

Χρονοσειρές

Υπολογιστική Εργασία

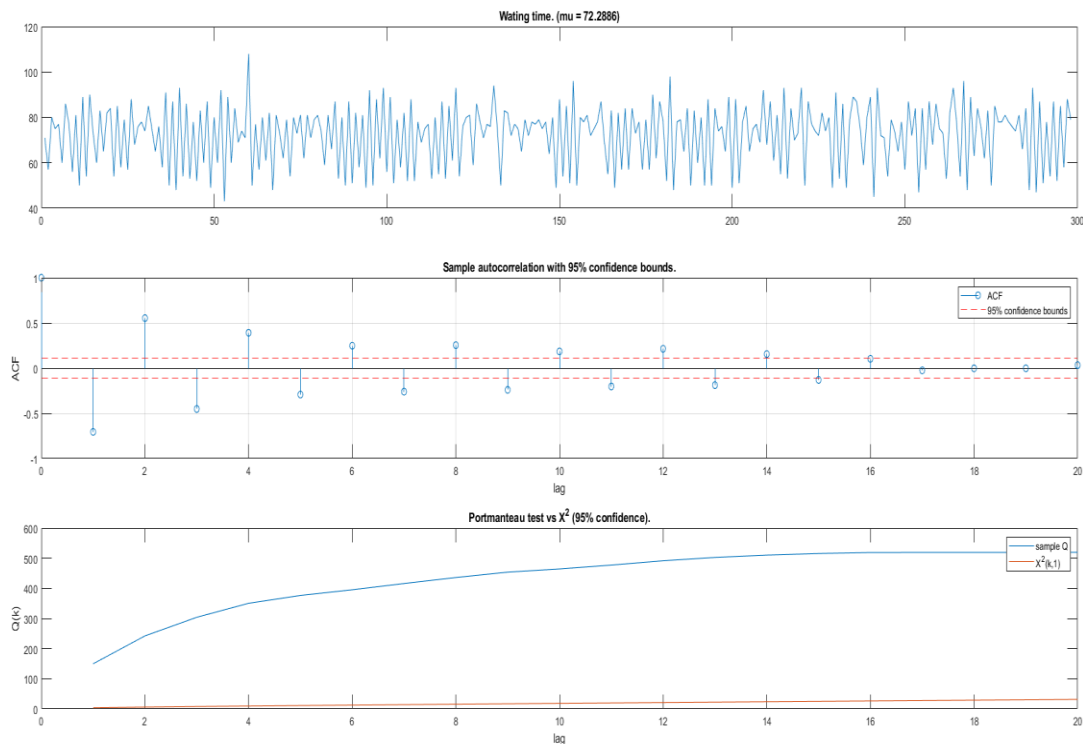
2022 - 2023

Ομάδα 6

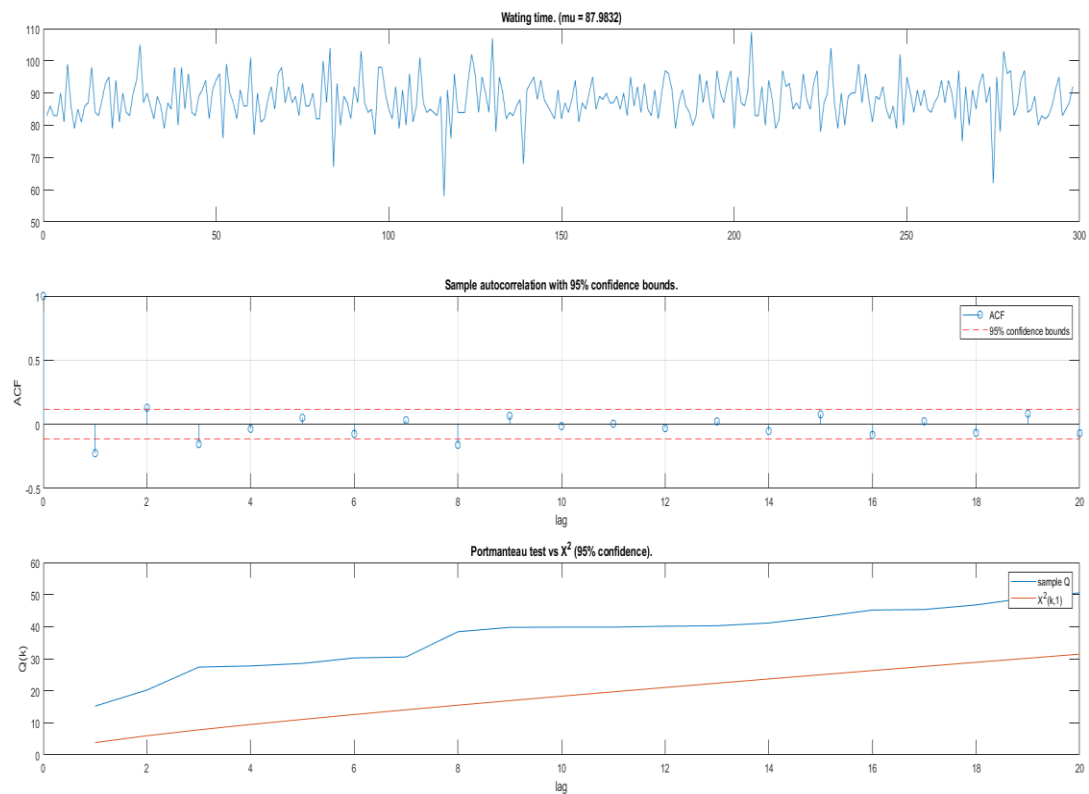
Γρηγορίου Στέργιος 9564

Πρώτο στάδιο:

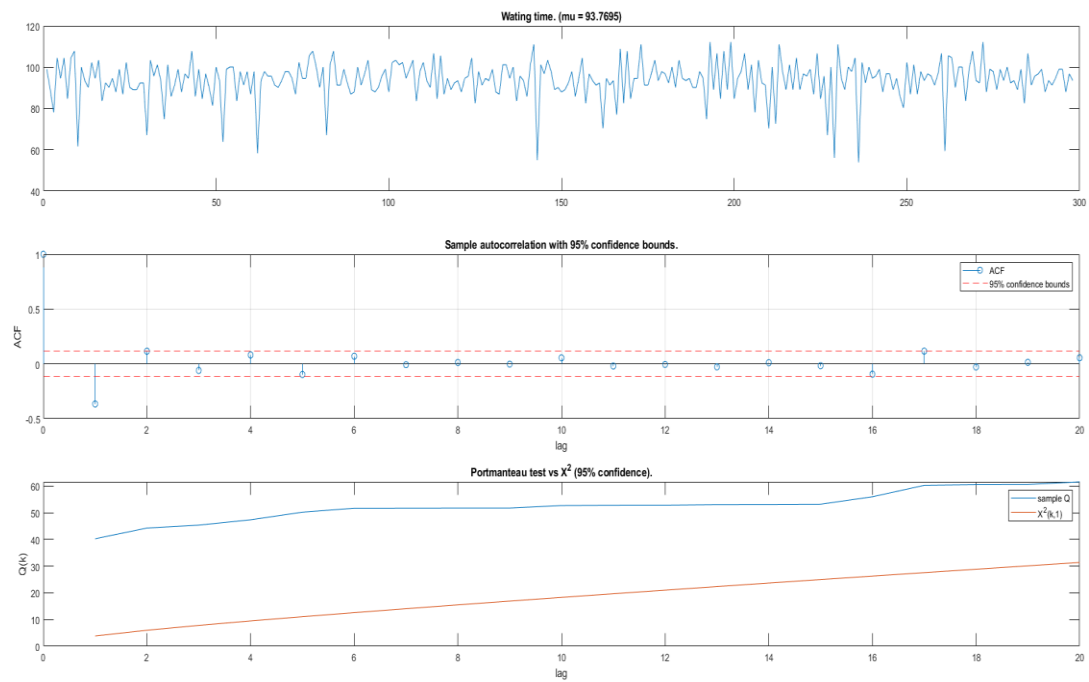
Αρχικά υπολογίζω τα διαγράμματα των 3 χρονοσειρών για 298 συνεχόμενους, αλλά τυχαίους, χρόνους αναμονής (όσο είναι και το μέγεθος της χρονοσειράς για το 1989). Η μέση τιμή είναι εμφανές πως παραμένει περίπου σταθερή σε κάθε μία από τις χρονοσειρές. Έχοντας τα διαγράμματα και τις μέσες τιμές τους, θα μπορούσαμε, να υποθέσουμε πως με την πάροδο των χρόνων η αναμονή ανάμεσα σε εξάρσεις του πίδακα έχει μια τάση αυξητική, αλλά τοπικά (στο χρόνο) μπορούμε να την θεωρήσουμε σταθερή (έχει πολύ μικρή κλίση). Παρατηρώ ότι ίσως υπάρχει κάποια περιοδικότητα, αλλά η εκφώνηση δηλώνει πως τις θεωρούμε στάσιμες. Οπότε δεν θα προσπαθήσω να βρω πιθανή περίοδο, αλλά θα ξεκινήσω την ανάλυση.



Εικόνα 1: Η χρονοσειρά, η αυτοσυσχέτιση της (για τις πρώτες 20 υστερήσεις), και έλεγχος portmanteau για την γραμμική ανεξαρτησία (1989).



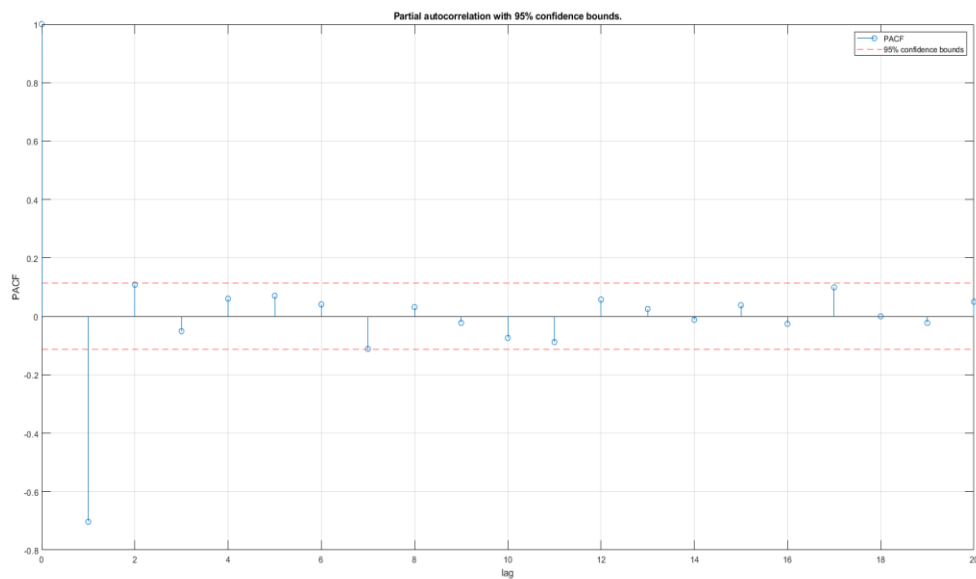
Εικόνα 2: Η χρονοσειρά, η αυτοσυσχέτιση της (για τις πρώτες 20 υστερήσεις), και έλεγχος portmanteau για την γραμμική ανεξαρτησία (2000).



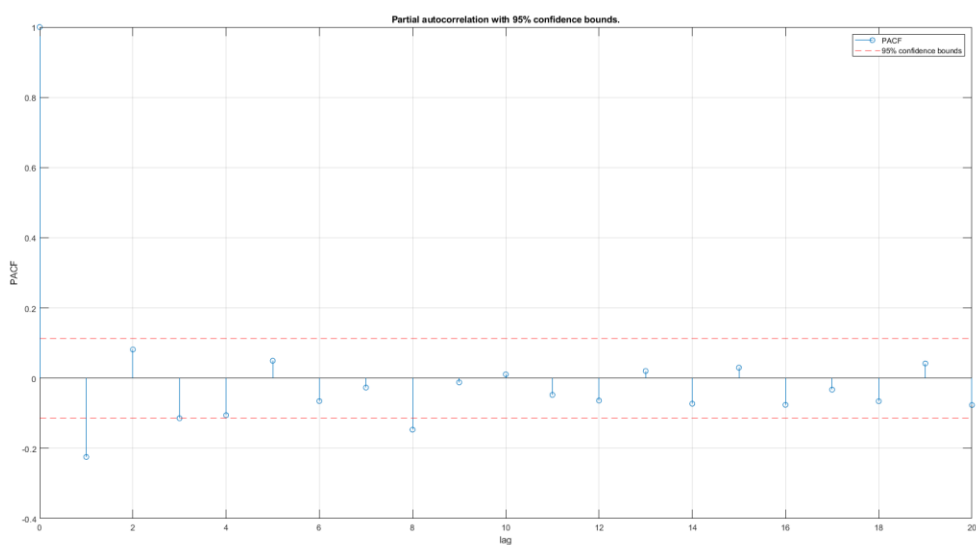
Εικόνα 3: Η χρονοσειρά, η αυτοσυσχέτιση της (για τις πρώτες 20 υστερήσεις), και έλεγχος portmanteau για την γραμμική ανεξαρτησία (2011).

Υπολογίζω, την αυτοσυσχέτιση της κάθε χρονοσειράς για 20 υστερήσεις και κάνω τα διαγράμματα τους με όρια εμπιστοσύνης 95%. Παρατηρώ πως και στις 3 χρονοσειρές πάνω από μία τιμή ξεπερνάει το όριο. Επομένως έχουμε απόρριψη πάνω από 5% για την μηδενική υπόθεση πως είναι λευκός θόρυβος. Συγκεκριμένα, 15 τιμές για το 1989 που φθίνει εναλλάξ εκθετικά, 4 τιμές για το 2000 που φθίνει αρκετά πιο γρήγορα και 3 τιμές για το 2011 (1,2,17) που φθίνει ακόμα πιο γρήγορα. Έπειτα υπολογίζω το στατιστικό Q των Ljung και Box, το οποίο επιβεβαιώνει την παρατήρησή μου πως οι χρονοσειρές έχουν σημαντικές αυτοσυσχετίσεις (υπάρχει γραμμική εξάρτηση) και σίγουρα δεν είναι λευκός θόρυβος.

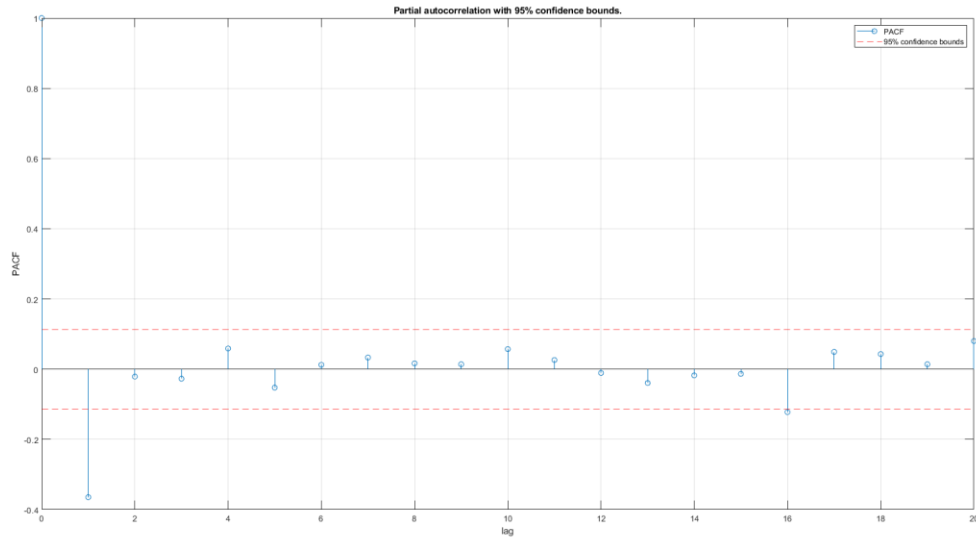
Το επόμενο βήμα είναι να υπολογίσω και την μερική αυτοσυσχέτιση και να εκτυπώσω το διάγραμμά της.



Εικόνα 4: Η μερική αυτοσυσχέτιση για τις πρώτες 20 υστερήσεις (1989).



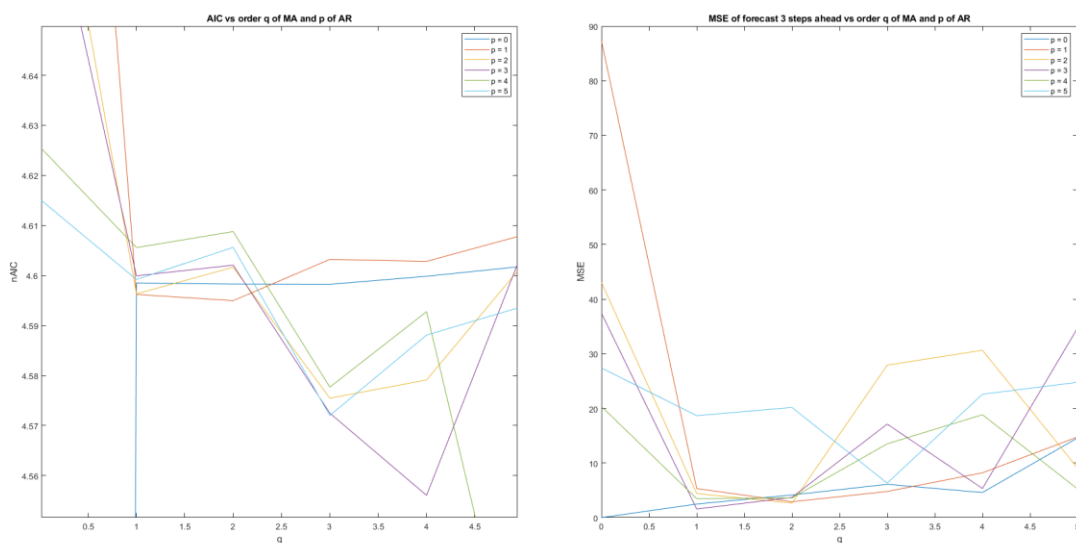
Εικόνα 5: Η μερική αυτοσυσχέτιση για τις πρώτες 20 υστερήσεις (2000).



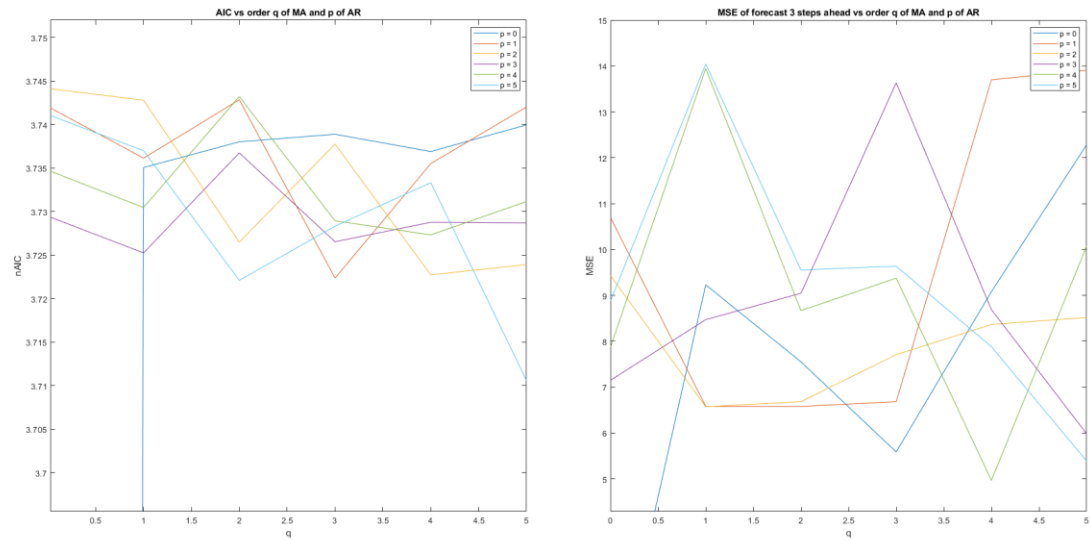
Εικόνα 6: Η μερική αυτοσυσχέτιση για τις πρώτες 20 υστερήσεις (2011).

Για την εύρεση του πιο κατάλληλου γραμμικού μοντέλου προσαρμογής και πρόβλεψης ακολούθησα την παρακάτω συλλογιστική.

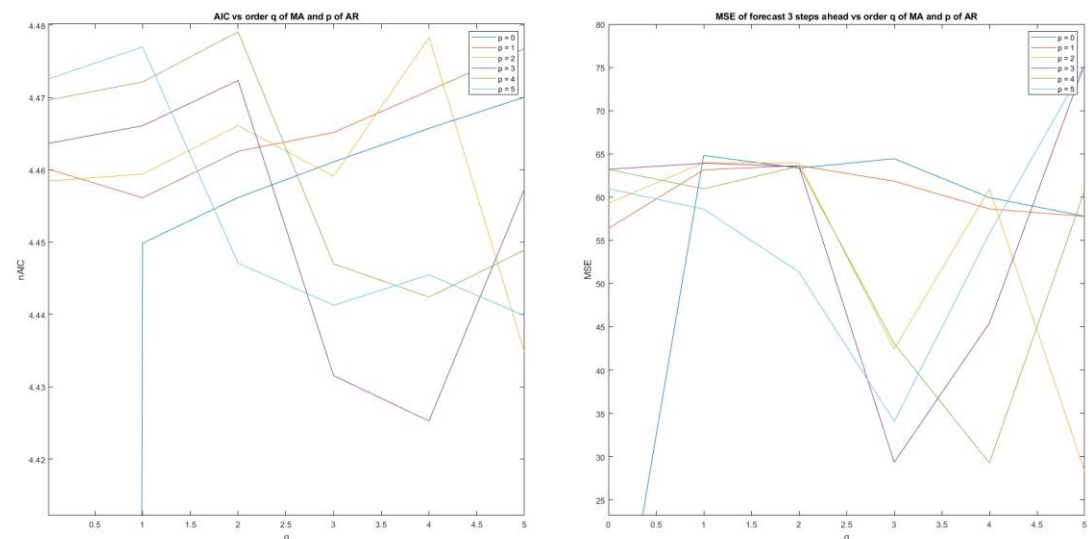
1. Χωρίζω την χρονοσειρά σε train και test παρατηρήσεις (90-10)(257 πρώτες-31 τελευταίες).
2. Προσαρμόζω όλα τα μοντέλα ARMA με p και q μικρότερο από 5 (για να αποφύγω την υπερπροσαρμογή, αποφεύγω να αυξήσω παραπάνω την τάξη) και με βάση το AIC, καθώς και την μερική αυτοσυσχέτιση για AR και την ολική για MA. Επιλέγω τα πιο καλά μοντέλα για κάθε χρονοσειρά. Επειδή το AIC δεν παρουσιάζει μεγάλη διακύμανση συμπληρωματικά, ελέγχω και την δυνατότητα πρόβλεψης στις πρώτες παρατηρήσεις του test set.
3. Δημιουργώ διαγράμματα για AIC και prediction MSE vs p, q .



Εικόνα 7: Αριστερά διάγραμμα AIC vs p, q και δεξιά το ίδιο για το MSE. (1989)



Εικόνα 8: Αριστερά διάγραμμα AIC vs p, q και δεξιά το ίδιο για το MSE. (2000)

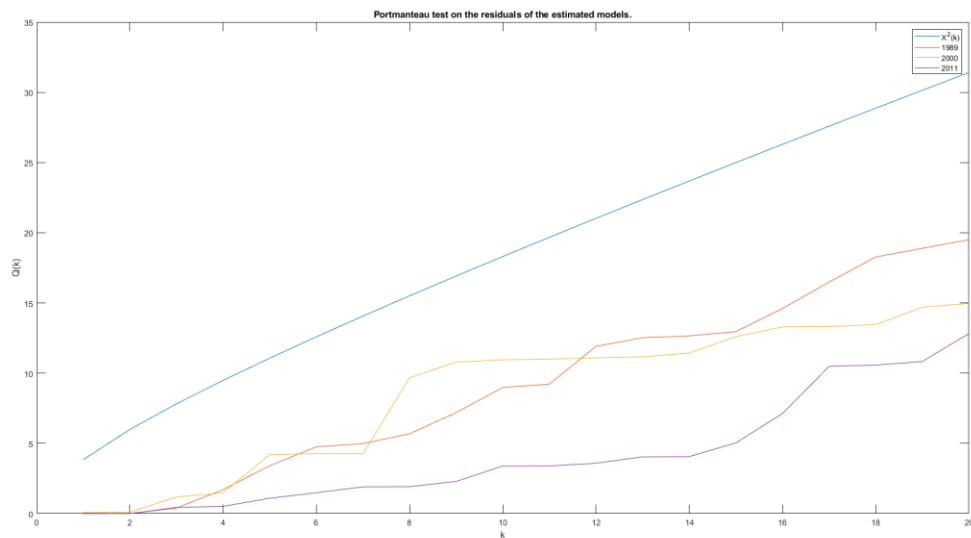


Εικόνα 9: Αριστερά διάγραμμα AIC vs p, q και δεξιά το ίδιο για το MSE. (2011)

Παρατηρώ πως τα 3 πιο καλά μοντέλα με βάση αυτές τις μετρικές και επιβραβεύοντας χαμηλότερης τάξης μοντέλα για κάθε έτος είναι αυτά που φαίνονται στον παρακάτω πίνακα:

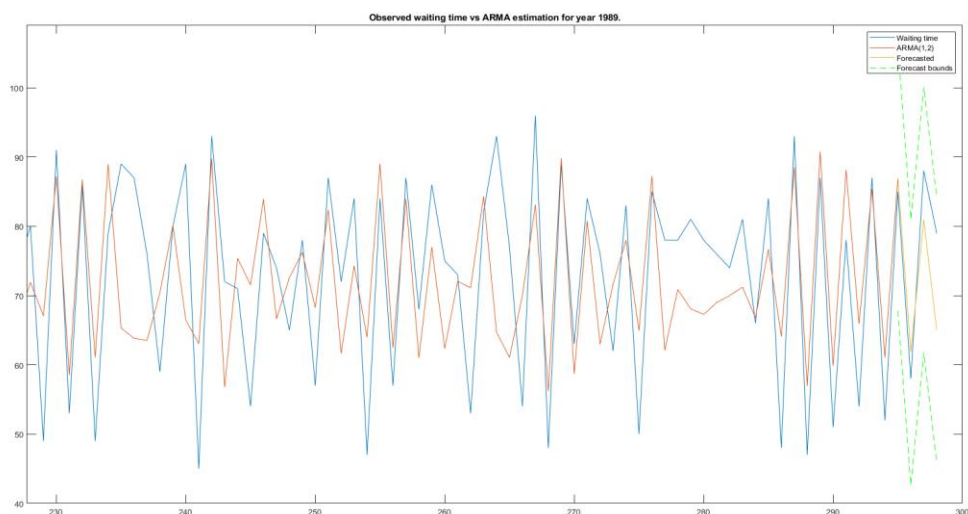
1989	2000	2011
AR(5)	AR(3)	AR(2)
ARMA(1,2)	ARMA(1,3)	ARMA(1,2)
ARMA(3,4)	ARMA(2,2)	ARMA(3,3)

Για να επιλέξω ανάμεσα στα επικρατέστερα, χρησιμοποίησα τις τελευταίες 9 παρατηρήσεις σαν τεστ σετ (που δεν είχαν χρησιμοποιηθεί ως τώρα) και για μετρική είχα το μέσο MSE προσαρμόζοντας το μοντέλο στις $n-9, n-6, n-3$ και προβλέποντας τις 3 επόμενες τιμές κάθε φορά. Για τα έτη 1989 και 2000 το επικρατέστερο ήταν το ARMA(1,2) ενώ για το 2011 το ARMA(3,3). Επειδή όμως οι τιμές του ARMA(1,3) ήταν πολύ κοντά στο καλύτερο επιλέχθηκε αυτό λόγω μικρότερης συνολικής τάξης. Τέλος, έκανα έλεγχο portmanteau στα υπόλοιπα του μοντέλου για να επιβεβαιώσω την καλή προσαρμογή. Τα αποτελέσματα φαίνονται παρακάτω και όντως τα υπόλοιπα δεν παρουσιάζουν κάποια γραμμική εξάρτηση.

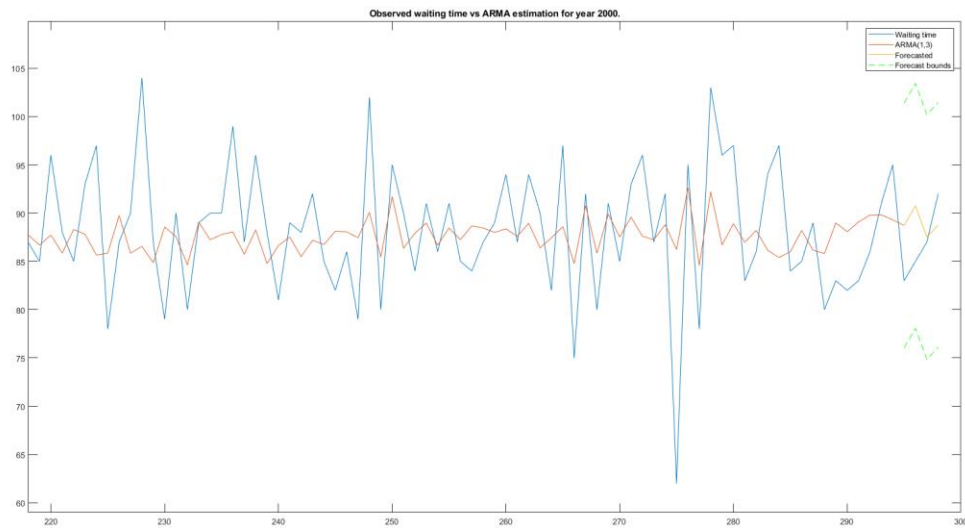


Εικόνα 10: Έλεγχος portmanteau στα υπόλοιπα των προσαρμοσμένων μοντέλων.

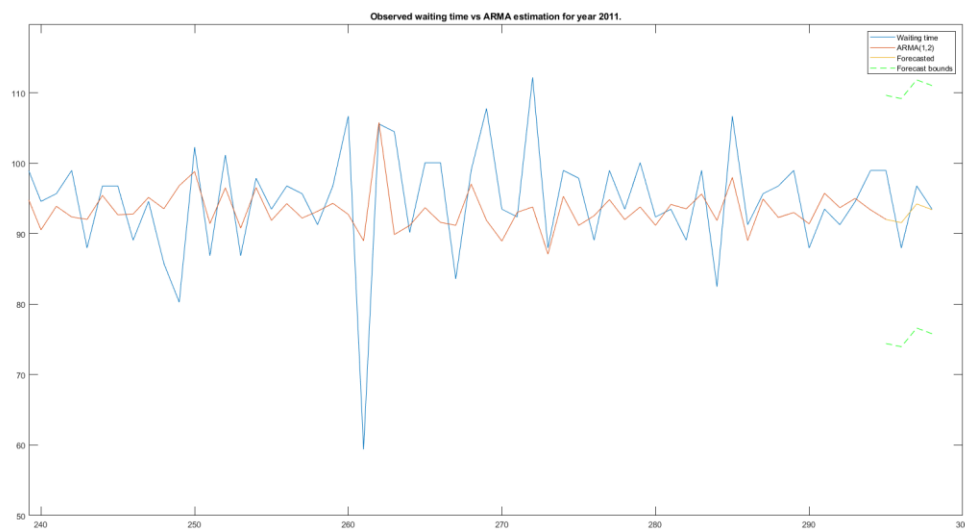
Συνοψίζοντας τα μοντέλα μοιάζουν μεταξύ τους στο πέρασμα των χρόνων, αλλά όπως φαίνεται και στα παρακάτω διαγράμματα (αν ίσως εξαιρέσουμε το 1989), δεν προσεγγίζουν πολύ καλά τις διακυμάνσεις των χρονοσειρών. Δεδομένης και της γραμμικής ανεξαρτησίας των υπολοίπων, αυτό ίσως υποδεικνύει κάποια ισχυρή μη γραμμική εξάρτηση που θα αναζητηθεί στο επόμενο στάδιο.



Εικόνα 11: Χρονοσειρά, προσαρμοσμένο γραμμικό μοντέλο $ARMA(1,2)$ και προβλέψεις για τις 3 τελευταίες τιμές με 95% όρια εμπιστοσύνης.(1989)



Εικόνα 12: Χρονοσειρά, προσαρμοσμένο γραμμικό μοντέλο $ARMA(1,2)$ και προβλέψεις για τις 3 τελευταίες τιμές με 95% όρια εμπιστοσύνης.(2000)



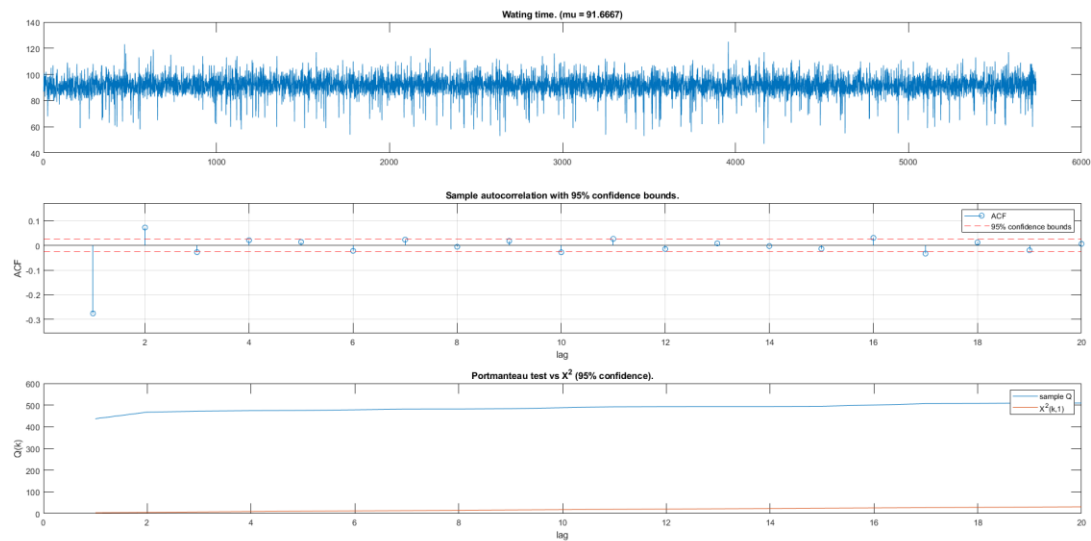
Εικόνα 13: Χρονοσειρά, προσαρμοσμένο γραμμικό μοντέλο $ARMA(1,3)$ και προβλέψεις για τις 3 τελευταίες τιμές με 95% όρια εμπιστοσύνης.(2011)

Ο κώδικας υλοποιήθηκε στο περιβάλλον της Matlab (έκδοση 2022b) και χρησιμοποιήθηκαν οι έτοιμες συναρτήσεις `parcorr` και `arima`. Αποτελείται από ένα κυρίως script και 6 βοηθητικές. Η λειτουργία τους εξηγείται στην αρχή κάθε αρχείου, ενώ στο κυρίως υπάρχουν επεξηγηματικά σχόλια.

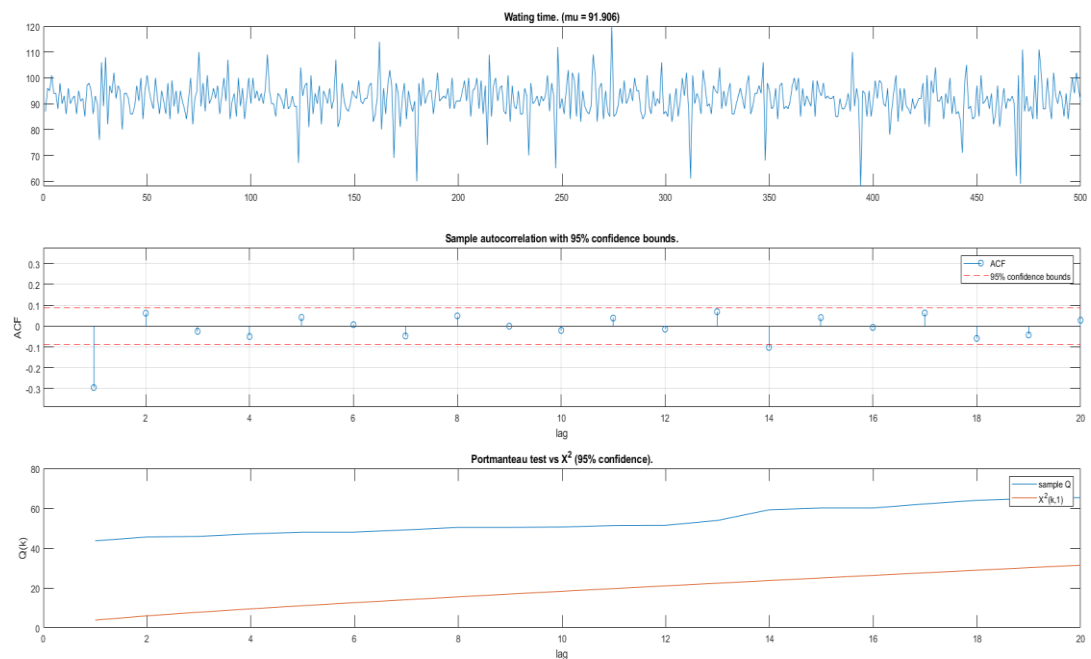
Στάδιο δεύτερο:

Ξεκινώντας την ανάλυση στο δεύτερο στάδιο. Ακολουθώ ακριβώς την ίδια συλλογιστική με πριν. Αρχικά υπολογίζω τα διαγράμματα των 2 χρονοσειρών (ολόκληρου του 2004 και μιας τυχαίας περιόδου 500 παρατηρήσεων). Αρχικά παρατηρώ ότι η μέση τιμή

των αναμονών είναι μεταξύ αυτής του 2000 και του 2011, κάτι που θα περιμέναμε από το πρώτο στάδιο. Έπειτα υπολογίζω τις αυτοσυσχετίσεις και διενεργώ έλεγχο portmanteau.



Εικόνα 24: Η χρονοσειρά, η αυτοσυσχέτιση της (για τις πρώτες 20 υστερήσεις), και έλεγχος portmanteau για την γραμμική ανεξαρτησία (2004-ολόκληρη)



Εικόνα 35: Η χρονοσειρά, η αυτοσυσχέτιση της (για τις πρώτες 20 υστερήσεις), και έλεγχος portmanteau για την γραμμική ανεξαρτησία (2004-500).

Παρατηρώ πως και οι δύο χρονοσειρές παρουσιάζουν σημαντικές αυτοσυσχετίσεις και όπως θα περιμέναμε δεν είναι λευκός θόρυβος.