**ΑΡΙΣΤΟΤΕΛΕΙΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΘΕΣΣΑΛΟΝΙΚΗΣ**

ΣΧΟΛΗ ΟΙΚΟΝΟΜΙΚΩΝ ΚΑΙ ΠΟΛΙΤΙΚΩΝ ΕΠΙΣΤΗΜΩΝ

ΤΜΗΜΑ ΟΙΚΟΝΟΜΙΚΩΝ ΕΠΙΣΤΗΜΩΝ

****

**ΕΡΓΑΣΙΑ ΜΕ ΘΕΜΑ ΤΗΝ ΑΝΑΛΥΣΗ ΧΡΟΝΟΣΕΙΡΩΝ**

ΜΑΘΗΜΑ:

**«Ανάλυση Χρονοσειρών »**

ΟΝΟΜΑΤΕΠΩΝΥΜΟ:

**Κωνσταντίνος Π. Γιαννακόπουλος**

ΑΕΜ:

**16051**

(Τμήμα Μαθηματικών)

Table of Contents

**1. ΕΙΣΑΓΩΓΗ3**

**2. Δεδομένα και γραφικές παραστάσεις4**

2.1 Dataset4

2.2 Γραφικές παραστάσεις6

**3. ΕΜΠΕΙΡΙΚΗ ΑΝΑΛΥΣΗ**8

3.1 Έλεγχος στασιμότητας 8

3.2 Έλεγχος συνολοκλήρωσης 12

3.3 Έλεγχος αιτιοτήτων κατά Granger 14

3.4 Εκτίμηση υποδείγματος ARIMA(p,d,q)15

3.5 Έλεγχος Ljung-Box19

4.3 Έλεγχος Jarque-Bera ………………………………………………………….…..21

**ΣΥΜΠΕΡΑΣΜΑΤΑ22**

**ΒΙΒΛΙΟΓΡΑΦΙΑ…………………………………………………………………………….24**

**1. ΕΙΣΑΓΩΓΗ**

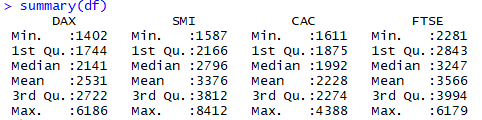
Η παρούσα εργασία γίνεται στα πλαίσια του μαθήματος της Ανάλυσης Χρονοσειρών. Στόχος της εργασίας είναι η πρακτική εξοικείωση με τις έννοιες και τις πρακτικές που αναφέρθηκαν κατά τη διάρκεια του εξαμήνου. Πιο συγκεκριμένα, λαμβάνοντας κάποια δεδομένα, τα οποία αποτελούν χρονικές σειρές, θα βρω το βέλτιστο υπόδειγμα που μοντελοποιεί τα δεδομένα, μέσω οπτικοποίησης αλλά και ελέγχων.

Στο 2ο κεφάλαιο θα αναλύσουμε το dataset και το γραμμικό υπόδειγμα σε ένα γενικό και θεωρητικό πλαίσιο. Στη συνέχεια, στο κεφάλαιο 3 θα αναλύσουμε τα δεδομένα, μέσω γραφικών παραστάσεων για να λάβουμε τις πρώτες ενδείξεις για τις μεταβλητές μας και τέλος θα βρεθεί και θα εκτιμηθεί το υπόδειγμα. Τέλος, θα καταλήξουμε σε συμπεράσματα σχετικά με την ανάλυση μας.

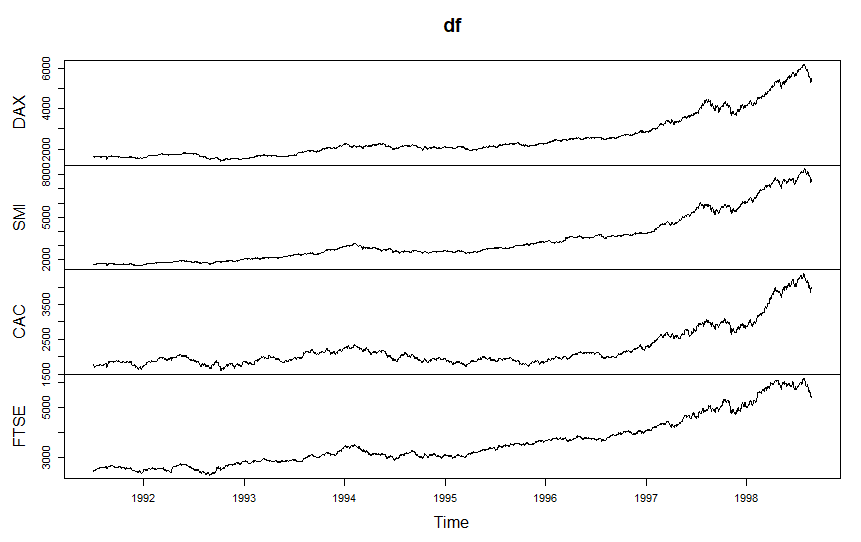
Η εργασία πραγματοποιήθηκε σε γλώσσα R (Rstudio).

**2.1 DATASET**

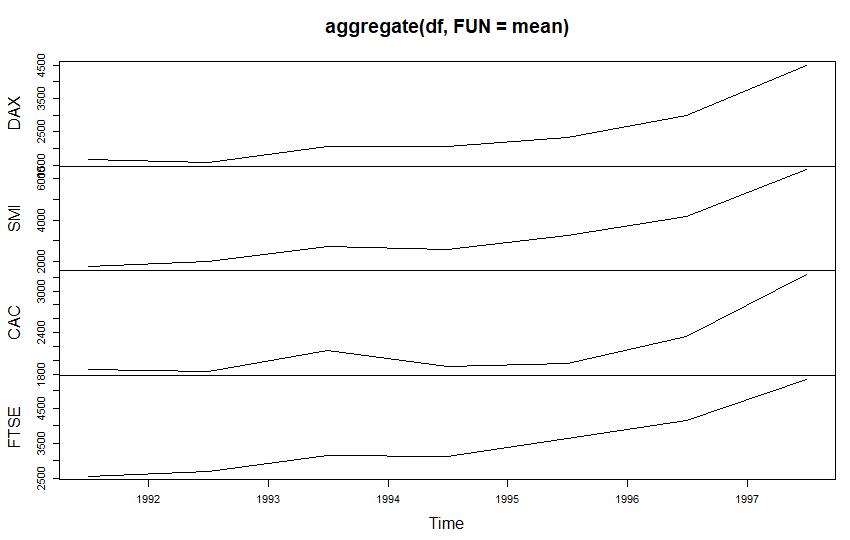
Τα δεδομένα που επιλέχθηκαν προέρχονται από τη βιβλιοθήκη της R. Συγκεκριμένα το dataset ονομάζεται EuStockMarkets και περιέχει τις ημερήσιες τιμές στις οποίες έκλεισαν 4 χρηματιστηριακοί δείκτες: DAX, FTSE, CAC και SMI από το 1991 έως το 1998.



Παρακάτω βλέπουμε τα διαγράμματα του dataset ως προς το χρόνο:



Εν συνεχεία, έχουμε ένα γράφημα των χρονοσειρών, παίρνοντας τη μέση τιμή κάθε έτους. Το αποτέλεσμα είναι πολύ πιο ομαλό από πριν και μας βοηθάει να δούμε τη “μεγαλύτερη εικόνα”.

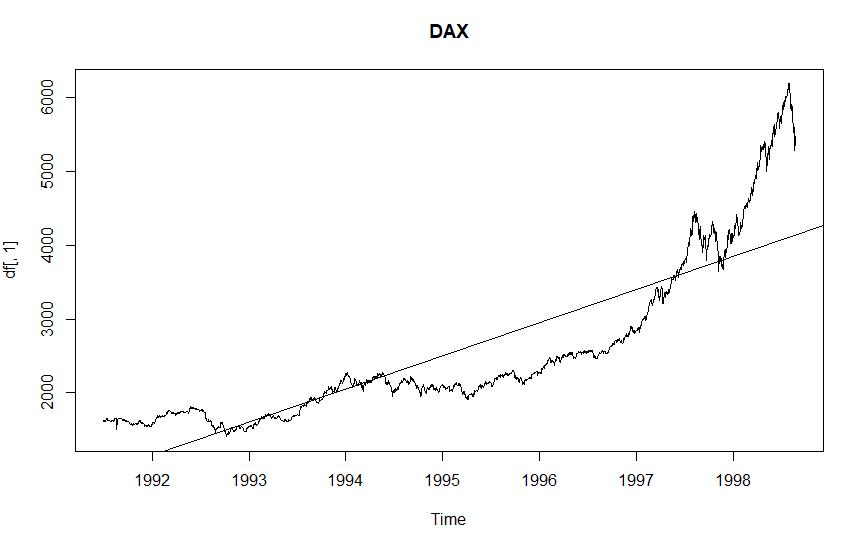


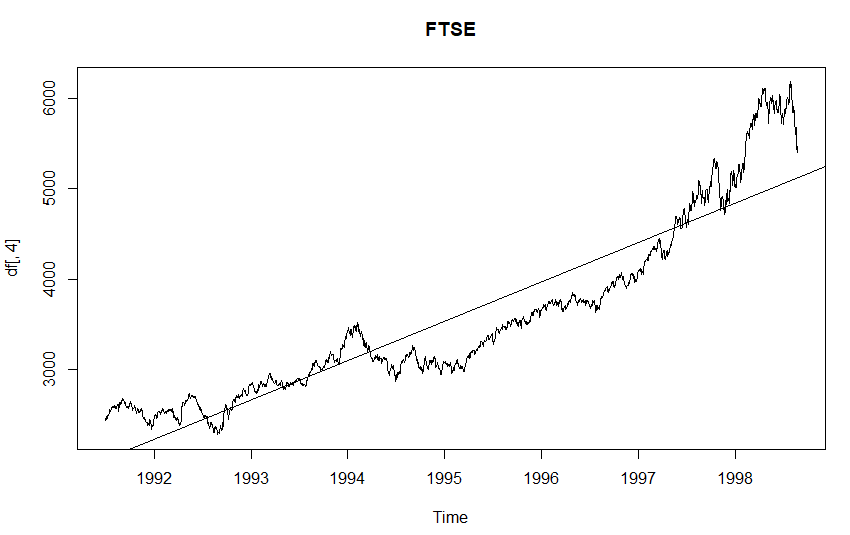
Σε μια ευρεία κλίμακα το μοτίβο φαίνεται να είναι παρόμοιο σε όλα τα γραφήματα. Αλλά ας τα εξετάσουμε λίγο πιο διεξοδικά.

**2.2 Γραφικές παραστάσεις**

Έγινε, λοιπόν, η επιλογή να ασχοληθώ με τους δείκτες DAX και FTSE. Τελικός στόχος είναι να βρεθεί κάποια σχέση που να τους μοντελοποιεί επαρκώς. Αυτή η σχέση/υπόδειγμα μπορεί να περιέχει lags των μεταβλητών, σφάλματα περασμένων χρονικών στιγμών και ενδεχομένως lags η μια της άλλης.

* Αρχικά, λοιπόν, έχουμε το γράφημα του δείκτη DAX με τη γραμμή παλινδρόμησης και στη συνέχεια του δείκτη FTSE.

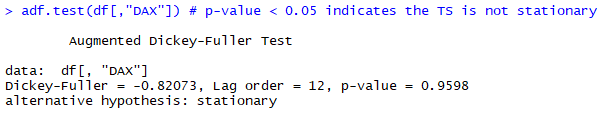




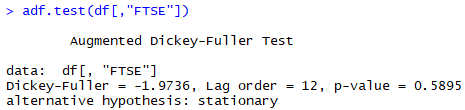
Με μια πιο λεπτομερή ματιά βλέπουμε ότι υπάρχουν μικρές διαφορές στα δύο γραφήματα.

**3.1 Έλεγχος στασιμότητας**

Μια ακόμη πληροφορία που έχουμε κοιτάζοντας τα γραφήματα είναι ότι οι χρονοσειρές δεν φαίνεται να είναι στάσιμες. Αυτό μας οδηγεί στη πραγματοποίηση ελέγχου ADF σε επίπεδο σημαντικότητας 0.05, ως προς τη στασιμότητα.



Για το δείκτη DAX η p-value είναι πολύ μεγάλη επομένως, θα πρέπει να δεχθούμε τη μηδενική υπόθεση, δηλαδή ότι δεν είναι στάσιμη.

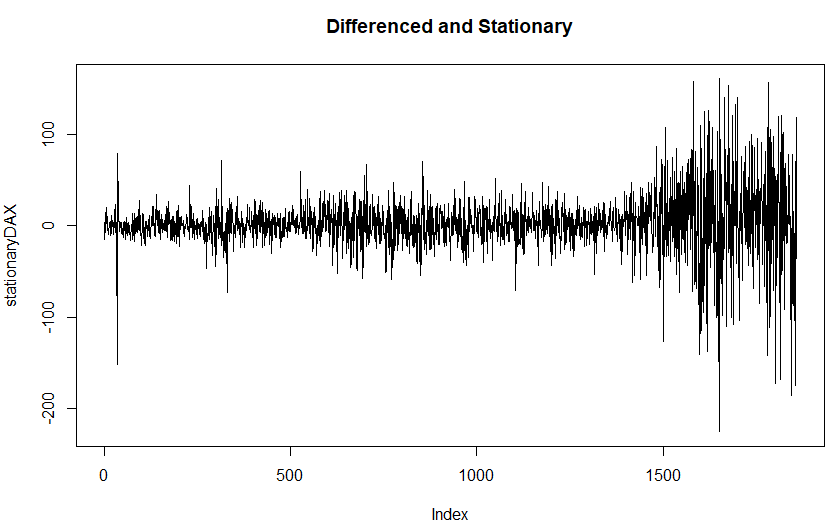


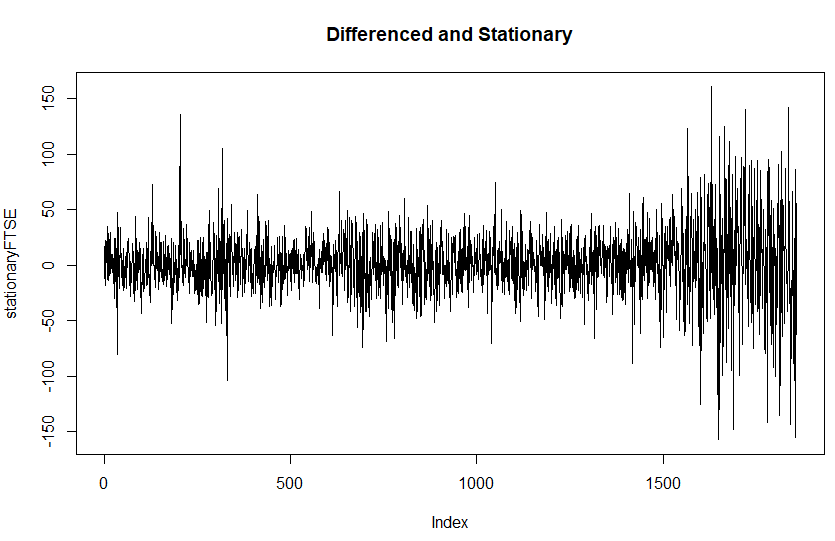
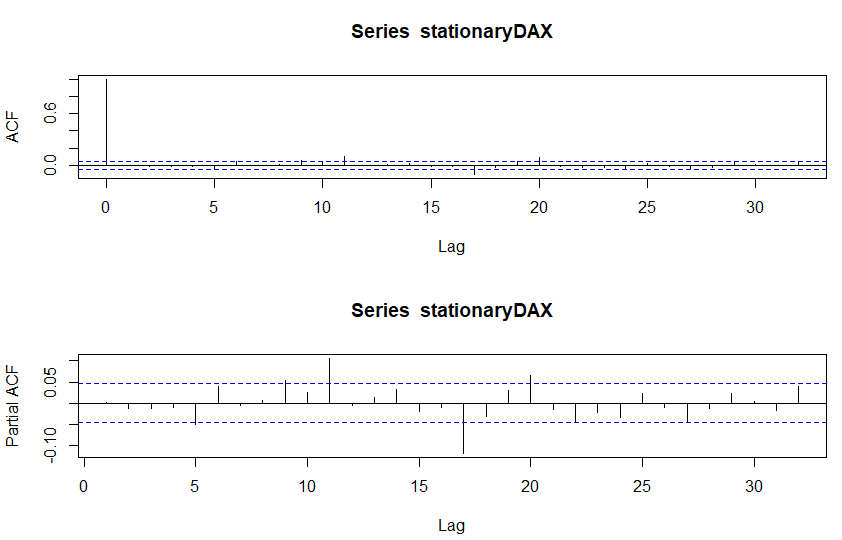
Αντίστοιχα και ο δείκτης FTSE είναι ολοκληρωμένη χρονική σειρά.

Στη συνέχεια, και πάλι με χρήση του ελέγχου ADF καταλήγουμε στο ότι και οι δύο χρονοσειρές είναι I(1), δηλαδή πρέπει να τις διαφορίσουμε μία φορά προκειμένου να γίνουν στάσιμες.

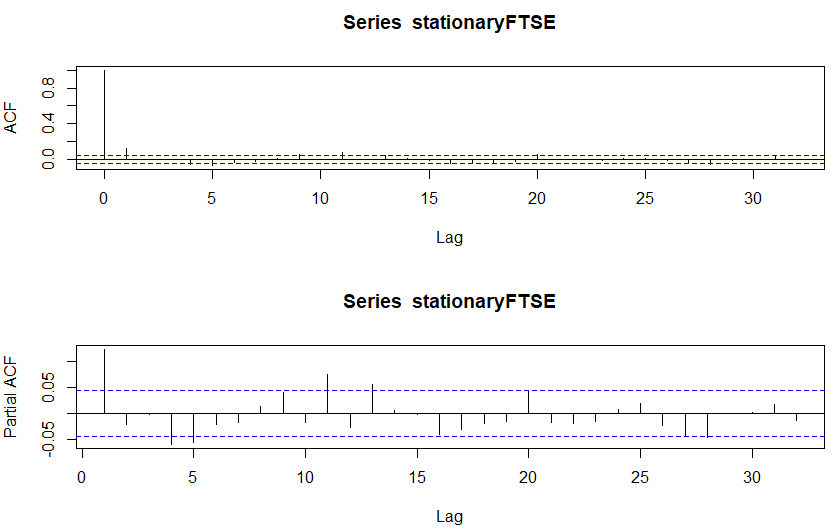
Η βοηθητική παλινδρόμηση περιέχει σταθερά και ένα γραμμικό όρο t (trend).

Αφού γίνει αυτό τα γραφήματα πλέον είναι τα παρακάτω:

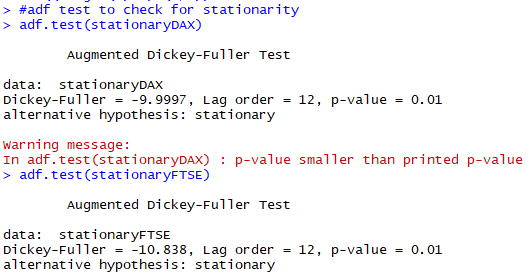




Παρακάτω, έχουμε τα γραφήματα αυτοσυσχέτισης και μερικής αυτοσυσχέτισης:



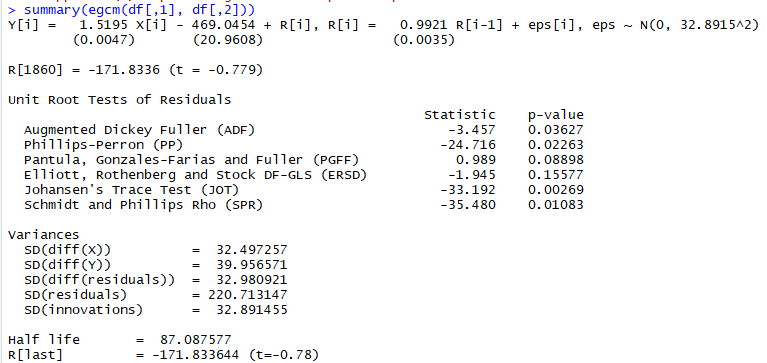
Παρατηρούμε ότι σε γενικές γραμμές δεν υπάρχει σημαντική αυτοσυχέτιση με τα lags κάθε χρονοσειράς, με ελάχιστες εξαιρέσεις.

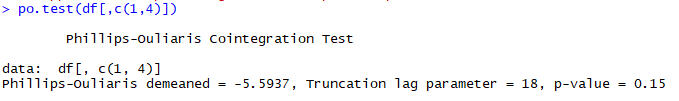
Επαναλαμβάνοντας τον έλεγχο στασιμότητας ADF παίρνουμε πολύ μικρές p-values, που συνηγορούν υπέρ της στασιμότητας.

**3.2 Έλεγχος συνολοκλήρωσης**

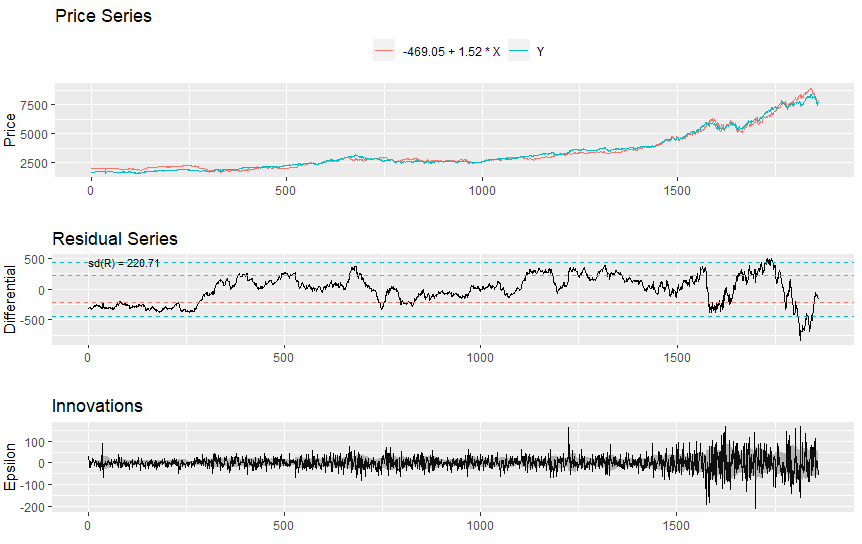
Στη συνέχεια, θα γίνει έλεγχος συνολοκλήρωσης για τις δύο χρονοσειρές. Γι’ αυτό θα χρησιμοποιηθεί η προσέγγιση των Engle και Granger.

Εδώ για να γίνει έλεγχος στασιμότητας των καταλοίπων χρησιμοποιούνται διάφορα test, μεταξύ των οποίων και το γνωστό μας ADF. H p-value είναι σημαντική και έτσι συμπεραίνουμε ότι δεν είναι στλασιμη, καθώς η μηδενική υπόθεση είναι ότι έχει μοναδιαία ρίζα.



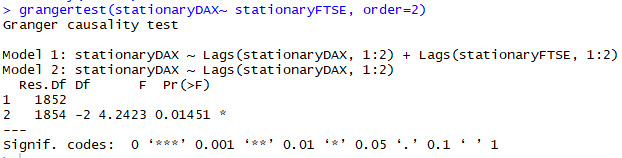
Για επαλήθευση χρησιμοποιείται και το Philips-Ouliaris test με μηδενική υπόθεση ότι δεν είναι συνολοκληρωμένες (εδώ χρησιμοποιείται η μέθοδος Engle-Granger δύο βημάτων). Η p-value=0.15 υποδεικνύει ότι δεν είναι συνολοκληρωμένες.

Και ένας οπτικός έλεγχος της συνολοκλήρωσης (residual seris).

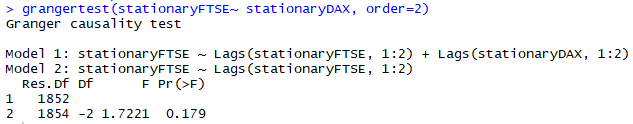


**3.3 Έλεγχος αιτιοτήτων κατά Granger**

Στη συνέχεια θα κάνουμε έλεγχο αιτιότητας κατά Granger. Έχουμε λοιπόν τις χρονοσειρές στις πρώτες διαφορές τους ώστε να είναι στάσιμες και όπως είδαμε δεν είναι συνολοκληρωμένες.



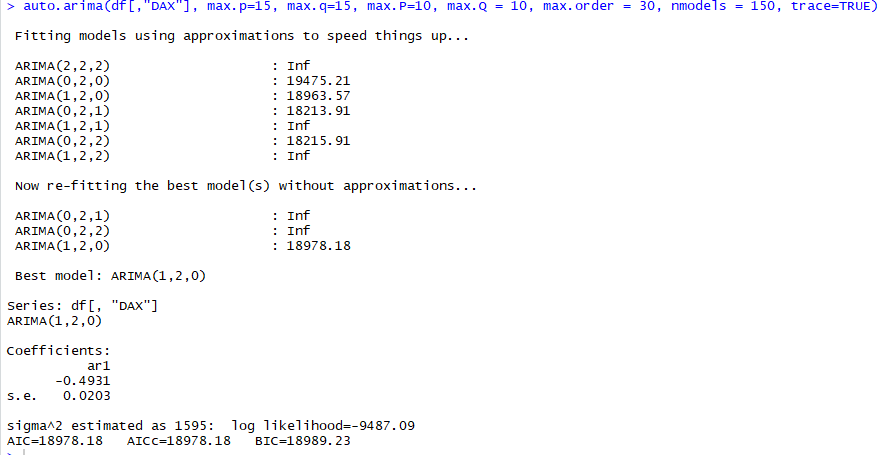
Χρησιμοποιούμε έλεγχο τάξης 2, δηλαδή υπόδειγμα με δύο lags για κάθε μεταβλητή. Η μηδενική υπόθεση είναι ότι οι συντελεστές είναι 0. Η p-value είναι μικρότερη του 0.05 επομένως απορρίπτουμε τη μηδενική. Άρα έχουμε μονόδρομη αιτιότητα FTSE -> DAX.

Ωστόσο, δεν ισχύει το αντίστροφο.

**3.4 Εκτίμηση υποδείγματος ARIMA(p,d,q)**

Θέλουμε να βρούμε το καλύτερο μοντέλο για τις χρονοσειρές μας. Αυτό μπορεί να περιέχει αυτοπαλίνδρομους όρους έως τάξης p, προηγούμενα σφάλματα έως τάξης q και φυσικά να είναι d ο αριθμός των διαφορών.

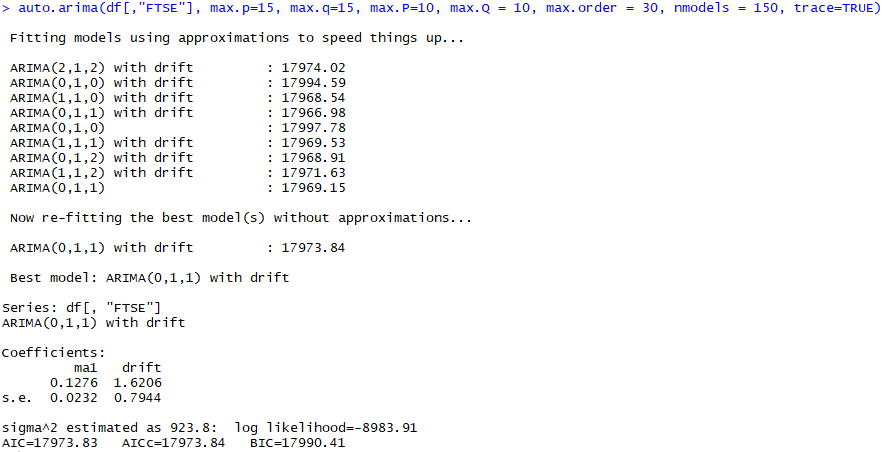
Η επιλογή του βέλτιστου μοντέλου γίνεται με βάση τα κριτήρια AIC και BIC.

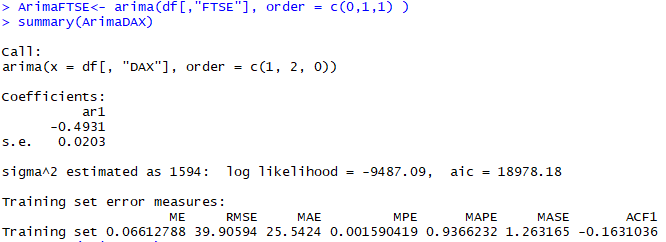
Αρχικά, για το δείκτη DAX:

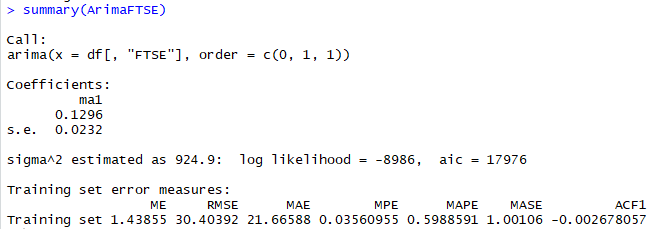
Η τάξη των p και q ελέγχεται έως το 15, ενώ το d έως 10.

Παρατηρούμε ότι το καλύτερο μοντέλο είναι τάξης (1,2,0). Περιέχει έναν αυτοπαλίνδρομο όρο και χρειάζεται να τη διαφορίσουμε 2 φορές. Εδώ υπάρχει μια αντίθεση με τα παραπάνω όπου είπαμε ότι είναι I(1). Aυτό οφείλεται στη χρήση διαφορετικού test ελέγχου.

Αντίστοιχη διαδικασία γίνεται και για τη FTSE, όπου το καλύτερο μοντέλο είναι τάξης (0,1,1), δηλαδή με ένα όρο σφάλμα τάξης 1. Επίσης εδώ έχουμε και όρο χρονικής τάσης.

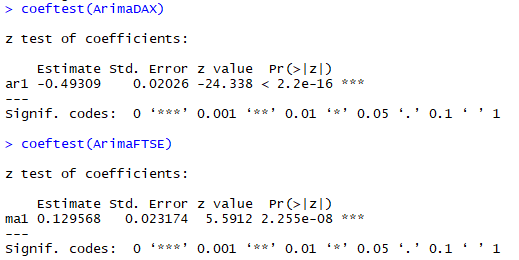


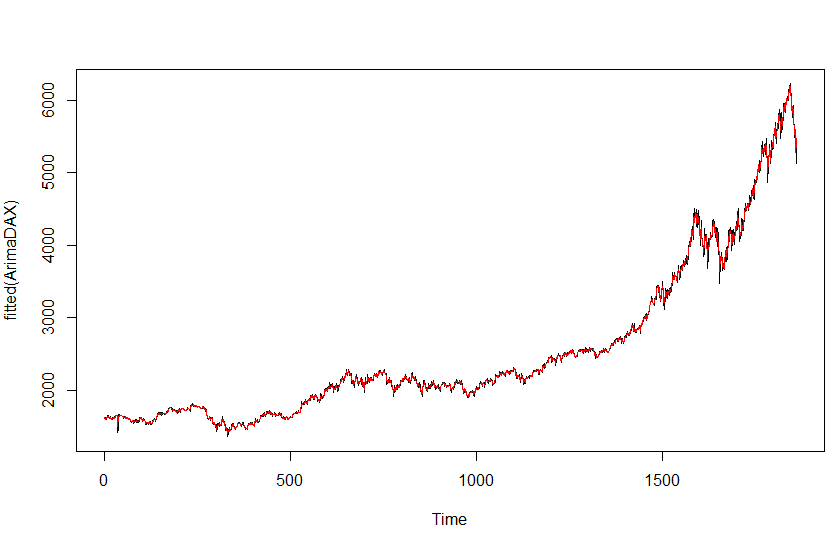
Κάνουμε λοιπόν fit αυτά τα μοντέλα στα δεδομένα μας και έχουμε τα παρακάτω αποτελέσματα:

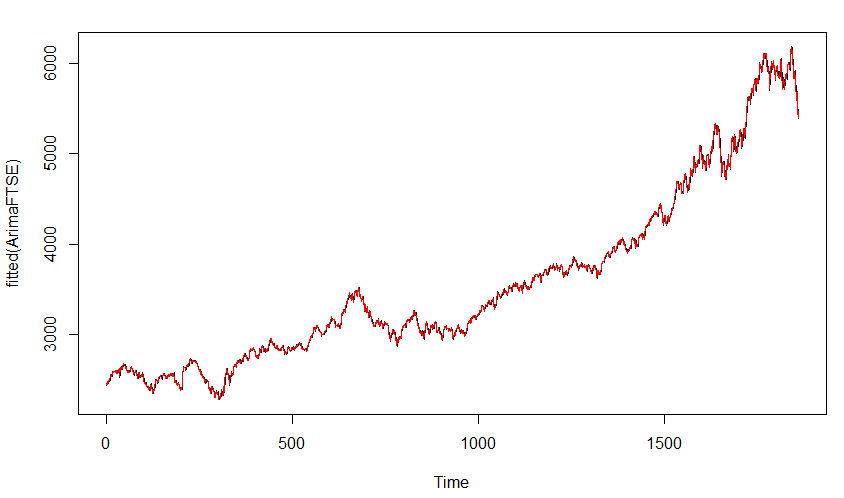


Εδώ αξίζει να παρατηρήσουμε την τιμή του συντελεστή και την τυπική του απόκλιση (s.e.). Έχουμε επίσης και μετρικές σφάλματος του μοντέλου στα δεδομένα (π.χ. RMSE και MAPE).

Πραγματοποιούμε έλεγχο σημαντικότητας των συντελεστών των υποδειγμάτων και όπως αναμέναμε είναι σημαντικοί.

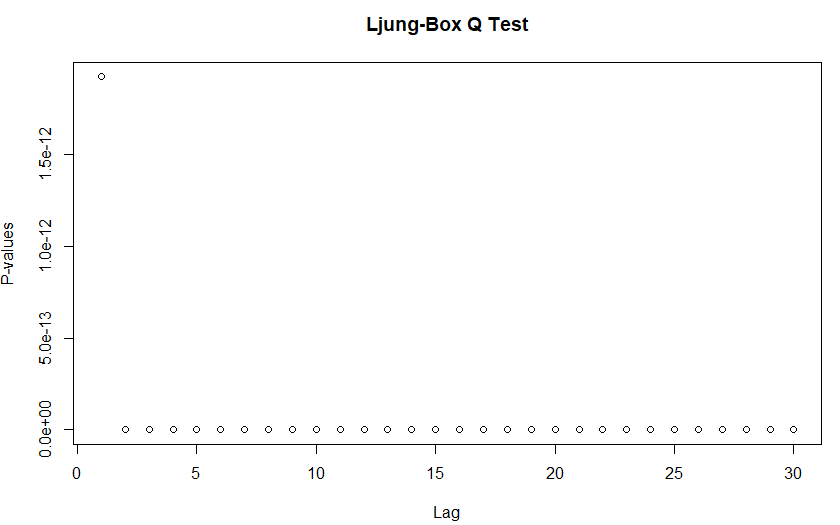


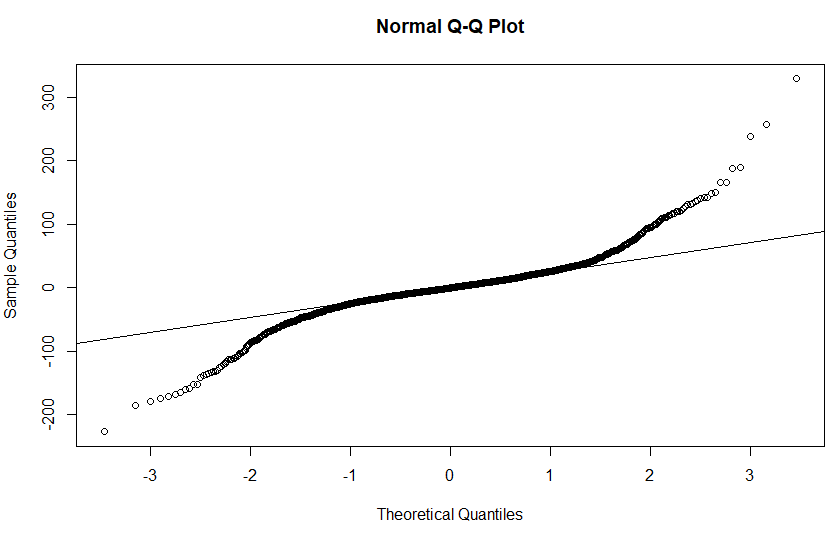
Στη συνέχεια έχουμε τα γραφήματα των υποδειγμάτων ARIMA που υπολογίσαμε (μαύρο) και την πραγματική χρονοσειρά (κόκκινο). (#1 ADX και #2 FTSE)

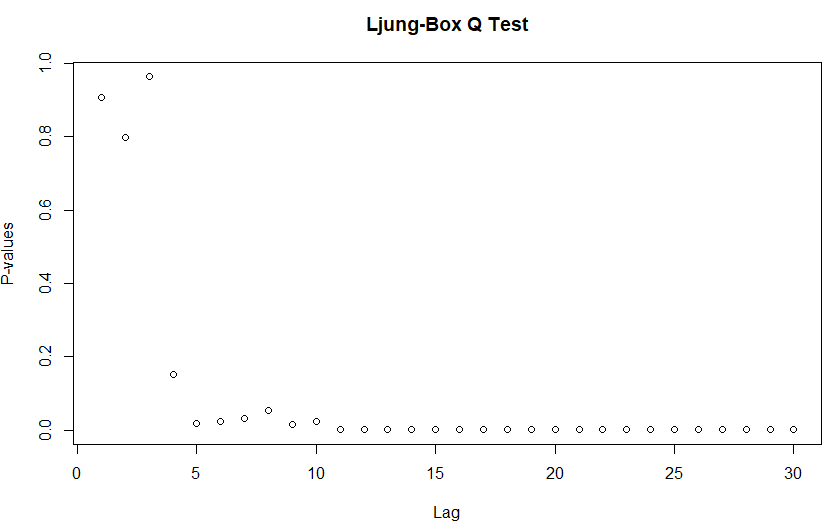


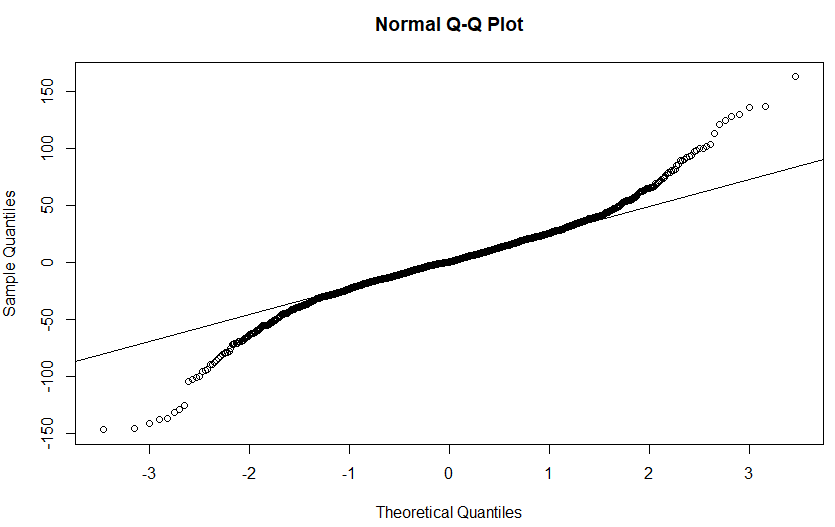
**3.5 Έλεγχος Ljung-Box**

Στη συνέχεια, πραγματοποιούμε τον έλεγχο Ljung-Box Portmanteau για να εξετάσουμε αν τα δεδομένα είναι ανεξάρτητα κατανεμημένα.

Για το δείκτη DAX



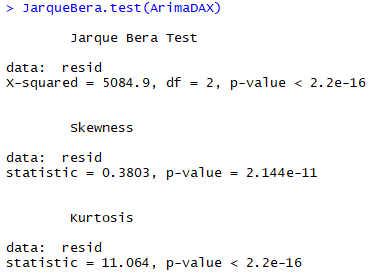
Για το δείκτη FTSE:

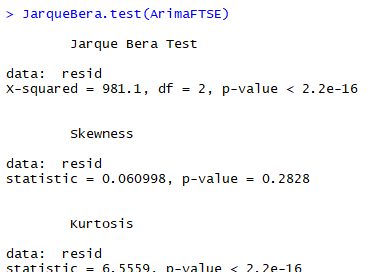


Ο έλεγχος είναι παραπλήσιος του Breusch-Godfrey. Παρατηρούμε ότι τα lags δεν είναι στατιστικώς σημαντικά, με λίγες εξαιρέσεις στο FTSE, ενώ στα Q-Q διαγράμματα τα σφάλματα είναι σε μια ευθεία κατά κύριο λόγο.

**3.6 Έλεγχος Jarque-Bera**

Στη συνέχεια πραγματοποιούμε έλεγχο Jarque-Bera για να ελέγξουμε την κανονικότητα των καταλοίπων. Ο έλεγχος υποδεικνύει ότι τα κατάλοιπα δεν είναι κανονικά κατανεμημένα και αυτό μπορεί να αποτελέσει πρόβλημα.



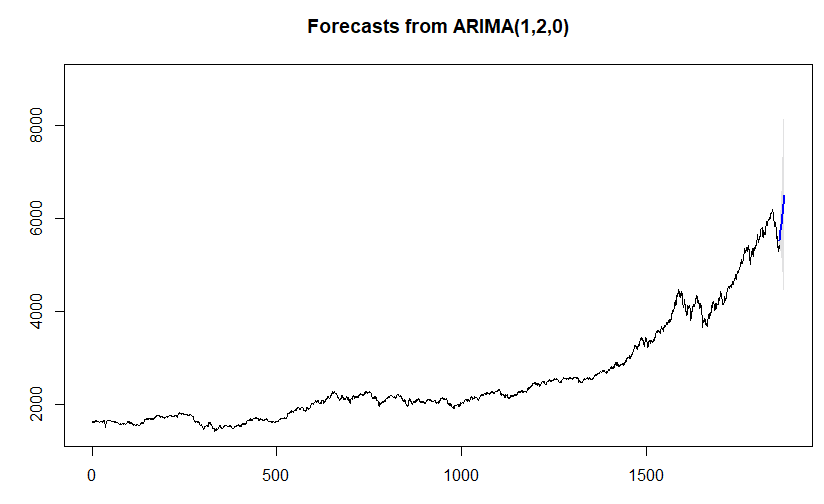


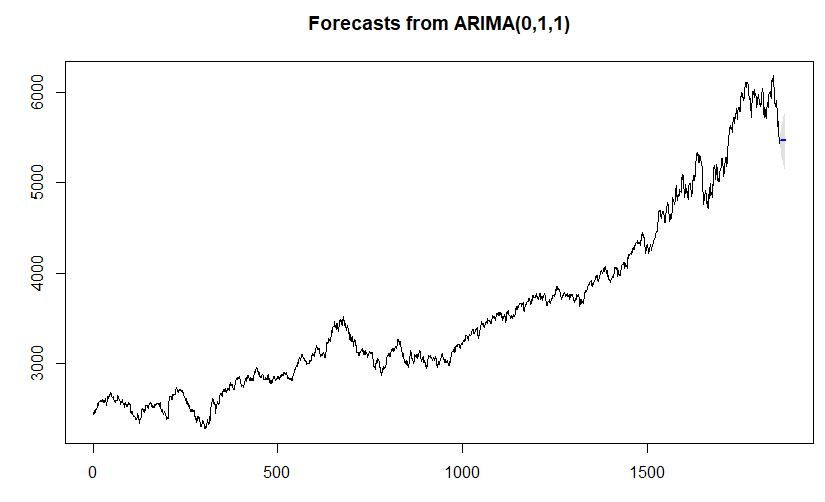
**Συμπεράσματα**

Έπειτα από πολλούς ελέγχους, οι παθογένειες του μοντέλου προκύπτουν από τα Ljung-Box και Jarque-Bera tests. Θα πρέπει ενδεχομένως να συμπεριλάβουμε περισσότερα lags προκειμένου να έχουμε ένα καλύτερο μοντέλο για τη χρονοσειρά.

Επίσης, δεν ασχοληθήκαμε με το αν παρουσιάζεται εποχικότητα την οποία θα έπρεπε να αφαιρέσουμε.

* Τέλος, θα μπορούσαμε να χρησιμοποιήσουμε τα μοντέλα αυτά για να κάνουμε προβλέψεις:





**Βιβλιογραφία**

[1] Φ. Κολυβά- Μαχαίρα Ε. Μπόρα- Σέντα, Στατιστική, Εκδόσεις Ζήτη, 2014

[2] Torsten Hothorn, Brian S. Everitt, A Handbook of Statistical Analyses Using R, second edition, Chapman and Hall, 2009

[3] https://www.analyticsvidhya.com/blog/2015/12/complete-tutorial-time-series-modeling/

[4] http://stats.stackexchange.com