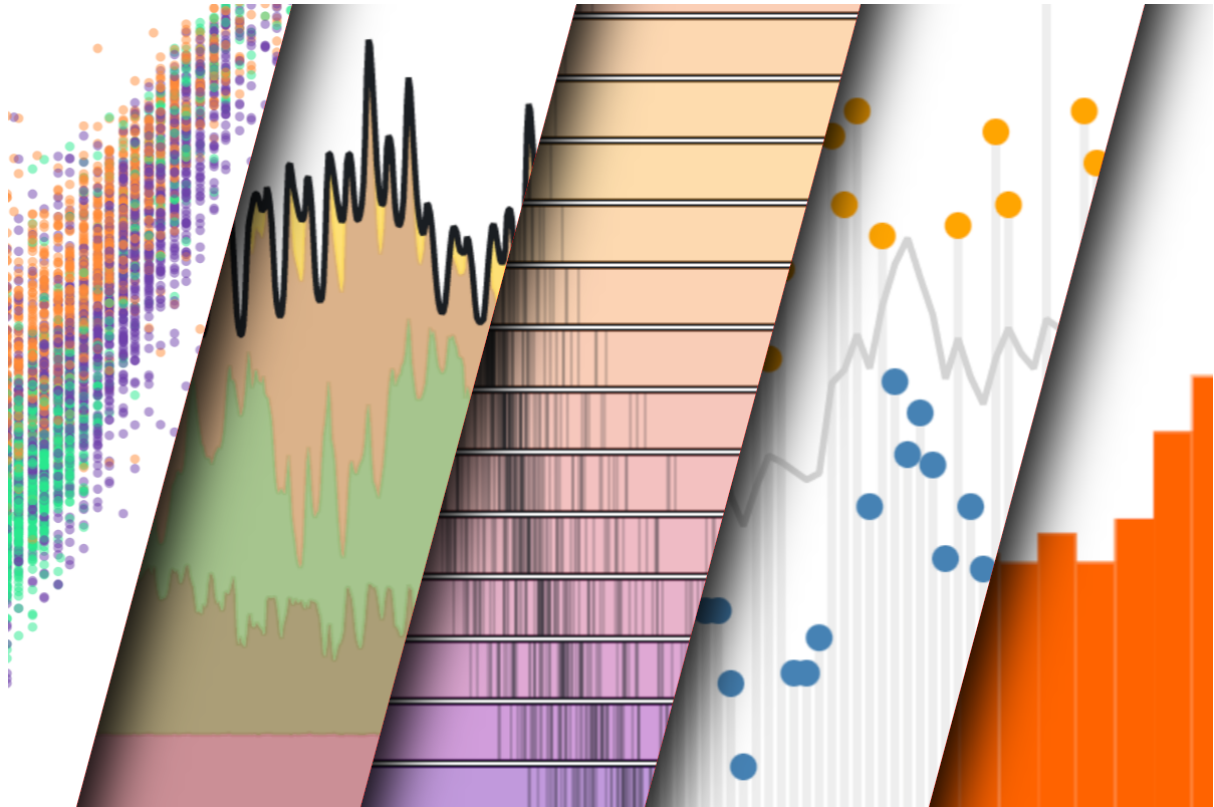


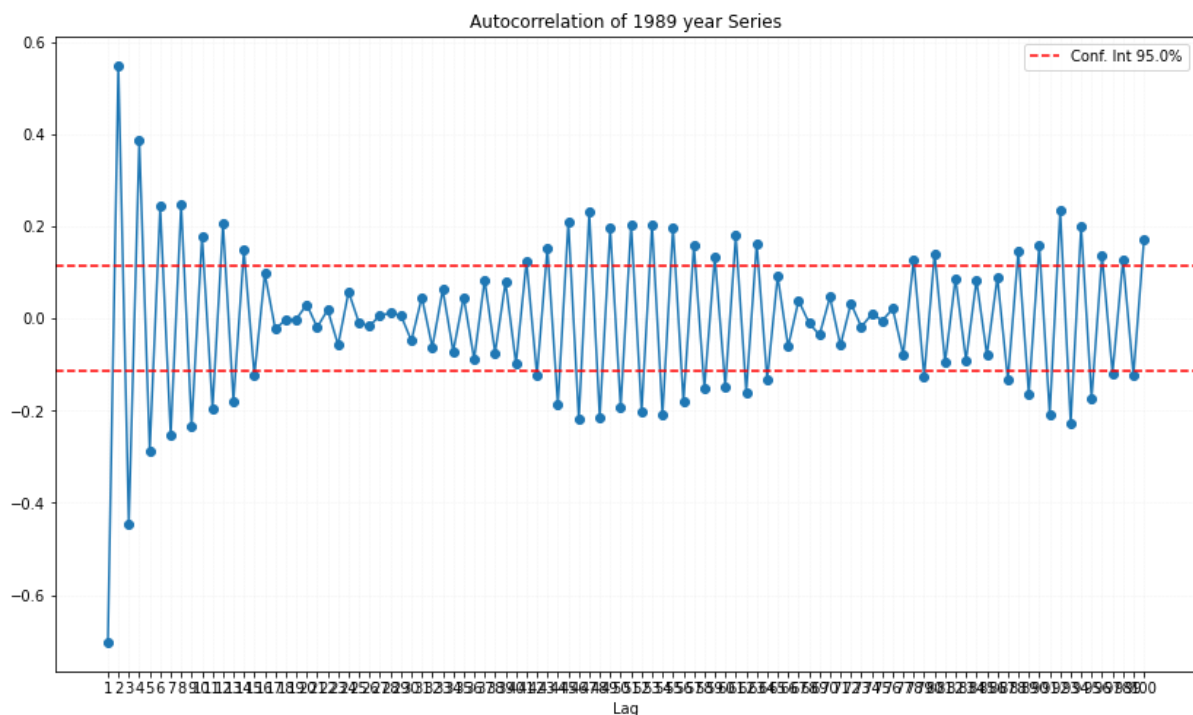
ΧΡΟΝΟΣΕΙΡΕΣ
ΥΠΟΛΟΓΙΣΤΙΚΗ ΕΡΓΑΣΙΑ
ΜΠΟΥΖΙΟΣ ΚΩΝΣΤΑΝΤΙΝΟΣ 8957



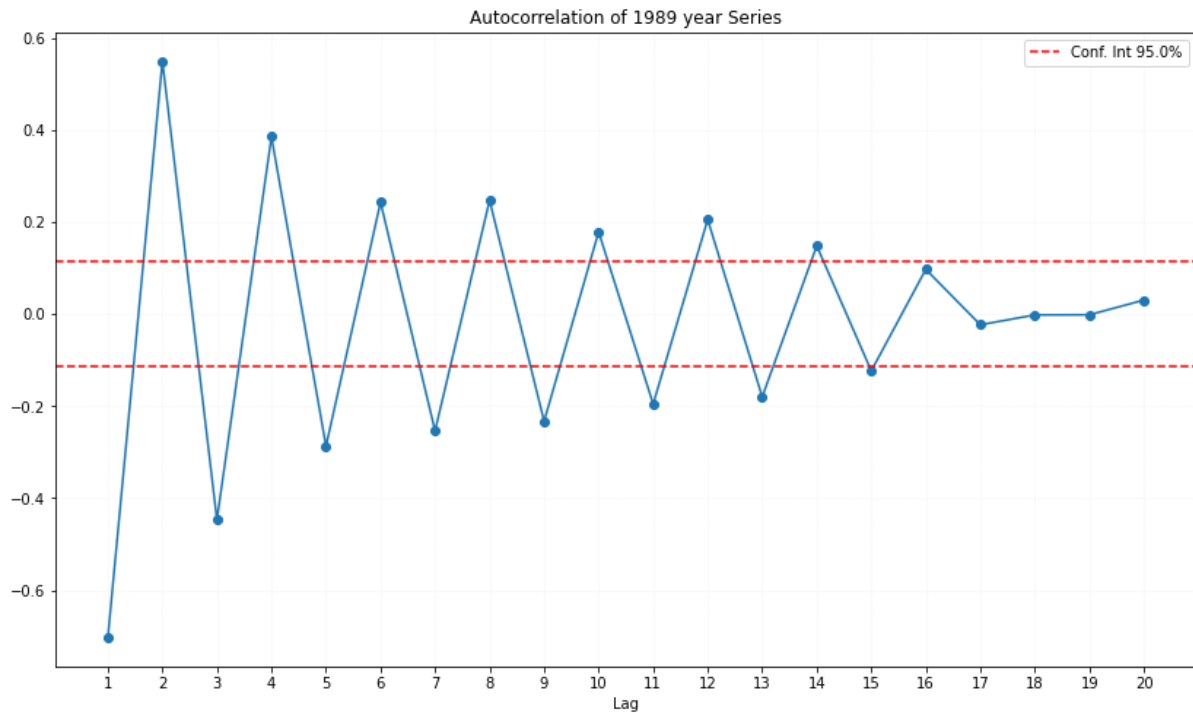
Πρώτο στάδιο - Γραμμική ανάλυση για τα έτη 1989, 2000 και 2011

Η χρονοσειρά που αντιστοιχεί στα δεδομένα του 1989 έχει 298 μετρήσεις. Οι άλλες δύο χρονοσειρές έχουν 739 δεδομένα. Όπως αναφέρεται στην εκφώνηση, γίνεται η τυχαία επιλογή 298 μετρήσεων από τις χρονοσειρές του 2000 και του 2011. Για λόγους απλότητας επιλέγω τις πρώτες 298 μετρήσεις.

Στο πρώτο ερώτημα ζητείται να εξαχθεί το συμπέρασμα αν η χρονοσειρά είναι λευκός θόρυβος ή διαφορετικά αν έχει σημαντικές αυτοσυσχετίσεις. Αυτό θα παρατηρηθεί από το διάγραμμα της αυτοσυσχέτισης (Διάγραμμα 1, Διάγραμμα 2).

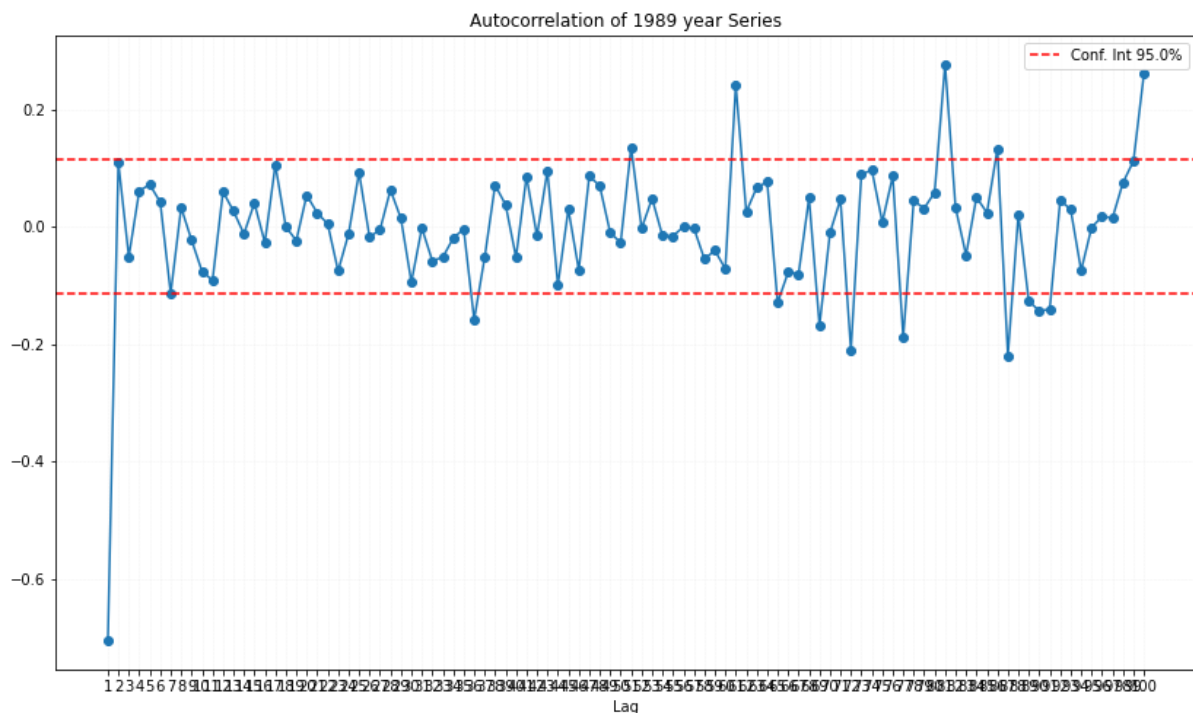


Διάγραμμα 1



Διάγραμμα 2

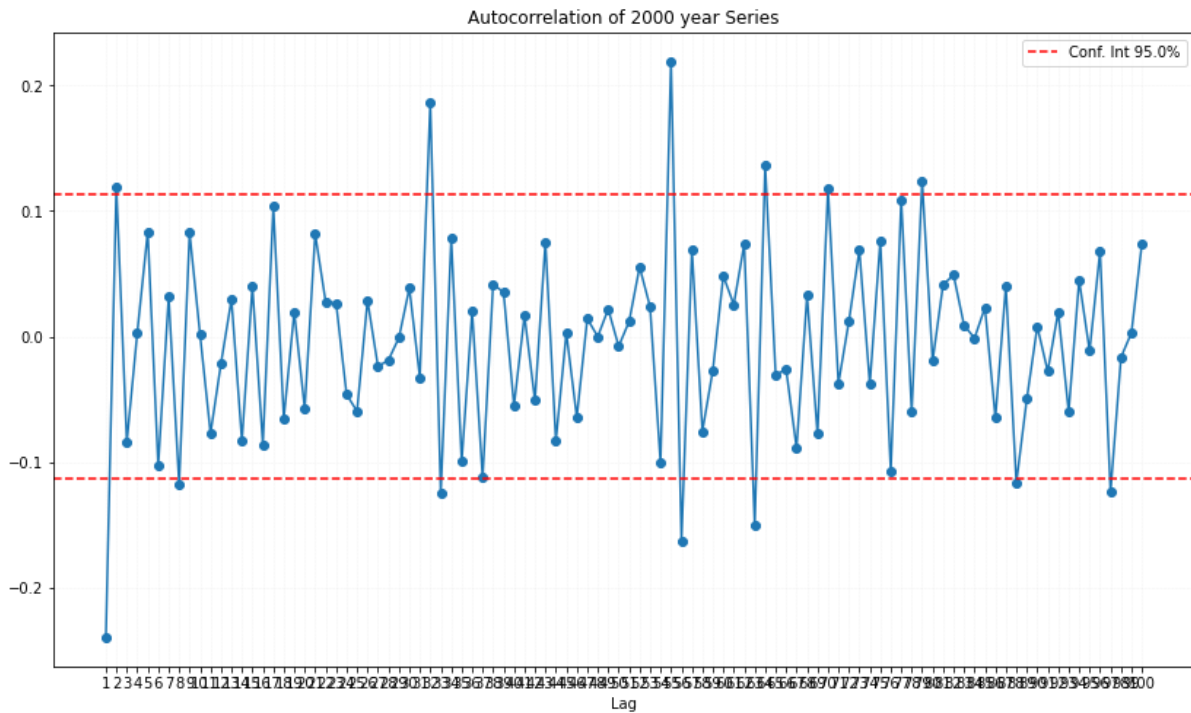
Τα δύο διαφορετικά διαγράμματα (Διάγραμμα 1, Διάγραμμα 2) είναι για διαφορετικό αριθμό lags, το πρώτο είναι για 100 και το δεύτερο για 20. Παρατηρούμε και στα δύο διαγράμματα υπάρχει σημαντική αυτοσυσχέτιση. Σε αυτό συμφωνεί και το διάγραμμα της μερικής αυτοσυσχέτισης(Διάγραμμα 3).



Διάγραμμα 3 (Partial Autocorrelation)

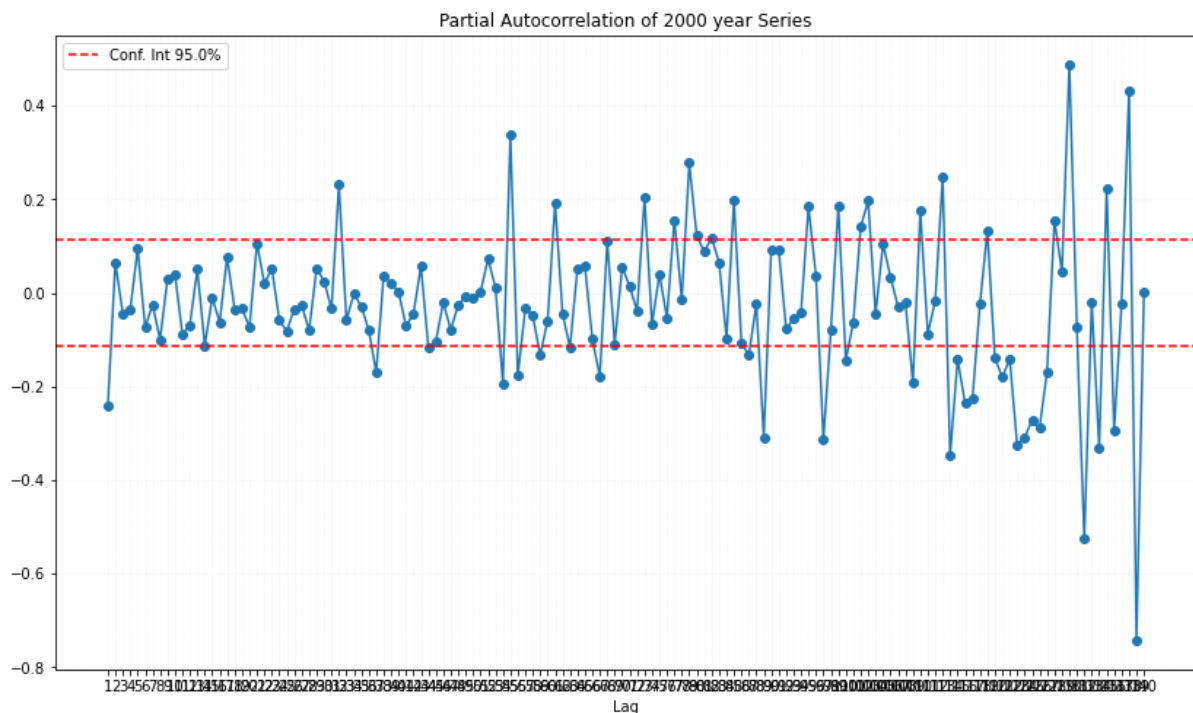
Στο διάγραμμα 3 της μερικής αυτοσυσχέτισης, παρατηρούμε ότι υπάρχει επίδραση των παλαιότερων παρατηρήσεων σε αρκετά μελλοντικές. Οπότε σύμφωνα με τα δύο διαγράμματα υπάρχει σημαντική αυτοσυσχέτιση και έτσι δεν είναι λευκός θόρυβος.

Ομοίως, στην μελέτη της χρονοσειράς των παρατηρήσεων του 2000 έγιναν τα διαγράμματα της αυτοσυσχέτισης και της μερικής αυτοσυσχέτισης ώστε να αποφανθούμε αν είναι λευκός θόρυβος ή όχι. Στο διάγραμμα 4 φαίνεται η αυτοσυσχέτιση για 100 lags.



Διάγραμμα 4

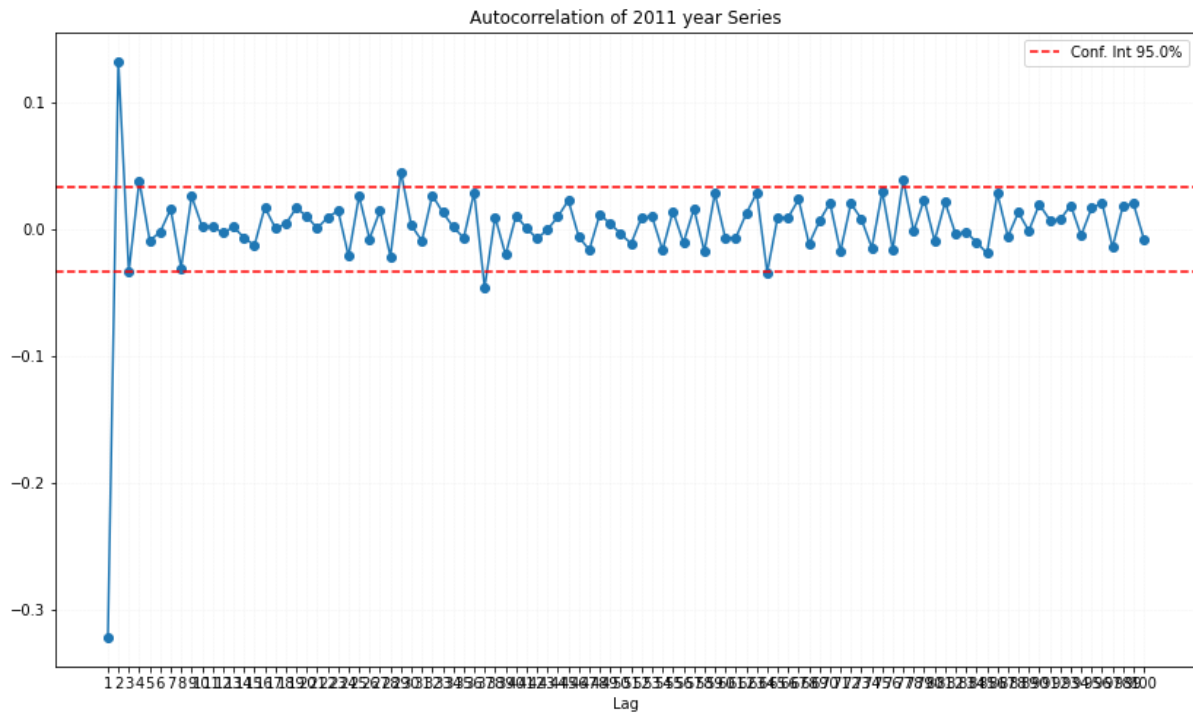
Στο διάγραμμα 4 παρατηρούμε ότι υπάρχει κάπως ισχυρή συσχέτιση μετά από 55 και 56 παρατηρήσεις, όπως επίσης και ότι φαίνεται και μετά από 32 παρατηρήσεις υπάρχει μια συσχέτιση. Βέβαια δεν γνωρίζουμε αν είναι κάτι τυχαίο, ή αν όντως υπάρχει κάτι που μετά από τόσες πολλές παρατηρήσεις συσχετίζεται. Για να πάρουμε κάποιο καλύτερο συμπέρασμα κάνουμε και το διάγραμμα της μερικής αυτοσυσχέτισης (Διάγραμμα 5).



Διάγραμμα 5

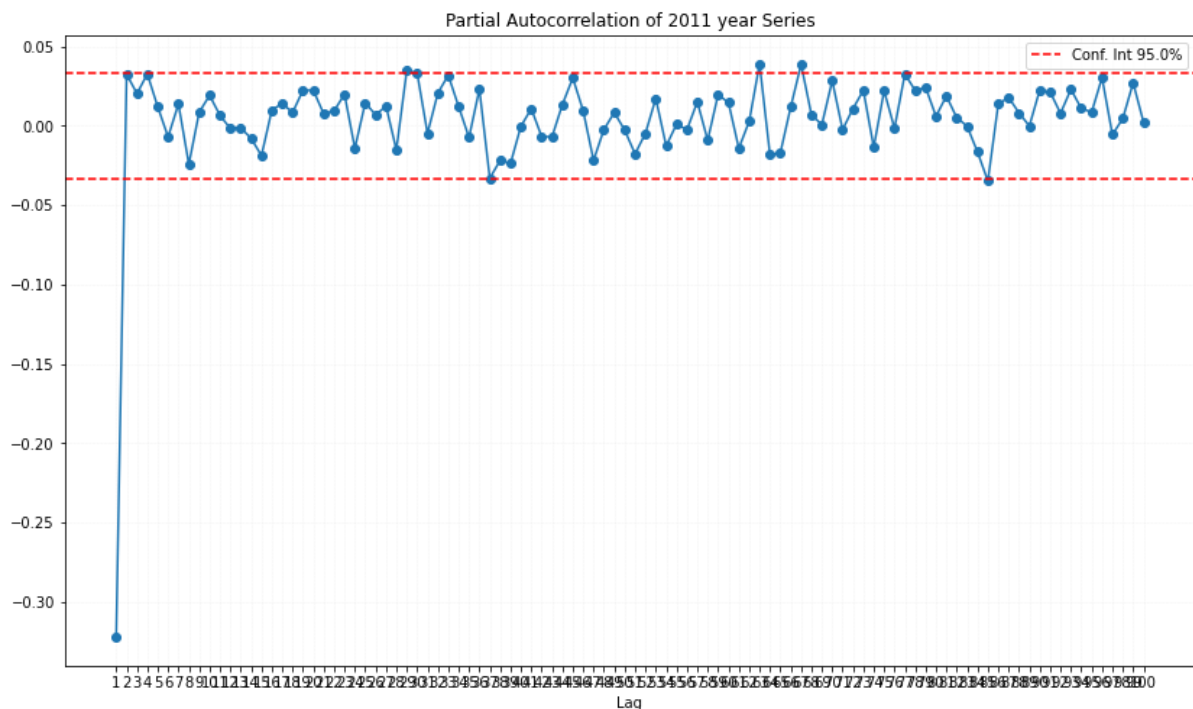
Σε αυτό το διάγραμμα 5 της μερικής αυτοσυσχέτισης, παρατηρούμε ότι αρκετά παλιές παρατηρήσεις έχουν επίδραση στο μέλλον. Αυτό ίσως είναι τυχαίο, ή όντως εκείνη την περίοδο με κάποιον τρόπο παρατηρήσεις αρκετά παλιές επιδρούσαν σε μελλοντικές. Κατά πάσα πιθανότητα, επειδή η επίδραση είναι πολύ μικρή, μπορεί να θεωρηθεί λευκός θόρυβος, αλλά θα γίνει η ανάλυση στο ερώτημα 2 για το καλύτερο μοντέλο ARMA.

Με παρόμοιο τρόπο κάνουμε και το διάγραμμα της αυτοσυσχέτισης της χρονοσειράς των μετρήσεων που πάρθηκαν το 2011. Είναι το διάγραμμα 6.



Διάγραμμα 6

Όπως φαίνεται στο Διάγραμμα 6 της αυτοσυσχέτισης, οι τιμές είναι μέσα στο διάστημα εμπιστοσύνης οπότε είναι πολύ πιθανό να είναι λευκός θόρυβος. Εξετάζουμε και το διάγραμμα της μερικής αυτοσυσχέτισης (Διάγραμμα 7).



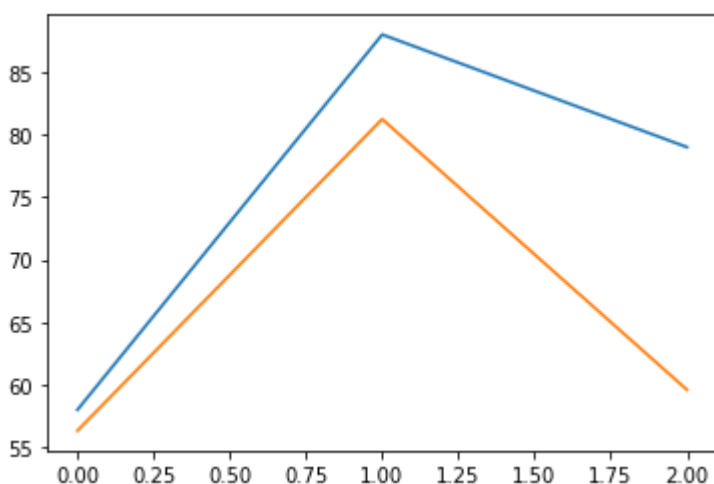
Διάγραμμα 7

Όπως φαίνεται κι από τα δύο διαγράμματα (6, 7) μπορούμε να αποφανθούμε ότι η χρονοσειρά των μετρήσεων του 2011 είναι λευκός θόρυβος.

Τα συμπεράσματά σχετικά με τις 3 χρονοσειρές είναι ότι διαφέρουν σε μεγάλο βαθμό. Στις παρατηρήσεις του 1989 υπάρχει μια πολύ ξεκάθαρη συσχέτιση μεταξύ των μετρήσεων, η οποία συσχέτιση μειώνεται κατά πολύ το 2000 και εξαφανίζεται σχεδόν εντελώς το 2011. Βέβαια επειδή το δείγμα είναι πολύ μικρό σε σχέση με όλες τις μετρήσεις από το 1989 μέχρι το 2011, δεν μπορούμε να οδηγηθούμε σε κάποιο ασφαλές συμπέρασμα. Αλλά υπάρχει η πιθανότητα να συνέβει κάτι το 1989 εκείνη την περίοδο ώστε να επηρεάστηκαν οι εξάρσεις με αυτόν τον τρόπο.

2.

Στην χρονοσειρά του 1989, που αποφανθήκαμε πριν ότι δεν είναι λευκός θόρυβος σίγουρα, ψάχνουμε το καλύτερο μοντέλο ARMA ώστε να γίνει όσο καλύτερη πρόβλεψη γίνεται. Αρχικά χωρίζουμε τις χρονοσειρές σε *train* και *test*. Στις παρατηρήσεις του *train* τοποθετούμε τις πρώτες 295 παρατηρήσεις, ενώ τις υπόλοιπες 3 παρατηρήσεις στο *test*. Βρίσκουμε λοιπόν, μέσω του χαμηλότερου AIC το καλύτερο μοντέλο με την βοηθητική συνάρτηση `auto_arima` της Python, το οποίο είναι το ARMA(1,1). Αρχικά, όπως λέει η εκφώνηση, κάνουμε πρόβλεψη για ένα χρονικό βήμα μπροστά, που είναι η 296 παρατήρηση. Η πραγματική τιμή της 296 παρατήρησης είναι 58. Η πρόβλεψη από το μοντέλο μας είναι η 56,33. Η πρόβλεψη για αυτό το χρονικό βήμα μπροστά είναι εξαιρετική. Κάνουμε το ίδιο και για 3 χρονικά βήματα μπροστά, το αποτέλεσμα είναι στο Διάγραμμα 8.



Διάγραμμα 8

Παρατηρούμε στο Διάγραμμα 8 ότι για τις 2 πρώτες προβλέψεις η πρόβλεψη είναι αρκετά ικανοποιητική, ενώ στο τρίτο χρονικό βήμα πρόβλεψης αρχίζει να χάνεται η ακρίβεια της πρόβλεψης. Το NRMSE για την πρόβλεψη δύο χρονικών βημάτων μπροστά είναι 0,328. Για τρία χρονικά βήματα μπροστά το NRMSE είναι 0,946. Φαίνεται ότι μετά από

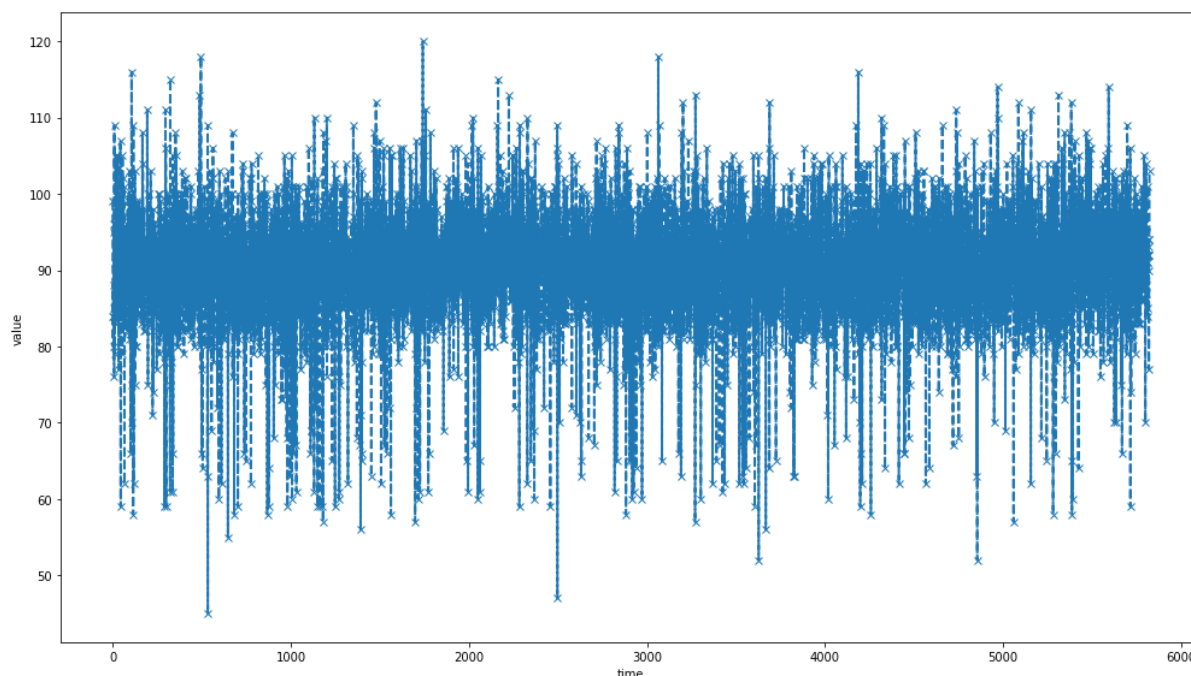
3 προβλέψεις το σφάλμα τριπλασιάζεται, ενώ στις πρώτες δύο προβλέψεις είναι αρκετά ικανοποιητικό.

Επίσης, έγινε μια δοκιμή και στις άλλες δύο χρονοσειρές, του 2000 και του 2011 ώστε να γίνει πρόβλεψη, αλλά η πρόβλεψη ήταν πολύ κακή, και έτσι επιβεβαιώνεται ότι είναι λευκός θόρυβος, τουλάχιστον για τις 298 μετρήσεις όπου έγινε η επιλογή των παρατηρήσεων.

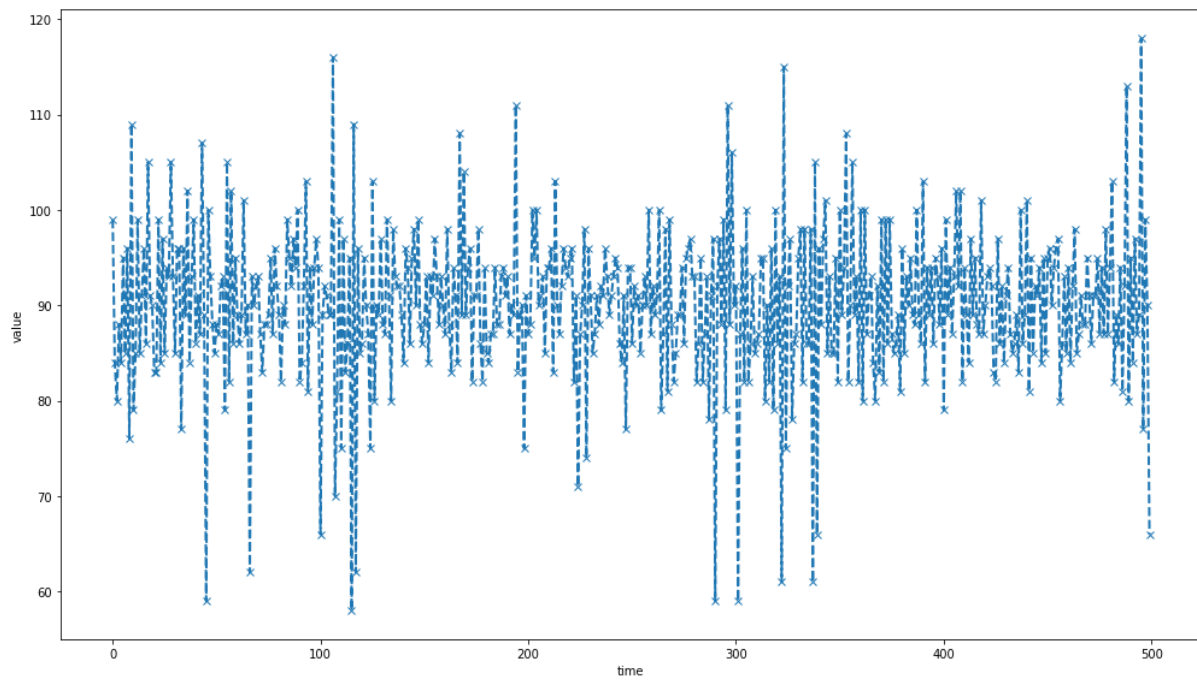
Δεύτερο στάδιο - Γραμμική και μη γραμμική ανάλυση

Η χρονοσειρά που αντιστοιχεί στο ΑΕΜ μου είναι οι παρατηρήσεις του 2007.

1. Αρχικά κάνουμε τα σχεδιαγράμματα των δύο χρονοσειρών. Στο σχεδιάγραμμα 9 είναι όλης της χρονοσειράς του 2007, ενώ στο σχεδιάγραμμα 10 είναι των πρώτων 500 παρατηρήσεων της χρονοσειράς του 2007.



Διάγραμμα 9



Διάγραμμα 10

2.

Το portmanteau test για όλες τις παρατηρήσεις του 2007 είναι το παρακάτω (Πίνακας 1)

	lb_stat	lb_pvalue
1	537.611312	6.238622e-119
2	631.316992	8.151864e-138
3	639.455078	2.815896e-138
4	643.993863	4.652729e-138
5	644.153385	5.810233e-137
6	644.225241	6.698050e-136
7	645.927541	3.113790e-135
8	646.035081	2.943278e-134
9	646.826487	1.847701e-133
10	648.114268	8.539397e-133
11	648.121255	7.035147e-132
12	649.482582	2.830566e-131
13	653.887912	2.442153e-131
14	657.389713	3.169848e-131
15	657.394582	2.209375e-130
16	657.473554	1.433038e-129
17	661.400944	1.371225e-129
18	663.921547	2.540990e-129
19	666.217220	5.120668e-129
20	666.221935	3.070016e-128
21	666.554729	1.528687e-127
22	667.002946	7.023367e-127
23	667.128070	3.687087e-126
24	669.626310	6.006605e-126
25	669.885075	2.833771e-125
26	671.703781	6.173989e-125
27	673.073947	1.640636e-124
28	673.747980	5.988527e-124
29	674.053408	2.563666e-123
30	674.735708	9.003165e-123

Πίνακας 1

Όσο και να αυξήσουμε το m στο portmanteau test, παίρνουμε παρόμοια αποτελέσματα, δηλαδή πολύ μικρότερα του 0,05. Έτσι μπορούμε να απορρίψουμε την υπόθεση της μηδενικής αυτοσυσχέτισης και να βγάλουμε το συμπέρασμα ότι δεν είναι ανεξάρτητα.

Παρόμοια κάνουμε και για τις 500 πρώτες μετρήσεις. Το portmanteau test φαίνεται στον πίνακα 2.

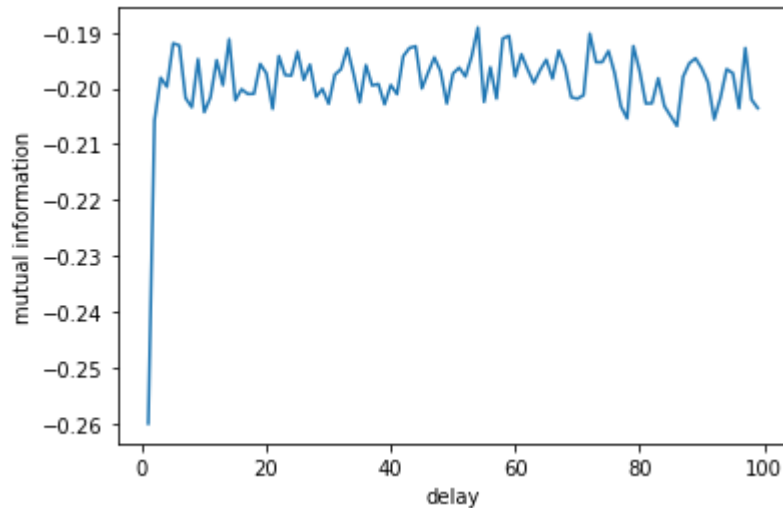
	lb_stat	lb_pvalue
1	71.930226	2.229427e-17
2	81.581825	1.926334e-18
3	81.608938	1.386374e-17
4	84.739334	1.722954e-17
5	87.601701	2.142428e-17
6	95.171458	2.546549e-18
7	99.315671	1.493605e-18
8	99.315747	5.890652e-18
9	100.160060	1.460482e-17
10	107.573693	1.644569e-18
11	115.288156	1.602025e-19
12	117.637396	1.826508e-19
13	120.025103	1.990513e-19
14	120.054593	6.139241e-19
15	120.873397	1.280803e-18
16	122.327718	1.957482e-18
17	128.427851	3.760431e-19
18	135.320972	5.085022e-20
19	137.509828	5.456586e-20
20	137.674577	1.395716e-19
21	137.698922	3.699429e-19
22	139.055732	5.401537e-19
23	139.347316	1.223201e-18
24	140.240163	2.104459e-18
25	140.450571	4.745428e-18
26	140.613837	1.070385e-17
27	141.303758	1.906390e-17
28	143.827185	1.576234e-17
29	143.871995	3.568781e-17
30	143.874009	8.078697e-17

Πίνακας 2

Συμπεραίνουμε ακριβώς το ίδιο όπως με τον Πίνακα 1. Έτσι μπορούμε να απορρίψουμε την υπόθεση της μηδενικής αυτοσυσχέτισης και να βγάλουμε το συμπέρασμα ότι δεν είναι ανεξάρτητα.

3.

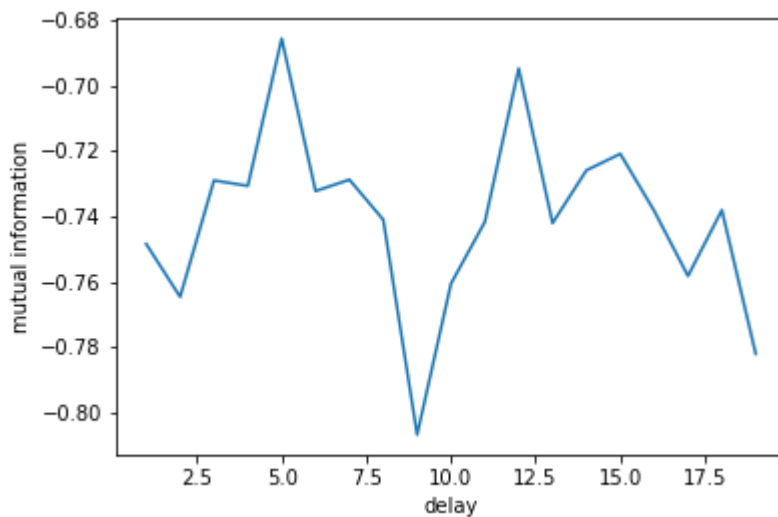
Από το κριτήριο της αμοιβαίας πληροφορίας, παίρνουμε το παρακάτω διάγραμμα 11.



Διάγραμμα 11

Όπως γνωρίζουμε η υστέρηση τ είναι το πρώτο τοπικό ελάχιστο της I (αμοιβαία πληροφορία). Οπότε για την χρονοσειρά που περιέχει όλες τις μετρήσεις $\tau=1$.

Για την χρονοσειρά που περιέχει τις 500 μετρήσεις έχουμε το διάγραμμα της αμοιβαίας πληροφορίας (Διάγραμμα 12).

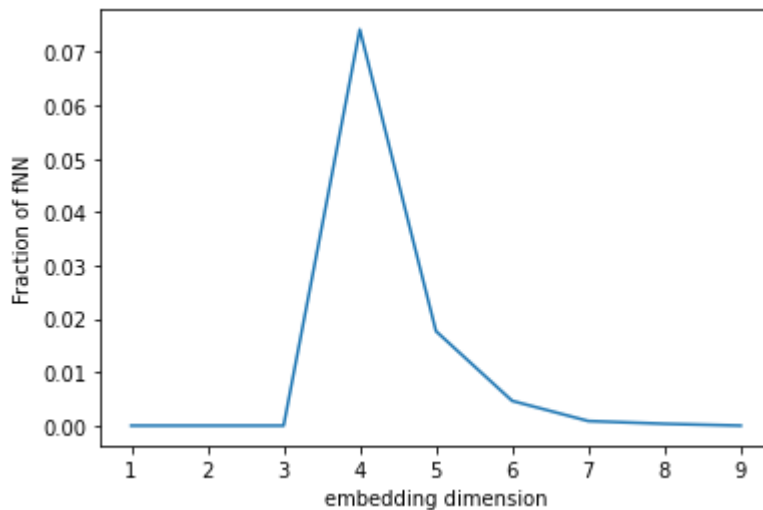


Διάγραμμα 12

Επειδή το πρώτο τοπικό ελάχιστο είναι η υστέρηση τ , παίρνουμε ως $\tau=2$.

4.

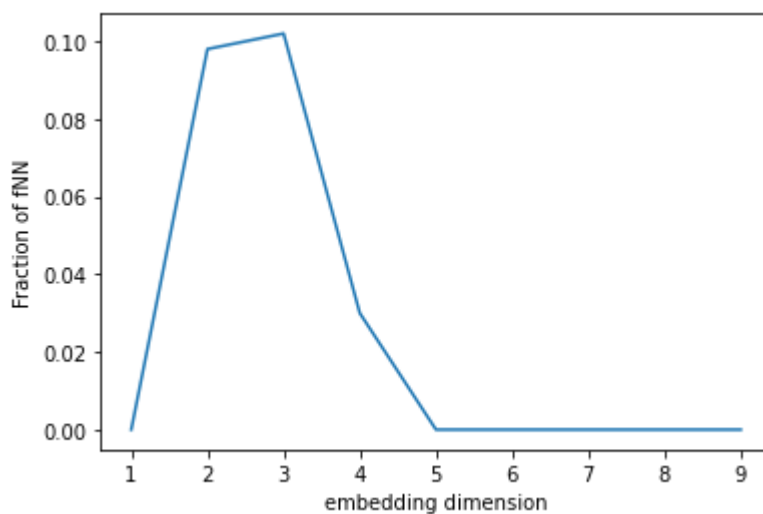
Εφαρμόζοντας το κριτήριο των ψευδών κοντινότερων γειτόνων για την χρονοσειρά που περιέχει όλες τις μετρήσεις του 2007 παίρνουμε το παρακάτω Διάγραμμα 13.



Διάγραμμα 13

Αυτό το διάγραμμα μας δείχνει ότι για διαστάσεις (1,2,3) και (7) μπορούμε να μοντελοποιήσουμε τα δεδομένα.

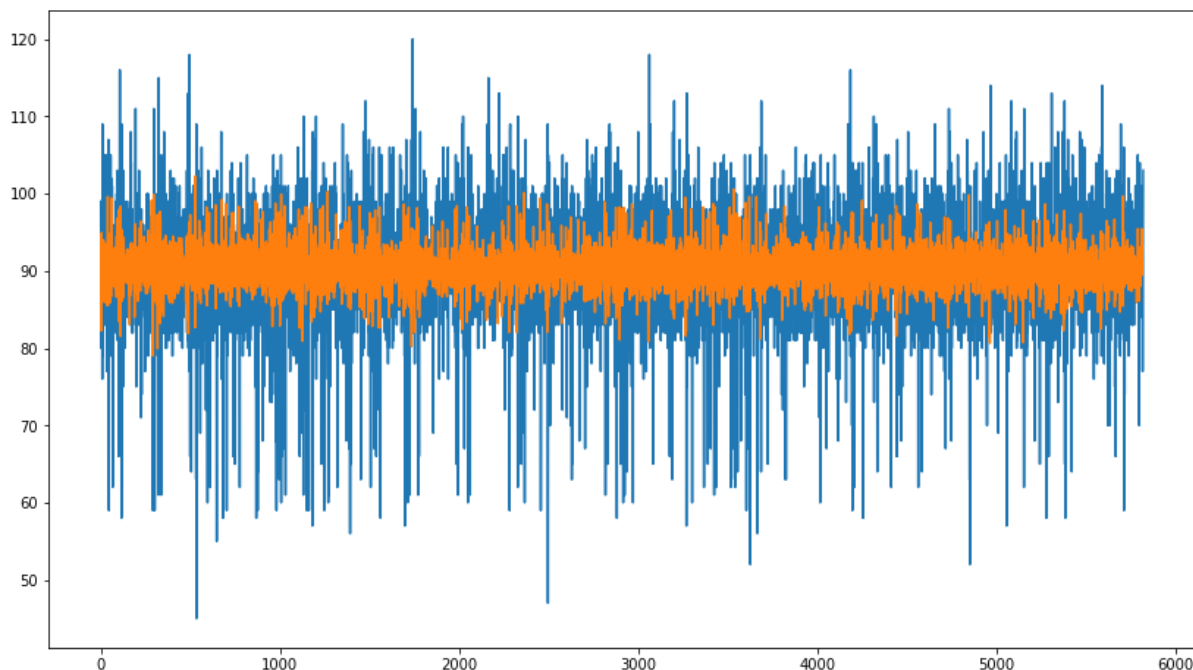
Κάνουμε το ίδιο και για την χρονοσειρά των 500 μετρήσεων και τα αποτελέσματα φαίνονται στο διάγραμμα 14.



Διάγραμμα 14

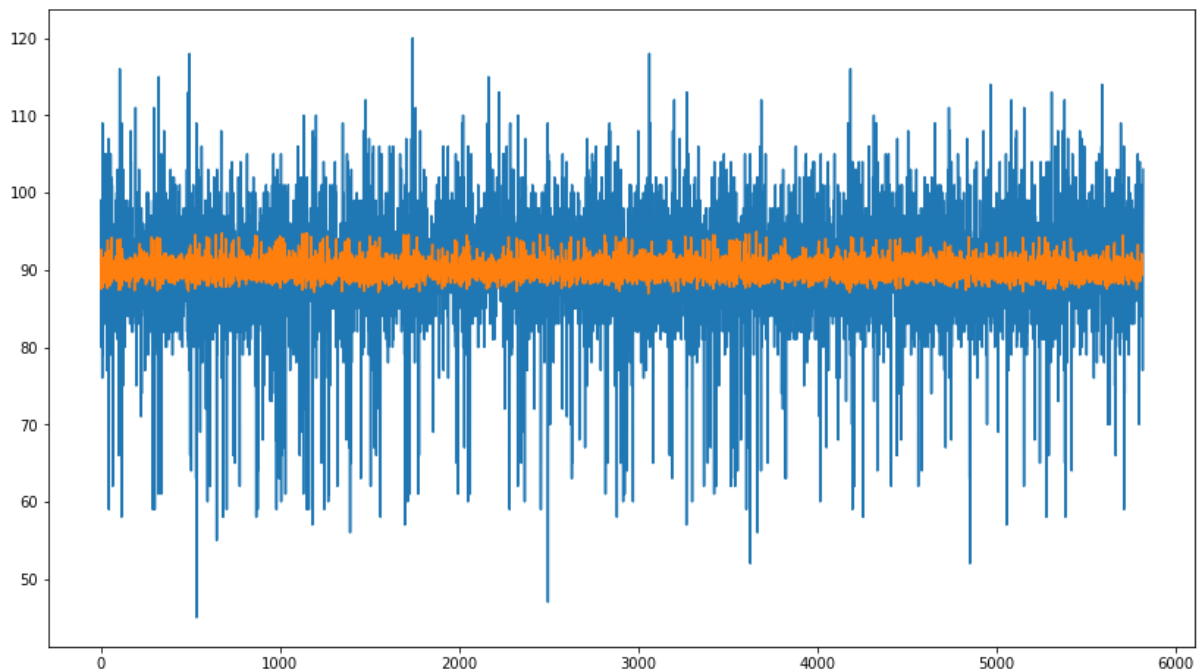
Από το Διάγραμμα 14 βγάζουμε ως συμπέρασμα ότι ο συντελεστής εμπύθισης μπορεί να είναι είτε 1 είτε 5.

5.



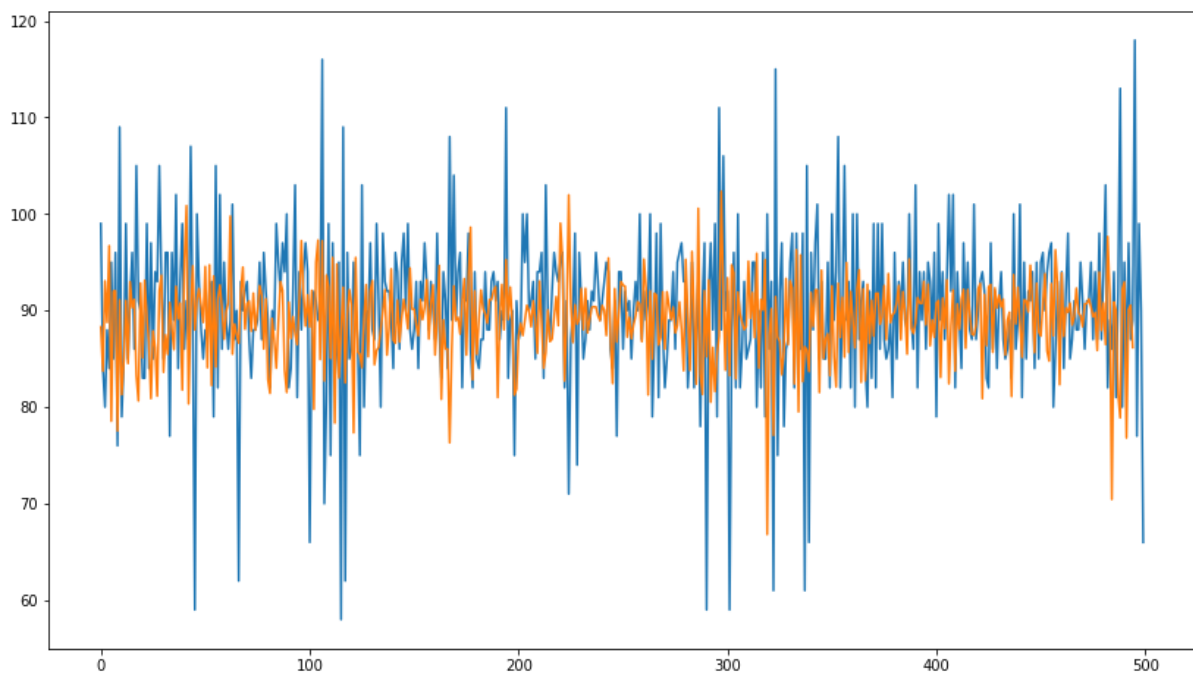
Διάγραμμα 15

Το Διάγραμμα 15 είναι το τοπικό γραμμικό μοντέλο και το Διάγραμμα 16 είναι το τοπικό μοντέλο μέσου όρου για την χρονοσειρά όλων των μετρήσεων του 2007. Η μπλε γραμμή είναι οι πραγματικές μετρήσεις ενώ η πορτοκαλί γραμμή οι fitted.

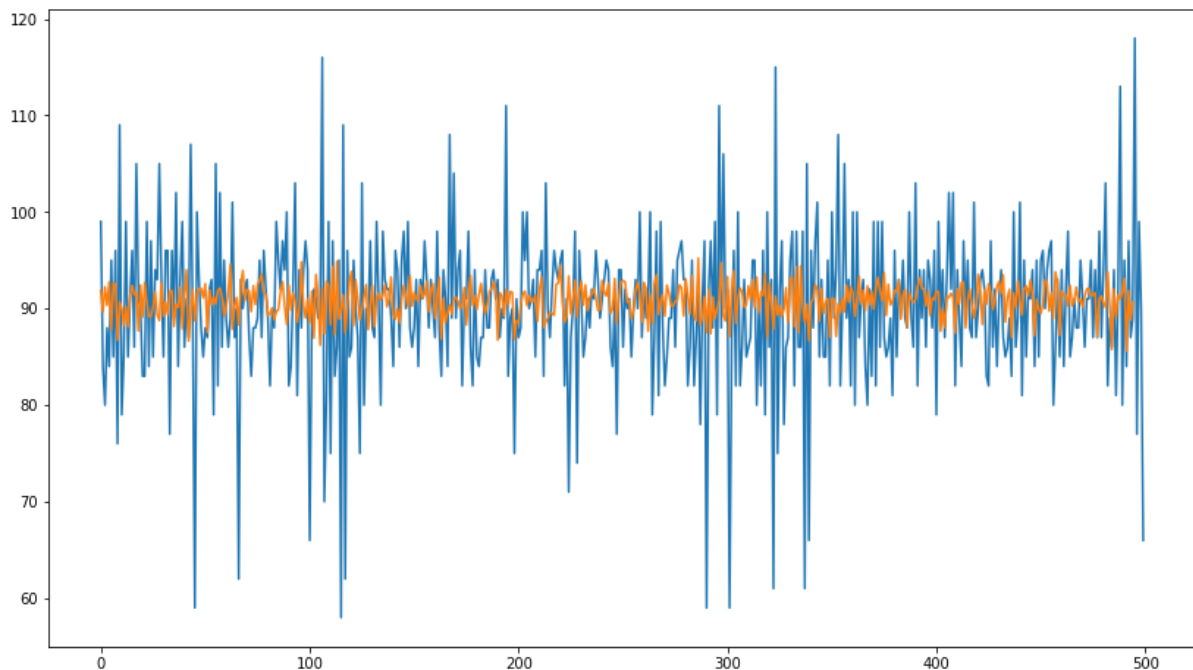


Διάγραμμα 16

Παρόμοια, τα διαγράμματα 17 και 18 είναι του τοπικού γραμμικού μοντέλου και του τοπικού μοντέλου μέσου όρου αντίστοιχα, των 500 μετρήσεων της χρονοσειράς του 2007. Προφάνως, η μπλε γραμμή είναι οι πραγματικές μετρήσεις ενώ η πορτοκαλί οι fitted.



Διάγραμμα 17



Διάγραμμα 18

Παρατηρούμε ότι και στις δύο περιπτώσεις, το τοπικό γραμμικό μοντέλο είναι λίγο καλύτερο.

6.

Εφαρμόζοντας το τοπικό γραμμικό μοντέλο για την χρονοσειρά όλων των μετρήσεων του 2007, βρίσκουμε σαν πρόβλεψη ενός χρονικού βήματος μπροστά την τιμή 90,17, ενώ η πραγματική τιμή είναι 103.

Κάνοντας το ίδιο με το τοπικό μοντέλο μέσου όρου βρίσκουμε 90,23, ενώ η πραγματική είναι 103.

Σε ό,τι αφορά την χρονοσειρά των 500 μετρήσεων, το τοπικό γραμμικό μοντέλο προβλέπει για ένα χρονικό βήμα μπροστά την τιμή 86 ενώ η πραγματική τιμή είναι η 66. Παρόμοια, το τοπικό μοντέλο μέσου όρου προβλέπει 90 ενώ η πραγματική τιμή είναι η 66.

Στο πρώτο κομμάτι της εργασίας, το γραμμικό μοντέλο που θεωρήσαμε τότε, για ένα χρονικό βήμα μπροστά έδινε πολύ καλό αποτέλεσμα που ήταν πολύ κοντά στην πραγματικότητα. Αυτό βέβαια μόνο για την

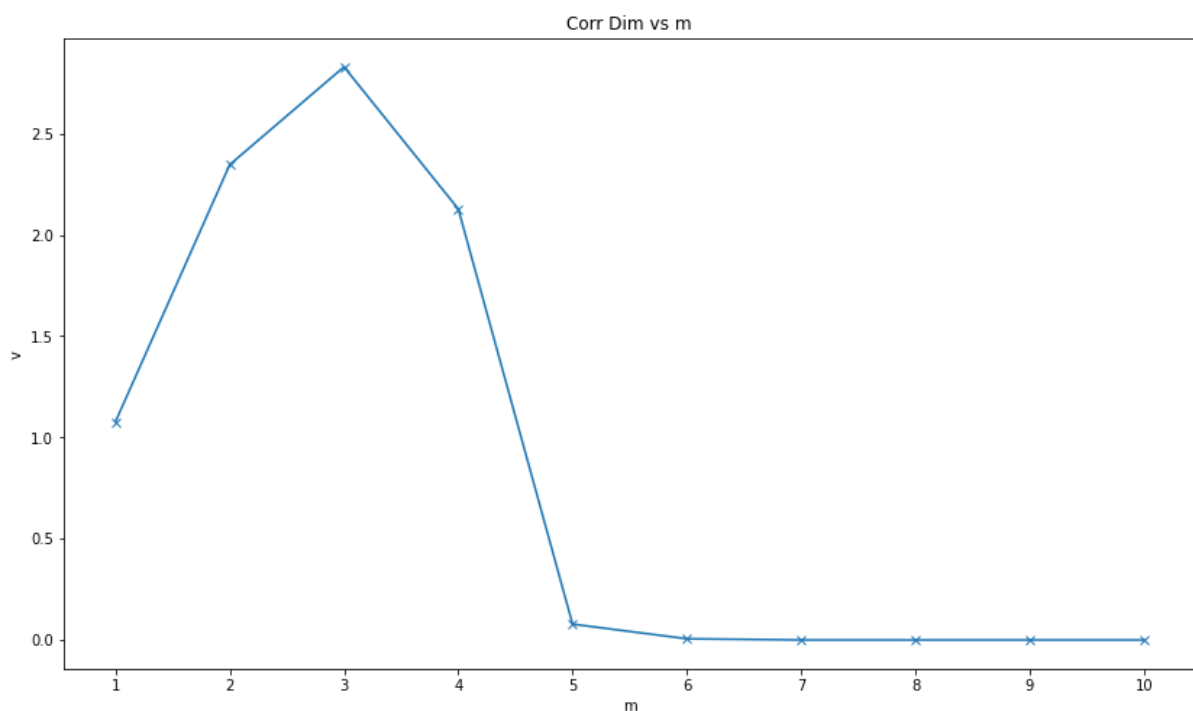
χρονοσειρά των μετρήσεων του 1989, για τις άλλες δύο χρονοσειρές τα αποτελέσματα ήταν τυχαία.

Φαίνεται ότι από την χρονοσειρά του 2007 οι μετρήσεις είναι τυχαίες, παρόμοιες με τις χρονοσειρές του 2000 και του 2011 και σε αντίθεση με ότι συνέβαινε το 1989.

Αν κάνουμε γραμμική ανάλυση για ολοκληρή την χρονοσειρά του 2007 βρίσκουμε το μοντέλο $ARMA(0,4)$, το οποίο για ένα χρονικό βήμα μπροστά προβλέπει 89, με την πραγματική τιμή να είναι 103. Η πρόβλεψη του είναι οριακά χειρότερη από αυτή των μη γραμμικών μοντέλων. Παρόμοια, για τις 500 μετρήσεις, το γραμμικό μοντέλο είναι το $ARMA(1,0)$ το οποίο προβλέπει τιμή 90 ενώ η πραγματική είναι 66. Πάλι οριακά το μη γραμμικό μοντέλο δίνει καλύτερη πρόβλεψη.

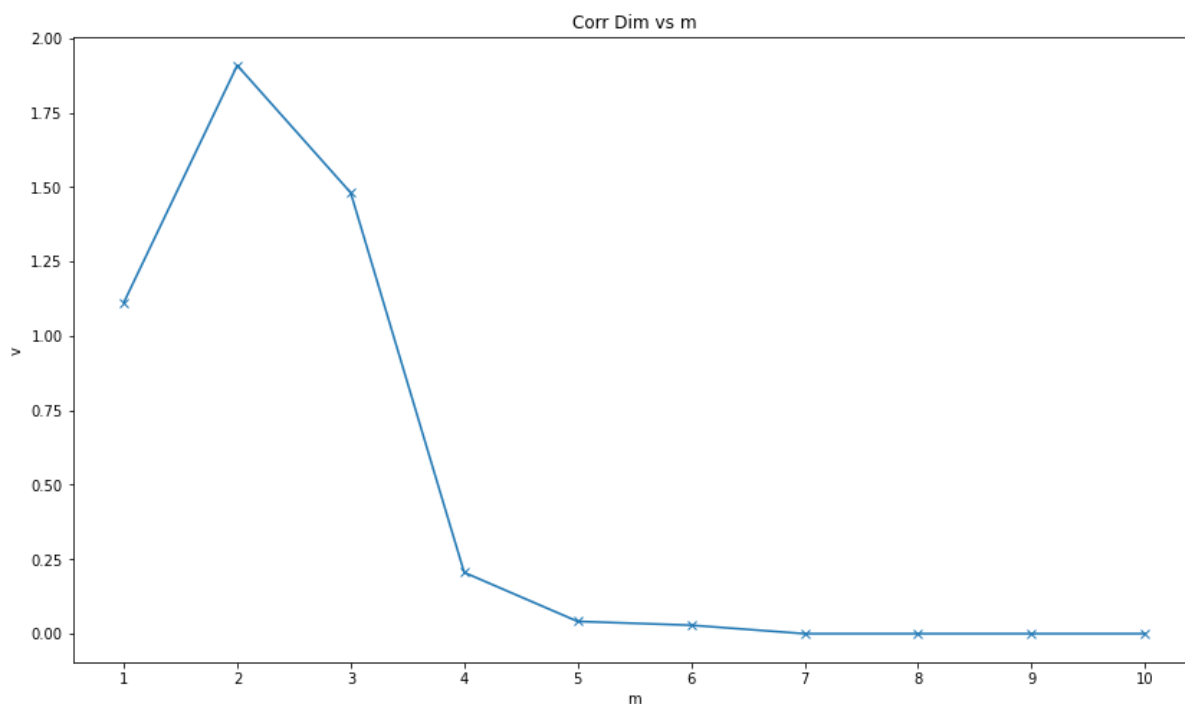
7.

Για την εκτίμηση της διάστασης συσχέτισης, κάνουμε το διάγραμμα συσχέτισης συναρτήσεως διάστασης. Αυτό φαίνεται στο παρακάτω Διάγραμμα 19.



Όπως φαίνεται, η διάσταση συσχέτισης είναι λίγο πάνω από το 0.

Στο Διάγραμμα 20 φαίνεται η διάσταση συσχέτισης για την χρονοσειρά των 500 μετρήσεων.



Διάγραμμα 20

Και σε αυτό το Διάγραμμα η διάσταση συσχέτισης είναι λίγο πάνω από το 0.

Από όλα αυτά τα μέτρα, φαίνεται ότι η χρονοσειρά του 2007 είναι πλήρως στοχαστική. Οι μόνες ομοιότητες είναι ότι οι τιμές συνήθως κυμαίνονται σε 80-90, χωρίς όμως αυτό να είναι νόμος.

Επίσης, φαίνεται ότι η χρονοσειρά 2007 είναι μη γραμμική.

Από το έτος 1989 οι παρατηρήσεις του 2007 διαφέρουν πάρα πολύ, ενώ από τα έτη 2000 και 2011 δεν διαφέρουν και τόσο. Ειδικά με το έτος 2000 έχει ομοιότητες.

