



Αριστοτέλειο Πανεπιστήμιο Θεσσαλονίκης
Πολυτεχνική Σχολή
Τμήμα Ηλεκτρολόγων Μηχανικών και Μηχανικών Υπολογιστών
Τομέας Ηλεκτρονικής και Υπολογιστών

ΧΡΟΝΟΣΕΙΡΕΣ (053)

Υπολογιστική Εργασία Ακαδημαϊκό έτος 2023/24

Πρόβλεψη και εκτίμηση του συστήματος που παράγει μια χρονοσειρά

Ομάδα 10

Κυπαρίσσης Κυπαρίσσης	10346	kyparkypar@ece.auth.gr
Αλεξανδρίδης Φώτιος	9953	faalexandr@ece.auth.gr

21 Μαΐου 2024

Περιεχόμενα

Περιεχόμενα.....	2
Κατάλογος Σχημάτων.....	3
Κατάλογος Πινάκων.....	3
Πρόλογος.....	4
Παραδοτέο Εργασίας.....	4
1 Βήματα Ανάλυσης.....	6
1.1 Βήμα 1° - Απαλοιφή της τάσης.....	6
1.1.1 Αφαιρώντας μια εκτιμώμενη συνάρτηση της τάσης.....	6
1.1.2 Πρώτες Διαφορές.....	8
1.2 Βήμα 2° - Διερεύνηση αν η χρονοσειρά είναι λευκός θόρυβος.....	8
1.2.1 Auto-correlation plots.....	9
1.2.2 Ljung-Box test.....	9
1.3 Βήμα 3° - Διερεύνηση του πιο κατάλληλου γραμμικού μοντέλου.....	10
1.4 Βήμα 4° - Εφαρμογή και μετρικές των γραμμικών μοντέλων.....	12
1.5 Βήμα 5° - Σύγκριση προβλέψεων και αξιολόγηση προσεγγίσεων.....	13
1.6 Βήμα 6° - Χρονοσειρές των υπολοίπων, από την προσαρμογή στις στάσιμες χρονοσειρές, των πιο κατάλληλων γραμμικών μοντέλων.....	15
1.7 Βήμα 7° - Έλεγχος αν η χρονοσειρά λευκού θορύβου είναι και iid.....	15
1.8 Βήμα 8° - Διερεύνηση του πιο κατάλληλου μη-γραμμικού μοντέλου για τη στάσιμη χρονοσειρά.....	17
1.9 Βήμα 9° - Εφαρμογή και μετρικές των μη-γραμμικών μοντέλων.....	18
1.10 Βήμα 10° - Σύγκριση προβλέψεων με τα δύο μη-γραμμικά μοντέλα από τις δύο προσεγγίσεις, σχολιασμός για τη δυνατότητα καλής πρόβλεψης της χρονοσειράς αυτής με μη-γραμμικά μοντέλα και σύγκριση με τις προβλέψεις με τα γραμμικά μοντέλα.....	18
1.11 Βήμα 11° - Σχολιασμός για το από ποιο σύστημα μπορεί να έχει προέλθει η χρονοσειρά.....	20
Βιβλιογραφία.....	21

Κατάλογος Σχημάτων

Σχήμα 1.0.1 - Χρονοσειρά που μελετάμε στην ομάδα 10.....	6
Σχήμα 1.1.1.1 - Χρονοσειρά που μελετάμε και η τάση μετά από χρήση φίλτρου κινούμενου μέσου όρου τάξης 5.....	7
Σχήμα 1.1.1.2 - Χρονοσειρά x με αφαίρεση μιας εκτιμώμενης συνάρτησης της τάσης (x_1).....	7
Σχήμα 1.1.2.1 - Χρονοσειρά x με απαλοιφή της τάσης μέσω πρώτων διαφορών (x_2).....	8
Σχήμα 1.2.1.1 - Auto-correlation plots των στάσιμων χρονοσειρών x_1, x_2	9
Σχήμα 1.2.2.1 - Ljung-Box test console output για τις χρονοσειρές x_1 και x_2	10
Σχήμα 1.3.1 - Partial Auto-correlation plots των στάσιμων χρονοσειρών x_1, x_2	11
Σχήμα 1.3.2 - Τιμές του Akaike Information Criterion (AIC) για το grid search του προσδιορισμού των βέλτιστων τιμών p και q του βέλτιστου ARMA μοντέλου των δύο προσεγγίσεων (x_1 και x_2).....	12
Σχήμα 1.3.3 - Console output της εύρεσης του πιο κατάλληλου γραμμικού μοντέλου για κάθε προσέγγιση (x_1 και x_2).....	12
Σχήμα 1.4.1 - NRMSE της πρόβλεψης με τα κατάλληλα γραμμικά μοντέλα για τα χρονικά βήματα μπροστά από 1 ως 5.....	13
Σχήμα 1.5.1 - Συγκρίση NRMSE των γραμμικών μοντέλων για τις προβλέψεις των δύο προσεγγίσεων για τα επόμενα 5 χρονικά βήματα.....	14
Σχήμα 1.6.1 - Χρονοσειρές των υπολοίπων (residuals) από την προσαρμογή στη στάσιμη χρονοσειρά των πιο κατάλληλων γραμμικών μοντέλων που βρήκαμε.....	15
Σχήμα 1.7.1 - Αμοιβαία πληροφορία των residuals των x_1 και x_2 για διάφορες τιμές χρονικής υστέρησης (lags).....	16
Σχήμα 1.7.2 - Console output του ελέγχου αν τα residuals των x_1 και x_2 είναι iid.....	16
Σχήμα 1.8.1 - Console output της εύρεσης του πιο κατάλληλου μη-γραμμικού μοντέλου για κάθε προσέγγιση (x_1 και x_2).....	18
Σχήμα 1.9.1 - NRMSE της πρόβλεψης με τα κατάλληλα μη-γραμμικά μοντέλα για τα χρονικά βήματα μπροστά από 1 ως 5.....	18
Σχήμα 1.10.1 - Σύγκριση προβλέψεων (NRMSE) με τα δύο μη-γραμμικά μοντέλα από τις δύο προσεγγίσεις για τα χρονικά βήματα μπροστά από 1 ως 5.....	19
Σχήμα 1.10.2 - Σύγκριση προβλέψεων (NRMSE) με τα δύο μη-γραμμικά μοντέλα και δύο γραμμικά μοντέλα από τις δύο προσεγγίσεις για τα χρονικά βήματα μπροστά από 1 ως 5.....	19

Κατάλογος Πινάκων

Πίνακας 1.5.1 - NRMSE των γραμμικών μοντέλων και των δύο προσεγγίσεων για τα επόμενα 5 χρονικά βήματα.....	14
---	----

Πρόλογος

Η εργασία αφορά ένα διαγωνισμό καλύτερης πρόβλεψης καθώς και καλύτερης εκτίμησης του συστήματος που παράγει τη χρονοσειρά. Οι χρονοσειρές προέρχονται από διαφορετικά συστήματα αλλά με τις ίδιες ιδιότητες (γραμμική στοχαστική διαδικασία ίδιου τύπου και τάξης ή/και μη-γραμμικό δυναμικό σύστημα ίδιας διάστασης και πολυπλοκότητας με επιπρόσθετη τάση).

Στο αρχείο `ContestData.dat` στην ιστοσελίδα του μαθήματος στο e-learning δόθηκε ένας πίνακας 600 x 50, δηλαδή 50 χρονοσειρές των 600 παρατηρήσεων η κάθε μια. Η ομάδα μας (Ομάδα 10) αναλύει παρακάτω την χρονοσειρά στην στήλη με τον αύξοντα αριθμό 10.

Η υλοποίηση της εργασίας έγινε σε  **MATLAB**.

Ως κύρια βιβλιογραφία για την υλοποίηση των MATLAB συναρτήσεων όσο και των σχολίων/συμπερασμάτων μας ήταν το υλικό του μαθήματος (τόσο οι σημειώσεις [2], όσο και οι κώδικες των εργαστηρίων).

Το  **GitHub Repository** της εργασίας αυτής βρίσκεται στο σύνδεσμο:

https://github.com/e-left/timeseries_analysis_project_2023

Παραδοτέο Εργασίας

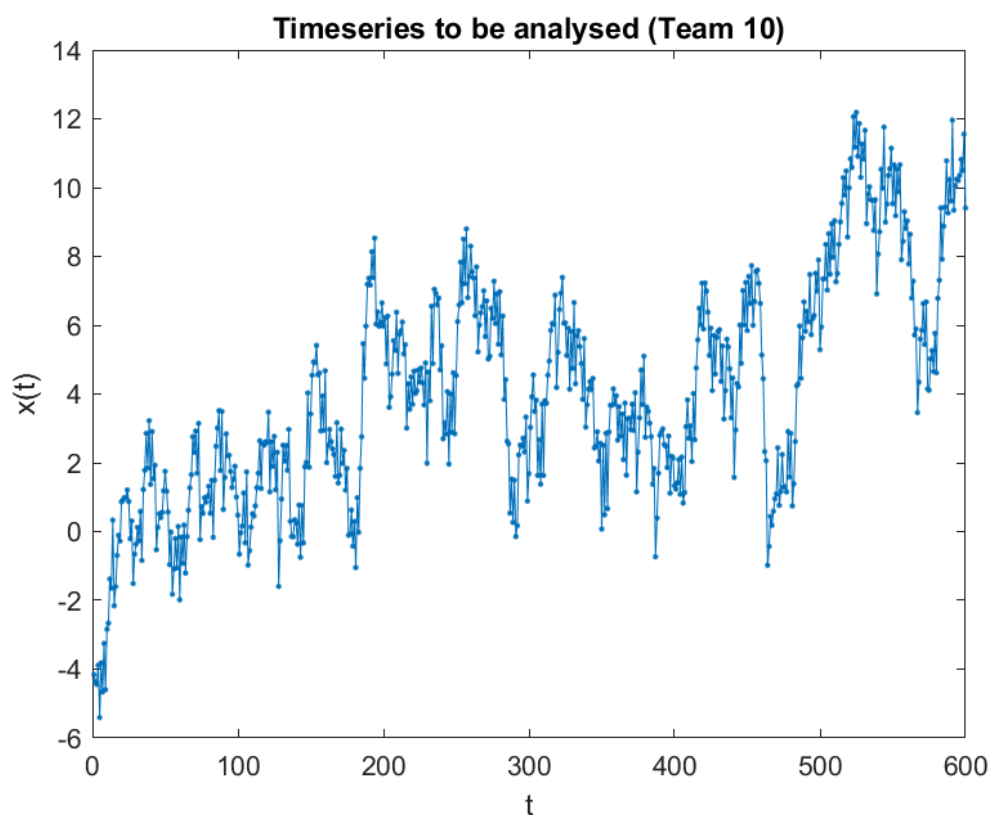
Παρακάτω αναφέρουμε τι αρχεία υπάρχουν στο παραδοτέο της εργασίας μας καθώς και πολύ σύντομα σχόλια για το τι είναι το καθένα (περισσότερα και αναλυτικότερα σχόλια βρίσκονται εντός των αρχείων κώδικα):

- **report.pdf** : Το αρχείο που περιέχει την αναφορά της εργασίας με τις αναλυτικές μας απαντήσεις στις ερωτήσεις της εκφώνησης.
- **main.m** : Το script που περιέχει τον κώδικα που εκτελεί όλα τα βήματα της εργασίας με την σειρά που ζητούνται και παράγει τα αποτελέσματα που παρουσιάζονται στην αναφορά (διαγράμματα, πίνακες, κ.λπ.).
- **timeseriesDetrending.m** : Η συνάρτηση που αφαιρεί την τάση από την χρονοσειρά με δύο τρόπους και χρησιμοποιεί και τις βοηθητικές συναρτήσεις `movingAverageSmoothing` και `movingAverageTrendEstimation` (για την 1η μέθοδο) και τις πρώτες διαφορές (για την 2η μέθοδο).
- **isTimeseriesWhiteNoise.m** : Η συνάρτηση που ελέγχει αν μια χρονοσειρά είναι λευκός θόρυβος.
- **findOptimalARMAModel.m** : Η συνάρτηση που βρίσκει το βέλτιστο μοντέλο ARMA(p, q) για μια χρονοσειρά.
- **predictARMAnrmse.m** : Η συνάρτηση που προβλέπει τις τιμές μιας χρονοσειράς με το μοντέλο ARMA(p, q) και υπολογίζει το NRMSE.

- **isTimeseriesIID.m** : Η συνάρτηση που ελέγχει αν μια χρονοσειρά είναι ανεξάρτητη και ομοιομορφα κατανεμημένη (iid).
- **findOptimalNonLinearModel.m** : Η συνάρτηση που βρίσκει το βέλτιστο μοντέλο μη γραμμικής πρόβλεψης για μια χρονοσειρά.
- **localpredictnrmse.m** : Η συνάρτηση που προβλέπει τις τιμές μιας χρονοσειράς με το μοντέλο μη γραμμικής πρόβλεψης και υπολογίζει το NRMSE.
- **localfitnrmse.m** : Η συνάρτηση που εκπαιδεύει το μοντέλο μη γραμμικής πρόβλεψης για μια χρονοσειρά.
- **calculateNRMSEForARMAWithTrend.m** : Η συνάρτηση που υπολογίζει το NRMSE για τις προβλέψεις του μοντέλου ARMA(p, q) με τάση.
- **calculateNRMSEforNonLinearWithTrend.m** : Η συνάρτηση που υπολογίζει το NRMSE για τις προβλέψεις του μοντέλου μη γραμμικής πρόβλεψης με τάση.
- **loadTimeseriesFromDataset.m** : Η συνάρτηση που φορτώνει την χρονοσειρά της ομάδας μας από το dataset.
- **nrmse.m** : Η συνάρτηση που υπολογίζει το NRMSE μεταξύ δύο διανυσμάτων.
- **myversion.m** : Η συνάρτηση που επιστρέφει την έκδοση του MATLAB που χρησιμοποιείται.
- **fitARMA.m** : Η συνάρτηση που εκπαιδεύει ένα μοντέλο ARMA(p, q) σε μια χρονοσειρά.
- **gridSearch_AIC.m** : Η συνάρτηση που υλοποιεί την αναζήτηση πλέγματος για την εύρεση του βέλτιστου μοντέλου ARMA(p, q) με βάση το κριτήριο AIC.
- **timeseriesPredTrending.m** : Η συνάρτηση που προσθέτει την τάση στις προβλέψεις του μοντέλου ARMA(p, q) με δύο τρόπους.
- **timeseriesPredTrendingNonLinear.m** : Η συνάρτηση που προσθέτει την τάση στις προβλέψεις του μοντέλου μη γραμμικής πρόβλεψης.
- **movingAverageSmoothing.m** : Η συνάρτηση που εφαρμόζει τον αλγόριθμο κινούμενου μέσου όρου σε μια χρονοσειρά.
- **movingAverageTrendEstimation.m** : Η συνάρτηση που εκτιμά την τάση μιας χρονοσειράς με τον αλγόριθμο κινούμενου μέσου όρου.

1 Βήματα Ανάλυσης

Πριν ξεκινήσουμε την ανάλυση μας σύμφωνα με τα βήματα της εκφώνησης της εργασίας, κάνουμε μια οπτικοποίηση της χρονοσειράς που έχουμε να μελετήσουμε για να πάρουμε μια πρώτη εικόνα για αυτή.



Σχήμα 1.0.1 - Χρονοσειρά που μελετάμε στην ομάδα 10

Την χρονοσειρά αυτή που θα μελετήσουμε την ονομάζουμε x .

1.1 Βήμα 1^ο - Απαλοιφή της τάσης

Εδώ κάνουμε απαλοιφή της τάσης (trend) από τη χρονοσειρά μας x με δύο τρόπους.

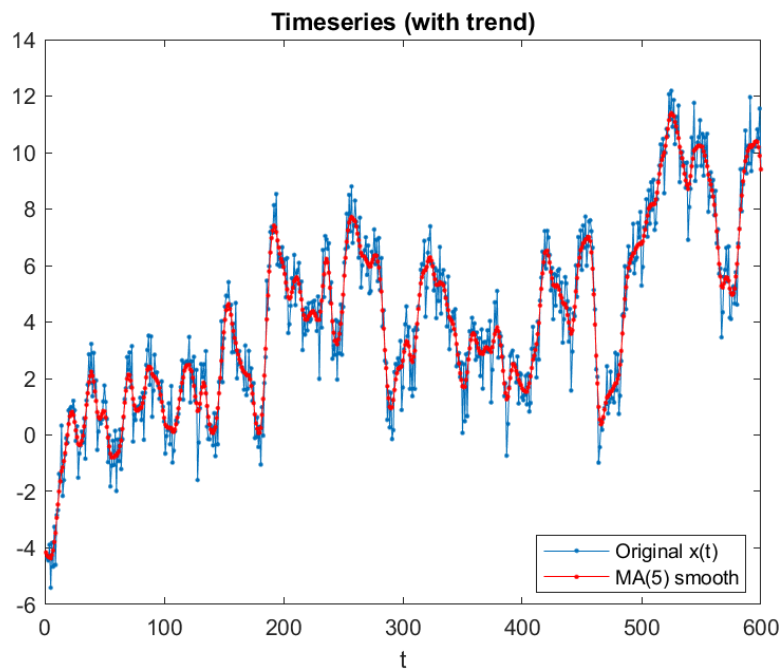
1.1.1 Αφαιρώντας μια εκτιμώμενη συνάρτηση της τάσης

Εδώ επιλέγουμε να κάνουμε απαλοιφή της τάσης είναι με τη **χρήση φίλτρου κινούμενου μέσου όρου τάξης 5**.

Επιλέξαμε την τιμή αυτή **διότι** από το σχήμα 1.0.1 είδαμε πως έχουμε να κάνουμε με μεταβολές σε μικρή χρονική κλίμακα. Επομένως έπρεπε να είναι και η τάξη μικρή, σύμφωνα με τις σημειώσεις του μαθήματος^[2].

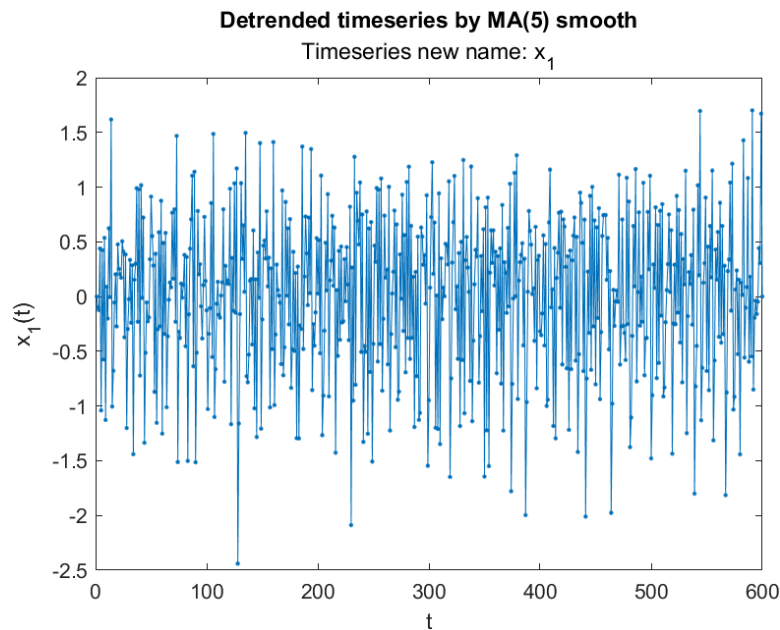
Η επιλογή της μικρής τάξης (ίση με 5) γίνεται για την απαλοιφή μεταβολών σε μικρότερη χρονική κλίμακα, όπως φαίνεται πως έχουμε και από το σχήμα 1.0.1. Άρα η τάξη πρέπει να είναι και αυτή αντίστοιχα μικρή.

Στο σχήμα 1.1.1.1 παρακάτω βλέπουμε, στο ίδιο σχήμα, την χρονοσειρά μας αλλά και την smoothed χρονοσειρά σύμφωνα με το φίλτρο MA(5)



Σχήμα 1.1.1.1 - Χρονοσειρά που μελετάμε και η τάση μετά από χρήση φίλτρου κινούμενου μέσου όρου τάξης 5

Αφαιρώντας, τώρα, από την χρονοσειρά μας x την χρονοσειρά που προέκυψε από την εφαρμογή του φίλτρου MA(5) παίρνουμε παρακάτω την ζητούμενη απαλλαγμένη από τάσεις (trends).



Σχήμα 1.1.1.2 - Χρονοσειρά x με αφαίρεση μιας εκτιμώμενης συνάρτησης της τάσης (x_1)

Εδώ, κοινή πρακτική είναι οι πρώτες και οι τελευταίες παρατηρήσεις να παραλείπονται ή θέτονται ίσες με τις αρχικές. Όμως, εδώ έγινε η χρήση της MATLAB συνάρτησης `filtfilt()` (όπως είδαμε και από τις έτοιμες MATLAB συναρτήσεις του εργαστηρίου) που διαχειρίζεται τις ακραίες αυτές τιμές (είναι $\text{floor}\left(\frac{5}{2}\right) = 2$

τιμές από “πάνω” και άλλες τόσες από κάτω). Έτσι, παίρνουμε μια χρονοσειρά πάλι μήκους 600.

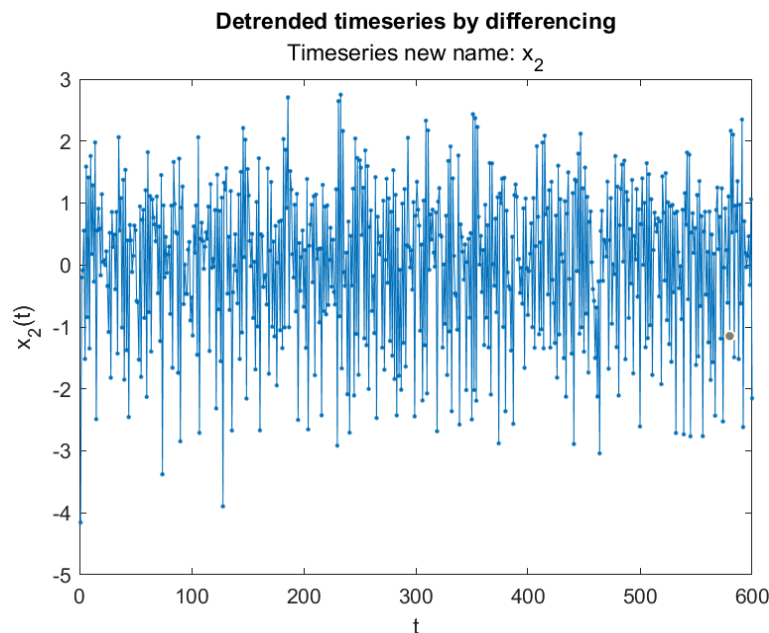
Την στάσιμη χρονοσειρά αυτή την ονομάζουμε x_1 .

1.1.2 Πρώτες Διαφορές

Εδώ ουσιαστικά η νέα χρονοσειρά (με απαλοιφή της τάσης) θα προκύψει παίρνοντας τις πρώτες διαφορές, δηλαδή από τη σχέση

$$\begin{aligned}x_2(i) &= x(i) - x(i-1), \quad i = 2, 3, \dots, 600 \\x_2(1) &= x(1)\end{aligned}\tag{1.1.2.1}$$

παίρνουμε την νέα στάσιμη χρονοσειρά:



Σχήμα 1.1.2.1 - Χρονοσειρά x με απαλοιφή της τάσης μέσω πρώτων διαφορών (x_2)

Για να έχουμε μήκος ίσο με 600 στη νέα στάσιμη χρονοσειρά, κρατήσαμε, στη νέα χρονοσειρά, ως πρώτη τιμή αυτή της αρχικής χρονοσειράς x .

Την στάσιμη χρονοσειρά αυτή την ονομάζουμε x_2 .

Η ανάλυση στη συνέχεια γίνεται ξεχωριστά για κάθε μια από τις δύο στάσιμες χρονοσειρές (x_1 και x_2) που προέκυψαν!

1.2 Βήμα 2° - Διερεύνηση αν η χρονοσειρά είναι λευκός θόρυβος

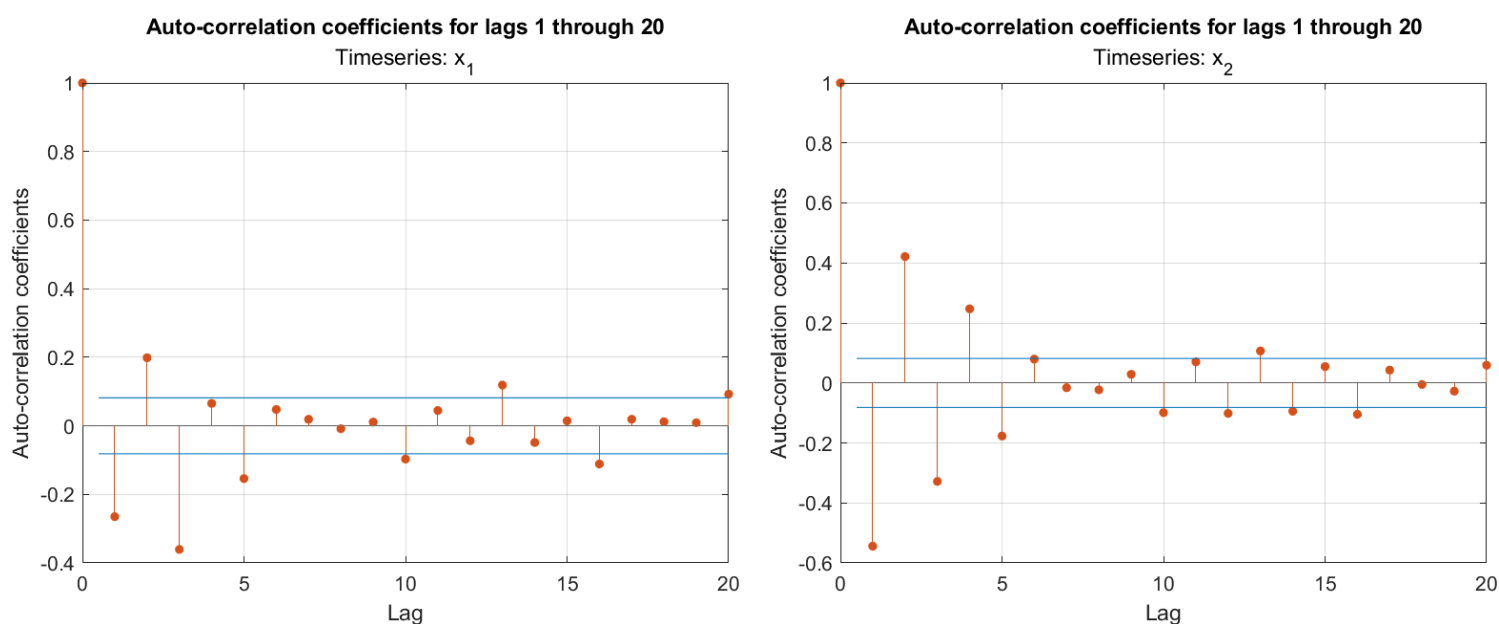
Η δοκιμή για λευκό θόρυβο είναι ένα από τα πρώτα πράγματα που πρέπει να κάνουμε για να αποφύγουμε να ξοδέψουμε χρόνο για την προσαρμογή μοντέλων σε σύνολα δεδομένων που δεν προσφέρουν ουσιαστική πληροφορία.

Θα εξετάσουμε δύο (2) τρόπους για να προσδιορίσουμε εάν η χρονοσειρά μας είναι στην πραγματικότητα, απλώς λευκός θόρυβος (white noise). Εφόσον, από την εκφώνηση δίνεται ότι έχουμε μια χρονοσειρά γραμμικής στοχαστικής διαδικασίας, είναι ικανοποιητικό να ελέγξουμε την ύπαρξη συσχέτισης, δηλαδή μη-μηδενικής αυτοσυσχέτισης. Θεωρητικά η αυτοσυσχέτιση της χρονοσειράς λευκού θορύβου είναι μηδενική.

1.2.1 Auto-correlation plots

Παρακάτω υπολογίζουμε τη συσχέτιση μεταξύ της χρονοσειράς με μια καθυστερημένη εκδοχή του εαυτού τους δίνοντας μας τα διαγράμματα αυτοσυσχέτισης (ACF), που χρησιμοποιούνται συχνά για να ελέγξουμε την τυχαιότητα σε μία χρονοσειρά.

Αυτή η τυχαιότητα διαπιστώνεται με τον υπολογισμό των αυτοσυσχετίσεων για τιμές δεδομένων σε ποικίλες χρονικές υστερήσεις (lags). Εάν είναι τυχαία, οι αυτοσυσχετίσεις θα πρέπει να είναι σχεδόν μηδενικές για όλους τους διαχωρισμούς χρονικής υστέρησης. Εάν δεν είναι τυχαία, τότε μία ή περισσότερες από τις αυτοσυσχετίσεις θα είναι σημαντικά μη μηδενικές.



Σχήμα 1.2.1.1 - Auto-correlation plots των στάσιμων χρονοσειρών x_1 , x_2

Παραπάνω γίνεται και plot του 95% confidence interval region και αν λοιπόν η χρονοσειρά είναι λευκός θόρυβος η αυτοσυσχέτιση για κάθε τιμή υστέρησης θα πρέπει να ανήκει στο διάστημα αυτό.

Μπορούμε να δούμε και στα δύο plots παραπάνω πως η χρονοσειρά (x_1 και x_2) περιέχει σημαντικά auto-correlations για διάφορες τιμές υστέρησης.

Επομένως, μπορούμε να πούμε πως η χρονοσειρά μας **ΔΕΝ είναι απλός λευκός θόρυβος**.

Επίσης αξίζει να σημειωθεί πως βλέπουμε πως η αυτοσυσχέτιση για υστέρηση ίση με 0 είναι πάντα 1 καθώς μια τιμή αυτοσυσχετίζεται πάντα τέλεια με τον εαυτό της.

Παρακάτω κάνουμε ακόμα ένα test για να επιβεβαιώσουμε το αποτέλεσμα αυτό.

1.2.2 Ljung-Box test

Πριν κάνουμε το test, πρέπει να δούμε για ποιές χρονικές υστερήσεις θα το τρέξουμε. Από τη βιβλιογραφία είδαμε πως:

- Για μεγάλες τιμές υστέρησης στο Ljung-Box test, η δοκιμή χάνει ισχύ όταν μια σημαντική συσχέτιση σε μια καθυστέρηση “χάνεται” από ασήμαντες συσχετίσεις σε άλλες καθυστερήσεις ενώ
- Για μικρές τιμές το test δεν ανιχνεύει αυτοσυσχετίσεις υψηλής τάξης.

Γι’ αυτό, οι Box, Jenkins, και Reinsel στην έρευνα τους^[1] πρότειναν την επιλογή της μέγιστης υστέρησης ίση με $\max\text{Lag} = \min(20, T-1)$, όπου T ο αριθμός των samples της χρονοσειράς.

Επίσης, και εδώ, επιλέχθηκε επίπεδο σημαντικότητας $\alpha = 0.05$.

Τα αποτελέσματα του test (όπως φαίνεται από την κονσόλα του MATLAB) για τις στάσιμες χρονοσειρές (x_1 και x_2) φαίνονται παρακάτω:

```
Ljung-Box test result
=====
The time series x1 is likely NOT white noise
The time series x2 is likely NOT white noise
fx >>
```

Σχήμα 1.2.2.1 - Ljung-Box test console output για τις χρονοσειρές x_1 και x_2

Βλέπουμε λοιπόν πως το Ljung-Box test **απορρίπτει τη μηδενική υπόθεση H_0** (για σημαντικότητα $\alpha = 0.05$) (H_0 = υπάρχει μηδενική αυτοσυσχέτιση στα residuals της χρονοσειράς) (δηλαδή η πιθανότητα που δίνει το test να έχουμε μηδενική αυτοσυσχέτιση είναι μικρότερη του 0.05). Άρα το αποτέλεσμα του test μας λέει πως η χρονοσειρά μας δεν είναι λευκός θόρυβος.

Βλέπουμε πως και οι δύο τρόποι παραπάνω συμφωνούν στα αποτελέσματα τους. Γι’ αυτό λοιπόν μπορούμε με ασφάλεια να πούμε πως η χρονοσειρά μας **δεν είναι λευκός θόρυβος**.

1.3 Βήμα 3^ο - Διερεύνηση του πιο κατάλληλου γραμμικού μοντέλου

Τώρα θέλουμε να βρούμε το βέλτιστο μικτό γραμμικό μοντέλο $\text{ARMA}(p, q)$, για καθεμία από τις προσεγγίσεις χρονοσειρών x_1, x_2 .

Σε ένα μοντέλο $\text{AR}(p)$, η μερική αυτοσυσχέτιση αναμένεται να είναι σημαντική σε καθυστερήσεις έως p . Από το Partial Auto-Correlation plot (PACF), αντίστοιχα, και παρατηρώντας μέχρι ποιά τιμή χρονικής υστέρησης (Lag) παραμένει significant η τιμή (εκτός της πρώτης τιμής που είναι πάντα = 1), μπορούμε να πάρουμε μια ιδέα για τη μέγιστη τιμή $\max(p)$ που έχει νόημα να κάνουμε διερεύνηση για το προσδιορισμό της τάξης p (εξασφαλίζοντας υπολογιστική αποτελεσματικότητα) του $\text{AR}(p)$ μέρος της μικτής διαδικασίας ARMA , που δίνει το βέλτιστο αποτέλεσμα.

Σε ένα μοντέλο $\text{MA}(q)$, η μερική αυτοσυσχέτιση αναμένεται να είναι σημαντική σε καθυστερήσεις έως q . Γνωρίζουμε έτσι πως από το Auto-Correlation plot και παρατηρώντας μέχρι ποιά τιμή χρονικής υστέρησης (Lag) παραμένει significant η τιμή (εκτός της πρώτης τιμής που είναι πάντα = 1), μπορούμε να πάρουμε μια ιδέα για τη μέγιστη τιμή $\max(q)$ που έχει νόημα να κάνουμε διερεύνηση για το προσδιορισμό της τάξης q (εξασφαλίζοντας υπολογιστική αποτελεσματικότητα) του $\text{MA}(q)$ μέρος της μικτής διαδικασίας ARMA , που δίνει το βέλτιστο αποτέλεσμα.

Παρατηρώντας τα Auto-Correlation plots στο σχήμα 1.2.1.1 αλλά και τα Partial Auto-Correlation plots των στάσιμων χρονοσειρών x_1, x_2 στο σχήμα 1.3.1, βλέπουμε:

Για την x_1 :

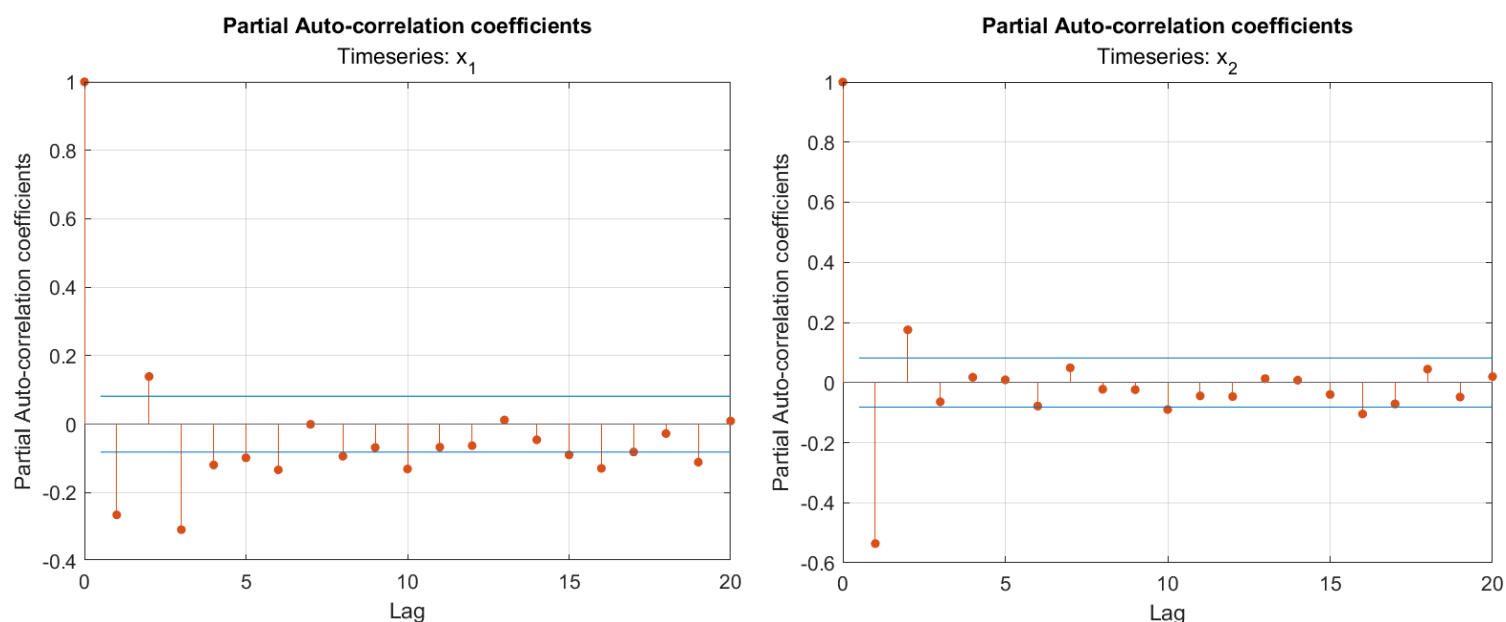
- Στο PACF plot, μέχρι την τιμή για lag = 3 η τιμή του partial auto-correlation παραμένει significant ενώ από εκεί και μετά φαίνεται να μειώνεται απότομα.
- Στο ACF plot, μέχρι την τιμή για lag = 3 η τιμή του auto-correlation παραμένει significant ενώ από εκεί και μετά φαίνεται να μειώνεται απότομα.

Επομένως έχει νόημα να ψάξουμε για την x_1 το βέλτιστο ARMA(p, q) μοντέλο για πιθανές τιμές $p \in \{0, 1, 2, 3\}$ και $q \in \{0, 1, 2, 3\}$.

Ομοίως, για την x_2 :

- Στο PACF plot, μέχρι την τιμή για lag = 5 η τιμή του partial auto-correlation παραμένει significant ενώ από εκεί και μετά φαίνεται να μειώνεται απότομα.
- Στο ACF plot, μέχρι την τιμή για lag = 2 η τιμή του auto-correlation παραμένει significant ενώ από εκεί και μετά φαίνεται να μειώνεται απότομα.

Επομένως έχει νόημα να ψάξουμε για την x_2 το βέλτιστο ARMA(p, q) μοντέλο για πιθανές τιμές $p \in \{0, 1, 2, 3, 4, 5\}$ και $q \in \{0, 1, 2\}$.

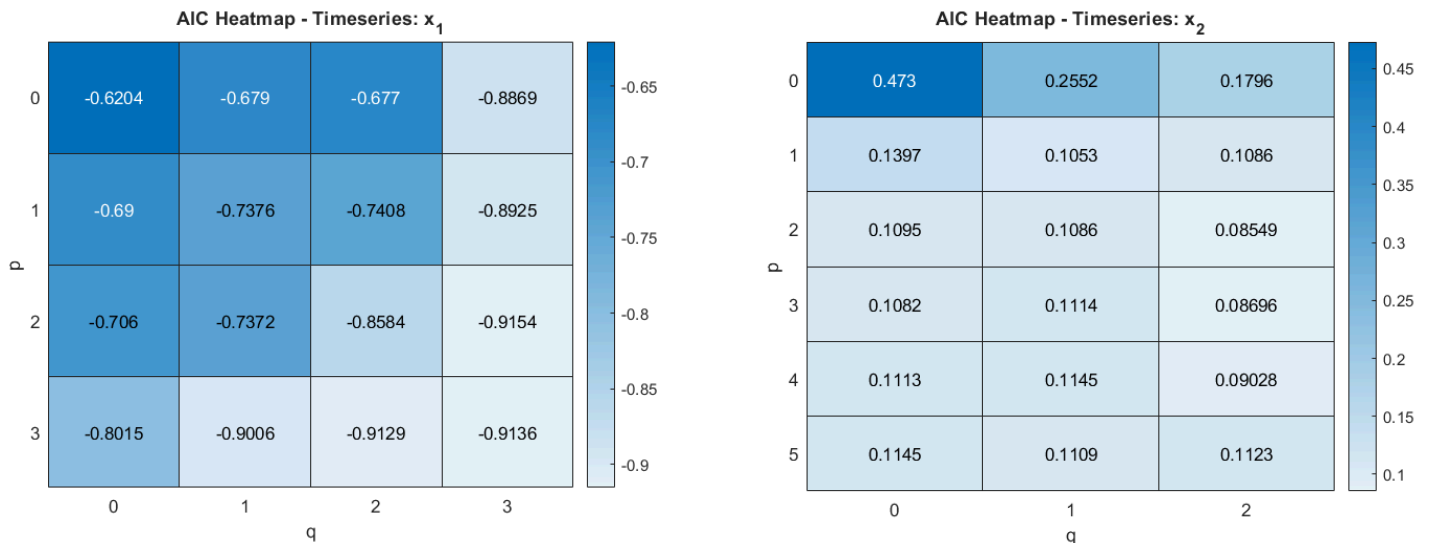


Σχήμα 1.3.1 - Partial Auto-correlation plots των στάσιμων χρονοσειρών x_1 , x_2

Για τον προσδιορισμό των τάξεων p και q του μοντέλου των δύο χρονοσειρών, χρησιμοποιούμε το *κριτήριο πληροφορίας του Akaike (Akaike Information Criterion - AIC)* εκτελώντας τα παρακάτω βήματα:

1. Αρχικά, υπολογίζουμε τις τιμές AIC, σύμφωνα με τον τύπο των σημειώσεων^[2], για κάθε πιθανό ζεύγος (p , q) των υποψήφιων βέλτιστων γραμμικών μοντέλων, όπως φαίνεται στο σχήμα 1.3.2 παρακάτω.
2. Στην συνέχεια, αφού γνωρίζουμε πως και για να είναι το γραμμικό μοντέλο κατάλληλο θα πρέπει η χρονοσειρά των υπολοίπων να είναι λευκός θόρυβος, ξεκινώντας από το υποψήφιο μοντέλο με την μικρότερη τιμή AIC (προς την μεγαλύτερη), υπολογίζουμε κάθε φορά ένα προσωρινό μοντέλο (με τις αντίστοιχες τιμές p και q).
3. Εφαρμόζουμε το προσωρινό μοντέλο στην χρονοσειρά.

4. Ελέγχουμε αν η χρονοσειρά των υπολοίπων είναι λευκός θόρυβος (όπως κάναμε και στο προηγούμενο ερώτημα (βλ. Βήμα 2^ο της εργασίας)).
 - a. Αν είναι, τερματίζουμε την διερεύνηση μας και λέμε πως βρήκαμε το βέλτιστο και κατάλληλο γραμμικό μοντέλο για την εκάστοτε στάσιμη χρονοσειρά.
 - b. Αν δεν είναι, παίρνουμε τις τιμές των p, q για την αμέσως μεγαλύτερη τιμή AIC επαναλαμβάνοντας την διαδικασία και πάμε στο βήμα 2.



Σχήμα 1.3.2 - Τιμές του Akaike Information Criterion (AIC) για το grid search του προσδιορισμού των βέλτιστων τιμών p και q του βέλτιστου ARMA μοντέλου των δύο προσεγγίσεων (x_1 και x_2)

```
Find best linear model
=====
ARMA(2, 3) is a good model for x1
Best linear model found for x1: ARMA(2, 3)
ARMA(2, 2) is a good model for x1
Best linear model found for x2: ARMA(2, 2)
fx >>
```

Σχήμα 1.3.3 - Console output της εύρεσης του πιο κατάλληλου γραμμικού μοντέλου για κάθε προσέγγιση (x_1 και x_2)

Τελικά, όπως φαίνεται και στο σχήμα 1.3.3, τα βέλτιστα γραμμικά μοντέλα είναι:

- **ARMA(2, 3) για την χρονοσειρά x_1 της προσέγγισης 1α)**
- **ARMA(2, 2) για την χρονοσειρά x_2 της προσέγγισης 1β)**

1.4 Βήμα 4^ο - Εφαρμογή και μετρικές των γραμμικών μοντέλων

Αρχικά, παρακάτω σε κάθε περίπτωση έγινε διαχωρισμός των δεδομένων σε:

- 75% (450 παρατηρήσεις) στο training set
- 25% (150 παρατηρήσεις) στο validation set

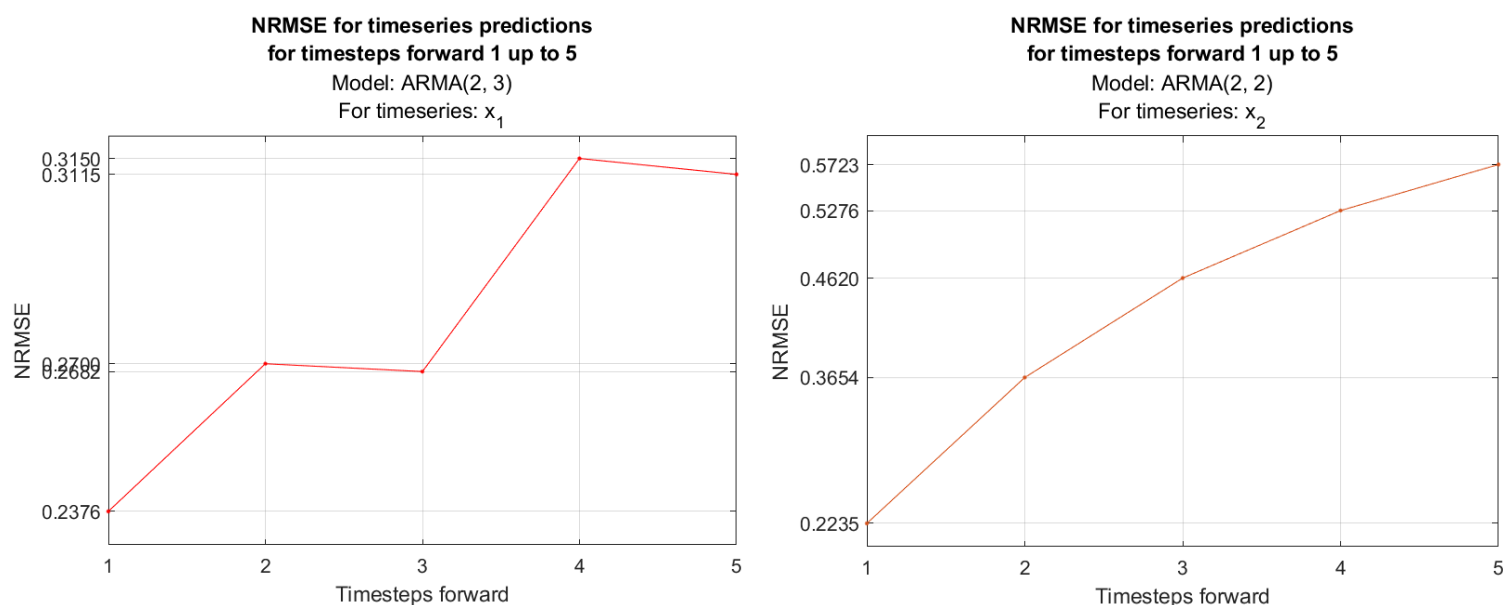
Στην συνέχεια, θέλουμε να εφαρμόσουμε τα μοντέλα ARMA που βρήκαμε στα δεδομένα εκπαίδευσης και να κάνουμε προβλέψεις στο validation set για τα επόμενα $T_{\max} = 5$ βήματα.

Επειδή όμως θέλουμε οι προβλέψεις αυτές αναφέρονται στην αρχική χρονοσειρά (που μας δόθηκε x) και όχι στις στάσιμες χρονοσειρές (x_1 και x_2), με χρήση συναρτήσεων που έχουν υλοποιηθεί στο μάθημα, υπολογίζουμε τις προβλέψεις των δύο μοντέλων ARMA που βρήκαμε στις αντίστοιχες στάσιμες χρονοσειρές.

Στην συνέχεια, για να κάνουμε αυτές τις προβλέψεις να αναφέρονται στην αρχική χρονοσειρά, πρέπει να προσθέσουμε πάλι την τάση που αφαιρέσαμε στο βήμα 1.

- Στην περίπτωση 1α) που χρησιμοποιήσαμε εξαμόλυνση κινούμενου μέσου, για κάθε χρονική στιγμή στο σύνολο αξιολόγησης, προσαρμόζουμε μια νοητή ευθεία στις αντίστοιχες τιμές του εξαμαλυσμένου κινούμενου μέσου και την προεκτίνουμε για τα επόμενα $T_{\max} = 5$ βήματα για να προβλέψουμε την τάση, να την προσθέσουμε στις προβλέψεις και να πάρουμε έτσι τις τελικές προβλέψεις (στην x).
- Στην περίπτωση 2α) που χρησιμοποιήσαμε πρώτες διαφορές, απλά αντιστρέφουμε τις πρώτες διαφορές για να πάρουμε τις τελικές προβλέψεις (από την x_2 στην x), αντιστρέφοντας απλά την εξίσωση (1.1.2.1) που χρησιμοποιήσαμε στο βήμα 1.

Τέλος, εφαρμόζοντας τον γνωστό τύπο του NRMSE (με βάση τις πραγματικές τιμές της x και τις προβλέψεις πάνω σε αυτήν), υπολογίζουμε το NRMSE για τις προβλέψεις των δύο μοντέλων ARMA για τα $T_{\max} = 5$ βήματα και παίρνουμε τα παρακάτω διαγράμματα του σχήματος 1.4.1:



Σχήμα 1.4.1 - NRMSE της πρόβλεψης με τα κατάλληλα γραμμικά μοντέλα για τα χρονικά βήματα μπροστά από 1 ως 5

1.5 Βήμα 5° - Σύγκριση προβλέψεων και αξιολόγηση προσεγγίσεων

Αρχικά, βλέπουμε και στις δύο προσεγγίσεις, όπως φαίνεται και από τα σχήματα 1.4.1 και 1.5.1, πως όσο προσπαθούμε να κάνουμε μια πιο μακροπρόθεσμη πρόβλεψη, το NRMSE αυξάνεται, το οποίο είναι και αναμενόμενο.

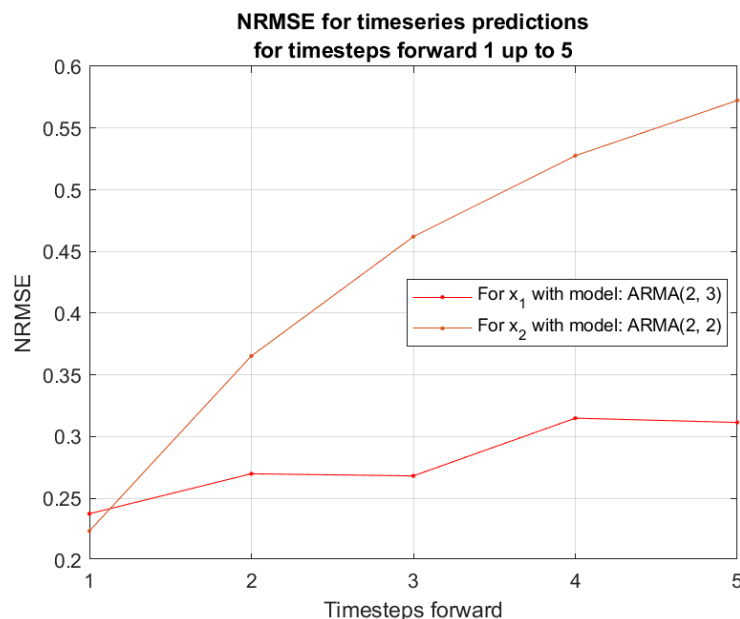
Τώρα, συγκρίνοντας τις προβλέψεις με τις δύο προσεγγίσεις, βλέπουμε πως η προσέγγιση 1α) (δηλαδή αυτή που απαλείψαμε την τάση αφαιρώντας μια

εκτιμώμενη συνάρτηση της τάσης παίρνοντας έτσι την στάσιμη χρονοσειρά x_1) φαίνεται να δίνει ένα **ελάχιστο χειρότερο αποτέλεσμα** (δηλαδή μεγαλύτερο NRMSE) **για το πρώτο επόμενο χρονικό βήμα** (σε σχέση με την προσέγγιση 1β), δηλαδή αυτή που απαλείψαμε την τάση παίρνοντας τις πρώτες διαφορές παίρνοντας έτσι την στάσιμη χρονοσειρά x_2)

Φαίνεται στην συνέχεια, όμως η **προσέγγιση 1α**) να δίνει συγκριτικά πολύ καλύτερα αποτελέσματα (σε σχέση με την προσέγγιση 1β που δίνει την χρονοσειρά x_2) καθώς το NRMSE **αυξάνεται αλλά με μικρότερο ρυθμό** (όπως φαίνεται και καλύτερα στο σχήμα 1.5.1 ή και στον πίνακα 1.5.1) .

Timesteps Forward	NRMSE με Προσέγγιση 1α) - Χρονοσειρά x_1	NRMSE με Προσέγγιση 1β) - Χρονοσειρά x_2
1	0.2376	0.2235
2	0.2682	0.3654
3	0.2789	0.4620
4	0.3115	0.5276
5	0.3150	0.5723

Πίνακας 1.5.1 - NRMSE των γραμμικών μοντέλων και των δύο προσεγγίσεων για τα επόμενα 5 χρονικά βήματα (με **bold** σημειώθηκε η μικρότερη τιμή κάθε φορά)

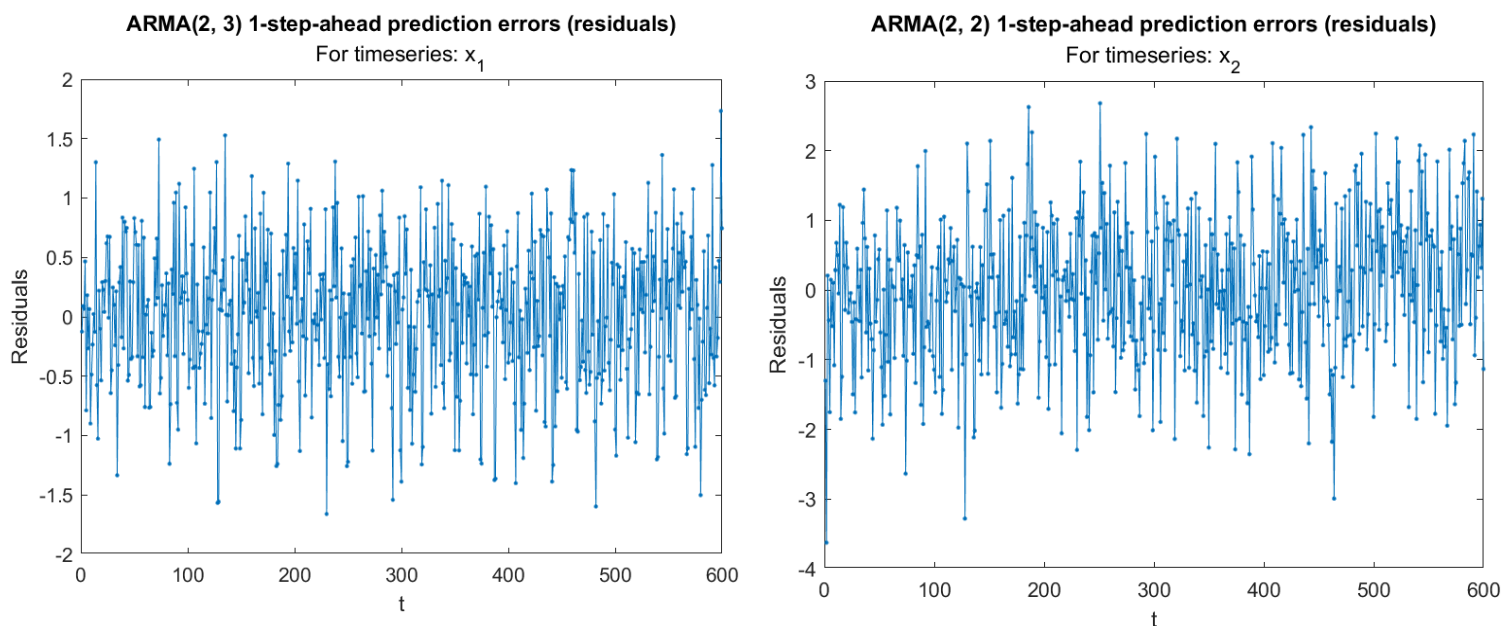


Σχήμα 1.5.1 - Συγκρίση NRMSE των γραμμικών μοντέλων για τις προβλέψεις των δύο προσεγγίσεων για τα επόμενα 5 χρονικά βήματα

Τέλος, φαίνεται πως γενικά μπορούμε να κάνουμε **καλές προβλέψεις** της χρονοσειράς αυτής **με γραμμικά μοντέλα** (καθώς το NRMSE παραμένει σχετικά μικρό και στις δύο προσεγγίσεις). Ιδιαίτερα, φαίνεται πως και με την χρήση σωστής προσέγγισης (εδώ η 1α)) και σωστής επιλογής του γραμμικού μοντέλου μπορούμε να κάνουμε καλές προβλέψεις της χρονοσειράς με γραμμικό μοντέλο έως και για τα επόμενα 5 χρονικά βήματα με σχετικά πολύ μικρό NRMSE (τιμές < 0.32).

1.6 Βήμα 6^ο - Χρονοσειρές των υπολοίπων, από την προσαρμογή στις στάσιμες χρονοσειρές, των πιο κατάλληλων γραμμικών μοντέλων

Παρακάτω, στο σχήμα 1.6.1, φαίνονται, για τις δύο προσεγγίσεις / χρονοσειρές 1α) / x_1 και 1β) / x_2 , οι ζητούμενες χρονοσειρές των υπολοίπων από την προσαρμογή στη στάσιμη χρονοσειρά του πιο κατάλληλου γραμμικού μοντέλου που βρήκαμε στο βήμα 3.



Σχήμα 1.6.1 - Χρονοσειρές των υπολοίπων (residuals) από την προσαρμογή στη στάσιμη χρονοσειρά των πιο κατάλληλων γραμμικών μοντέλων που βρήκαμε

Η ανάλυση μας στη συνέχεια θα γίνει ξεχωριστά για κάθε μια από τις δύο στάσιμες χρονοσειρές χωρίς γραμμικές συσχετίσεις.

1.7 Βήμα 7^ο - Έλεγχος αν η χρονοσειρά λευκού θορύβου είναι και iid

Οι έλεγχοι που κάνουμε για να δούμε αν μια χρονοσειρά είναι iid είναι οι εξής:

Αρχικά, ελέγχουμε αν η χρονοσειρά είναι λευκός θόρυβος. Αν δεν είναι, τότε δεν μπορεί να είναι iid.

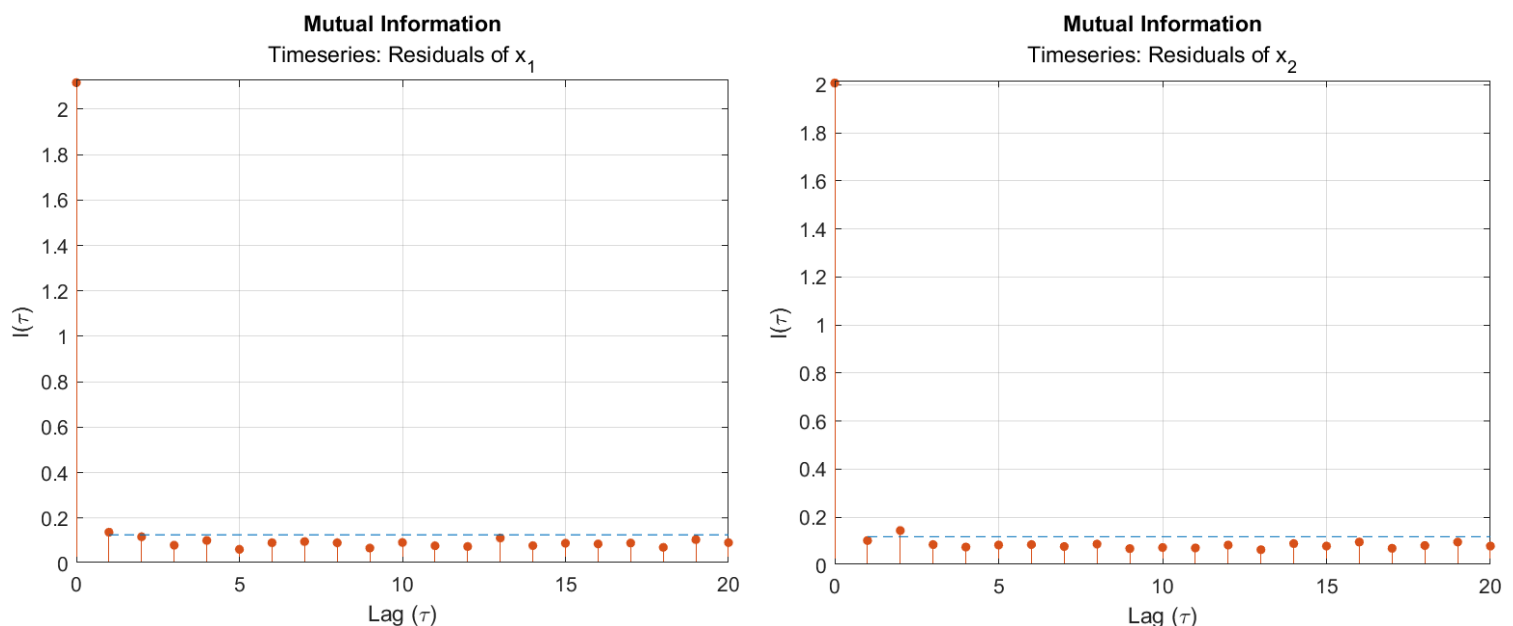
Στη συνέχεια, υπολογίζουμε το μη-γραμμικό μέτρο της Αμοιβαίας Πληροφορίας (Mutual Information - MI) για όλα τα lags. Αν η χρονοσειρά είναι iid, τότε το MI για όλα τα lags πρέπει να είναι ασήμαντο, δηλαδή θα πρέπει να είναι κάτω από ένα κατώφλι. Επιλέγουμε το κατώφλι ως το μέσο MI συν το Z-score για το επιθυμητό επίπεδο σημαντικότητας.

Επιλέξαμε το επίπεδο σημαντικότητας να είναι 95% (δηλαδή $\alpha = 0.05$).

Τέλος, ελέγχουμε αν η πρώτη διαφορά της χρονοσειράς είναι ένας τυχαίος περίπατος (Random Walk). Αν είναι, τότε η χρονοσειρά είναι iid^[2].

Αυτό το κάνουμε με τον έλεγχο του variance ratio test. Αν η χρονοσειρά είναι iid, τότε ο έλεγχος δεν απορρίπτει τη μηδενικής υπόθεσης H_0 ότι η χρονοσειρά είναι ένας τυχαίος περίπατος. Αν απορριφθεί η μηδενική υπόθεση, τότε η χρονοσειρά δεν είναι iid.

Σε αυτό τον έλεγχο περάσαμε ως argument την υπόθεση πως τα innovations (δηλαδή τα errors) είναι iid. Αυτό σημαίνει ότι τα errors είναι ανεξάρτητα και έχουν την ίδια κατανομή με σταθερή διασπορά. Έτσι, η απόρριψη της μηδενικής υπόθεσης θα μας πει πως δεν είναι iid αυτά και άρα και η χρονοσειρά εισόδου.



Σχήμα 1.7.1 - Αμοιβαία πληροφορία των residuals των x_1 και x_2 για διάφορες τιμές χρονικής υστέρησης (lags)

Τρέχοντας τον κώδικα, βλέπουμε πως ούτε τα residuals των x_1 ούτε τα residuals των x_2 είναι iid, καθώς όπως φαίνεται από το διάγραμμα της αμοιβαίας πληροφορίας στο σχήμα 1.7.1, υπάρχει μια τιμή για κάποιο lag που βρίσκεται πάνω από το κατώφλι που έχουμε θέσει. Συνεπώς, δεν μπορούμε να θεωρήσουμε τα residuals ως iid (με αυτό το test).

Για επιβεβαίωση, αφού τρέξουν και τα άλλα test μας, παίρνουμε το console output που φαίνεται και παρακάτω στο σχήμα 1.7.2:

```
x1 residuals are NOT iid
```

```
x2 residuals are NOT iid
```

Σχήμα 1.7.2 - Console output του ελέγχου αν τα residuals των x_1 και x_2 είναι iid

Επομένως, μπορούμε να πούμε πως οι χρονοσειρές λευκού θορύβου **ΔΕΝ είναι και iid**, δηλαδή άρα υπάρχουν μη-γραμμικές συσχετίσεις.

1.8 Βήμα 8° - Διερεύνηση του πιο κατάλληλου μη-γραμμικού μοντέλου για τη στάσιμη χρονοσειρά

Δεδομένης της καλής απόδοσης των γραμμικών μοντέλων για την πρόβλεψη των χρονοσειρών x_1 και x_2 , και την υπόδειξη που μας δίνεται, αναζητούμε το βέλτιστο τοπικά γραμμικό μη γραμμικό μοντέλο για την πρόβλεψη των χρονοσειρών.

Για τα μοντέλα αυτά, υπάρχουν τέσσερις παράμετροι που πρέπει να βρεθούν:

- ο χρόνος καθυστέρησης τ
- η διάσταση εμβύθισης m
- η παράμετρος αποκοπής q
- ο αριθμός των γειτονικών σημείων $N_{\text{neighbors}}$

Για την εύρεση των κατάλληλων παραμέτρων, χρησιμοποιείται μια διαδικασία ανάλογη του grid search. Κατά την αναζήτησή, πρέπει να έχουμε κατα νού τους ακόλουθους δύο περιορισμούς:

- $0 \leq q \leq m$
- $m \leq N_{\text{neighbors}}$

Έτσι, παρέχουμε στην συνάρτηση **findOptimalNonLinearModel** εύρη αναζήτησης για όλες τις παραμέτρους εκτός της q , και υλοποιούμε την αναζήτηση με τέσσερις εμφωλευμένες for loops. Η πιο εσωτερική που αναζητεί για την βέλτιστη παράμετρο q αναζητεί στο εύρος $[1, m]$.

Σε κάθε επανάληψη της πιο εσωτερικής for loop, χρησιμοποιούμε την δοσμένη συνάρτηση **localfitnrsm** για να δημιουργήσουμε και να εκπαιδεύσουμε ένα μη γραμμικό τοπικά γραμμικό μοντέλο για πρόβλεψη με ένα βήμα στο μέλλον, και να υπολογίσουμε το NRMSE του. Επίσης, υπολογίζουμε την χρονοσειρά των υπολοίπων μετά την πρόβλεψη. Αφού ελέγξουμε αν αυτή είναι i.i.d. με συνάρτηση που υπολογίστηκε σε προηγούμενο βήμα, κρατάμε το μοντέλο αυτό ως κατάλληλο.

Μετά από όλες τις επαναλήψεις, αναζητούμε από τα κατάλληλα μοντέλα αυτό με το χαμηλότερο σφάλμα NRMSE, και το επιστρέφουμε στον χρήστη.

Με την εκτέλεση του κώδικα αυτού, λαμβάνουμε τα παρακάτω αποτελέσματα για τα ακόλουθα εύρη παραμέτρων, όπως φαίνεται και από το console output στο σχήμα 1.8.1:

Για την χρονοσειρά x_1 :

- $\tau : [1, 15]$
- $m : [1, 10]$
- $N_{\text{neighbors}} : [10, 20]$

Για την χρονοσειρά x_2 :

- $\tau : [1, 5]$
- $m : [10, 20]$
- $N_{\text{neighbors}} : [20, 30]$

Είναι σημαντικό να τονιστεί ότι αναζητούμε το βέλτιστο μοντέλο για πρόβλεψη στο άμεσο μελλοντικό βήμα και όχι για παραπάνω μελλοντικά βήματα, όπως ασχολούμαστε στην υπόλοιπη εργασία.

Find the best non linear (local linear) model

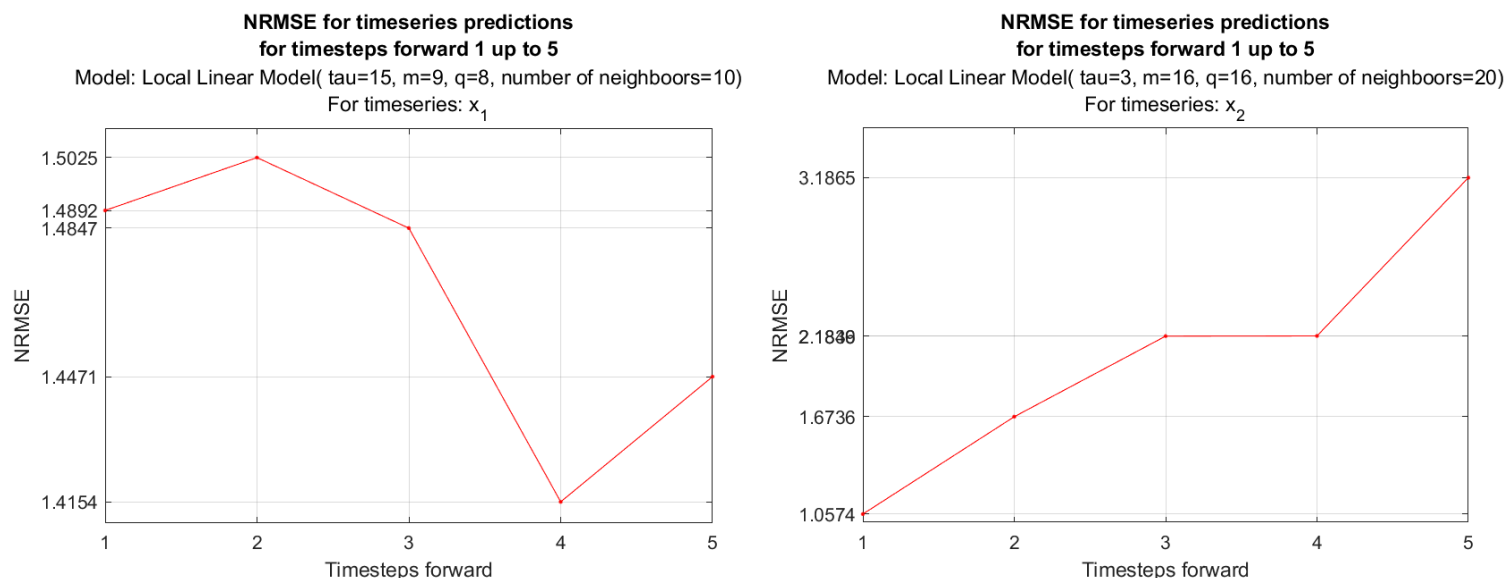
Optimal model for timeseries x1 has following parameters: tau=15, m=9, q=8, number of neighbors=10 with fitting error NRMSE=1.927258
Optimal model for timeseries x2 has following parameters: tau=3, m=16, q=16, number of neighbors=20 with fitting error NRMSE=2.851745

Σχήμα 1.8.1 - Console output της εύρεσης του πιο κατάλληλου μη-γραμμικού μοντέλου για κάθε προσέγγιση (x_1 και x_2)

Επιπρόσθετα, γνωρίζουμε ότι υπάρχουν και πιο προχωρημένες μέθοδοι εύρεσης των παραμέτρων, αλλά για τους σκοπούς της εργασίας η απλή αυτή μέθοδος μας εξυπηρετεί ικανοποιητικά.

1.9 Βήμα 9^ο - Εφαρμογή και μετρικές των μη-γραμμικών μοντέλων

Το παράδοξο που παρατηρούμε στα μη γραμμικά μοντέλα, και στο σχήμα 1.9.1, που βρέθηκαν στο προηγούμενο βήμα, και συγκεκριμένα στο πρώτο εκ των δύο, είναι ότι το σφάλμα πρόβλεψης NRMSE μειώνεται, με τον αριθμό των βημάτων πρόβλεψης μπροστά. Το μοντέλο που εκπαιδεύτηκε στην χρονοσειρά x_2 ακολουθεί μια πιο τυπική συμπεριφορά, με το σφάλμα να αυξάνεται με ομαλό ρυθμό όσο μεγαλώνουν τα βήματα πρόβλεψης.



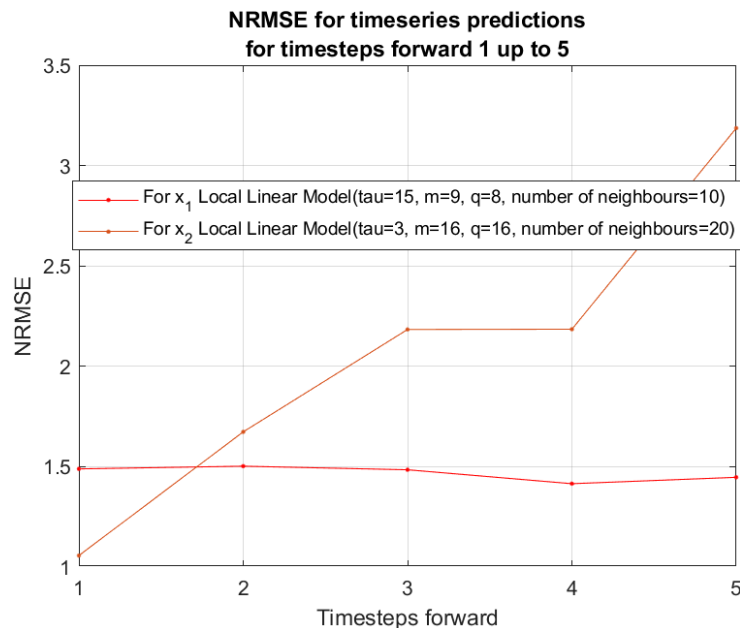
Σχήμα 1.9.1 - NRMSE της πρόβλεψης με τα κατάλληλα μη-γραμμικά μοντέλα για τα χρονικά βήματα μπροστά από 1 ως 5

1.10 Βήμα 10^ο - Σύγκριση προβλέψεων με τα δύο μη-γραμμικά μοντέλα από τις δύο προσεγγίσεις, σχολιασμός για τη δυνατότητα καλής πρόβλεψης της χρονοσειράς αυτής με μη-γραμμικά μοντέλα και σύγκριση με τις προβλέψεις με τα γραμμικά μοντέλα.

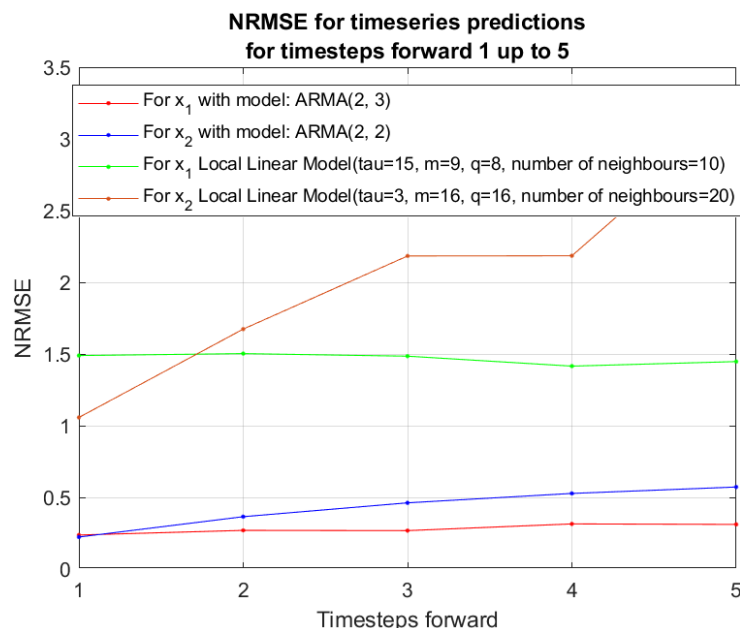
Από την σύγκριση των δύο μη-γραμμικών μοντέλων, παρατηρούμε ότι το δεύτερο μοντέλο (της x_2) δίνει σημαντικά καλύτερα αποτελέσματα πρόβλεψης (από αυτό της x_1) για βήματα πρόβλεψης μεγαλύτερα από 1. Αυτό οφείλεται στο ότι το μοντέλο λαμβάνει υπόψη του μεγαλύτερο αριθμό γειτονικών σημείων (20 έναντι

10), το οποίο του δίνει καλύτερη δυνατότητα εξαγωγής long spanning μοτίβων, που είναι καίριο για την πρόβλεψη πολλά βήματα μελλοντικά.

Επίσης, γενικά βλέπουμε πως τα μη-γραμμικά μοντέλα δίνουν πολύ υψηλές τιμές NRMSE (> 1), και στις δύο προσεγγίσεις, και συνεπώς μπορούμε να καταλάβουμε εύκολα πως **δεν υπάρχει η δυνατότητα καλής πρόβλεψης** της χρονοσειράς αυτής με μη-γραμμικά μοντέλα



Σχήμα 1.10.1 - Σύγκριση προβλέψεων (NRMSE) με τα δύο μη-γραμμικά μοντέλα από τις δύο προσεγγίσεις για τα χρονικά βήματα μπροστά από 1 ως 5



Σχήμα 1.10.2 - Σύγκριση προβλέψεων (NRMSE) με τα δύο μη-γραμμικά μοντέλα και δύο γραμμικά μοντέλα από τις δύο προσεγγίσεις για τα χρονικά βήματα μπροστά από 1 ως 5

Συγκρίνοντας τα μη-γραμμικά μοντέλα με τα γραμμικά μοντέλα παίρνουμε, από το σχήμα 1.10.2 βλέπουμε εύκολα ότι τα δύο **γραμμικά μοντέλα υπερνικούν των μη-γραμμικών για οποιοδήποτε αριθμό βημάτων** πρόβλεψης από 1 έως 5.

Ανάλογα με την εφαρμογή μας και την επιθυμητή ακρίβεια, μπορούμε να επιλέξουμε οποιοδήποτε από τα μοντέλα. Για μικρό βήμα πρόβλεψης (1) μπορούμε να επιλέξουμε το μη-γραμμικό μοντέλο που εκπαιδεύτηκε στην χρονοσειρά x_2 . Παρόλα αυτά, τα γραμμικά μοντέλα είναι κατά πολύ καλύτερα, με περίπου το 25% του σφάλματος των αντίστοιχων μη γραμμικών ανάλογα την περίπτωση, λιγότερη υπολογιστική πολυπλοκότητα, και σημαντικά χαμηλότερο αριθμό παραμέτρων, οπότε δεν κρίνεται σε καμία περίπτωση προτιμότερη η χρήση των μη γραμμικών μοντέλων έναντι των γραμμικών.

1.11 Βήμα 11^ο - Σχολιασμός για το από ποιο σύστημα μπορεί να έχει προέλθει η χρονοσειρά

Από τα συγκριτικά αποτελέσματα των μοντέλων πρόβλεψης στο σχήμα 1.10.2, παρατηρούμε ότι τα αμιγώς γραμμικά μοντέλα αποδίδουν καλύτερα από μη γραμμικά μοντέλα όπως τα τοπικά γραμμικά (πολύ χαμηλότερο σφάλμα πρόβλεψης), καταλήγουμε στο συμπέρασμα ότι το σύστημα από το οποίο προέρχεται η χρονοσειρά είναι **γραμμικό**.

Μάλιστα, ο χαμηλός αριθμός των παραμέτρων που απαιτούνται για την ικανοποιητική πρόβλεψη μας οδηγεί στο να υποθέσουμε ότι το σύστημα είναι σχετικά **απλό**. Καθώς από τα διαγράμματα αυτοσυσχέτισης βλέπουμε να υπάρχει σχετικά υψηλή αυτοσυσχέτιση, και ομαλή μείωσή της σταδιακά, το σύστημα πιθανότατα έχει **χαμηλή διαστασιμότητα**, καθώς ένας μικρός αριθμός από μεταβλητές φαίνεται να επαρκεί για την περιγραφή του συστήματος.

Τέλος, βασιζόμενοι στο ότι τόσο οι έλεγχοι λευκού θορύβου όσο και i.i.d. είναι αρνητικοί, πιθανότατα το σύστημα είναι **αιτιοκρατικό** και όχι στοχαστικό.

Μη στοχαστικά απλά γραμμικά συστήματα χαμηλής διαστασιμότητας με ανοδική τάση (από την αρχική χρονοσειρά), μπορούν να χρησιμοποιηθούν για να μοντελοποιήσουν την αύξηση του πληθυσμού ενός τόπου, ή κατά συνέπεια την αύξηση των αναγκών του πληθυσμού αυτού του τόπου ή της οικονομικής ανάπτυξης.

Βιβλιογραφία

- [1] Box, George E. P., Gwilym M. Jenkins, and Gregory C. Reinsel. Time Series Analysis: Forecasting and Control. 3rd ed. Englewood Cliffs, NJ: Prentice Hall, 1994
- [2] Κουγιουμτζής Δημήτρης, «ΑΝΑΛΥΣΗ ΧΡΟΝΟΣΕΙΡΩΝ», Πανεπιστημιακές Σημειώσεις, Αριστοτέλειο Πανεπιστήμιο Θεσσαλονίκης, Τμήμα Ηλεκτρολόγων Μηχανικών & Μηχανικών Υπολογιστών, Θεσσαλονίκη