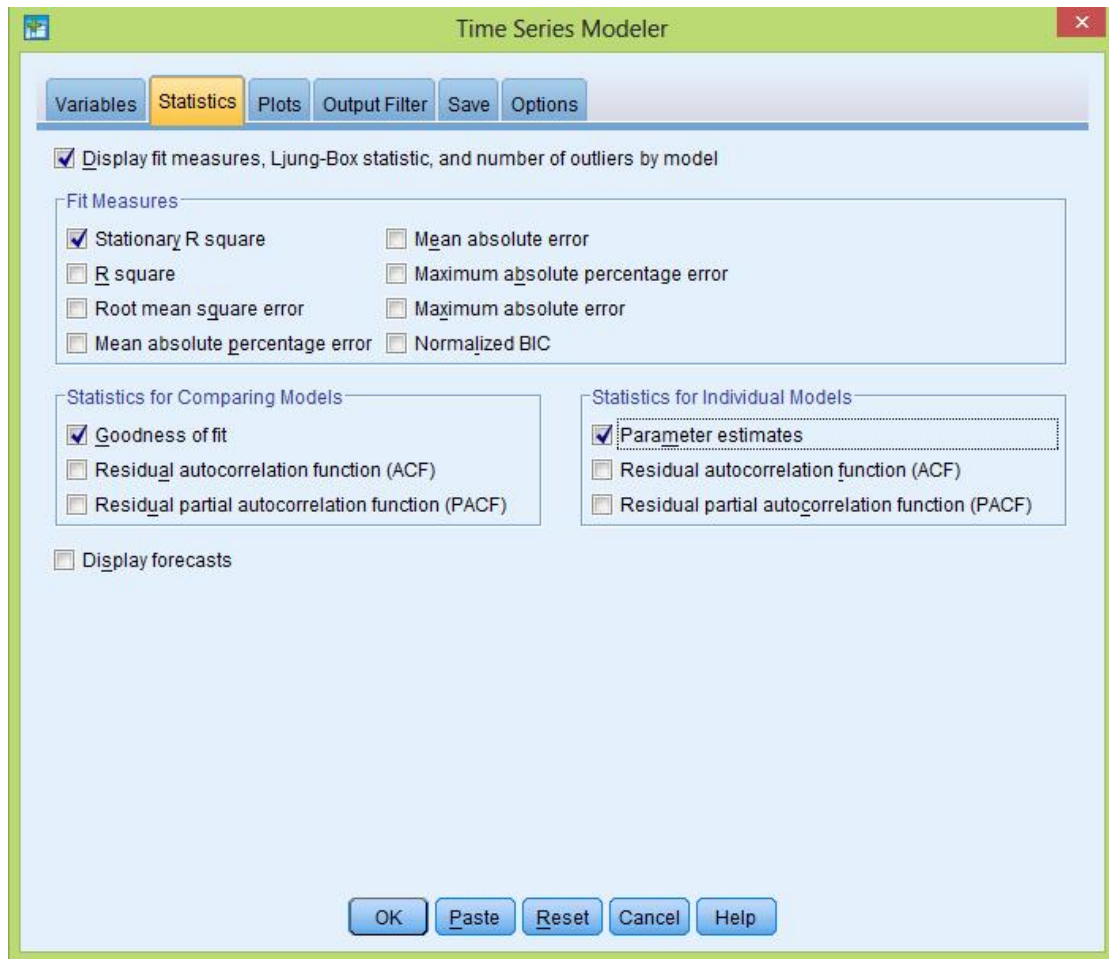
Εικόνα 12 : Δημιουργία μοντέλου ARIMA

Στα κριτήρια που επιλέχθηκαν δηλώνονται οι τιμές των p, d, q του μοντέλου και έπειτα το πεδίο OK.



Εικόνα 13 : Δημιουργία μοντέλου ARIMA #2

Αφότου πραγματοποιηθεί η προαναφερθέντα διεργασία επιλέγεται από το αρχικό μενού το πεδίο Statistics και σε το νέο παράθυρο επιλέγεται το πεδίο Estimate Parameters.



Εικόνα 14 : Δημιουργία μοντέλου ARIMA #3

ΜΟΝΤΕΛΟ ARIMA (1,0,0)

Εδώ είναι εμφανές ότι το α ισούται με 3032660,562 το $\alpha_1=0,263$ και το $\theta_1 =-0,111$

Εδώ είναι επίσης εμφανές πως το sig της σταθεράς είναι ίσο με 0 άρα είναι στατιστικά σημαντικό ενώ τα υπόλοιπα μέτρα λόγω του ότι οι συντελεστές sig τους είναι μεγαλύτεροι του 0,05 δεν είναι στατιστικά σημαντικά. Λόγο της μη στατιστικής σημαντικότητας όλων των παραμέτρων δεν είναι σωστή η επιλογή του για πρόβλεψη.

Έτσι θα πραγματοποιηθεί ξανά η διαδικασία ARIMA αφαιρώντας έναν από τους δύο παραμέτρους και πιο συγκεκριμένα το θ αφού έχει μεγαλύτερο pvalue.

ARIMA Model Parameters

				Estimate	SE	t	Sig.
Έσοδα-	Έσο	No	Constant	3032660,	36890,	82,20	,000
Model_1	δα	Transformat		562	304	8	
		ion	AR Lag 1	,263	,360	,731	,468
			MA Lag 1	-,111	,368	-,301	,764

Πίνακας των παραμέτρων του μοντέλου Arima (1,0,1)

Εδώ είναι εμφανές ότι το α ισούται με 3031704,347 το $\alpha_1=0,362$.

Εδώ είναι επίσης εμφανές πως το sig της σταθεράς είναι ίσο με 0 άρα είναι στατιστικά σημαντικό ενώ και το sig του AR είναι στατιστικά σημαντικό αφού είναι μικρότερο του 0,05. Λόγο της στατιστικής σημαντικότητας όλων των παραμέτρων η επιλογή του μοντέλου για πρόβλεψη είναι σωστή.

Η ευθεία που δημιουργείται εν τέλει είναι η $y_t = 3031704,347 + 0,362y_{t-1}e_t$

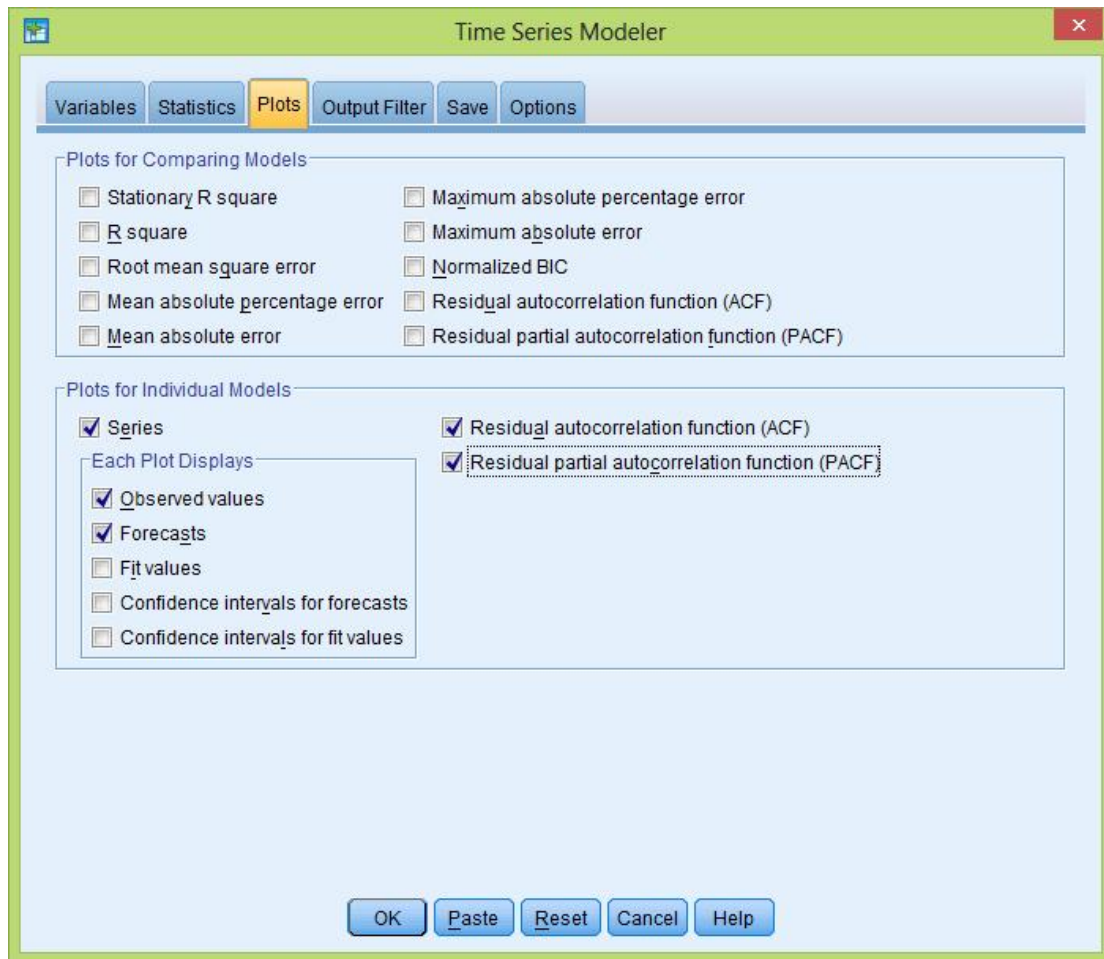
ARIMA Model Parameters

				Estimate	SE	t	Sig.
Έσοδα- Model_1	Έσο δα	No Transformat ion	Constant	3031704,	37971,	79,84	,000
			AR Lag 1	,362	,123	2,952	,005

Πίνακας των παραμέτρων του μοντέλου Arima (1,0,0)

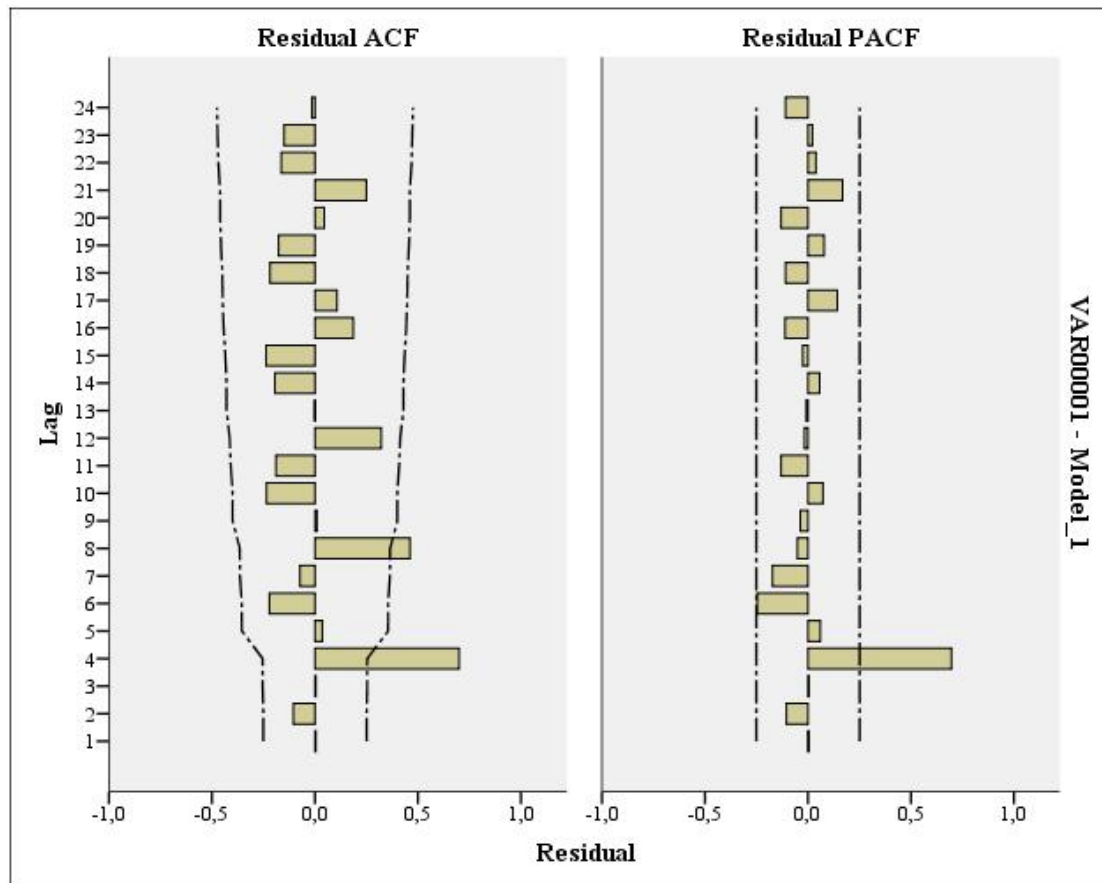
Αφού επιλέχθηκε το μοντέλο καλό είναι να πραγματοποιηθεί διαγνωστικός έλεγχος για το εάν το μοντέλο εξηγεί επαρκώς τα δεδομένα. Αυτό γίνεται με τα κριτήρια της στασιμότητας, της τυχαίας κατανομής και της κανονικής κατανομής. Για τα δύο πρώτα αρκεί να αποδειχθεί ότι τα e_t είναι λευκός θόρυβος. Εάν στα γραφήματα ACF & PACF υπάρχουν μηδενικές τιμές αυτό ισχύει.

Μέσω της διαδικασίας Analyze → Forecasting → Create models (εικόνα 9) και έπειτα επιλέγοντας το πεδίο Plots και τα πεδία ACF και PACF όπως στην εικόνα 13, εμφανίζονται στο αρχείο του output τα γραφήματα που θα αναλυθούν.



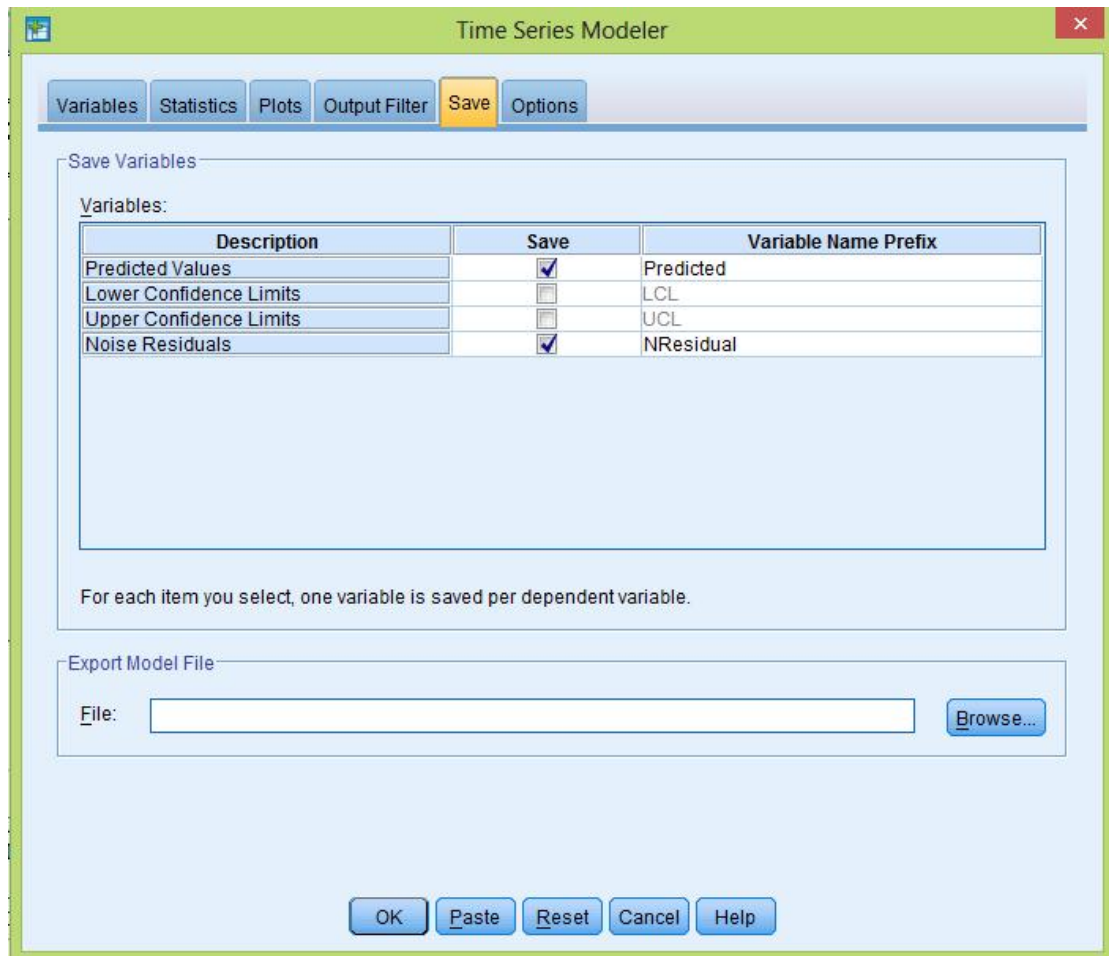
Εικόνα 15 : Διαγνωστικός έλεγχος των σφαλμάτων

Εδώ είναι εμφανές πως οι τιμές βρίσκονται στο μηδέν συνεπώς υπάρχει λευκός θόρυβος. Η ύπαρξη μιας τιμής εκτός είναι πιθανό αποτέλεσμα της εποχικότητας.



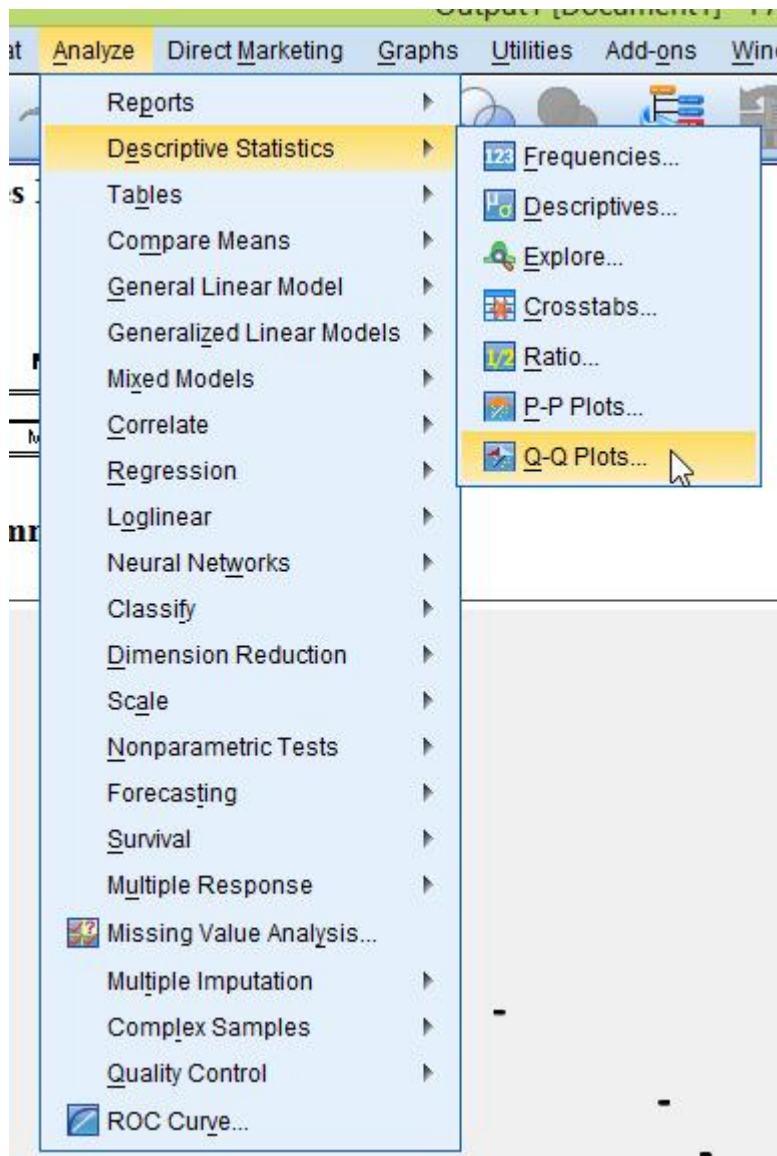
Γράφημα λευκού θορύβου του μοντέλου ARIMA (1,0,0)

Στη συνέχεια ακολουθεί η διερεύνηση της κανονικότητας του δείγματος, αυτό γίνεται από τη διαδικασία Analyze à Forecasting à Create models (εικόνα 9) και έπειτα επιλέγονται τα δύο πεδία Predicted values και Noise Residuals από το πεδίο Save (εικόνα 14). Μέσω αυτής της διαδικασία δημιουργούνται στο αρχείο των δεδομένων νέες μεταβλητές, οι οποίες θα χρησιμοποιηθούν στη συνέχεια

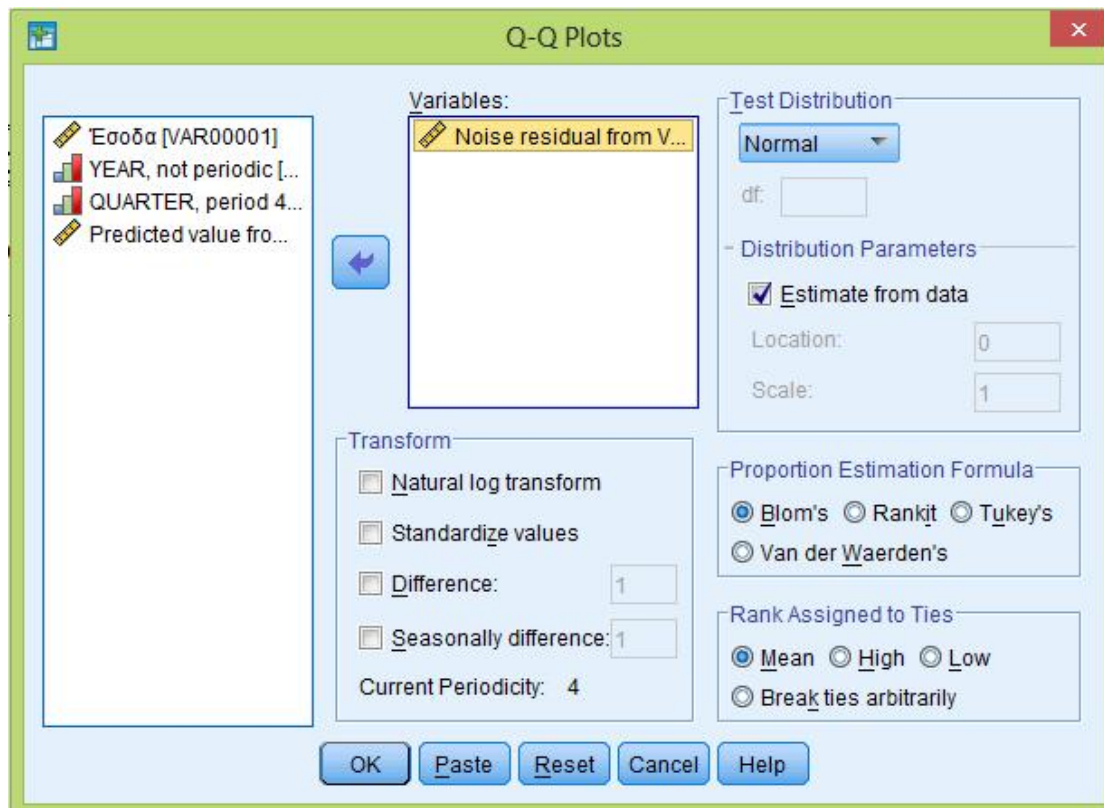


Εικόνα 16 : Διερεύνηση της κανονικότητας του μοντέλου ARIMA (1,0,0)

Αφού δημιουργήθηκαν οι νέες μεταβλητές πραγματοποιείται η διαδικασία Analyze → Descriptive Statistics → Q-Q plots (εικόνα 15) όπου εισάγεται στο πεδίο Variables η μεταβλητή Noise Residuals (εικόνα 16).

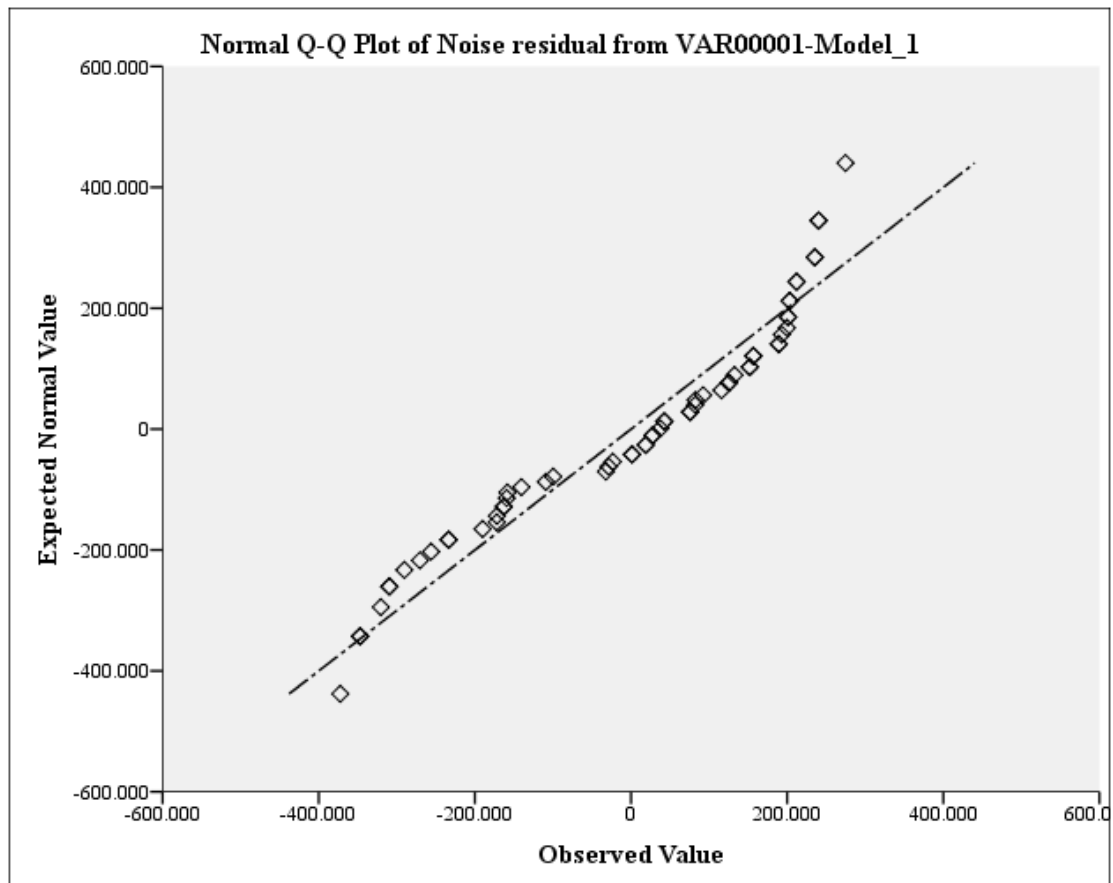


Εικόνα 17 : Διερεύνηση της κανονικότητας του μοντέλου ARIMA (1,0,0) #2



Εικόνα 18 : Διερεύνηση της κανονικότητας του μοντέλου ARIMA (1,0,0) #3

Από το διάγραμμα που δημιουργήθηκε καθίσταται εμφανές πως αφού η τιμές βρίσκονται κοντά στη ευθεία γραμμή υπάρχει κανονική κατανομή.



Διάγραμμα διερεύνησης της κανονικότητας

Λαμβάνοντας υπόψη ότι τα κριτήρια επαληθεύονται πλέον μπορεί να πραγματοποιηθεί πρόβλεψη για κάποια από τις επόμενες χρονικές περιόδους.

Αυτό πραγματοποιείται μέσω της διαδικασίας Analyze → Forecasting → Create Models και από το πεδίο Options επιλέγεται το πεδίο First case after the end of estimated period through a specific date. Εικόνα

Time Series Modeler

Variables Statistics Plots Output Filter Save Options

Forecast Period

First case after end of estimation period through last case in active dataset
 First case after end of estimation period through a specified date

Date:

Year	Quarter
2017	1

User-Missing Values

Treat as invalid
 Treat as valid

Confidence Interval Width (%): 95

Prefix for Model Identifiers in Output: Model

Maximum Number of Lags Shown in ACF and PACF Output: 24

OK Paste Reset Cancel Help

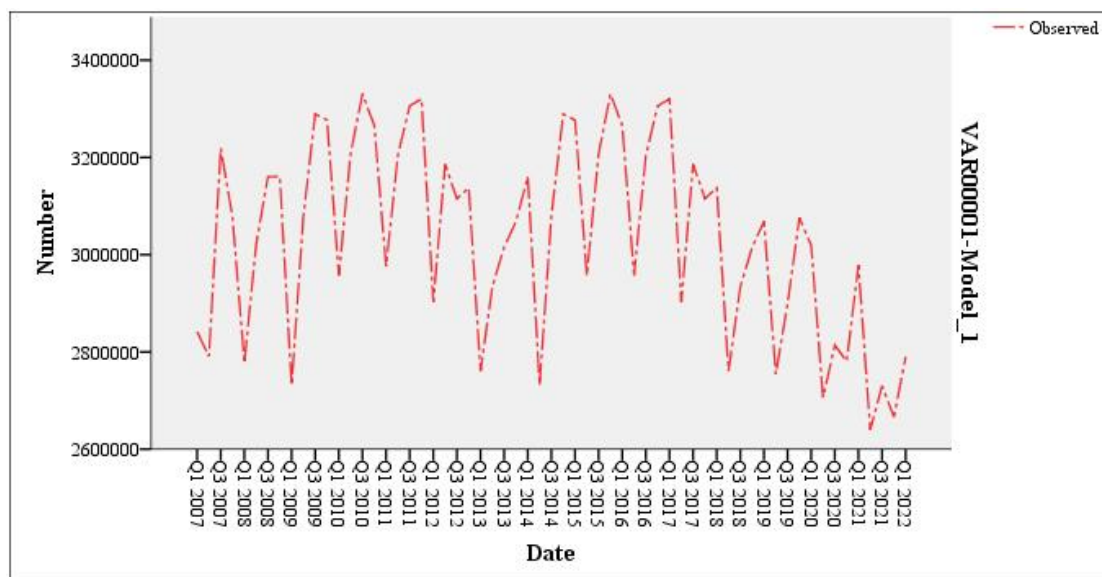
Πρόβλεψη για τα έσοδα του ΚΤΕΛ Ηλείας.

Αφού πραγματοποιηθεί η προαναφερθείσα διαδικασία στο αρχείο των δεδομένων πλέον υπάρχουν τιμές οι οποίες είναι το αποτέλεσμα της πρόβλεψης και για τα επόμενα έτη.

41	3320456	2017	1	Q1 2017
42	2902634	2017	2	Q2 2017
43	3185954	2017	3	Q3 2017
44	3115145	2017	4	Q4 2017
45	3137669	2018	1	Q1 2018
46	2760481	2018	2	Q2 2018
47	2934616	2018	3	Q3 2018
48	3015369	2018	4	Q4 2018
49	3068402	2019	1	Q1 2019
50	2754552	2019	2	Q2 2019
51	2899037	2019	3	Q3 2019
52	3075882	2019	4	Q4 2019
53	3019026	2020	1	Q1 2020
54	2706555	2020	2	Q2 2020
55	2814064	2020	3	Q3 2020
56	2780844	2020	4	Q4 2020
57	2978295	2021	1	Q1 2021
58	2639864	2021	2	Q2 2021
59	2730422	2021	3	Q3 2021
60	2666406	2021	4	Q4 2021
61	2790134	2022	1	Q1 2022

Προβλεπόμενα έσοδα για τα επόμενα έτη των ΚΤΕΛ Ηλείας.

Διαγραμματικά είναι εμφανές πως προβλέπεται μείωση των εσόδων για τα επόμενα έτη.



Διάγραμμα των εσόδων ως προς τον χρόνο

ΜΟΝΤΕΛΟ ARIMA (0,0,1)

Αφαιρώντας το p πλέον στο μοντέλο το α ισούται με 3034440,212 και το $\theta_1 = -0,323$. Εδώ είναι επίσης εμφανές πως το σ^2 της σταθεράς είναι ίσο με 0 άρα είναι στατιστικά σημαντικό ενώ και το σ^2 του MA είναι στατιστικά σημαντικό αφού είναι μικρότερο του 0,05. Λόγω της στατιστικής σημαντικότητας όλων των παραμέτρων η επιλογή και αυτού του μοντέλου για πρόβλεψη είναι σωστή.

ARIMA Model Parameters

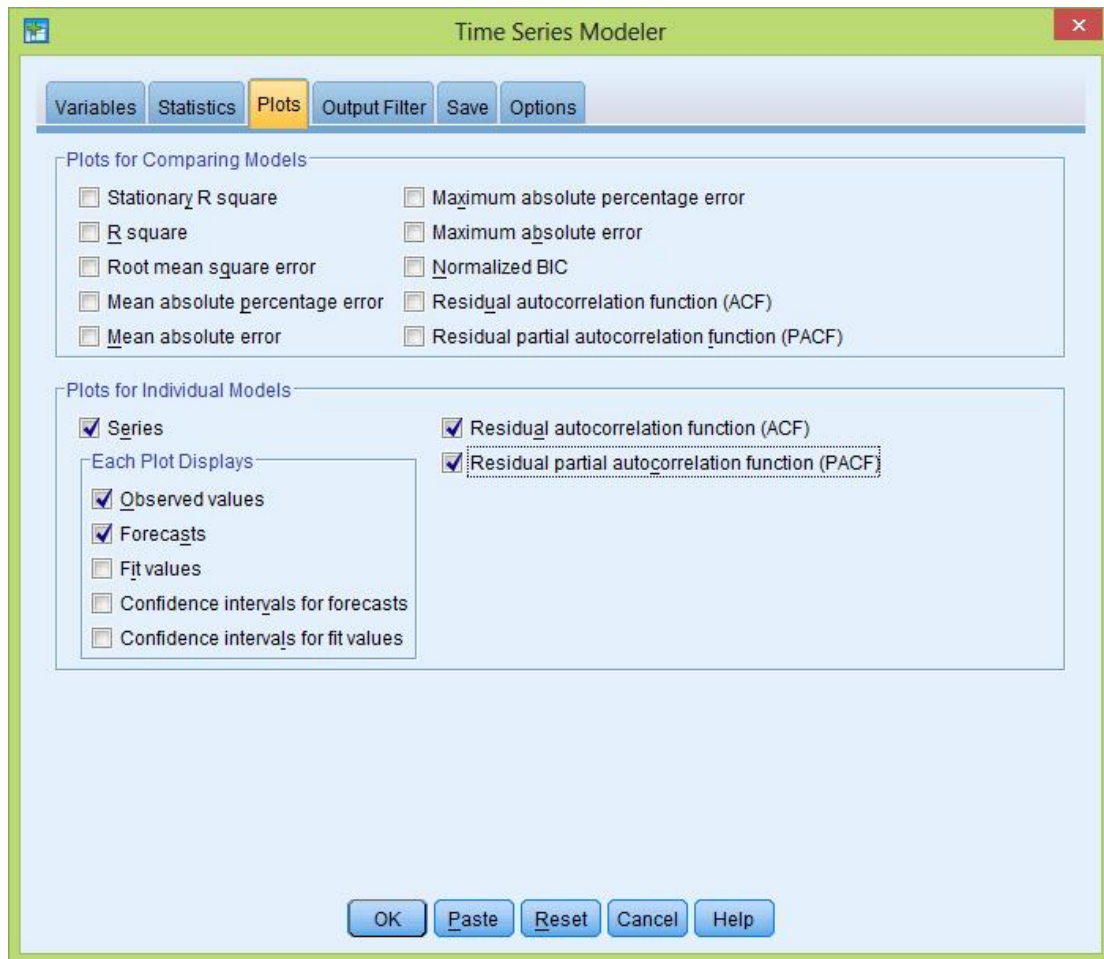
			Estimate	SE	t	Sig.
Έσοδα-	Έσοδα No	Constant	3034440,212	32371,773	93,737	,000
Model_1	Transformation MA	Lag 1	-,323	,124	-2,614	,011

Πίνακας των παραμέτρων του μοντέλου Arima (0,0,1

Η ευθεία που δημιουργείται εν τέλει είναι η $y_t = 3034440,212 + 0,323y_{t-1} + e_t$

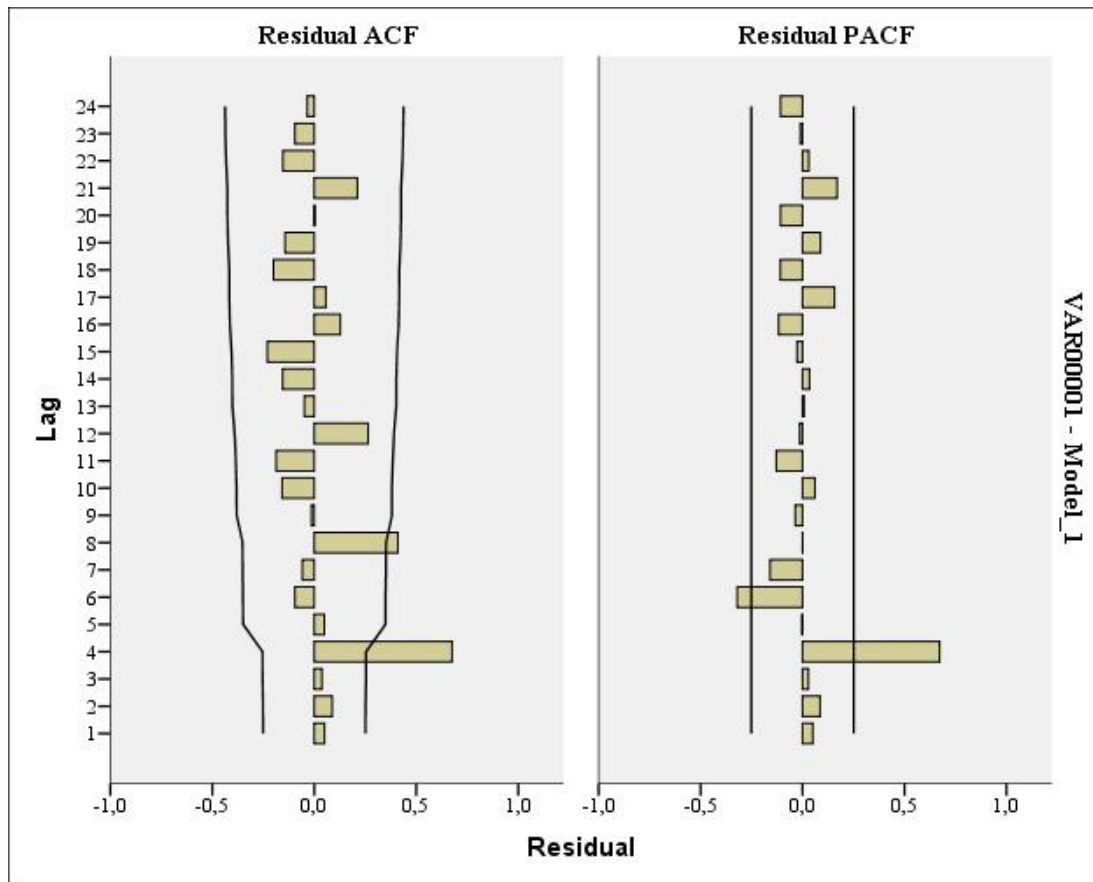
Αφού επιλέχθηκε το μοντέλο καλό είναι να πραγματοποιηθεί διαγνωστικός έλεγχος για το εάν το μοντέλο εξηγεί επαρκώς τα δεδομένα. Αυτό γίνεται με τα κριτήρια της στασιμότητας, της τυχαίας κατανομής και της κανονικής κατανομής. Για τα δύο πρώτα αρκεί να αποδειχθεί ότι τα e_t είναι λευκός θόρυβος. Εάν στα γραφήματα ACF & PACF υπάρχουν μηδενικές τιμές αυτό ισχύει.

Μέσω της διαδικασίας Analyze → Forecasting → Create models (εικόνα 9) και έπειτα επιλέγοντας το πεδίο Plots και τα πεδία ACF και PACF όπως στην εικόνα 13, εμφανίζονται στο αρχείο του output τα γραφήματα που θα αναλυθούν.



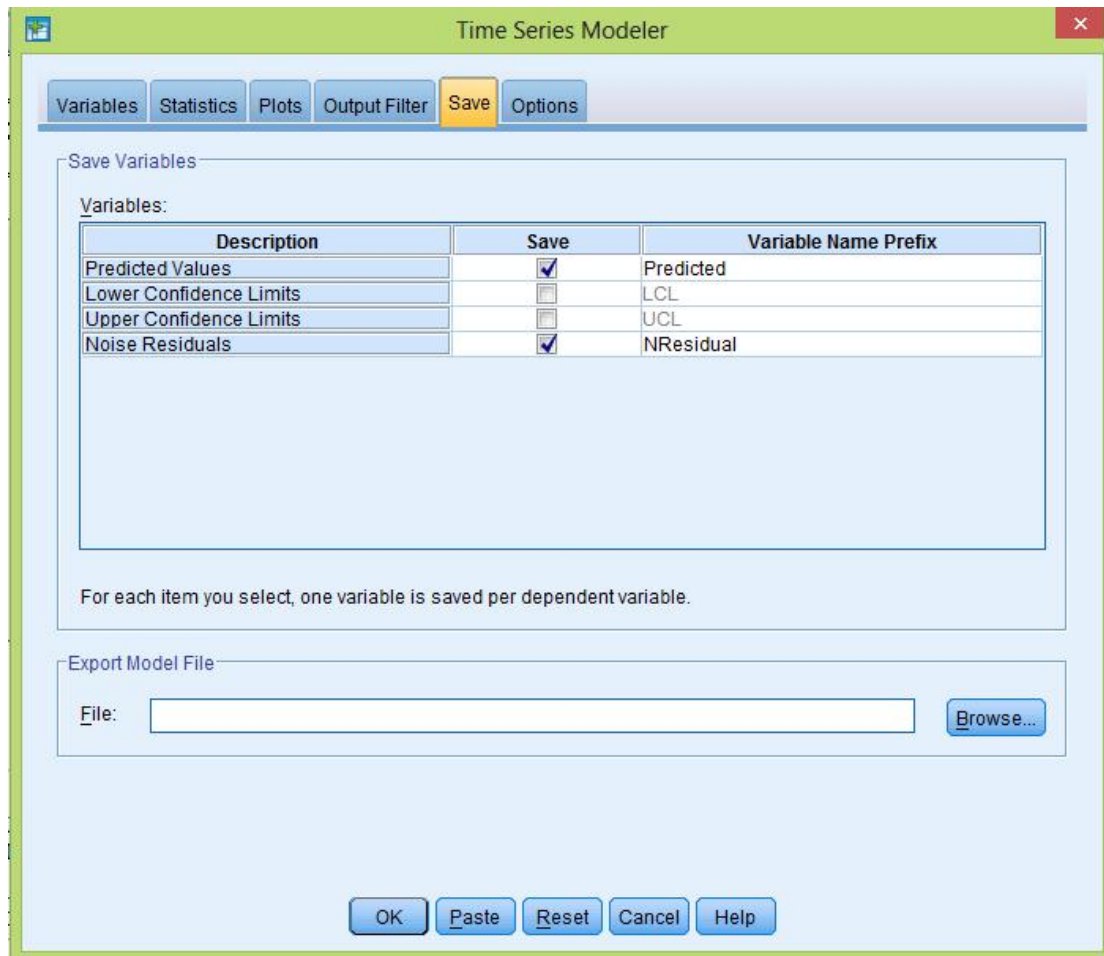
Εικόνα 1 : Διαγνωστικός έλεγχος των σφαλμάτων

Εδώ είναι εμφανές πως οι τιμές βρίσκονται στο μηδέν συνεπώς υπάρχει λευκός θόρυβος. Η ύπαρξη μιας τιμής εκτός είναι πιθανό αποτέλεσμα της εποχικότητας.



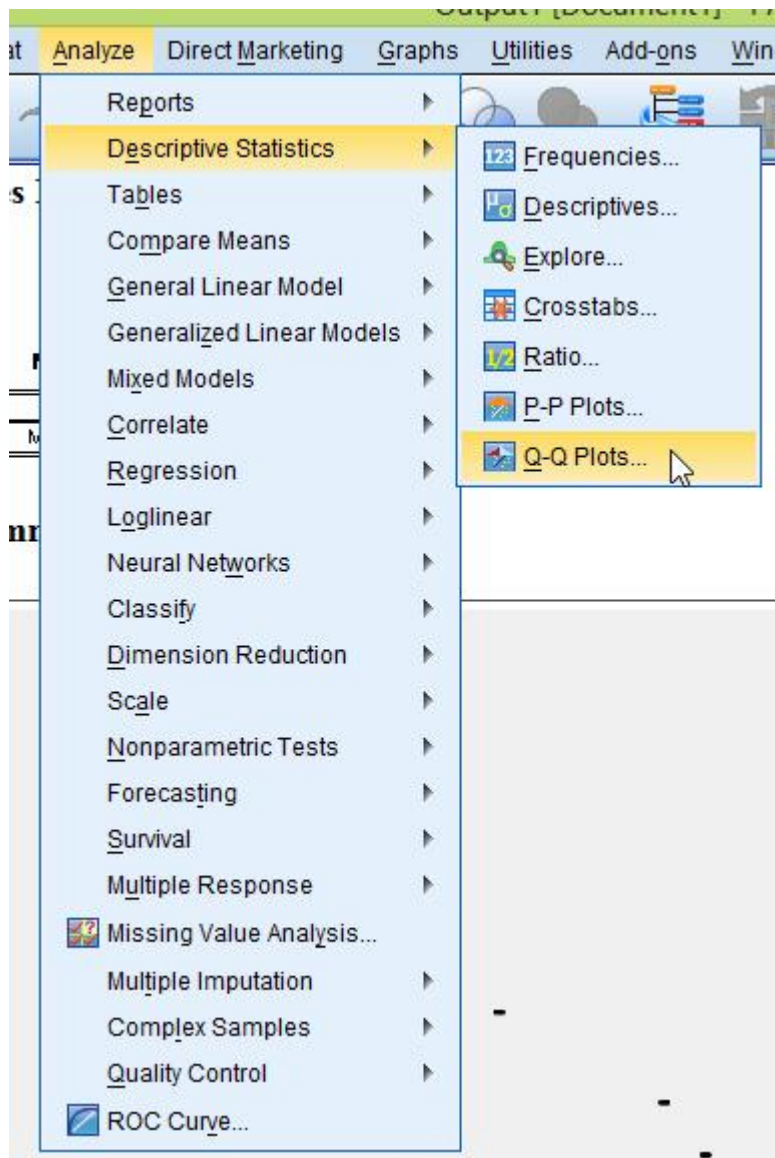
Γράφημα λευκού θορύβου του μοντέλου ARIMA (0,0,1)

Στη συνέχεια ακολουθεί η διερεύνηση της κανονικότητας του δείγματος, αυτό γίνεται από τη διαδικασία Analyze à Forecasting à Create models (εικόνα 9) και έπειτα επιλέγονται τα δύο πεδία Predicted values και Noise Residuals από το πεδίο Save (εικόνα 14). Μέσω αυτής της διαδικασίας δημιουργούνται στο αρχείο των δεδομένων νέες μεταβλητές, οι οποίες θα χρησιμοποιηθούν στη συνέχεια

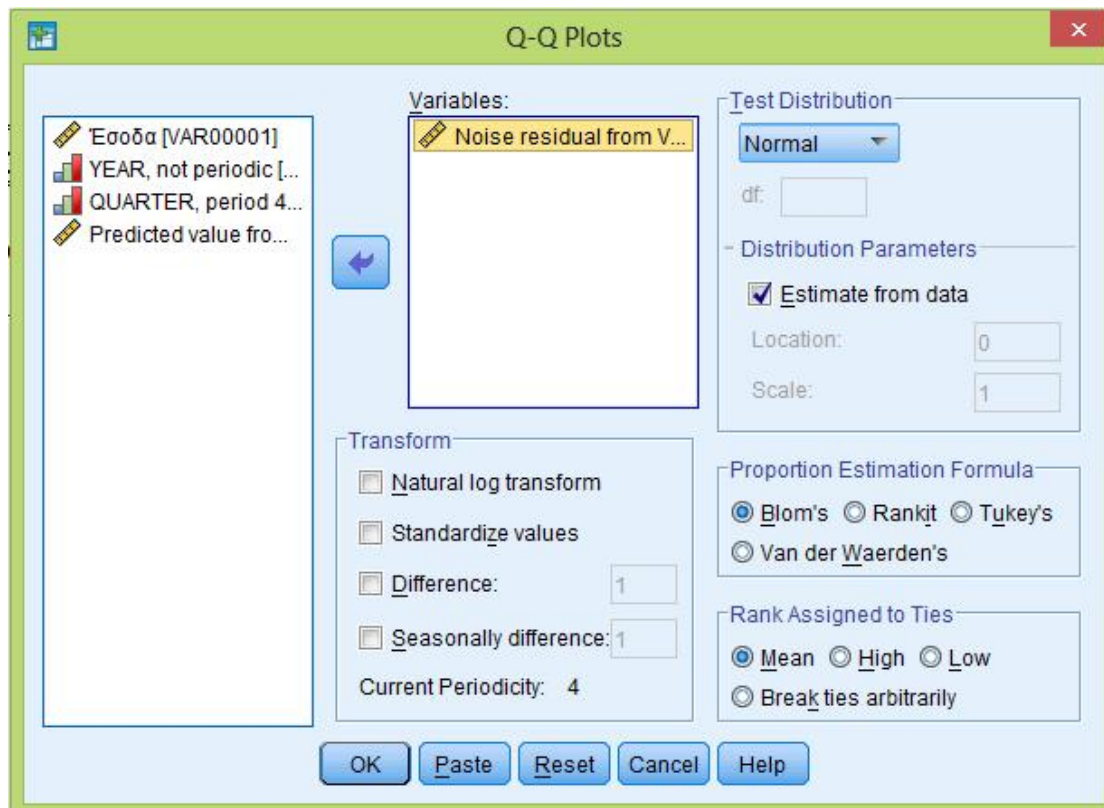


Εικόνα 2 : Διερεύνηση της κανονικότητας του μοντέλου ARIMA (1,0,0)

Αφότου δημιουργήθηκαν οι νέες μεταβλητές πραγματοποιείται η διαδικασία Analyze à Descriptive Statistics à Q-Q plots (εικόνα 15) όπου εισάγεται στο πεδίο Variables η μεταβλητή Noise Residuals (εικόνα 16).

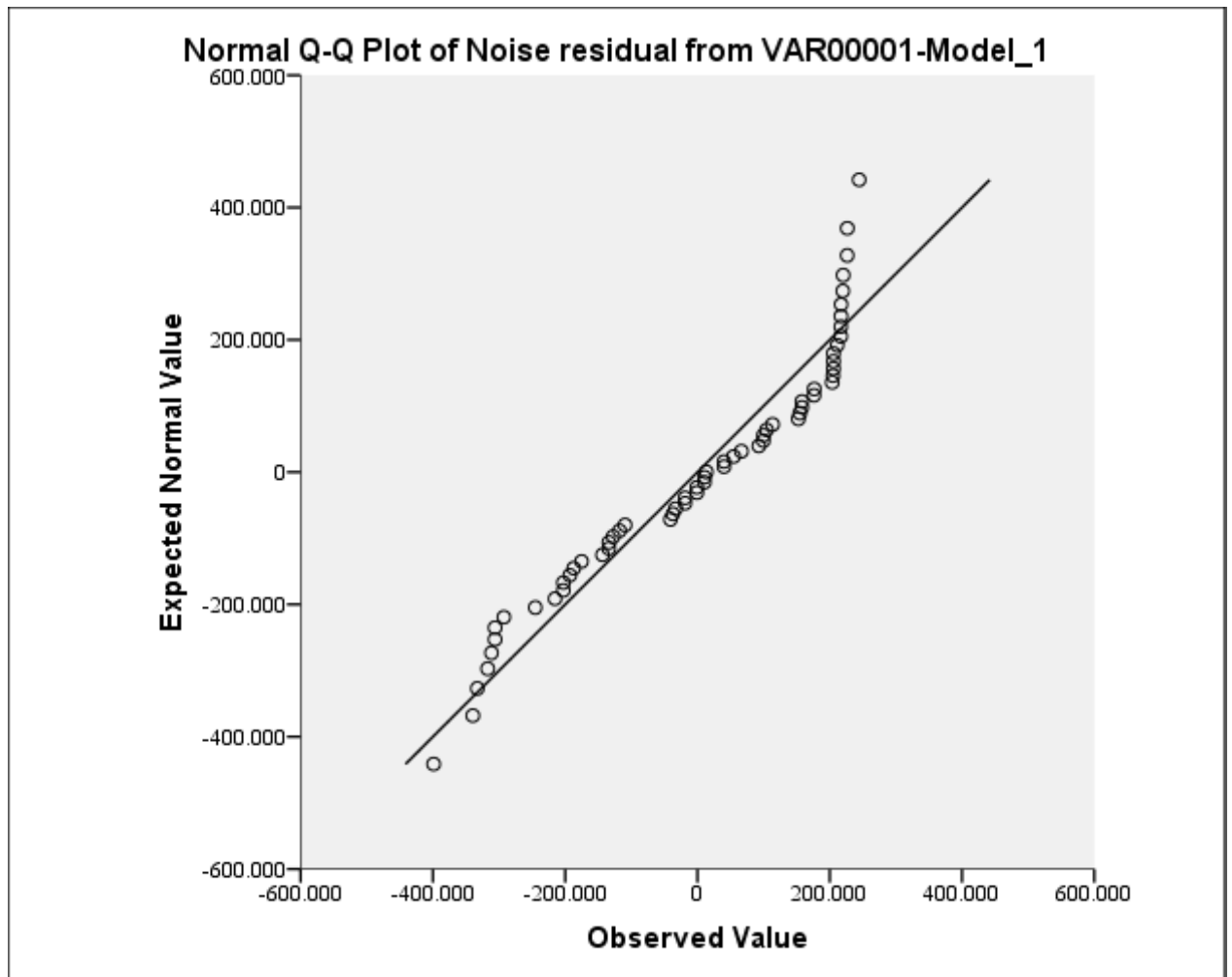


Εικόνα 3 : Διερεύνηση της κανονικότητας του μοντέλου
ARIMA (0,0,1) #2



Εικόνα 4 : Διερεύνηση της κανονικότητας του μοντέλου ARIMA (0,0,1) #3

Από το διάγραμμα που δημιουργήθηκε καθίσταται εμφανές πως αφού η τιμές βρίσκονται κοντά στη ευθεία γραμμή υπάρχει κανονική κατανομή.



Διάγραμμα διερεύνησης της κανονικότητας

Λαμβάνοντας υπόψη ότι τα κριτήρια επαληθεύονται πλέον μπορεί να πραγματοποιηθεί πρόβλεψη για κάποια από τις επόμενες χρονικές περιόδους.

Αυτό πραγματοποιείται μέσω της διαδικασίας Analyze → Forecasting → Create Models και από το πεδίο Options επιλέγεται το πεδίο First case after the end of estimated period through a specific date.

Time Series Modeler

Variables Statistics Plots Output Filter Save Options

Forecast Period

First case after end of estimation period through last case in active dataset
 First case after end of estimation period through a specified date

Date:

Year	Quarter
2017	1

User-Missing Values

Treat as invalid
 Treat as valid

Confidence Interval Width (%): 95

Prefix for Model Identifiers in Output: Model

Maximum Number of Lags Shown in ACF and PACF Output: 24

OK Paste Reset Cancel Help

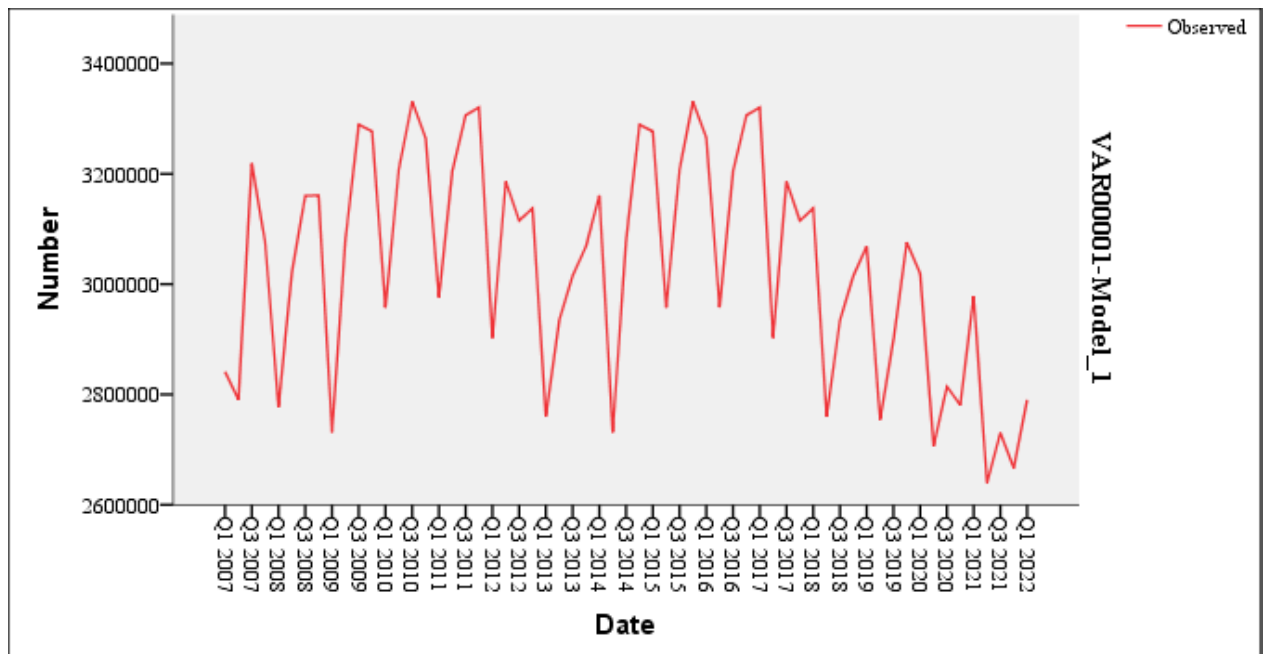
Πρόβλεψη για τα έσοδα του ΚΤΕΛ Ηλείας.

Αφότου πραγματοποιηθεί η προαναφερθείσα διαδικασία στο αρχείο των δεδομένων πλέον υπάρχουν τιμές οι οποίες είναι το αποτέλεσμα της πρόβλεψης και για τα επόμενα έτη.

45	3137669	2018	1	Q1 2018
46	2760481	2018	2	Q2 2018
47	2934616	2018	3	Q3 2018
48	3015369	2018	4	Q4 2018
49	3068402	2019	1	Q1 2019
50	2754552	2019	2	Q2 2019
51	2899037	2019	3	Q3 2019
52	3075882	2019	4	Q4 2019
53	3019026	2020	1	Q1 2020
54	2706555	2020	2	Q2 2020
55	2814064	2020	3	Q3 2020
56	2780844	2020	4	Q4 2020
57	2978295	2021	1	Q1 2021
58	2639864	2021	2	Q2 2021
59	2730422	2021	3	Q3 2021
60	2666406	2021	4	Q4 2021
61	2790134	2022	1	Q1 2022

Προβλεπόμενα έσοδα για τα επόμενα έτη των ΚΤΕΛ Ηλείας.

Διαγραμματικά είναι εμφανές πως προβλέπεται μείωση των εσόδων για τα επόμενα έτη.



Διάγραμμα των εσόδων ως προς τον χρόνο

Παρατηρώντας τα δύο μοντέλα καθίσταται εμφανές πως μπορούν και τα δύο να χρησιμοποιηθούν και τα δύο για πρόβλεψη βάσει των ιδιοτήτων που τα διέπουν παρόλα αυτά το πιο σωστό μοντέλο για πρόβλεψη θεωρείται το ARIMA (1,0,0) βάσει του ότι λόγω της μη σημαντικότητας πρέπει να αφαιρεθεί το q .

ΣΥΜΠΕΡΑΣΜΑΤΑ

Χρονοσειρά (time serie) είναι ένα σύνολο παρατηρήσεων $1, 2, \dots, t, y, y, y$ όπου ο δείκτης T παριστάνει ισαπέχοντα χρονικά σημεία ή διαστήματα. Τα χρονικά διαστήματα μπορεί να είναι (έτος, μήνας, ημέρα, εβδομάδα, ώρα, κ.α).

Οι παρατηρήσεις $1, 2, \dots, t, y, y, y$ είναι συγκεκριμένες τιμές των τυχαίων μεταβλητών $1, 2, \dots, Y, Y, Y, T$ και είναι μέρος μόνο μιας άπειρης ακολουθίας τυχαίων μεταβλητών και αναφέρονται στην έννοια του δείγματος ενώ οι τυχαίες μεταβλητές στην έννοια του πληθυσμού. Απαραίτητη συνθήκη για τη μελέτη χρονολογικών σειρών είναι η ύπαρξη δεδομένων (data). Δεδομένα χρονολογικών σειρών συναντούμε σε πολλές επιστήμες (οικονομικές, κοινωνικές, φυσική, ιατρική κ.α.)

Μερικά παραδείγματα είναι οι μηνιαίες πωλήσεις μιας επιχείρησης, οι τιμές ενός αγαθού ανά τρίμηνο, οι δείκτες των μετοχών στο ΧΑΑ. Βασικό χαρακτηριστικό κάθε χρονολογικής σειράς είναι η εξάρτηση μεταξύ των διαδοχικών τιμών της. Αντικείμενο μελέτης του κλάδου των χρονολογικών σειρών είναι η φύση της αλληλεξάρτησης που υπάρχει μεταξύ των παρατηρήσεων και χωρίζεται σε δύο μέρη.

Το πρώτο περιλαμβάνει την ανάλυση των ιδιοτήτων της σειράς έτσι ώστε να προσδιοριστούν τα χαρακτηριστικά που διέπουν τη συμπεριφορά της. Αυτό γίνεται με τη χρονική όσο και τη φασματική ανάλυση (spectral analysis). Το δεύτερο μέρος με το οποίο θα ασχοληθούμε και περισσότερο αφορά τα υποδείγματα χρονολογικών σειρών (time series models). Στα συγκεκριμένα υποδείγματα απώτερος σκοπός είναι η δημιουργία προβλέψεων και συνεπακόλουθα η μείωση της αβεβαιότητας και η καλύτερη εκτίμηση διαφόρων γεγονότων.

Αντικειμενικός σκοπός της μελέτης χρονολογικών σειρών είναι η χρησιμοποίησή τους στη διενέργεια προβλέψεων. Η πρόβλεψη των μελλοντικών τιμών μιας μεταβλητής μπορεί να γίνει με διάφορες μεθόδους. Οι μέθοδοι αυτοί διαφέρουν ως προς την πολυπλοκότητα, την ταχύτητα, το κόστος υπολογισμού τους καθώς επίσης και από τη διαθεσιμότητα των απαραίτητων δεδομένων. Σε γενικές γραμμές, οι μέθοδοι πρόβλεψης (forecast methods) διαχωρίζονται σε υποκειμενικές ή ποιοτικές (subjective or qualitative) και σε αντικειμενικές ή ποσοτικές (objective or quantitative). Οι υποκειμενικές μέθοδοι πρόβλεψης, οι οποίες δεν θα μας απασχολήσουν ιδιαίτερα, γίνονται κυρίως από έμπειρους επιστημονικούς αναλυτές οι οποίοι χρησιμοποιούν τις γνώσεις τους σε συνδυασμό με την κρίση τους, για την

ασφαλή διεξαγωγή συμπερασμάτων. Πρακτικά ,δεν δίνουν βάρος σε μαθηματικές ή στατιστικές μεθόδους.

Μια αρκετά διαδεδομένη ποιοτική μέθοδος είναι η μέθοδος των Δελφών, η οποία χρησιμοποιείται αρκετά στις επιχειρήσεις και βασίζεται στη συγκέντρωση πληροφοριών από ειδικές ομάδες εμπειρογνομόνων. Αντιθέτως ,οι αντικειμενικές μέθοδοι πρόβλεψης βασίζονται σε κάποιο μαθηματικό ή στατιστικό υπόδειγμα και σε ποσοτικά δεδομένα (model based forecasts). Τα υποδείγματα αυτά χωρίζονται σε δύο μεγάλες κατηγορίες, τα αιτιατά (causal) και τα μη αιτιατά (non-causal).

Στα υποδείγματα χρονολογικών σειρών η πρόβλεψη στηρίζεται αποκλειστικά και μόνο στις προηγούμενες τιμές της ίδιας χρονολογικής σειράς που θέλουμε να προβλέψουμε. Δηλαδή, προβλέπουμε τη μελλοντική συμπεριφορά μιας χρονοσειράς όχι σε συνάρτηση άλλων σειρών αλλά εξετάζοντας την προηγούμενη συμπεριφορά της δηλαδή το “ιστορικό” της. Τα υποδείγματα χρονολογικών σειρών χωρίζονται σε καθοριστικά υποδείγματα (deterministic models) τα οποία βασίζονται σε απλές μαθηματικές μορφές (υποδείγματα κινητών μέσων όρων, εκθετικής εξομάλυνσης και τάσης) αλλά και σε στοχαστικά υποδείγματα (stochastic models) όπως το υπόδειγμα μορφής Box-Jenkins. Με κάποια από τα παραπάνω υποδείγματα θα ασχοληθούμε διεξοδικά στη συνέχεια. Τα πλεονεκτήματα των χρονολογικών υποδειγμάτων σε σχέση με τα οικονομετρικά είναι πως είναι έχουν χαμηλότερο κόστος διενέργειας προβλέψεων και είναι λιγότερο πολύπλοκα. Αντιθέτως, βασικό τους μειονέκτημα είναι πως δεν στηρίζονται σε κάποια θεωρία που να εξηγεί πώς διαμορφώνονται οι τιμές της χρονολογικής σειράς.

ΒΙΒΛΙΟΓΡΑΦΙΑ

Mitchell, D.B., (1982) Should the Consumer Price Index Determine Wages?, California Management Review Vol. 25, Fall 1982, pp. 19

Song H., Triantis, J. (2003) Building Alliance With Clients: Key To Success In Foracasting, The journal of Business Forecasting, Winter 2002-03, pp 2-6

Weatherford, L., Kimes, S. (2003) A Comparison of Forecasting methods for hotel revenue management, International Journal of Forecasting, 19, pp 413-414

Βάμβουκας, Γ. (2007). Σύγχρονη Οικονομετρία: Ανάλυση και Εφαρμογές, Οικονομικό Πανεπιστήμιο Αθηνών, Αθήνα

Βενέτης, Ι. (2009). Εισαγωγικές διαλέξεις στην Οικονομετρία, Γκιούρδας Εκδοτική, Αθήνα

Ζάρκος, Σ., Κωνσταντόπουλος, Ι., (1999) Έρευνες Οικονομικής Συγκυρίας και Δείκτες Βραχυχρόνιας Πρόβλεψης, Δελτίο Οικονομικό και Στατιστικό της εθνικής Τράπεζας της Ελλάδος, Νοέμβριος 1999, σελ. 17-19

Σαριαννίδης, Ν., Κοντέος, Γ., Λαζαρίδης, Θ. (2013). Στατιστική και Οικονομετρία, Κοζάνη.