Μέθοδοι και Τεχνικές Ανάλυσης Χρονολογικών Σειρών

1. ***Εισαγωγή***

*Στην συγκεκριμένη εργασία θα ασχοληθούμε με την εξαγωγή των δεδομένων τα όποια βρίσκονται σε ένα xml file, το όποιο παρέχετε από το* [*http://dblp.uni-trier.de*](http://dblp.uni-trier.de) *όπου είναι ένα computer science bibliography. Τα δεδομένα που θέλουμε να εξάγουμε από το συγκεκριμένο data set περιέχει το πλήθος των άρθρων σε περιοδικά , διατριβές ,δημοσιεύσεις κ.α. που έχουν γίνει στο κλάδο του computer science ανά έτος. Στην συνεχεία θα ασχοληθούμε με τη ανάλυση διαφορών μοντέλων χρονολογικών σειρών όπου ως χρονολογικές σειρές ορίζεται ένα σύνολο παρατηρήσεων μιας μεταβλητής οι οποίες έχουν ληφθεί σε ίσα χρονικά διαστήματα .Για τα συγκεκριμένα δεδομένα χρησιμοποιήθηκαν μοντέλα χρονολογικών σειρών όπως την στοχαστική προσέγγιση ARIMA ,εκθετική εξομάλυνση καθώς και το υπόδειγμα τυχαίας διαδρομής . Δηλαδή μοντέλα τα όποια έχουν υψηλό accuracy ,καθώς και να έχουν την ικανότητα για σωστή πρόβλεψη του μέλλοντος. Οι χρονολογικές σειρές αναλύονται σύμφωνα με την Τάση, Εποχικότητα, Κυκλικότητα, Τυχαιότητα και αναπτύσσεται η πρόβλεψη από τη σύνθεση αυτών των συνιστωσών. Γενικά οι μέθοδοι προβλέψεις μπορούν διαχωριστούν σε ποιοτικές και ποσοτικές. Οι ποιοτικές προβλέψεις στηρίζονται σε ποιοτικά δεδομένα , όπου οι επιστήμονες αναλυτές χρησιμοποιούν την εμπειρία ή την κρίση τους αλλά και στατιστικές μεθόδους (DELPHI) ενώ οι ποσοτικές προβλέψεις στηρίζονται μόνο σε στατιστικές μεθόδους και αφορούν ποσοτικές μεταβλητές. Στην συγκεκριμένη εργασία όπως είδαμε και παραπάνω θα ασχοληθούμε με ποσοτικές προβλέψεις*

1. ***Ανάλυση Χρονολογικών Σειρών***

*Στην ανάλυση ποσοτικών χρονολογικών σειρών έχουμε δυο ειδή μοντέλων, το προσθετικό και το πολλαπλασιαστικό μοντέλο, όπου το προσθετικό είναι της μορφής :*

*Καθώς και το πολλαπλασιαστικό :*

*Όπου είναι η τιμή της μεταβλητής, είναι η εποχιακή συνιστώσα , η κυκλική συνιστώσα και είναι η τυχαία συνιστώσα .Επίσης μια σημαντική έννοια για την ανάλυση χρονολογικών σειρών είναι η στασιμότητα. Μια στοχαστική*

*διαδικασία χαρακτηρίζεται ως στάσιμη όταν οι στατιστικές της ιδιότητες δεν επηρεάζονται από μια μεταβολή στην αρχή του χρόνου . Δηλαδή , οι στατιστικές ιδιότητες των Ν παρατηρήσεων με αρχή το t είναι οι ίδιες με τις στατιστικές ιδιότητες των Ν παρατηρήσεων με αρχή την περίοδο t+k . Γενικά μια χρονολογική σειρά θα είναι στάσιμη αν ο μέσος και η διακύμανση της δεν μεταβάλλονται με το χρόνο και η συνδιακύμανση μεταξύ των τιμών της σε δυο χρονικά σημεία και όχι από τον ίδιο το χρόνο. Αν μια χρονολογική σειρά είναι στάσιμη, τότε για όλα τα t θα ισχύουν :*

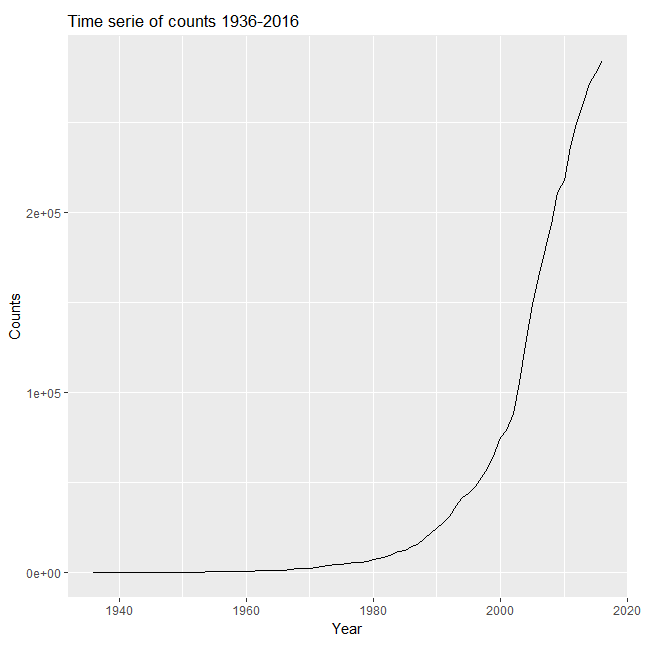
* *E()= ,*
* *var()=E*
* *cov = cov*

*όπου οι δυο πρώτες συνθήκες υποδηλώνουν σταθερό μέση και σαθρή διακύμανση . Η τρίτη δηλώνει ότι η συνδιακύμανση μεταξύ δυο οποιονδήποτε τιμών της που απέχουν κ περιόδους (αυτοσυνδιακυμαση) είναι συνάρτηση μόνο του k ,δηλαδή της χρονικής υστέρησης ή προήγησης των δυο αυτών τιμών .*

1. ***Επεξεργασία Δεδομένων***

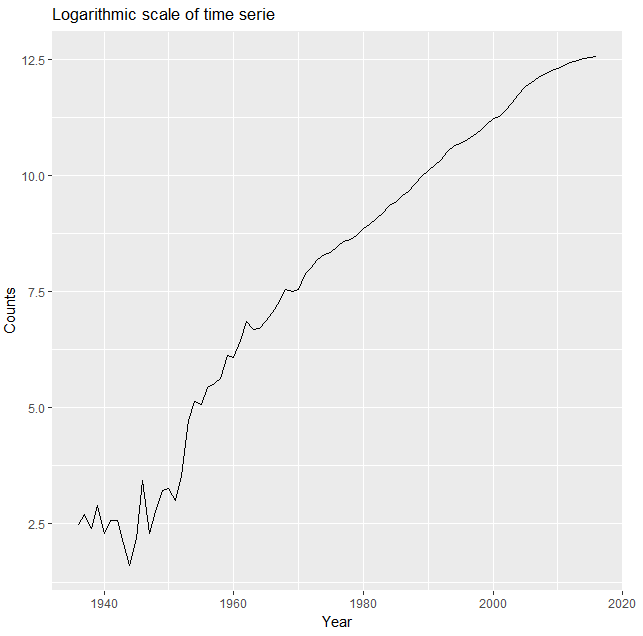
*Τα δεδομένα ήταν κοντά στα 2GB πράγμα που σημαίνει πως μια γλωσσά προγραμματισμού όπως R, Python κ.α. είναι δύσκολο και όχι τόσο καλή λύση να ανοίξουν το συγκεκριμένο αρχείο. Τα δεδομένα ανοιχτήκαν από το command line ενός Ubuntu Server καθώς και η επεξεργασία των δεδομένων που θα χρησιμοποιήσουμε ,διότι στο xml αρχείο που είχαμε, περιείχε περίπου 55 εκατομμύρια γραμμές εκ των οποίων εμάς μας ενδιέφεραν μόνο οι 4.092,332 γραμμές όπου είχαν την χρονολογία. Αφού κάναμε εξαγωγή ένα νέο αρχείο με της γραμμές που περιέχουν την χρονολογία έπειτα συνεχίσαμε διαβάζοντας το αρχείο αυτό στην R και υπολογίσαμε τις συχνότητες εμφάνισης κάθε χρονιάς .Έπειτα συνεχίσαμε με την ανάλυση της χρονολογικής σειράς καθώς και την πρόβλεψη της.*

***Διάγραμμα******3.1***

**

*Στο παραπάνω διάγραμμα βλέπουμε τα δεδομένα μας και παρατηρούμε ότι υπάρχει μια αυξητική τάση. Έχουμε βγάλει εκτός μερικές χρονιές (2017 έως 2019) για τον λόγο ότι υπήρχε μια καθοδική τάση, πράγμα το όποιο έχουμε υπόθεση ότι είναι εσφαλμένο και δεν αντανακλά την αληθινή εικόνα . Αυτό μπορεί να οφείλεται στο ότι δεν έχουν περαστεί ακόμα όλα τα δεδομένα για αυτά τα χρόνια .Κάτι που δεν μπούμε να δούμε στο παραπάνω διάγραμμα είναι μετατοπίσεις το όποιο οφείλεται στην κλίμακα που έχουμε απεικονίσει . Μια καλή κλίμακα για να φάνουνε οι μετατοπίσεις αυτές, είναι η λογαριθμική κλίμακα .*

***Διάγραμμα 3.2***

******

*Στο παραπάνω διάγραμμα παρατηρούμε αρκετές μετατοπίσεις από το 1936 έως 1970 περίπου τα όποια πριν φαινόντουσαν να έχουν μια σταθερή αυξητική τάση.*

1. ***ARIMA***

***Αυτοπαλινδρομο Υπόδειγμα***

*Αυτοπαλινδρομο υπόδειγμα p τάξης συμβολίζετε με AR(p) και εκφράζετε από την σχέση :*

*Όπου με είναι η σταθεροί παράμετροι και ο όποιος μετρά τυχαία σφάλματα. Στο αυτοπαλινδονουμενο υπόδειγμα , η εξαρτώμενη μεταβλητή παλινδρομείται στης προηγούμενες τιμές της . Η τάξη του αυτοπαλινδρονουμενου υποδείγματος συμβολίζεται με p και προσδιορίζει το μήκος της υστέρησης .*

***Υπόδειγμα Κινητού Μέσου***

*Ένα υπόδειγμα κινητού μέσου q τάξης συμβολίζεται με MA(q) και εκφράζετε από την σχέση :*

*Όπου με οι σταθεροί παράμετροι και τα τυχαία σφάλματα .Στην διαδικασία κινητού μέσου η χρονολογική σειρά θεωρείται ότι δημιουργείται ως ένας σταθμικός μέσος τυχαίων σφαλμάτων των q προηγουμένων περιόδων .*

*Ένα αυτοπαλινδρομο ολοκληρωμένο υπόδειγμα κινητού μέσου ARIMA(p,d,q) προκύπτει από τον συνδυασμό των αυτοπαλινδρομον διαδικασιών ΑR(p) και των διαδικασιών κινητού μέσου MA(q), όπου d είναι ο αριθμός των διαφόρων που απαιτούνται προκείμενου να μετατραπεί η σειρά σε στάσιμη.*

***Έλεγχος Στασιμότητας***

*Κάνοντας τον έλεγχο στασιμότητας* Dickey-Fuller tests και Phillips-Perron Unit Root Test στα *δεδομένα μας. Όπου* Το κριτήριο των Dickey-Fuller βασίζεται στον ακόλουθο έλεγχο:

:α=1 Η γίνεται δεκτή αν η t στατιστική *συνάρτηση* του συντελεστή α είναι μικρότερη από την t στατιστική *συνάρτηση* των Dickey-Fuller (αντίστοιχα για το Phillips-Perron).αν δεν έχουμε ένδειξης να απορριφτεί η τότε η σειρά δεν είναι στάσιμη

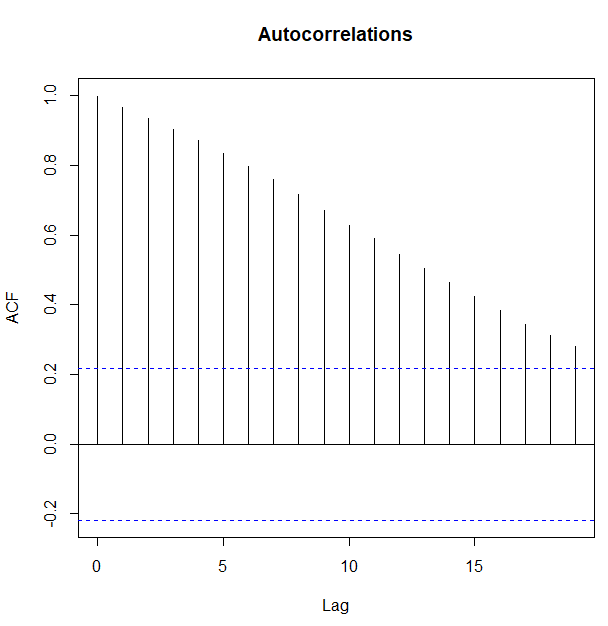
:α<1 Τότε συμπεραίνουμε ότι έχουμε μοναδιαία ρίζα και άρα η σειρά είναι στάσιμη

**Πινάκας 4.1**

|  |  |  |
| --- | --- | --- |
| *Έλεγχος* | *P-value* | *Στατιστική συνάρτηση-Τ* |
| Dickey-Fuller | *0.8379* | *-1.3596* |
| Phillips-Perron Unit Root | *0.737* | *-6.4766* |

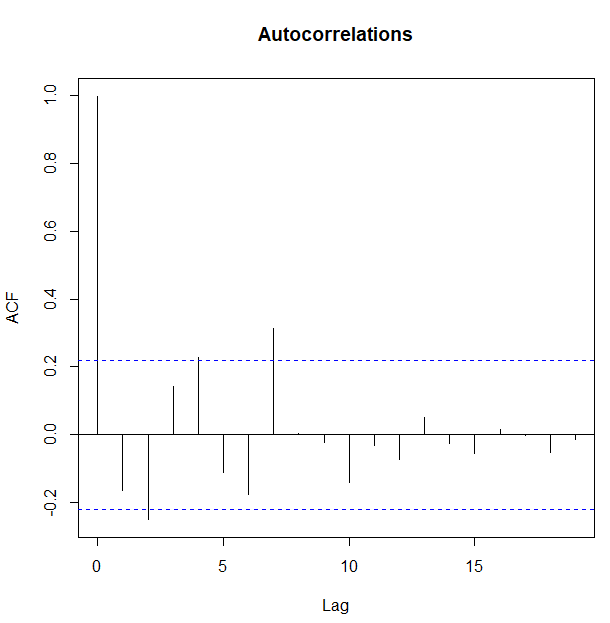
Και τα δύο τεστ συμφωνούν στο γεγονός μη επαρκούς ένδειξης απόρριψης της μηδενικής υπόθεσης, η χρονοσειρά ενδέχεται να έχει μοναδιαία ρίζα. Αυτό προκύπτει διότι το p-value είναι μεγαλύτερο από το επίπεδο σημαντικότητας 5% που ορίζουμε εμείς ότι θέλουμε να έχουμε ως σφάλμα. Από τους δυο αυτούς έλεγχους συμπεραίνουμε ότι η χρονοσειρα μας δεν είναι στάσιμη όποτε δεν πληρούνται οι προϋποθέσεις για το μοντέλο ARIMA.Περά από τους δυο αυτούς τους έλεγχους μπορούμε να διαπιστώσουμε και από την συνάρτηση αυτοσυσχέτισης (correlogram) που δίνετε στο διάγραμμα 4.1. Οι αυτοσυσχετισεις ‘AC’ φθίνουν με πολύ αργό ρυθμό που δηλώνει μη στασιμότητα. . Μια τεχνική για να γίνει η χρονολογική σειρά στάσιμη είναι το να πάρουμε τις πρώτες διαφορές της σειράς , δηλαδή ορίζουμε ως πρώτες διαφορές .

**Διάγραμμα 4.1**

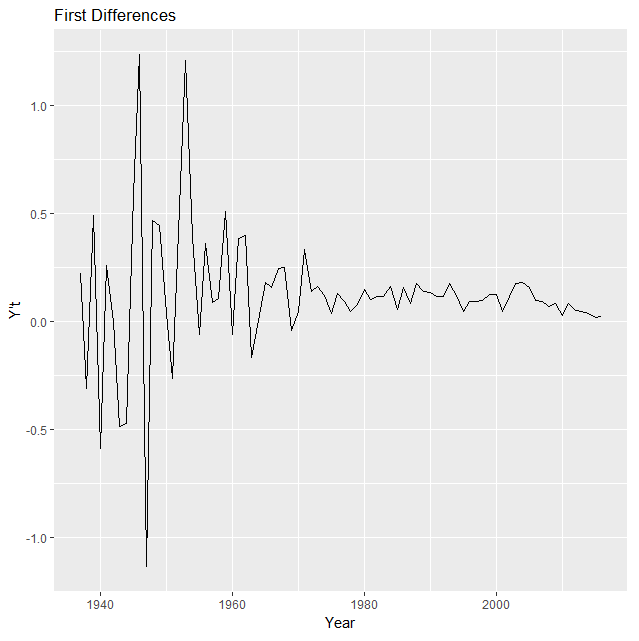


Αφού υπολογίσουμε τις διαφορές αυτές και ξανά κάνουμε τους δυο ελέγχους τότε παρατηρούμε πως δεν θα δεχτούμε τις μηδενικές υπόθεσης διότι τα p-value είναι 0.04731 και μικρότερα από 0.01 αντίστοιχα για τους δυο έλεγχους και οι στατιστικές συναρτήσεις είναι -3.502, -82.097 ,κάτι που φαίνετε και από το διάγραμμα αυτοσυσχέτισης παρακάτω. Επόμενος αφού έχουμε πλέον μια στάσιμη χρονολογική σειρά μπορούμε να συνεχίσουμε με την εκτίμηση των παραμέτρων του μοντέλου ARIMA(p,1,q).

**Διάγραμμα 4.2**

****

**Διάγραμμα 4.3** Η χρονοσειρά με πρώτες διαφορές .



***Μέθοδος Box-Jenkins***

*Η μέθοδος Box-Jenkins περιγράφεται από τρία βασικά σταδία :*

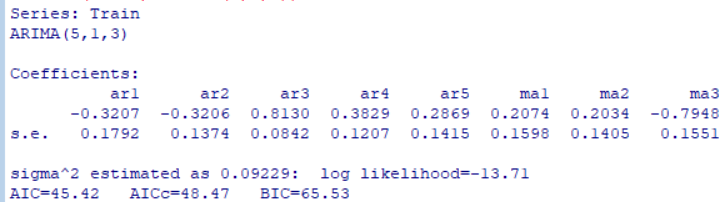
1. *Ταυτοποίηση*
2. *Εκτίμηση*
3. *Διαγνωστικός Έλενος*

*Στο στάδιο της ταυτοποίησης γίνεται ο καθαρισμός των τιμών p,d,q του μοντέλου ARIMA με βάση της πληροφορίας που παρέχει το δείγμα. Στην αρχή καθορίζεται ο αριθμός των διαφόρων d που απαιτούνται ώστε να γίνει η σειρά στάσιμη .Προφανώς η εκτίμηση για την τιμή του d είναι μηδέν διότι τι δουλεία αυτή την κάναμε εμείς προηγούμενος . Για τον έλεγχο στασιμότητας εξετάζεται η δειγματική συνάρτηση αυτοσυσχέτισης. Μετά το στάδιο της ταυτοποίησης έπεται το στάδιο της εκτίμησης όπου εκεί εκτιμούνται το p,d,q όπου έχουμε αναφέρει .Αφού ολοκληρωθεί το στάδιο της εκτίμησης συνεχίζει στο στάδιο του διαγνωστικού έλεγχου του οποίου ο σκοπός είναι να χρησιμοποιηθεί το μοντέλο για μελλοντικές πρόβλεψεις .Αρχικά διαχωρίσουμε τα δεδομένα μας σε δεδομένα εκπαίδευσης και δεδομένα έλεγχου, ,με δεδομένα εκπαίδευσης να είναι τις χρονολογίες [1938,2007] και τα δεδομένα έλεγχου [2008,2015] . Παρατηρούμε πως τα δεδομένα μας αρχικά είχαν ως αφετηρία το 1936 και ήταν μέχρι και το 2016 αλλά λόγο των δευτέρων διαφορών που πήραμε μερικές χρονιές εξαφανιστήκαν .*

*Από την μέθοδο Box –Jenkins πήραμε ως εκτίμηση τον (p,d,q)=(5,1,3)*

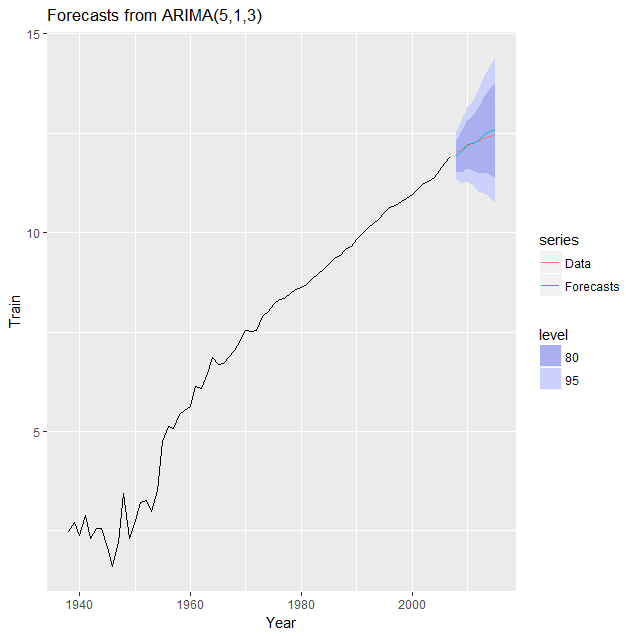
*Με συντελεστές του μοντέλου τους βλέπουμε στον παρακάτω πινάκα .*

***Πινάκας 4.2***

******

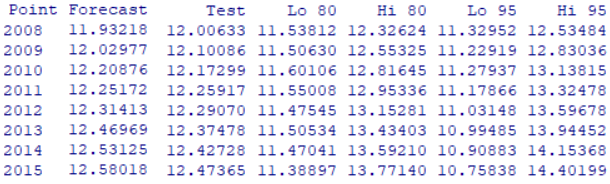
*Στο παρακάτω διάγραμμα βλέπουμε την χρονολογική σειρά μαζί με την πρόβλεψη όπου έγινε ,καθώς και τα πραγματικά δεδομένα για τις χρονιές της πρόβλεψης . Με μπλε χρώμα βλέπουμε το διάστημα εμπιστοσύνης που μας δίνετε για την πρόβλεψη.*

***Διάγραμμα. 4.2***

**

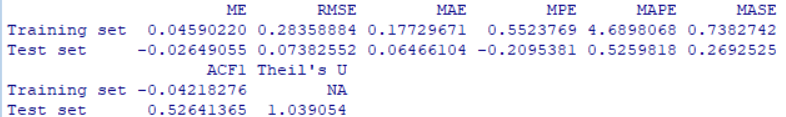
*Στον πινάκα 4.3 βλέπουμε τα αριθμητικά αποτελέσματα που πήραμε από την πρόβλεψη μας ,τις πραγματικές τιμές καθώς και δυο διάστημα εμπιστοσύνης, το ένα για 95% και ένα για 80% .*

***Πίνακας 4.3*** *Αποτελέσματα εκτίμησης 2008-2015*

**

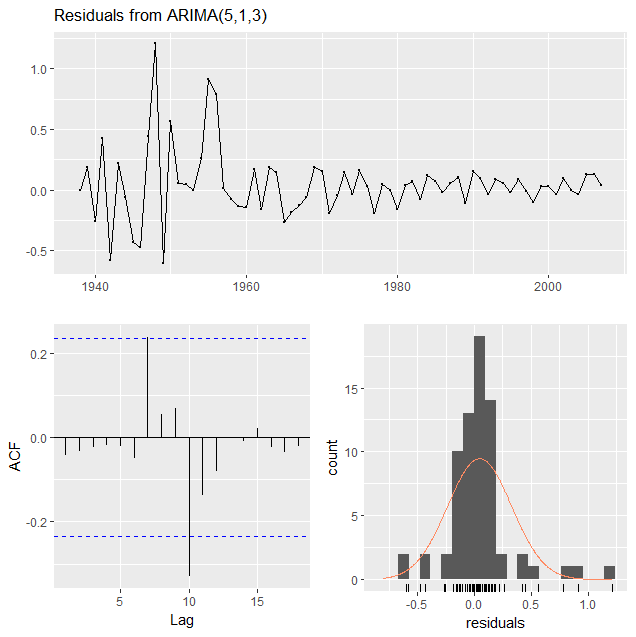
*Παρατηρούμε πως η πρόβλεψη του μοντέλου είναι αρκετά ακριβής σε σύγκριση με τα πραγματικά δεδομένα .* Η αξιοπιστία μετράται µε µια αντικειμενική συνάρτηση (πχ RMSE root mean squared error), απόκλιση της πρόβλεψης από την πραγματική τιµή στην περίοδο πρόβλεψης

***Πίνακας 4.4***

**

*Από ότι βλέπουμε η τιμή του RMSE είναι αρκετά μικρή και για το Train set αλλά και για το test set όποτε η ακρίβεια μας είναι αρκετά μεγάλη για τον λόγο ότι η απόκλιση από της πραγματικές τιμές είναι αρκετά μικρή ποσότητα.*

***Διάγραμμα 4.3***

**

*Από ότι παρατηρούμε από το παραπάνω διάγραμμα τα κατάλοιπα της παλινδρόμησης ακολούθου περίπου κανονική κατανομή, για την ακρίβεια έχει πιο «παχιές ουρές», πράγμα αναμενόμενο για μια χρονολογική σειρά .*

1. ***Εκθετική Εξομάλυνση***

Οι μέθοδοι εκθετικής εξομάλυνσης βασίζονται στην εκθετική μείωση της βαρύτητας που δίνεται στα στοιχεία των προηγούμενων περιόδων. Αυτές οι μέθοδοι συνήθως χρησιμοποιούνται σε περιπτώσεις όπου ο χρονικός που θα ορίσουμε για την πρόβλεψη είναι σχετικά μικρός ενώ δεν υπάρχουν διαθέσιμες πληροφορίες για την αιτιοκρατική σχέση που συνδέει την προς πρόβλεψη μεταβλητή και τους ανεξάρτητους παράγοντες που την επηρεάζουν. Επίσης σημαντικό είναι ότι η εκθετική εξομάλυνση χαρακτηρίζονται από την εξομάλυνση των τυχαίων διακυμάνσεων που μπορεί να παρουσιάζουν τα διάφορα στοιχεία των χρονοσειρών (οριζόντιο, τάσης, εποχικό και κυκλικό).

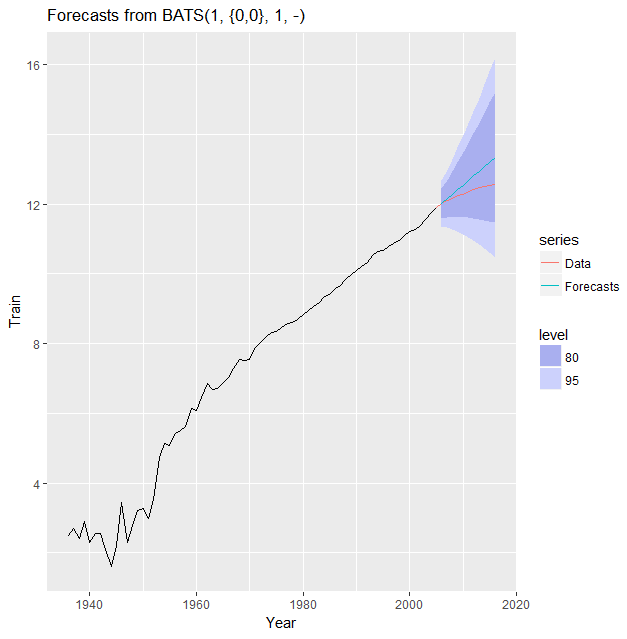
είναι η πραγματική ζήτηση την περίοδο t

είναι η πρόβλεψη για την επόμενη περίοδο

α=Σταθερά εξομάλυνσης (μεταξύ 0 και 1), συνήθως μεταξύ 0,01 και 0,3

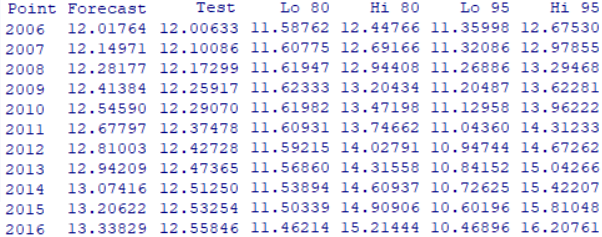
*Στην συγκεκριμένο πρόβλημα χρησιμοποίησα το μοντέλο* Exponential Smoothing State Space Model With Box-Cox Transformation, ARMA Errors, Trend And Seasonal Components.Το Box-Cox είναι μια οικογένεια μετασχηματισμών μέσα στους οποίους μπορούμε να προσδιορίσουμε τον πλέον κατάλληλο μετασχηματισμό για την μεταβλητή Y ώστε να εξομαλύνουμε η και να εξαλείψουμε (κάποιες φορές) τις αποκλίσεις του μοντέλου. χρησιμοποιείτε σε περίπτωσης που διαπιστώνεται έλλειψη κανονικότητας των σφαλμάτων .

***Διάγραμμα 5.1*** *Τα δεδομένα σε λογαριθμική κλίμακα*

**

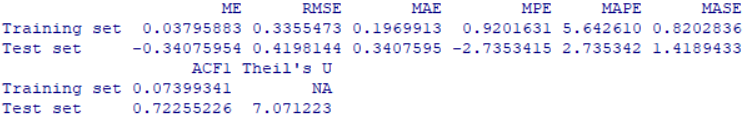
*Από ότι βλέπουμε από το παραπάνω διάγραμμα το μοντέλο αυτό είναι καλό διότι η παράβλεψη μας είναι αρκετά ακριβής αν και έχει μια αυξητική τάση γενικά ενώ στα πραγματικά δεδομένα δεν ισχύει άρα μπορεί και να μην είναι καλό μοντέλο για μακροχρόνια πρόβλεψη. Από ότι φαίνεται και από τον παρακάτω πινάκα έχει υψηλή ακρίβεια .*

***Πινάκας 5.1*** *Αποτελέσματα εκτίμησης 2006-2016*

**

*Στον παραπάνω πινάκα βλέπουμε τα αριθμητικά αποτελέσματα που έχουμε από την πρόβλεψη του μοντέλου καθώς επίσης, όπως και προηγουμένως δίνονται και δυο διαστήματα εμπιστοσύνης (80% και 95 %) που απεικονίζονται και στο διάγραμμα 5.1 με τις το γαλάζιο σκούρο και γαλάζιο ανοιχτό αντίστοιχα 0.3355 – 0.4198*

***Πίνακας 5.2***

**

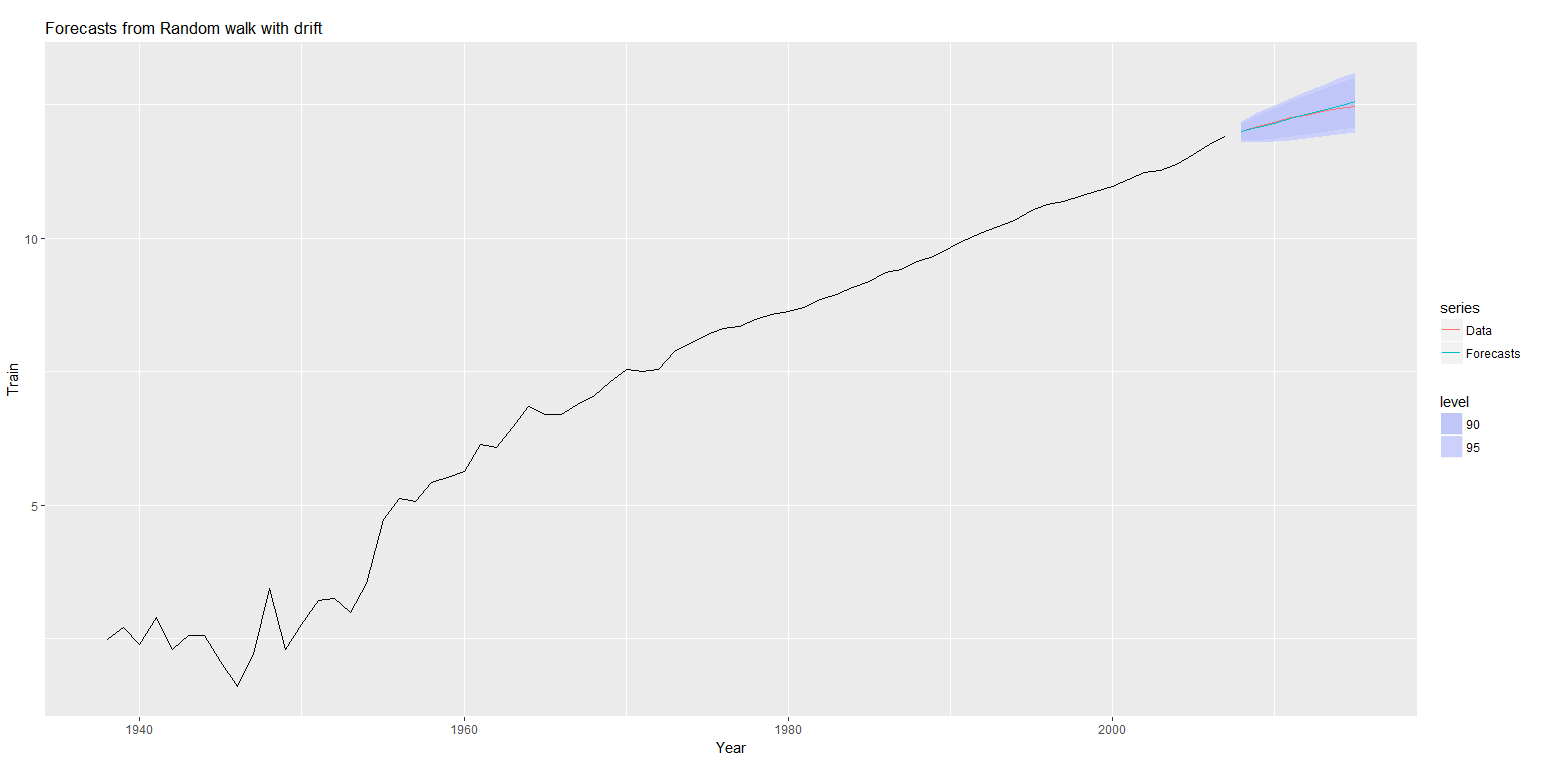
Η απόκλιση της πρόβλεψης από την πραγματική τιµή στην περίοδο πρόβλεψης είναι ικανοποιητική αλλά μεγαλύτερη από το μοντέλο ARIMA που κάναμε προηγούμενος .

1. ***Υπόδειγμα Τυχαίας Διαδρομής***

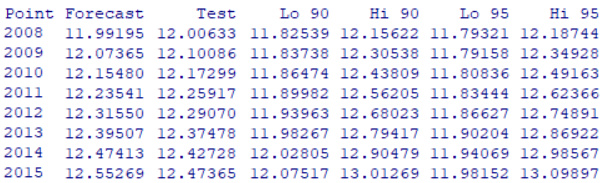
*Σε ένα υπόδειγμα τυχαίας διαδρομής κάθε τιμή της χρονολογικής σειράς, έστω προκύπτει από την αμέσως προηγούμενη της με την προσθήκη ενός τυχαίου σφάλματος , δηλαδή (1), Με τυχαία σφάλματα . στο υπόδειγμα αυτό παρατηρούμε ότι οι διαδοχικές μεταβολές των τιμών είναι ανεξάρτητες τυχαίες μεταβλητές. Δηλαδή (2) .Ισχύει επίσης ότι (3) . Η εξίσωση αυτή αποτελεί και τη γενική λύση της εξίσωσης διαφορών πρώτης τάξης που αντιπροσωπεύει η τυχαία διαδρομή (2).*

*Η σειρά της τυχαίας διαδρομής (1) έχει σταθερό μέσο καθώς .Η διακύμανση της και οι συνδιακυμάνσεις των τιμών της δεν παραμένουν σταθερές διαχρονικά και άρα η είναι μη στάσιμη .*

***Διάγραμμα 6.1***

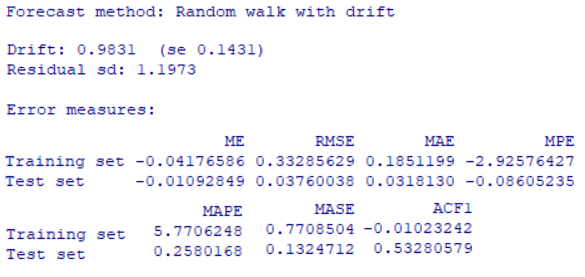
**

***Πινάκας 6.1*** *Αποτελέσματα εκτίμησης 2008-2015*

**

*Από ότι παρατηρούμε από τον παρακάτω πινάκα αποτελεσμάτων της πρόβλεψης που έγινε με την μέθοδο του τυχαίου περιπάτου , Η πρόβλεψη που έγινε είναι αρκετά ακριβής σε σχέση με όλες της προηγούμενες που έχουν γίνει έως τώρα κάτι που φαίνεται και από το διάγραμμα 6.1 .*

***Πινάκας 6.2***

**

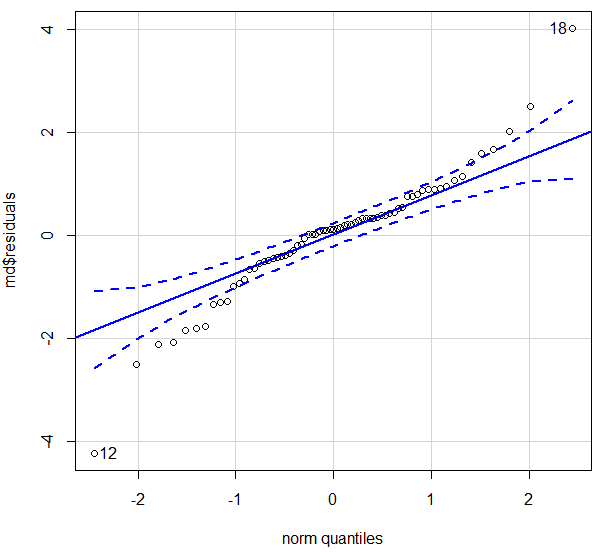
*Το υπόδειγμα του τυχαίου περιπάτου που εκτελέσαμε μας έβγαλε και μια σταθερά (drift=0.9831 ) . Αν ένα υπόδειγμα της τυχαίας διαδρομής περιλαμβάνει και τον σταθερό ορό αυτόν τότε έχουμε υπόδειγμα τυχαίας διαδρομής με σταθερά . Όπου β (drift) είναι η σταθερά της παρακάτω εξίσωσης .*

*(4)*

*Είναι σημαντικό να αναφέρουμε ότι ,σε σύγκριση με το υπόδειγμα τυχαίας διαδρομής , οι πρώτες διαφορές στο υπόδειγμα με σταθερά είναι εν μέρει στοχαστικές και εν μέρει σταθερές .*

*(5)*

***Διάγραμμα 4.2*** *qqplot καταλοίπων*

**

*Παρατηρούμε ότι και σε αυτό το μοντέλο τα σφάλματα έχουν πολύ έντονα παχιές ουρές κάτι το όποιο αναμένετε γιατί στις περισσότερες χρονολογικές σειρές παρατηρείτε αυτό το φαινόμενο, και θυμίζουν περισσότερο ότι ακλουθούνε την t-student κατανομή και όχι κανονική.*

1. ***Αποτελέσματα Μοντέλων***

*Στο παρακάτω πινάκα βλέπουμε τα αποτελέσματα του RMSE για τα 3 μοντέλα και διαπιστώνουμε ότι το μοντέλο με την καλύτερη ικανότητα πρόβλεψης για το μέλλον είναι το υπόδειγμα τυχαίας διαδρομής με σφάλμα για το test set να είναι 0.03760 αρκετά μικρότερο από όλα τα αλλά υποδείγματα .*

***Πινάκας 7.1***

|  |  |  |  |
| --- | --- | --- | --- |
| *Υπόδειγμα* | *ARIMA* | *Εκθετική εξομάλυνση* | *Τυχαίας διαδρομής* |
| *Train Set* | 0.2835 | *0.3355* | *0.3328* |
| *Test set* | 0.0738 | *0.4198* | *0.03760* |

1. ***Εξαγωγή δεδομένων***

*Στο ερώτημα που έχει γίνει στο ένα τα δεδομένα ήταν πολλαπλασιασμένα επί (10,100,1000,100000,…) μια λύση θα ήταν εφικτή για την ανάλυση των δεδομένων είναι να χρησιμοποιηθεί κάποιο εργαλείο για Big data όπως για παράδειγμα το Spark το όποιο είναι κατάλληλο για την ανάλυση μεγάλου όγκου δεδομένων. Τα δεδομένα για αυτήν την εργασία τα διαβάσαμε με το command line ενός VM ubundu server όπως έχουμε αναφέρει και στο πρώτο κεφάλαιο με χαρακτηριστικά 4GB Ram x1 Cpu και 20 GB χωρητικότητα . Έγιναν πειράματα να αναπαράγουμε 2 φορές τα δεδομένα με ένα copy-paste από το command line και να ξανά κάνουμε την ιδία διαδικασία ώστε να δούμε την διάφορα του χρόνου στα 55 σχεδόν εκατομμύρια με σύγκριση τα 111 εκατομμύρια δεδομένα. Το ίδιο πείραμα προσπαθήσαμε να το κάνουμε και για 4 φορές τα δεδομένα μας αλλά το μηχάνημα δεν τα κατάφερε να το κάνει .Τα κανονικά δεδομένα έκανε να τα ανοίξει 8.40 δευτερόλεπτα ενώ τα διπλασιαζόμενα δεδομένα έκανε 1.30 λεπτά ,άρα καταλαβαίνουμε ότι υπάρχει μεγάλη διάφορα . Επόμενος σε παράλληλο σύστημα όπως το Spark θα ήταν μια πολύ καλή λύση.*

***Βιβλιογραφία***

***[1].*** *Δημελη Σ. ; ( 2003 ) ; Σύγχρονες μέθοδοι ανάλυσης χρονολογικών σειρών.* ***[2].*** *Μαρκοπουλος Α. , Ντεντης Ι. Παρασκευοπουλος Ν ; (2015) ; Η χρήση της μεθοδολογίας box-jenkins στην ανάλυση χρονοσειρων.****[3].***  *Εμιρης Δ. ; (2012) ; Προβλέψεις .* ***[4].*** *Κουντουρη Φ. ; (2008 ) ; Χρονολογικες σειρες.* ***[5].*** *Γιώργος Θεοδώρου , Δημήτρης Κουγιουμτζής ; (2009 ) ; Μοντέλων Χρονοσειρών και Πρόβλεψη.* ***[6].*** *Σαριαννίδης Ν. ; Οικονομετρία ; Ανάλυση Χρονολογικών Σειρών ;* ***[7].*** *Τεχνικές προβλέψεων & έλεγχου; Μάθημα θεωρίας στάσιμες διαδικασίες - υποδείγματα .*