
Αναφορά Υπολογιστικής Εργασίας
στο μάθημα των Χρονοσειρών

Ζήσου Χαρίλαος
ΑΕΜ: 9213

Καρατζάς Μιχάλης
ΑΕΜ: 9137

Ανάλυση ιστορικών δεδομένων ηλεκτρικής ενέργειας της Ιταλίας

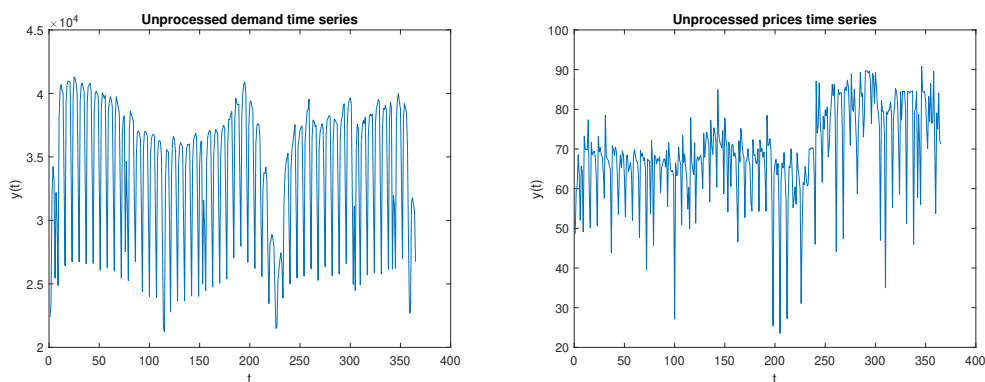
0.1 Γραμμική ανάλυση

Η ομάδα μας είχε τον αριθμό 7, επομένως κληθήκαμε να αναλύσουμε δεδομένα στις 08:00 η ώρα για όλη την επικράτεια. Η ανάλυση χωρίστηκε σε 2 αρχεία στο matlab ξεχωριστά για demand και price, κάθε ένα από τα οποία περιέχει τη γραμμική και μη γραμμική μελέτη.

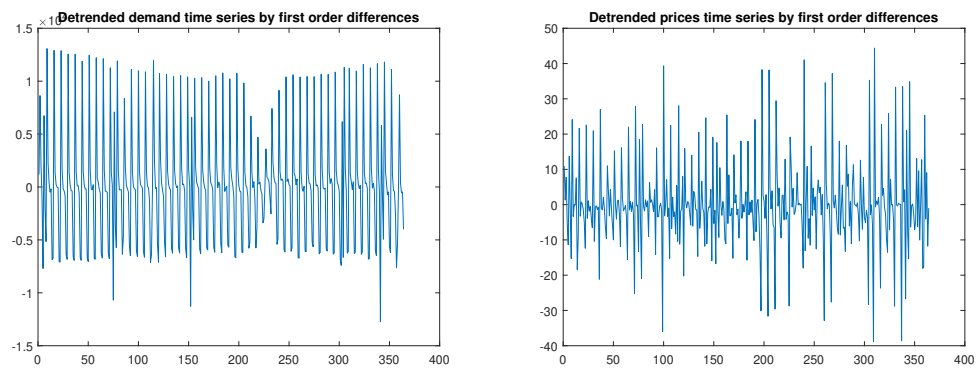
0.1.1 Τάση και περιοδικότητα

Όσον αφορά την τάση, δοκιμάζοντας προσαρμογή με πολυώνυμο και moving average φίλτρο, δεν μείναμε ικανοποιημένοι από τα αποτελέσματα και αποφασίσαμε να χρησιμοποιήσουμε τις πρώτες διαφορές, λαμβάνοντας έτσι μια χρονοσειρά εντελώς απαλλαγμένη από την τάση (βλ. σχήματα 1 και 2).

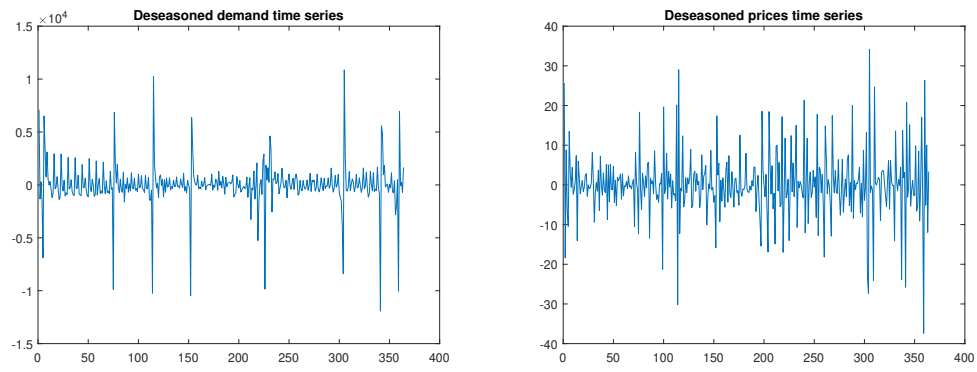
Για την περιοδικότητα χρησιμοποιήσαμε τη συνάρτηση `seasonalcomponents` με περίοδο 7, στοχεύοντας στην αφαίρεση του εβδομαδιαίου κύκλου (βλ. σχήμα 3).



Σχήμα 1

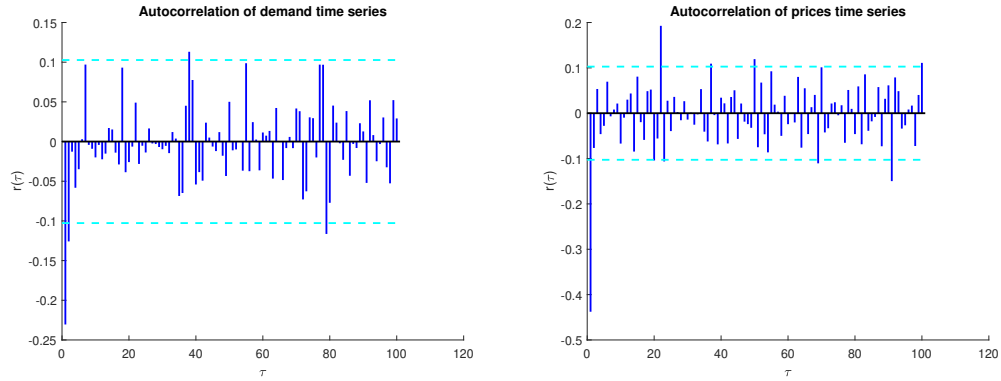


Σχήμα 2

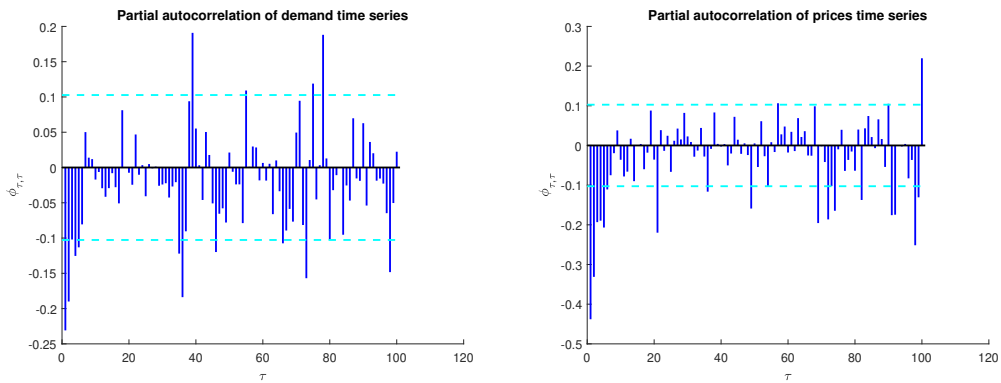


Σχήμα 3

0.1.2 Συνάρτηση αυτοσυσχέτισης - Έλεγχος λευκού θορύβου

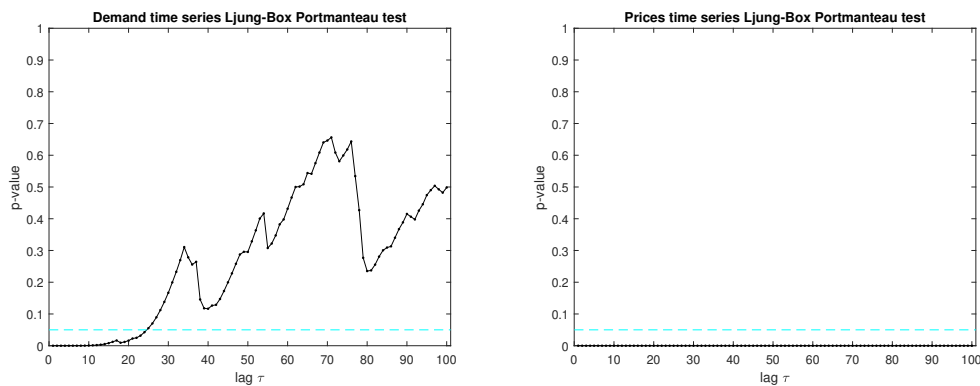


Σχήμα 4



Σχήμα 5

Από το σχήμα 4, λόγω των αυτοσυσχετίσεων για $\tau = 1, 2$ και $\tau = 1$ αντίστοιχα, αλλά λόγω και των άλλων τιμών που ξεπερνούν τα όρια σημαντικότητας, δεν μπορούμε να θεωρήσουμε πως η χρονοσειρά που απομένει είναι λευκός θόρυβος. Μάλιστα ένα πιο σωστό τεστ που θα μπορούσαμε να κάνουμε είναι το Ljung-Box Portmanteau test το οποίο δίνεται παρακάτω στο σχήμα 6:

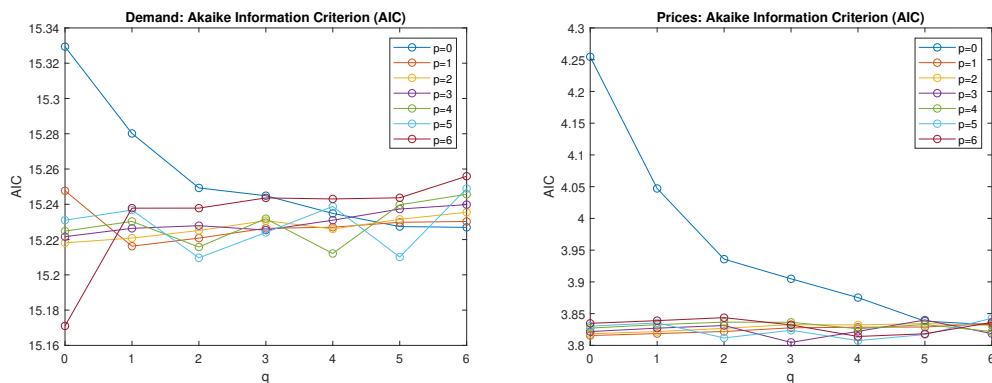


Σχήμα 6

Και από αυτό το τεστ συμπεραίνουμε πως το επίπεδο $p = 0.05$ δεν ξεπερνιέται παρα μόνο για μεγάλες υστερήσεις στην μια περίπτωση και καθόλου στην άλλη, επομένως μπορούμε να απορρίψουμε την μηδενική υπόθεση ότι οι χρονοσειρές είναι λευκός θόρυβος και άρα έχει νόημα να προχωρήσουμε στο επόμενο βήμα και να προσπαθήσουμε να προσαρμόσουμε κάποιο γραμμικό μοντέλο.

0.1.3 Προσαρμογή γραμμικού μοντέλου ARMA

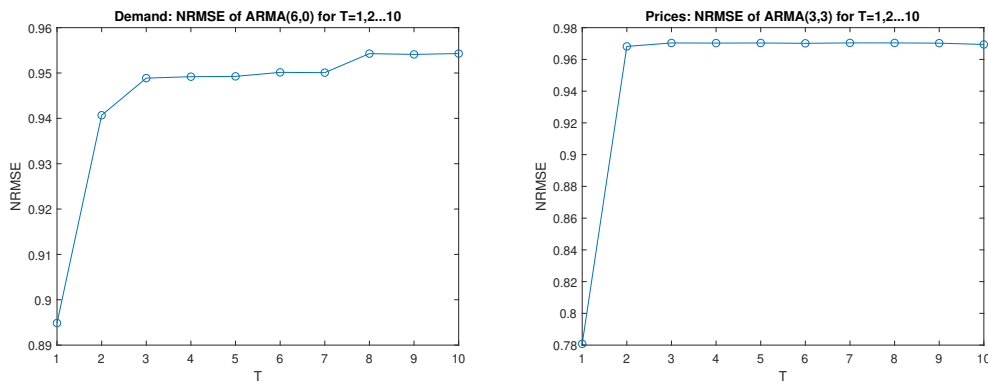
Σύμφωνα με τη θεωρία, για την επιλογή ενός κατάλληλου μοντέλου ARMA δεν μπορούμε να βασιστούμε στις συναρτήσεις αυτοσυσχέτισης η μερικής αυτοσυσχέτισης, αλλά πρέπει να υπολογίσουμε κάποιο άλλο κριτήριο όπως το κριτήριο του Akaike (AIC) που χρησιμοποιήσαμε εμείς στο σχήμα 7.



Σχήμα 7

Προσοχή, ο κώδικας για το AIC είναι σε comments , για να μην πάρει ώρα σε περίπτωση που τον τρέξετε!

Επομένως, τα μοντέλα που χρησιμοποιήσαμε ήταν $ARMA(6,0) = AR(6)$ για demand time series με συντελεστές $\phi_0 = 25.7958, \phi_1 = -0.3163, \phi_2 = -0.2671, \phi_3 = -0.1727, \phi_4 = -0.15, \phi_5 = -0.1551, \phi_6 = -0.0805, SDz = 2.0088 \cdot 10^3, AIC = 15.2269$ και $ARMA(3,3)$ για prices time series με $\phi_0 = 0.1666, \phi_1 = -0.805, \phi_2 = -0.9384, \phi_3 = 0.0817, \theta_1 = -0.0678, \theta_2 = -0.2797, \theta_3 = 0.8155, SDz = 6.647213, AIC = 3.804742$, καθώς ήταν αυτά με την χαμηλότερη τιμή του κριτηρίου. Τα σφάλματα που παράγουν τα μοντέλα για έως και 10 βήματα μπροστά, δίνονται παρακάτω στο σχήμα 8:



Σχήμα 8

Με βάση όλα τα παραπάνω, παρατηρούμε πως οι χρονοσειρές ζήτησης και τιμής έχουν παρόμοιες συναρτήσεις αυτοσυσχέτισης και μερικής αυτοσυσχέτισης (βλ. σχήματα 4 και 5), και πως καμία από τις 2 δεν μπορεί να θεωρηθεί λευκός θόρυβος. Ωστόσο τα μοντέλα που προσαρμόσαμε φαίνεται να διαφέρουν, αν και έχουν ένα σχετικά παρόμοιο σφάλμα προσαρμογής, 0.89 για demand και λίγο καλύτερο 0.78 για prices (για προβλέψεις ενός βήματος), τα οποία φυσικά τείνουν στην μονάδα καθώς αυξάνονται τα βήματα (σχήμα 8).

0.1.4 Προβλέψεις επόμενης ημέρας

Για τις προβλέψεις της επόμενης ημέρας σπάσαμε τις χρονοσειρές σε 5 διακριτά σημεία ώστε να χρησιμοποιούμε συγκεκριμένο ποσοστό των δεδομένων ως training data (70, 75, 80, 85, 90%) και τα υπόλοιπα ως validation data.

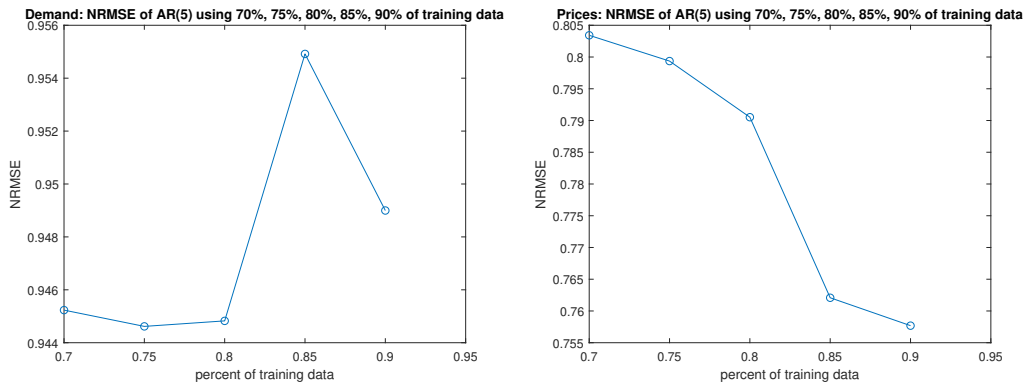


Figure 9

Από το σχήμα 9, καθώς μετακινούμε το σημείο του διαχωρισμού, παρατηρούμε πως δεν έχουμε ίδια αποτελέσματα για τις 2 χρονοσειρές. Στη χρονοσειρά της ζήτησης, καθώς αυξάνεται το σύνολο εκμάθησης, παρόλο που στην αρχή φαίνεται να υπάρχει μια μικρή μείωση, εν τέλει το σφάλμα πρόβλεψης αυξάνεται, γεγονός που υποδηλώνει over-fitting ενώ στη χρονοσειρά των τιμών, παρατηρούμε πως συνεχώς μειώνεται.

0.2 Μη-γραμμική ανάλυση

Για των υπολογισμό των υπολοίπων(residuals) κάθε χρονοσειράς, χρησιμοποιήσαμε την συνάρτηση *predict*. Συγκεκριμένα

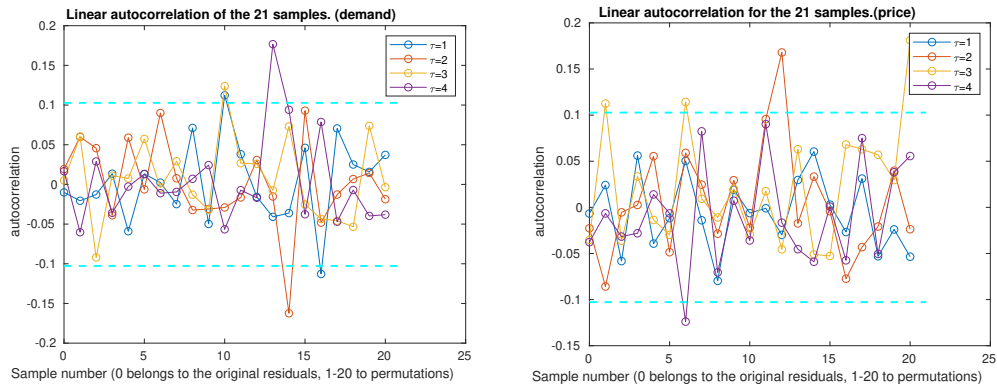
$$residuals = deseasoned - predict(armamodel, deseasoned)$$

όπου *armamodel* είναι το γραμμικό μοντέλο που προσαρμόσαμε στο ερώτημα 3 της γραμμικής ανάλυσης, ενώ *deseasoned* η απαλλαγμένη από τάσεις και περιοδικότητα χρονοσειρά.

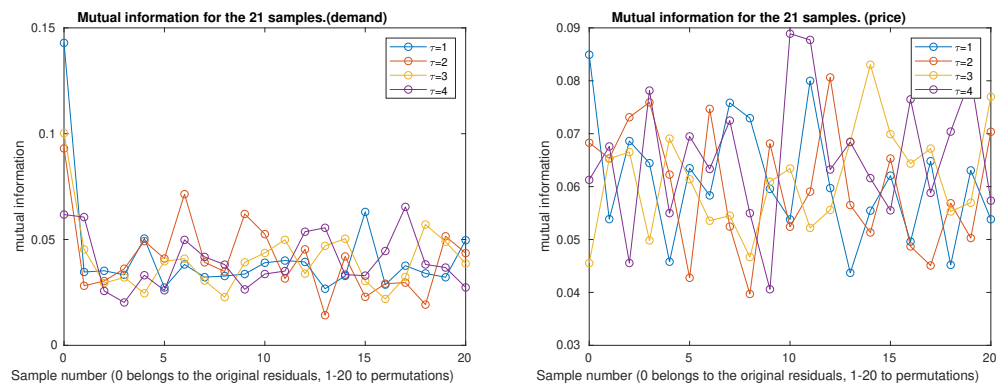
Οι 20 iid χρονοσειρές καθώς δημιουργήθηκαν με τυχαία αντιμετάθεση, δεν περιέχουν καμία γραμμική ή μη γραμμική αυτοσυσχέτιση (μηδενική τιμή για γραμμική αυτοσυσχέτιση και αμοιβαία πληροφορία), γι' αυτό τις λέμε και *iid*. Μετρώντας διάφορα γραμμικά και μη-γραμμικά χαρακτηριστικά μεγέθη για την αρχική χρονοσειρά υπολοίπων και συγκρίνοντάς τα με τις κατανομές των *iid* θα μπορέσουμε να αποφανθούμε αν κι αυτή είναι ανεξάρτητη ή αν περιέχει κάποια επιπλέον πληροφορία, γραμμική η μη.

0.2.1 Εκτίμηση γραμμικών και μη γραμμικών στατιστικών

Αρχικά, υπολογίσαμε τις συναρτήσεις της γραμμικής αυτοσυσχέτισης και της αμοιβαίας πληροφορίας για τις πρώτες 4 υστερήσεις στα σχήματα 10 και 11.

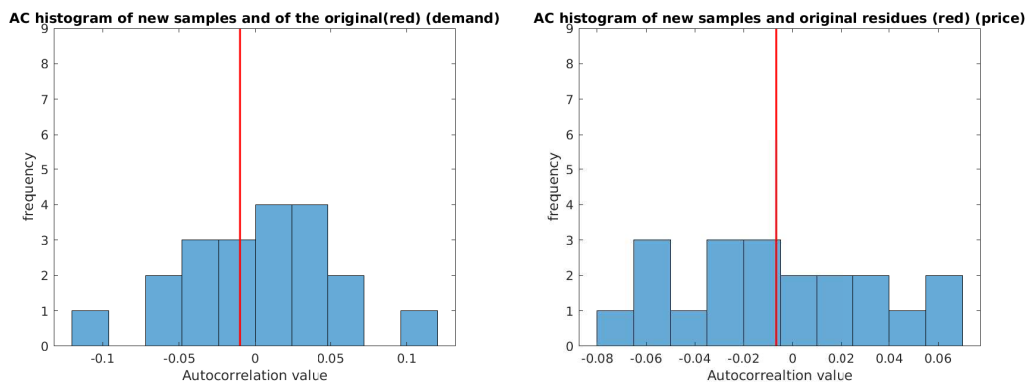


Σχήμα 10



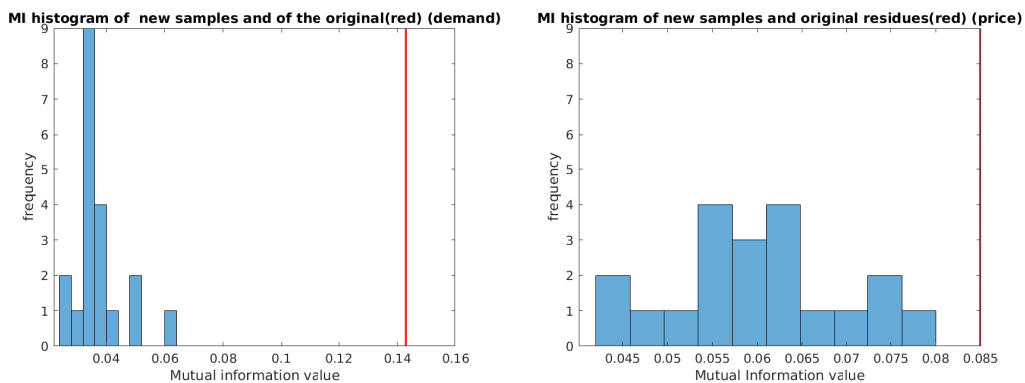
Σχήμα 11

Τα αντίστοιχα ιστογράμματα, για την περίπτωση $\tau = 1$ βρίσκονται στα σχήματα 12 και 13, όπου με κόκκινη γραμμή απεικονίζεται η τιμή των πραγματικών υπολοίπων ενώ με μπλε η κατανομή των μεταθέσεων



Σχήμα 12

Από τα παραπάνω ιστογράμματα, βλέπουμε ότι τόσο στην περίπτωση της ζήτησης όσο και της τιμής, η πραγματική τιμή της αυτοσυσχέτισης των υπολοίπων είναι εντός, και μάλιστα σχεδόν στο κέντρο της κατανομής της αυτοσυσχέτισης των μεταθέσεων, κάτι που μας δείχνει ότι και στις 2 περιπτώσεις στα υπόλοιπα δεν υπάρχει κάποια γραμμική συσχέτιση. Αυτό ουσιαστικά αποτελεί απόδειξη στο ότι η προσαρμογή του μοντέλου στο 3ο ερώτημα της γραμμικής ανάλυση ήταν επιτυχής, καθώς στα υπόλοιπα δεν υπάρχει καμία γραμμική πληροφορία.



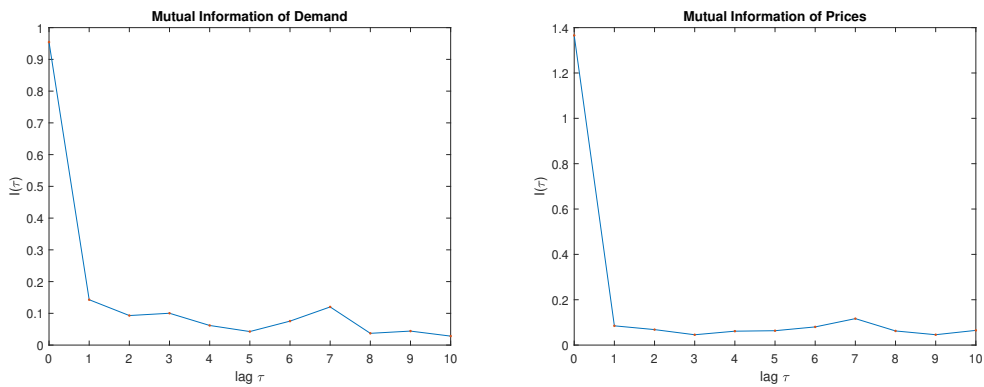
Σχήμα 13

Από την άλλη πλευρά βέβαια στο ιστόγραμμα της συνάρτησης της αμοιβαίας πληροφορίας (βλ. σχήμα 13), στη χρονοσειρά της ζήτησης, η τιμή της συνάρτησης αμοιβαίας πληροφορίας είναι αρκετά δεξιότερα της κατανομής της αμοιβαίας πληροφορίας των 20 iid μεταθέσεων. Ομοίως στην περίπτωση της χρονοσειράς της τιμής, βλέπουμε ότι είναι στα δεξιά. Συνεπώς, συνειδητοποιούμε ότι αφού στα υπόλοιπα παραμένει κάποια συσχέτιση και αυτή δεν είναι γραμμική, αναγκα-

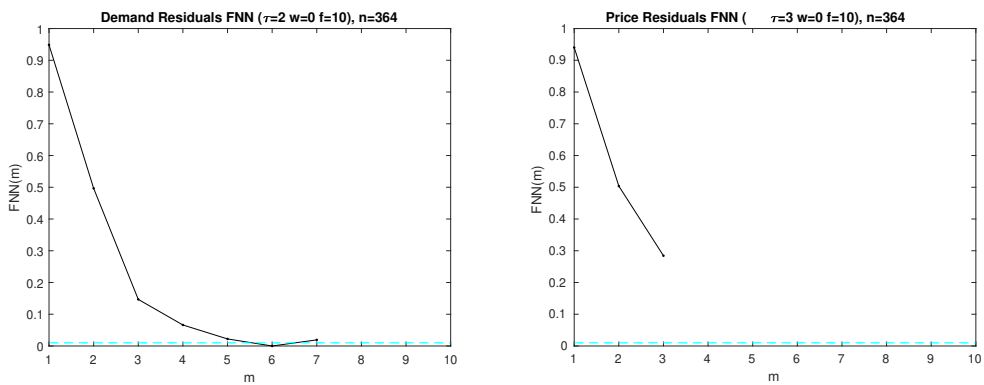
στικά είναι μη-γραμμική. Άρα το σύστημα μας έχει σίγουρα και μη-γραμμικό κομμάτι.

0.2.2 False Nearest Neighbors (FNN)

Εφαρμόσαμε επίσης ανάλυση και για το ποσοστό FNN για να βρούμε την διάσταση εμβύθισης m . Για να επιλέξουμε την κατάλληλη υστέρηση τ , εντοπίσαμε τα ελάχιστα στη συνάτηση αμοιβαίας πληροφορίας των υπολοίπων, η οποία δίνεται στο σχήμα 14.



Σχήμα 14



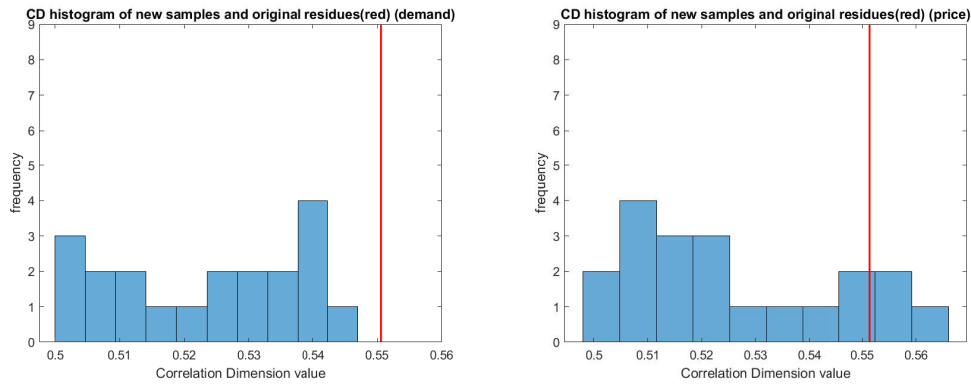
Σχήμα 15

Παρατηρώντας τα 2 διαγράμματα του σχήματος 14, επιλέξαμε $\tau = 2$ για τη ζήτηση και $\tau = 3$ για τη τιμή. Στην περίπτωση της ζήτησης φαίνεται και η επιλογή $\tau = 5$ ως κατάλληλη, αλλά λόγω του τρόπου με τον οποίο προέκυψε η αρχική μας χρονοσειρά (αραιή δειγματοληψία κάθε 24 ώρες), θεωρήσαμε πως

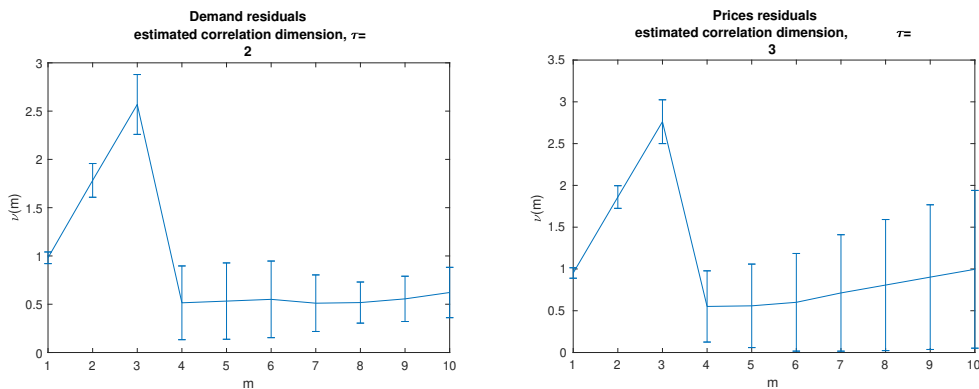
ένα τόσο μεγάλο τ θα αφαιρούσε χρήσιμη πληροφορία. Υπολογίζοντας λοιπόν το μέγεθος FNN και παρατηρώντας το σχήμα 15, επιλέξαμε ως κατάλληλη διάσταση εμφύθισης $m = 6$ για τη ζήτηση και $m = 4$ για την τιμή. Παρατηρούμε λοιπόν πως για την ανακατασκευή του ελκυστή μας, χρειάζεται και για τις 2 περιπτώσεις μεγάλο m (ακόμα περισσότερο για τη ζήτηση), άρα το σύστημά μας χαρακτηρίζεται ως υψηλής διάστασης και άρα υψηλής πολυπλοκότητας.

0.2.3 Διάσταση συσχέτισης (Correlation Dimension)

Ομοίως από τα ιστογράμματα στο σχήμα 16 παρατηρούμε ότι στην μία περίπτωση είναι εκτός, ενώ στην άλλη τείνει προς την άκρη, άρα έχει νόημα να υπολογίσουμε και τη διάσταση συσχέτισης από τα διαγράμματα στο σχήμα 17.



Σχήμα 16



Σχήμα 17

Από τα παραπάνω βλέπουμε ότι η διάσταση συσχέτισης ν και στις δύο χρονοσειρές των υπολοίπων είναι λίγο μεγαλύτερη του 0.5 και σαφώς μικρότερη

του 1, και θα λέγαμε ότι είναι κάπου στο $[0.5, 0.6]$ με κάποια αβεβαιότητα φυσικά. Με βάση την μη ακέραια τιμή της διάστασης συσχέτισης, μπορούμε να χαρακτηρίσουμε το σύστημα μας ως χαοτικό.

0.2.4 Τελικός χαρακτηρισμός συστήματος

Με βάση όλη την παραπάνω ανάλυση μπορούμε να βγάλουμε κάποια χρήσιμα συμπεράσματα για τα συστήματα που παράγουν τις χρονοσειρές μας. Παρατηρήσαμε, αρχικά, ότι και τα 2 χαρακτηρίζονται από μια αιτιοκρατική τάση και εποχικότητα-περιοδικότητα. Αφαιρώντας τες, καταφέραμε να προσαρμόσουμε γραμμικό στοχαστικό μοντέλο ARMA, που στη χρονοσειρά της ζήτησης είχε μεγάλο αιτιοκρατικό μέρος $AR(6)$ και ως στοχαστικό μέρος μόνο τον θόρυβο, ενώ στη χρονοσειρά της τιμής $ARMA(3, 3)$ είχε ίδιας τάξης αιτιοκρατικό και στοχαστικό μέρος. Στην πρώτη περίπτωση είχαμε να κάνουμε με σύστημα μεγάλης μνήμης (σχεδόν μία εβδομάδα) ενώ στη δεύτερη λίγο πιο μικρής μνήμης (περίπου μισή εβδομάδα). Η ύπαρξη μνήμης, φανερώνεται επίσης και από τις γραμμικές και μη γραμμικές συσχετίσεις κάποια βήματα πίσω. Ωστόσο, τα γραμμικά μοντέλα, δεν ήταν ικανοποιητικά για να περιγράψουν ολόκληρο το σύστημά μας, αλλά μόνο ένα μέρος αυτού, καθώς διαπιστώσαμε πως τα υπόλοιπα χαρακτηρίζονταν από μη-γραμμική συμπεριφορά. Τα ιστογράμματα που υπολογίσαμε, μας έδειξαν ότι οι πραγματικές τιμές των στατιστικών της χρονοσειράς σε σχέση με τις κατανομές των τιμών των *iid* χρονοσειρών διέφεραν, επομένως δεν ήταν τυχαίες, αλλά είχε νόημα να τις συμπεριλάβουμε στην ανάλυσή μας. Έτσι συνδυάσαμε την αυτοσυσχέτιση και την αμοιβαία πληροφορία, οι οποίες μας φανέρωσαν μη-γραμμικές συσχετίσεις. Όσον αφορά την διάσταση εμβύθισης, βρήκαμε να είναι και στις 2 περιπτώσεις σχετικά μεγάλη (6 και 4), επομένως χαρακτηρίσαμε τα συστήματά μας ως υψηλής διάστασης και πολυπλοκότητας. Τέλος, από τη διάσταση συσχέτισης που υπολογίστηκε κοντά στο 0.55, μη-ακέραιο ποσό, μπορέσαμε να χαρακτηρίσουμε τα συστήματά μας ως χαοτικά.

0.3 Επίλογος

Τα αποτελέσματα που προέκυψαν, θεωρούμε ότι κινούνται σε λογικά πλαίσια, καθώς έχουμε να κάνουμε με μια πραγματική χρονοσειρά που εξαρτάται από πολλούς παράγοντες, άρα αναμένουμε ένα τέτοιο σύστημα να είναι χαοτικό και μάλιστα υψηλής πολυπλοκότητας. Σίγουρα έχει ένα στοχαστικό μέρος, καθώς συμμετέχει ο θόρυβος, ωστόσο ένα κομμάτι του το θεωρήσαμε αιτιοκρατικό, γεγονός που απλοποίησε την ανάλυσή μας.