

ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΕΙΡΑΙΩΣ

ΣΧΟΛΗ ΒΙΟΜΗΧΑΝΙΑΣ ΚΑΙ ΝΑΥΤΙΛΙΑΣ
ΤΜΗΜΑ ΝΑΥΤΙΛΙΑΚΩΝ ΣΠΟΥΔΩΝ

ΣΧΟΛΗ ΝΑΥΤΙΚΩΝ ΔΟΚΙΜΩΝ
ΤΜΗΜΑ ΝΑΥΤΙΚΩΝ ΕΠΙΣΤΗΜΩΝ



ΤΜΗΜΑ ΝΑΥΤΙΛΙΑΚΩΝ ΣΠΟΥΔΩΝ

ΠΡΟΓΡΑΜΜΑ ΔΠΜΣ 'Διοίκηση στη Ναυτική Επιστήμη και Τεχνολογία'

Ανάλυση και Πρόβλεψη Χρονοσειρών με Εφαρμογές σε Θέματα Ναυτιλίας

Σούντρης Σταύρος

Διπλωματική Εργασία

που υποβλήθηκε στο Τμήμα Ναυτιλιακών Σπουδών
του Πανεπιστημίου Πειραιώς ως μέρος των
απαιτήσεων για την απόκτηση του Μεταπτυχιακού
Διπλώματος Ειδίκευσης στην 'Διοίκηση
στη Ναυτική Επιστήμη και Τεχνολογία'

Πειραιάς

Φεβρουάριος 2022



Σούντρης Σταύρος,
“Ανάλυση και Πρόβλεψη Χρονοσειρών
με Εφαρμογές σε Θέματα Ναυτιλίας”

Επιβλέπων:

Γαλάνης Γεώργιος

Καθηγητής ΣΝΔ

Διευθυντής Τομέα Μαθηματικών και Εργαστηρίου Μαθηματικής Μοντελοποίησης και
Εφαρμογών της Σχολής Ναυτικών Δοκίμων

Ανδρουλάκης Εμμανουήλ

Ε.Ε.ΔΙ.Π. ΙΙ ΣΝΔ

Συνεργάτης του Εργαστηρίου Μαθηματικής Μοντελοποίησης και Εφαρμογών της Σχολής
Ναυτικών Δοκίμων

ΔΗΛΩΣΗ ΑΥΘΕΝΤΙΚΟΤΗΤΑΣ / ΖΗΤΗΜΑΤΑ COPYRIGHT

Το άτομο το οποίο εκπονεί την Διπλωματική Εργασία φέρει ολόκληρη την ευθύνη προσδιορισμού της δίκαιης χρήσης του υλικού, η οποία ορίζεται στην βάση των εξής παραγόντων: του σκοπού και χαρακτήρα της χρήσης (εμπορικός, μη κερδοσκοπικός ή εκπαιδευτικός), της φύσης του υλικού που χρησιμοποιεί (τμήμα του κειμένου, πίνακες, σχήματα, εικόνες ή χάρτες), του ποσοστού και της σημαντικότητας των πιθανών συνέπειων αυτής στην αγορά ή στη γενικότερη αξία του υπό copyright κειμένου.

© 2022 – All rights reserved

Copyright © Σταύρος Σούντρης, 2022

Με επιφύλαξη παντός δικαιώματος.



*Σούντρης Σταύρος,
“Ανάλυση και Πρόβλεψη Χρονοσειρών
με Εφαρμογές σε Θέματα Ναυτιλίας”*

Η παρούσα Διπλωματική Εργασία εγκρίθηκε ομόφωνα από την Τριμελή Εξεταστική Επιτροπή που ορίστηκε από τη ΕΔιΕ του ΔΠΜΣ σύμφωνα με τον Κανονισμό Λειτουργίας του ΔΠΜΣ ‘Διοίκηση στη Ναυτική Επιστήμη και Τεχνολογία’.

Τα μέλη της Επιτροπής ήταν:

- Γεώργιος Γαλάνης, Καθηγητής Σχολής Ναυτικών Δοκίμων
- Εμμανουήλ Ανδρουλάκης, ΕΕΔΙΠ ΙΙ Σχολής Ναυτικών Δοκίμων
- Δήμητρα Κουλουμπού, Μέλος Εκπαιδευτικού Προσωπικού με Σύμβαση, Σχολή Ναυτικών Δοκίμων

Η έγκριση της Διπλωματικής Εργασίας από το Τμήμα Ναυτιλιακών Σπουδών του Πανεπιστημίου Πειραιώς δεν υποδηλώνει αποδοχή των γνωμών του συγγραφέα.»



Σούντρης Σταύρος,
“Ανάλυση και Πρόβλεψη Χρονοσειρών
με Εφαρμογές σε Θέματα Ναυτιλίας”

Πρόλογος

Η διπλωματική αυτή εργασία εκπονήθηκε στα πλαίσια των ερευνητικών δραστηριοτήτων του Δι-ιδρυματικού Μεταπτυχιακού Προγράμματος Σπουδών με τίτλο “Διοίκηση Στη Ναυτική Επιστήμη Και Τεχνολογία” του Τμήματος Ναυτιλιακών Σπουδών του Πανεπιστημίου Πειραιώς και του Τμήματος Ναυτικών Επιστημών της Σχολής Ναυτικών Δοκίμων κατά το κύκλο σπουδών 2021-2022.

Ευχαριστίες

“Θα ήθελα να ευχαριστήσω όλους όσους με βοήθησαν στην ολοκλήρωση της διπλωματικής εργασίας, η οποία σημαίνει και το τέλος του κύκλου σπουδών μου στο συγκεκριμένο μεταπτυχιακό τίτλο, το οποίο αποτέλεσε την αρχή σε ένα πολύ ενδιαφέρον μονοπάτι και σημαίνει πολλά για εμένα και την εξέλιξή μου και πρωτίστως να ευχαριστήσω ολόθερμα τους Καθηγητές μου κ. Γεώργιο Γαλάνη και κ. Εμμανουήλ Ανδρουλάκη για την ευκαιρία που μου δόθηκε να ασχοληθώ με το αντικείμενο των χρονοσειρών, για την καθοδήγησή τους με πολύτιμες συμβουλές και την καταλυτικής σημασίας βοήθεια καθ’ όλη τη διάρκεια της συνεργασίας μας.”

Αφιέρωση

“Αφιερώνω τη συγκεκριμένη εργασία και ό,τι έχω καταφέρει μέχρι στιγμής στους γονείς μου, Χρίστο και Παναγιώτα, και στην αδερφή μου, Ιωάννα που είναι δίπλα μου σε όλες τις δύσκολες στιγμές και ξέρω ότι θα είναι εκεί ό,τι και να χρειαστεί και στην Άννα για τη συνεχή υποστήριξη και αγάπη της αυτά τα δύο χρόνια.”



Περίληψη

Η ανάλυση και πρόβλεψη χρονοσειρών αποτελεί μία σημαντική γνωστική περιοχή που εξελίχθηκε μέσω της επιστήμης των Μαθηματικών και τη δημιουργία μοντέλων και μεθόδων παραγωγής προβλέψεων.

Η συγκεκριμένη διπλωματική εργασία πραγματεύεται την ανάλυση και πρόβλεψη χρονοσειρών. Η ανάλυση και πρόβλεψη μέσω χρονοσειρών είναι μείζονος σημασίας σε πολλούς τομείς της σύγχρονης κοινωνίας όπως είναι τα οικονομικά και η μηχανική. Επίσης, για τις εταιρείες του Ναυτιλιακού κλάδου αποτελεί ένα σημαντικό βοήθημα για την αποτελεσματική σχεδίαση και λήψη αποφάσεων, είτε πρόκειται για παράδειγμα για μια απόφαση κατασκευής νέου πλοίου, δηλαδή μια μακροπρόθεσμη κεφαλαιουχική επένδυση, είτε για την επιλογή είδους ναύλωσης ενός πλοίου μια συγκεκριμένη χρονική περίοδο.

Οι χρονοσειρές αποτελούν ένα πεδίο που έλκει το ενδιαφέρον της επιστήμης και των επιχειρήσεων. Στόχος της εργασίας είναι η βιβλιογραφική ανασκόπηση των βασικών μεθόδων ανάλυσης και πρόβλεψης χρονοσειρών, η πρακτική εφαρμογή τους με χρήση του προγράμματος MATLAB, καθώς και η παραγωγή προβλέψεων με μοντέλα ARIMA για τις εξαγωγές των κυριότερων λιμένων της Ελλάδας σε μεσοπρόθεσμο ορίζοντα.



Σούντρης Σταύρος,
“Ανάλυση και Πρόβλεψη Χρονοσειρών
με Εφαρμογές σε Θέματα Ναυτιλίας”

Abstract

Time Series Analysis and Forecasting is an important cognitive area that has evolved through the science of Mathematics and the creation of methodologies and models for the production of forecasts.

This Master thesis deals with Time Series Analysis and Forecasting, which is of major importance in many fields of modern society such as economics and engineering. Also, companies of the shipping cluster use it as an important aid for effective planning and decision-making, whether it is for example a decision to build a new ship, which is a long-term capital investment, or to choose a ship's charter type for a specific time period.

Time series is a field that attracts the interest of both science and business. The aim of this master thesis is a literature review of the basic methods of Time Series Analysis and Forecasting, a practical implementation using MATLAB, as well as a medium term forecast production using ARIMA models for Greece's main ports exports.



Πίνακας Περιεχομένων

Πρόλογος.....	4
Ευχαριστίες	4
Αφιέρωση.....	4
Περίληψη	5
Abstract	6
Πίνακας Περιεχομένων.....	7
Πίνακας Εικόνων.....	9
Κεφάλαιο 1: Εισαγωγή	11
1.1 Αντικείμενο της Εργασίας.....	11
Κεφάλαιο 2 : Βασικά χαρακτηριστικά των χρονοσειρών.....	12
2.1 Ορισμός της χρονοσειράς.....	12
2.2 Κατηγορίες χρονοσειρών.....	12
2.3 Βασικά στοιχεία χρονοσειρών	13
Κεφάλαιο 3 : Εισαγωγή στην Ανάλυση Χρονοσειρών.....	21
3.1 Στόχοι της Ανάλυσης Χρονοσειρών.....	21
3.2 Η στοχαστική διαδικασία.....	21
3.3 Ανάλυση Τάσης.....	21
3.4 Ανάλυση Εποχικότητας.....	24
Κεφάλαιο 4 : Πρόβλεψη, πλάνο και στόχοι.....	26
4.1 Αποφασίζοντας τι θα προβλέψουμε.....	26
4.2 Δεδομένα και μέθοδοι πρόβλεψης.....	26
4.3 Τα βασικά βήματα σε μια διαδικασία πρόβλεψης.....	28
Κεφάλαιο 5 : Τεχνικές Πρόβλεψης	29
5.1 Εισαγωγή στις βασικές αρχές των μοντέλων χρονοσειρών.....	29
5.2 Αυτοπαλίνδρομο Μοντέλο (AR ή Autoregressive)	29
5.3 Μοντέλο Κινητού Μέσου Όρου (MA ή Moving Average).....	30
5.4 Αυτοπαλίνδρομο Μοντέλο Κινητού Μέσου Όρου (ARMA ή Autoregressive Moving Average).....	31
5.5 Μοντέλο ARIMA ή Autoregressive Integrated Moving Average	32
Κεφάλαιο 6 : Ανάλυση και Πρόβλεψη Χρονοσειρών με χρήση MATLAB.....	34



Σούντρης Σταύρος,
“Ανάλυση και Πρόβλεψη Χρονοσειρών
με Εφαρμογές σε Θέματα Ναυτιλίας”

6.1	Εισαγωγή – Στόχοι των προβλέψεων	34
6.2	Παρουσίαση δεδομένων	36
6.3	Ανάλυση χρονοσειρών με Γραμμικά Μοντέλα	38
6.4	Παρουσίαση αποτελεσμάτων και διαγραμμάτων προβλέψεων	47
6.5	Μέτρηση Απόδοσης Μοντέλων Πρόβλεψης	49
Κεφάλαιο 7 :	Συμπεράσματα – Επίλογος.....	52
Κεφάλαιο 8:	Βιβλιογραφία.....	53



Πίνακας Εικόνων

Εικόνα 1. Παράδειγμα ντετερμινιστικού μοντέλου χρονοσειράς (οι τιμές εξαρτώνται σε μεγάλο βαθμό από το παρελθόν).....	13
Εικόνα 2. Χρονοσειρά που εμφανίζει τυχαιότητα και μερική θετική αυτοσυσχέτιση.....	13
Εικόνα 3. Η τάση στη χρονοσειρά απεικονίζεται με πορτοκαλί χρώμα.....	14
Εικόνα 4. Εποχική συνιστώσα της χρονοσειράς.....	14
Εικόνα 5. Τυχαία συνιστώσα χρονοσειράς (white noise).....	15
Εικόνα 6. Το ΑΕΠ της Ελλάδας την τελευταία δεκαετία.....	15
Εικόνα 7. Το καθαρό εθνικό διαθέσιμο εισόδημα στην Ελλάδα για τα έτη 1995-2020 (με διακεκομμένη η τάση της χρονοσειράς).....	16
Εικόνα 8. Συνολική Μηνιαία Αξία Εισαγωγών προέλευσης Ενδοκοινοτικού Εμπορίου στην Ελλάδα (Ιανουάριος 2019 - Αύγουστος 2021).....	17
Εικόνα 9. Η τιμή της μετοχής της Google κατά το 2013.....	17
Εικόνα 10. Διάγραμμα Αυτοσυσχέτισης τριμηνιαίου ποσοστού αλλαγής κατανάλωσης στις ΗΠΑ.....	19
Εικόνα 11. Διάγραμμα Μερικής Αυτοσυσχέτισης τριμηνιαίου ποσοστού αλλαγής κατανάλωσης στις ΗΠΑ.....	20
Εικόνα 12. Παράδειγμα διαγράμματος στάσιμης χρονοσειράς.....	22
Εικόνα 13. Παράδειγμα διαγράμματος μη στάσιμης χρονοσειράς.....	23
Εικόνα 14. Συνάρτηση Αυτοσυσχέτισης χρονοσειράς EXPORTVOLUMEDiff.....	24
Εικόνα 15. (γ) Ο εκτιμημένος ετήσιος κύκλος για την GICP. (δ) Η χρονοσειρά που προκύπτει από την αφαίρεση του εκτιμημένου ετήσιου κύκλου από τη χρονοσειρά στο (β).....	24
Εικόνα 16. Πίνακας μεθόδων πρόβλεψης.....	27
Εικόνα 17. Δεδομένα που χρησιμοποιήθηκαν για την ανάλυση.....	36
Εικόνα 18. Απεικόνιση της χρονοσειράς EXPORTVOLUME (ποσότητα φορτωθέντων εμπορευμάτων εξωτερικού στους κυριότερους λιμένες Ελλάδος). Στον άξονα Y, η ποσότητα των εμπορευμάτων σε χιλιάδες τόνους. Στον άξονα X, ο αύξων αριθμός παρατήρησης (δηλαδή το κάθε τρίμηνο με έναρξη το Γ' τρίμηνο 2004).....	37
Εικόνα 19. Γραφική παράσταση της ExportVolume.....	37
Εικόνα 20. Συνάρτηση αυτοσυσχέτισης της χρονοσειράς ExportVolume.....	38
Εικόνα 21. Γραφική παράσταση χρονοσειράς ExportVolume1Diff (σε εκατ. τόνους).....	39
Εικόνα 22. Συνάρτηση αυτοσυσχέτισης της χρονοσειράς ExportVolume1Diff.....	41
Εικόνα 23. Συνάρτηση μερικής αυτοσυσχέτισης της χρονοσειράς ExportVolume1Diff.....	41
Εικόνα 24. Αναπαράσταση σφαλμάτων μοντέλου ARIMA(1,1,0).....	43



Σούντρης Σταύρος,
“Ανάλυση και Πρόβλεψη Χρονοσειρών
με Εφαρμογές σε Θέματα Ναυτιλίας”

Εικόνα 25. Συνάρτηση αυτοσυσχέτισης των σφαλμάτων του μοντέλου ARIMA(1,1,0).....	43
Εικόνα 26. Αναπαράσταση σφαλμάτων μοντέλου ARIMA (3,1,0).....	45
Εικόνα 27. Συνάρτηση αυτοσυσχέτισης των σφαλμάτων του μοντέλου ARIMA (3,1,0).	46
Εικόνα 28. Απεικόνιση πρόβλεψης μοντέλου ARIMA(3,1,0) σε διάστημα εμπιστοσύνης 95%.....	48
Εικόνα 29. Απεικόνιση πρόβλεψης μοντέλου ARIMA(1,1,0) σε διάστημα εμπιστοσύνης 95%.....	49



Κεφάλαιο 1: Εισαγωγή

1.1 Αντικείμενο της Εργασίας

Στόχος αυτής της διπλωματικής εργασίας είναι αρχικά μια βιβλιογραφική έρευνα και ανασκόπηση των θεωρητικών στοιχείων και βασικών εννοιών Χρονοσειρών, καθώς και των βασικών αρχών και μεθοδολογιών Ανάλυσης και Πρόβλεψης Χρονοσειρών. Η τελευταία αποτελεί μια μεθοδολογία ιδιαίτερα σημαντικής χρησιμότητας, η οποία χρησιμοποιείται σε δεδομένα του Ναυτιλιακού Κλάδου και βοηθά στην ορθότερη λήψη αποφάσεων, είτε πρόκειται για χάραξη μακροπρόθεσμης στρατηγικής είτε βραχυπρόθεσμων ενεργειών.

Αρχικά παρουσιάζεται μια εισαγωγή στο πεδίο των χρονοσειρών. Ορίζεται η χρονοσειρά σαν έννοια και αναλύονται τα στοιχεία από τα οποία αποτελείται συζητώντας επίσης σχετικές κατηγορίες βάσει των διαφόρων χαρακτηριστικών τους και για την καλύτερη κατανόηση δίνονται παραδείγματα στατιστικών δεδομένων τα οποία αποτελούν χρονοσειρά.

Στο τρίτο κεφάλαιο γίνεται εισαγωγή στην ανάλυση χρονοσειρών και των επιμέρους συνιστωσών τους, καθώς και στο στόχο που αποσκοπεί, ώστε να εξαχθούν συμπεράσματα για τα χαρακτηριστικά τους και τον τρόπο με τον οποίο θα πρέπει να προετοιμαστούν, ώστε να γίνει χρήση των μοντέλων πρόβλεψης, καθώς και συζήτηση για το ποια μοντέλα πρόβλεψης είναι καταλληλότερα για παραγωγή προβλέψεων για κάθε είδος χρονοσειράς.

Το τέταρτο κεφάλαιο αποτελεί θεωρητική εισαγωγή στην επιστήμη της πρόβλεψης, όπου θα αναλυθεί η νοοτροπία και τα βασικά βήματα της διαδικασίας που πρέπει να λαμβάνονται υπόψη, ενώ θα αναφερθούν και οι βασικές μέθοδοι πρόβλεψης. Εν συνεχεία, εστιάζουμε στην πρόβλεψη χρονοσειρών και στις μεταβλητές που την καθορίζουν.

Στο πέμπτο κεφάλαιο αναλύονται οι βασικότερες υφιστάμενες τεχνικές πρόβλεψης χρονοσειρών και παρουσιάζονται εκτενώς τα μοντέλα και οι μεθοδολογίες που χρησιμοποιούνται.

Μετά την θεωρητική ανάλυση όλων των επιμέρους στοιχείων που αναφέρθηκαν γίνεται πρόβλεψη μεσοπρόθεσμου ορίζοντα σε δεδομένα που ανακτήθηκαν από την Ελληνική Στατιστική Αρχή και αφορούν τις εξαγωγές των κυριότερων λιμένων της Ελλάδας χρησιμοποιώντας το υπολογιστικό περιβάλλον MATLAB, το οποίο αποτελεί το εργαλείο για την πρακτική εφαρμογή των παραπάνω μεθόδων και για την παραγωγή προβλέψεων με χρήση πραγματικών δεδομένων.

Πραγματοποιήθηκε συγκεκριμένα πρακτική εφαρμογή των μεθοδολογιών αναπτύσσοντας ολοκληρωμένα αυτοπαλίνδρομα μοντέλα κινητού μέσου όρου (ARIMA). Τέλος, επιχειρήθηκε μία αξιολόγηση της απόδοσης των μοντέλων που χρησιμοποιήθηκαν για την εξαγωγή συμπερασμάτων σε σχέση με τα διαθέσιμα δεδομένα.



Κεφάλαιο 2 : Βασικά χαρακτηριστικά των χρονοσειρών

2.1 Ορισμός της χρονοσειράς

Ως χρονοσειρά ορίζεται μια ιεραρχημένη ως προς το χρόνο ακολουθία παρατηρήσεων μιας μεταβλητής που λαμβάνονται σε τακτά χρονικά διαστήματα. Η ανάλυση μιας χρονοσειράς στοχεύει στην εύρεση των χαρακτηριστικών που συμβάλουν στην κατανόηση της συμπεριφοράς της μεταβλητής και επιτρέπουν τη δημιουργία στατιστικών μοντέλων που θα προβλέπουν επαρκώς τις μελλοντικές τιμές του υπό εξέταση χαρακτηριστικού.

Το εύρος μεθόδων μέσω της στατιστικής που χρησιμοποιούνται για την ανάλυση τους λέγεται Ανάλυση Χρονοσειρών.

Οι παρατηρήσεις μιας χρονοσειράς ενδεχομένως να εκφράζουν διαφόρων ειδών δεδομένα οποιουδήποτε φυσικού μεγέθους, όπως οικονομικά, καιρικά φαινόμενα κτλ. Αυτός είναι ουσιαστικά και ο λόγος που έχουν τεράστιο εύρος εφαρμογής. Όταν, δε, υπάρχει δυνατότητα ταυτόχρονης παρατήρησης περισσότερου του ενός μεγέθους στο ίδιο σύστημα, τότε έχουμε τις ονομαζόμενες πολλαπλές ή πολυδιάστατες χρονοσειρές.

2.2 Κατηγορίες χρονοσειρών

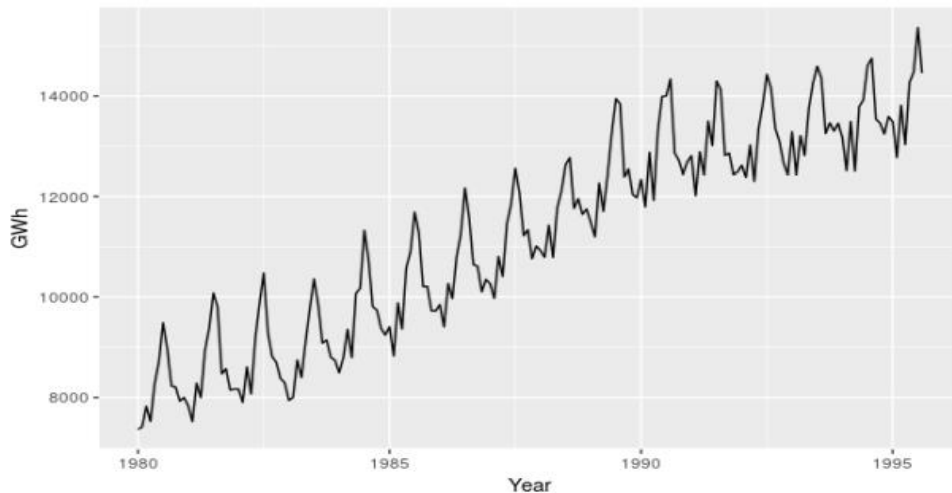
Οι χρονοσειρές κατηγοριοποιούνται βάσει συγκεκριμένων χαρακτηριστικών τους και η κατηγοριοποίησή τους αποτελεί ένα αυτοτελές μέρος γύρω από το οποίο έχει αναπτυχθεί μια ολόκληρη επιστημονική μεθοδολογία, ωστόσο η συγκεκριμένη εργασία δεν θα πραγματεύεται το συγκεκριμένο θέμα, οπότε θα αναφερθούμε ενδεικτικά σε κάποια βασικά χαρακτηριστικά τους.

Οι χρονοσειρές διακρίνονται σε:

1. **Συνεχείς και σε Διακριτές.** Συνεχείς είναι οι χρονοσειρές των οποίων οι τιμές παρατηρούνται συνεχώς ή σε συνεχή διαστήματα και διακριτές είναι εκείνες των οποίων οι τιμές παρατηρούνται σε ορισμένα χρονικά διαστήματα.
2. **Γραμμικές και μη Γραμμικές.** Γραμμικές είναι αυτές όπου, κάθε σημείο δεδομένων μπορεί να θεωρηθεί ως γραμμικός συνδυασμός προηγούμενων ή μελλοντικών τιμών. Οι μη γραμμικές χρονικές σειρές δημιουργούνται από μη γραμμικές εξισώσεις.

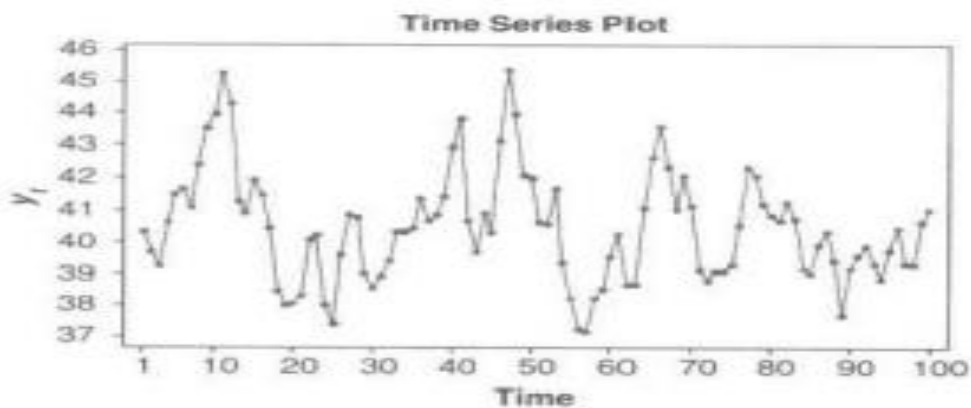
Η κατηγοριοποίηση των χρονοσειρών μπορεί επίσης να βασίζεται στον τρόπο προσδιορισμού των μελλοντικών τιμών τους, δηλαδή κατά πόσο εξαρτώνται από τις προηγούμενες διαδοχικές παρατηρήσεις που τις απαρτίζουν. Όταν οι παρατηρήσεις δεν είναι ανεξάρτητες μεταξύ τους, τότε αυτοσυσχετίζονται και οι επόμενες τιμές μπορούν να προσδιοριστούν σε ένα ποσοστό βάσει των παρελθοντικών. Αυτές οι χρονολογικές σειρές παρουσιάζουν θετική αυτοσυσχέτιση, καθώς οι προηγούμενες τιμές τους τείνουν να ακολουθούνται από παρόμοιες τιμές (Αθηνιώτης, 2019).

Ένα σύστημα με τα παραπάνω χαρακτηριστικά λέγεται ντετερμινιστικό. Παρακάτω παρατίθεται μια χρονοσειρά, της οποίας οι παρατηρήσεις εμφανίζουν θετική αυτοσυσχέτιση, καθώς οι τιμές της κυμαίνονται μεταξύ ενός σταθερού ορίου με την πάροδο του χρόνου (Hyndman & Athanasopoulos, 2018).



Εικόνα 1. Παράδειγμα ντετερμινιστικού μοντέλου χρονοσειράς (οι τιμές εξαρτώνται σε μεγάλο βαθμό από το παρελθόν)

Ωστόσο σε άλλες χρονοσειρές, οι μελλοντικές τιμές δεν βασίζονται μόνο στις παρελθοντικές, αλλά εμπεριέχουν και την τυχαιότητα, οπότε σε αυτή τη περίπτωση έχουμε τις λεγόμενες **στοχαστικές χρονοσειρές**, όπως η παρακάτω της οποίας οι τιμές φαίνεται να περιέχουν το στοιχείο της τυχαιότητας, αλλά είναι αυτοσυσχετιζόμενες (παρακάτω θα εξηγήσουμε αναλυτικά τι σημαίνει αυτοσυσχέτιση) σε κάποιο βαθμό με τις προηγούμενες τιμές τους.



Εικόνα 2. Χρονοσειρά που εμφανίζει τυχαιότητα και μερική θετική αυτοσυσχέτιση

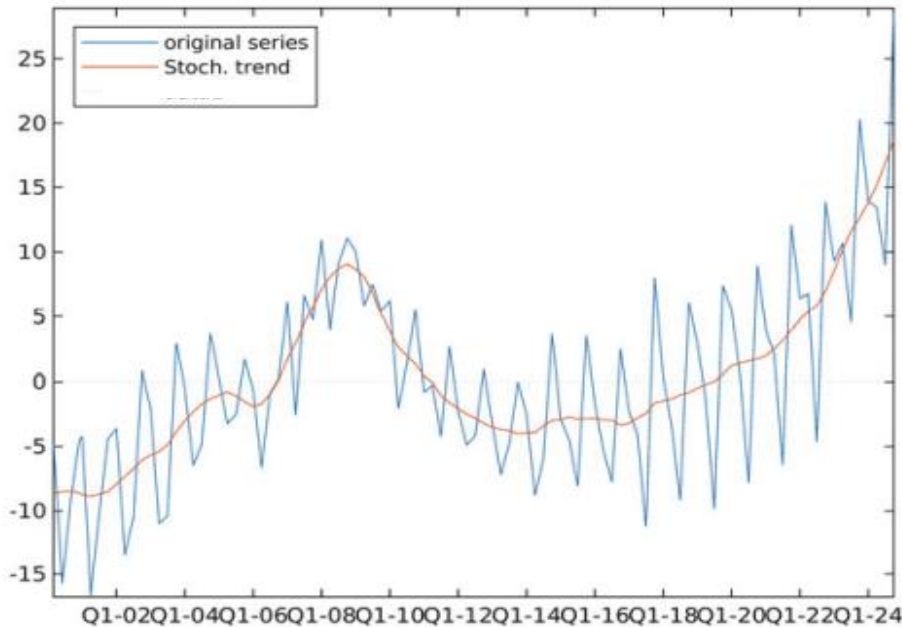
Οι χρονοσειρές αποτελούνται από τέσσερα συστατικά στοιχεία: Τάση, κυκλική μεταβλητότητα, εποχική ή περιοδική μεταβλητότητα και τυχαιά ή ακανόνιστη μεταβλητότητα. Τα στοιχεία αυτά θα αναλυθούν στο παρακάτω μέρος του κεφαλαίου και θα μας απασχολήσουν στη συνέχεια στο κομμάτι της ανάλυσης (Montgomery et al, 2008).

2.3 Βασικά στοιχεία χρονοσειρών

Όπως αναφέρθηκε, οι χρονοσειρές μπορεί να αποτελούνται από τις εξής τέσσερις συνιστώσες:

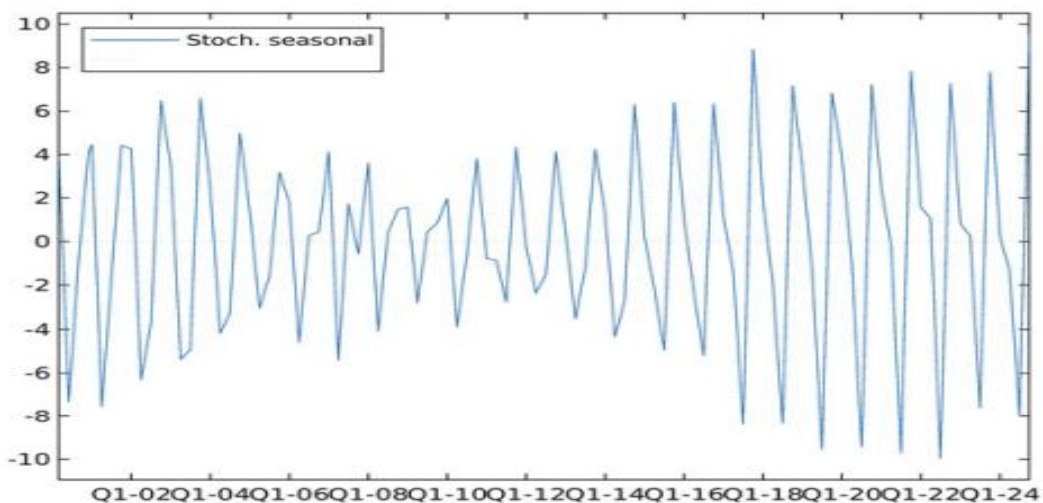


- **Τάση**, η οποία δείχνει τη κατεύθυνση της χρονοσειράς μακροπρόθεσμα. Μπορεί να αυξάνεται, να μειώνεται ή να παραμένει σταθερή για μεγάλο χρονικό διάστημα.



Εικόνα 3. Η τάση στη χρονοσειρά απεικονίζεται με πορτοκαλί χρώμα.

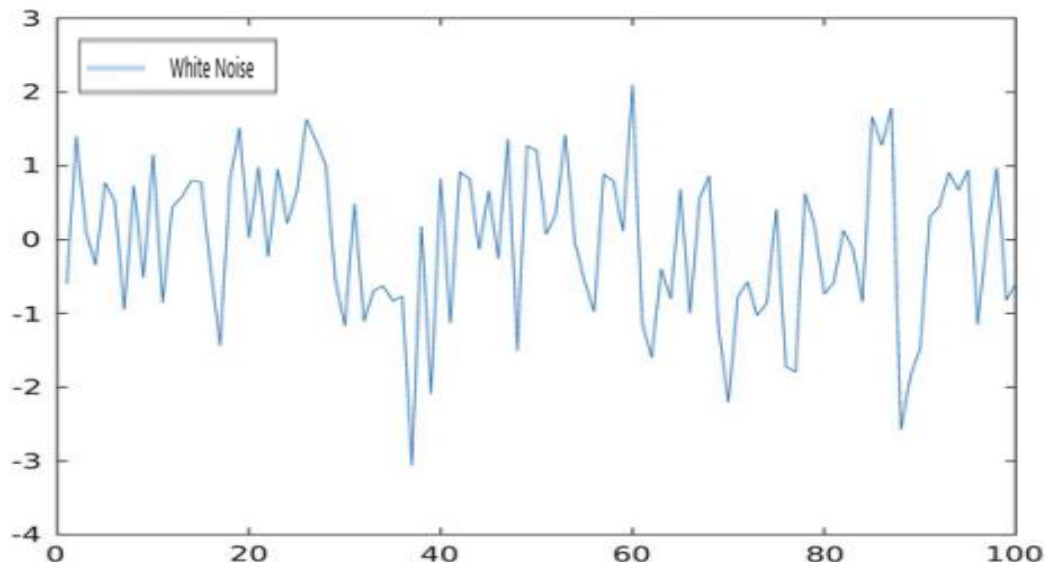
- **Κυκλική μεταβλητότητα**, η οποία περιγράφει τις μεσοπρόθεσμες αλλαγές που προκαλούνται από συνθήκες, που επαναλαμβάνονται σε κύκλους χωρίς όμως σταθερή συχνότητα. Οι περισσότερες από τις οικονομικές και χρηματοπιστωτικές χρονοσειρές παρουσιάζουν κάποιο είδος κυκλικής μεταβλητότητας.
- **Εποχική μεταβλητότητα ή Περιοδική μεταβλητότητα**, η οποία εμφανίζεται όταν μια χρονοσειρά επηρεάζεται από εποχικούς παράγοντες πχ. μια εποχή του χρόνου ή μία μέρα της εβδομάδας. Είναι σταθερή με γνωστή συχνότητα.



Εικόνα 4. Εποχική συνιστώσα της χρονοσειράς

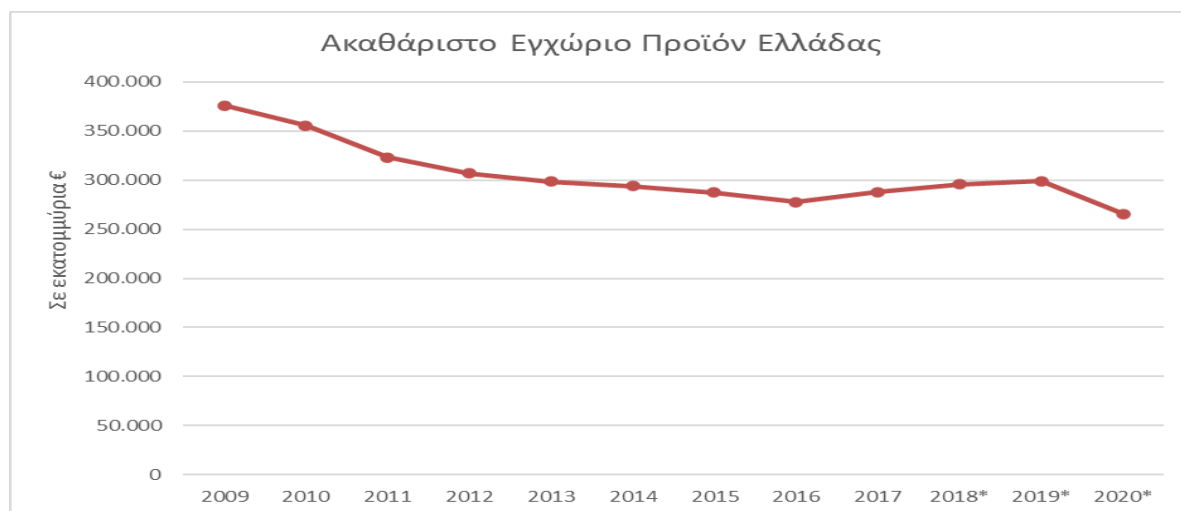


- **Τυχαία μεταβλητότητα ή Ακανόνιστη μεταβλητότητα**, η οποία οδηγεί σε αφύσικες τιμές που προκαλούνται από απρόβλεπτες επιρροές, οι οποίες δεν είναι κανονικές και επίσης δεν επαναλαμβάνονται με κάποιο συγκεκριμένο τρόπο.



Εικόνα 5. Τυχαία συνιστώσα χρονοσειράς (white noise)

Τα γραφήματα χρονοσειρών μπορούν να αποκαλύψουν μοτίβα όπως τυχαιότητα, τάσεις, μεταβολές επιπέδου, περιόδους ή κύκλους, ασυνήθιστες παρατηρήσεις ή συνδυασμό μοτίβων. Μοτίβα τα οποία βρίσκονται συνήθως στα δεδομένα των χρονοσειρών θα αναλυθούν στο Κεφάλαιο 3. Παρακάτω θα δώσουμε μερικά παραδείγματα δεδομένων ώστε να κατανοήσουμε τα βασικά στοιχεία μιας χρονοσειράς. Η παρακάτω χρονοσειρά είναι το ΑΕΠ της Ελλάδος από το 2009 έως το 2020. Βλέπουμε ότι τα δεδομένα εμφανίζουν μια μακροχρόνια αρνητική τάση, δηλαδή παρουσιάζουν μια συνεχόμενη μείωση.

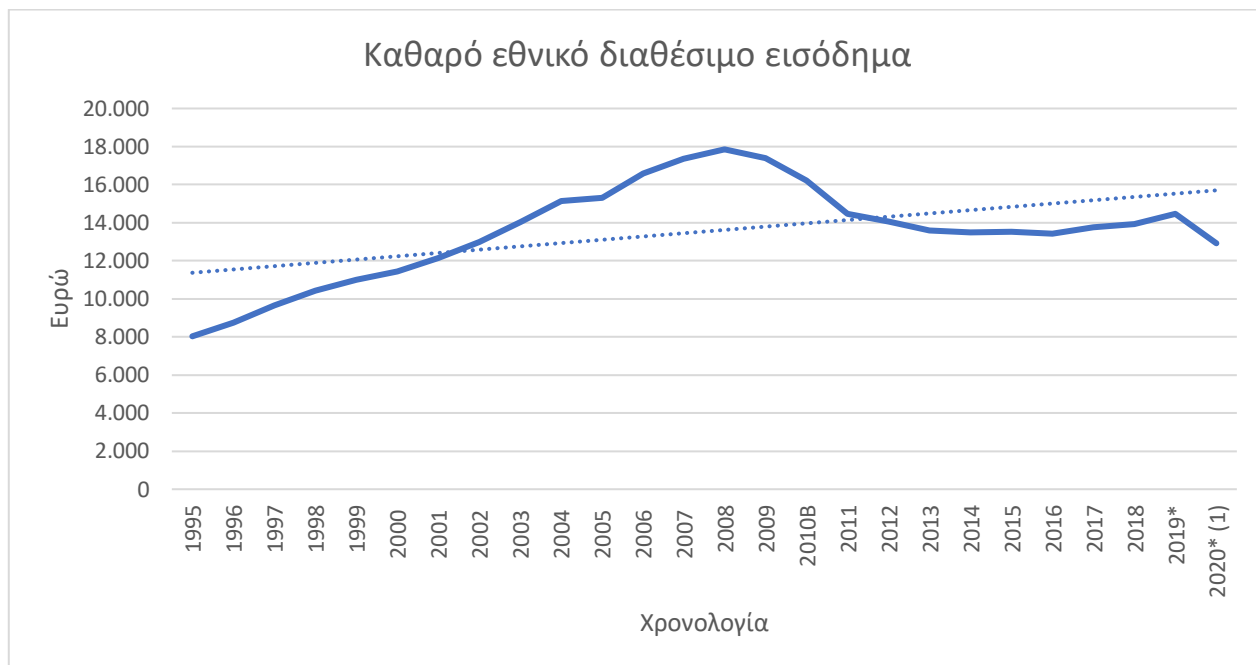


Εικόνα 6. Το ΑΕΠ της Ελλάδας την τελευταία δεκαετία¹

¹ Τα στοιχεία της χρονοσειράς ανακτήθηκαν από την Ελληνική Στατιστική Αρχή στο σύνδεσμο <https://www.statistics.gr/el/statistics/-/publication/SEL84/->



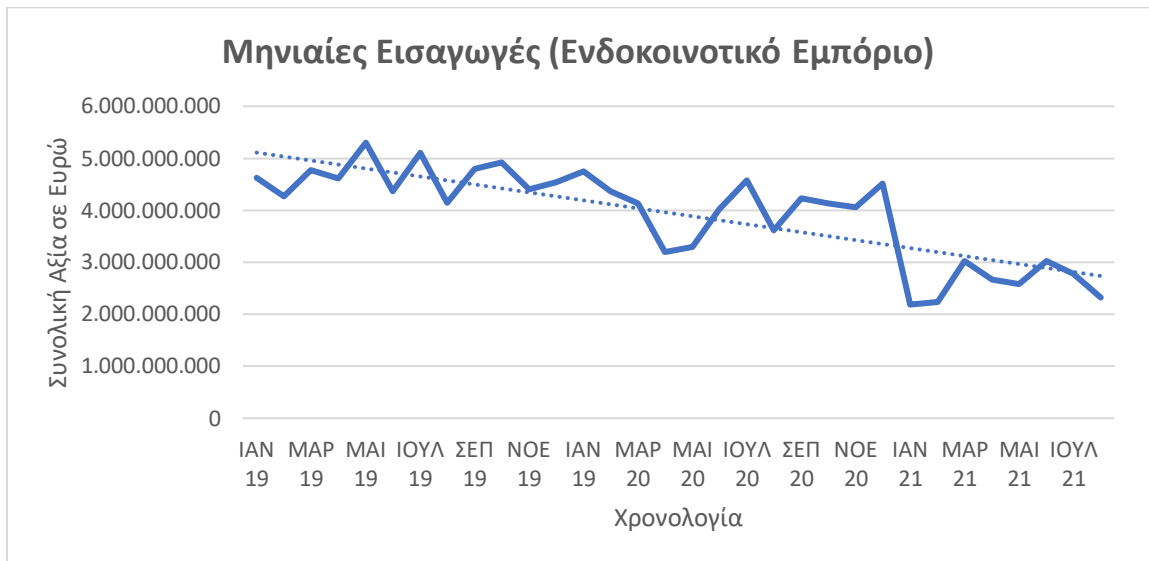
Αντίστοιχα στο επόμενο γράφημα εμφανίζεται το καθαρό εθνικό διαθέσιμο εισόδημα στην Ελλάδα, όπου παρουσιάζεται μια αυξητική μακροχρόνια τάση.



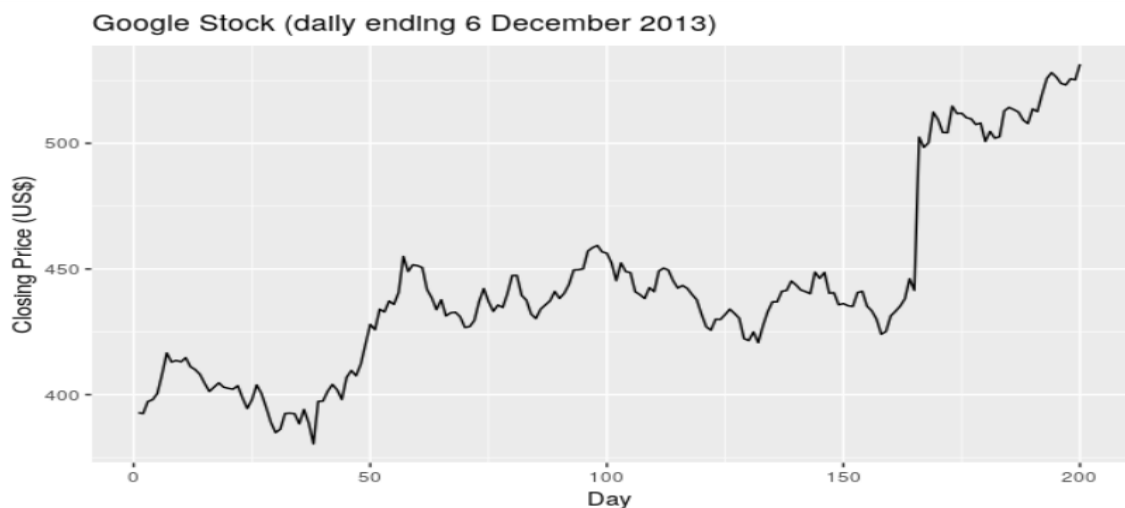
Εικόνα 7. Το καθαρό εθνικό διαθέσιμο εισόδημα στην Ελλάδα για τα έτη 1995-2020 (με διακεκομμένη η τάση της χρονοσειράς)²

Άλλες χρονοσειρές ενδέχεται να εμφανίζουν μιας μορφής κυκλικότητα ή εποχικότητα. Στο γράφημα που ακολουθεί παρουσιάζονται τα μηνιαία στοιχεία εισαγωγών της Ελλάδας που προέρχονται από εμπόριο εντός Ευρωπαϊκής Ένωσης για τα τελευταία τρία χρόνια. Η χρονοσειρά εμφανίζει στοιχεία κυκλικότητας, καθώς παρατηρείται άνοδος και πτώση της αξίας ανά χρονικά διαστήματα. Η χρονοσειρά εμφανίζει ταυτόχρονα και τάση μείωσης.

² Τα στοιχεία της χρονοσειράς ανακτήθηκαν από την Ελληνική Στατιστική Αρχή στο σύνδεσμο <https://www.statistics.gr/el/statistics/-/publication/SEL33/->



Εικόνα 8. Συνολική Μηνιαία Αξία Εισαγωγών προέλευσης Ενδοκοινοτικού Εμπορίου στην Ελλάδα (Ιανουάριος 2019 - Αύγουστος 2021)³



Εικόνα 9. Η τιμή της μετοχής της Google κατά το 2013

Οι μετοχές αποτελούν χρονοσειρές, των οποίων συνήθως η ανάλυση βασίζεται στη θεωρία του τυχαίου περιπάτου, δηλαδή μιας τυχαίας και απρόβλεπτης πορείας που καθιστά οποιαδήποτε μέθοδο πρόβλεψης μάταια σε μακροπρόθεσμο ορίζοντα, και αυτό γιατί εμφανίζουν ως βασικό συστατικό την τυχαιότητα, καθώς επηρεάζονται από διάφορα τυχαία γεγονότα.

Βασικό χαρακτηριστικό στην ανάλυση χρονοσειρών είναι η στασιμότητα.

Στάσιμη ονομάζεται η χρονοσειρά με χαρακτηριστικά που δεν εξαρτώνται από το χρόνο της παρατήρησης.

³ Τα στοιχεία της χρονοσειράς ανακτήθηκαν από την Ελληνική Στατιστική Αρχή στο σύνδεσμο <https://www.statistics.gr/el/statistics/-/publication/SFC02/2021-M11>



Οπότε, μια χρονοσειρά με τάση, ή εποχικότητα δεν μπορεί να είναι στάσιμη. Η τάση και η εποχικότητα επηρεάζουν τη τιμή της χρονοσειράς σε διαφορετικούς χρόνους. Από την άλλη, μια χρονοσειρά λευκού θορύβου είναι στάσιμη, γιατί δεν έχει σημασία πότε θα την παρατηρήσει κανείς, μοιάζει ίδια οποιαδήποτε χρονική στιγμή. Μερικές φορές μπορεί να είναι δύσκολο να διακριθεί η στασιμότητα μιας χρονοσειράς. Μια χρονοσειρά που παρουσιάζει κυκλικότητα, χωρίς τάση και εποχικότητα, θεωρείται στάσιμη, διότι οι κύκλοι δεν έχουν σταθερή συχνότητα, οπότε πριν την παρατήρηση υπάρχει αβεβαιότητα για τη θέση των κύκλων στη χρονοσειρά.

Σημαντική έννοια στις χρονοσειρές είναι η υστέρηση και ο συντελεστής υστέρησης (Lag operator). Η **υστέρηση** είναι η χρονική μετατόπιση μιας μεταβλητής κάποιες περιόδους ή παρατηρήσεις πίσω. Για παράδειγμα, για ένα σύνολο δεδομένων Y_1, Y_2, \dots, Y_n , τα Y_2 και Y_7 έχουν $\text{lag} = 7 - 2 = 5$. Στην ανάλυση χρονοσειρών, ο συντελεστής υστέρησης (L) επιδρά σε μια χρονοσειρά για την παραγωγή του προηγούμενου στοιχείου ως εξής:

$$L^i y_t = y_{t-i}, \text{ όπου } y_t \text{ μια χρονοσειρά με } t = 0, 1, 2, \dots, i$$

Σημαντική έννοια είναι η τάξη (order) ενός μοντέλου χρονοσειρών. Η τάξη ενός μοντέλου ορίζεται ως το πλήθος των παρελθοντικών τιμών ή παραγόντων, οι οποίες συμβάλλουν σημαντικά στη διαμόρφωση της τρέχουσας τιμής και λαμβάνονται υπόψη σε ένα μοντέλο πρόβλεψης.

Η αυτοσυσχέτιση και η μερική αυτοσυσχέτιση είναι έννοιες πολύ χρήσιμες στην ανάλυση και πρόβλεψη χρονοσειρών. Η αυτοσυσχέτιση μπορεί να χαρακτηριστεί ως «μνήμη» σε μια χρονοσειρά που σημαίνει ότι οι τιμές των δεδομένων σχετίζονται μεταξύ τους. Στόχος της ανάλυσης είναι να δημιουργηθεί μια συνάρτηση αυτοσυσχέτισης, η οποία εκμεταλλεύομενη τις τιμές να περιγράφει όσο το δυνατό καλύτερα τη σχέση των δεδομένων μεταξύ τους, πως δηλαδή οι προηγούμενες τιμές της χρονοσειράς επηρεάζουν τις επόμενες, έτσι ώστε να έχουμε όσο το δυνατό λιγότερο «θόρυβο» ή αλλιώς τυχαιότητα. Ουσιαστικά με τη βοήθειά της δημιουργείται ένα αυτοπαλίνδρομο μοντέλο για τη χρονοσειρά.

Η αυτοσυσχέτιση χρησιμοποιείται στην ανάλυση χρονοσειρών για δύο λόγους. Πρώτον, για να εντοπιστεί η μη τυχαιότητα στα δεδομένα, δηλαδή να ποσοτικοποιήσουμε τη «μνήμη» στη χρονοσειρά. Δεύτερον, εάν τα δεδομένα δεν είναι τυχαία χρησιμοποιείται για να ταυτοποιήσουμε ένα κατάλληλο μοντέλο που θα περιγράφει τη χρονοσειρά. Η αυτοσυσχέτιση είναι η συσχέτιση μιας χρονοσειράς και της ίδιας χρονοσειράς σε διαφορετικό χρόνο. Η χρονική διαφορά αυτή ορίζεται ως υστέρηση ή αλλιώς lag όπως αναφέρθηκε παραπάνω.

Τα διαγράμματα αυτοσυσχέτισης είναι ένα εργαλείο που χρησιμοποιείται συνήθως για τον έλεγχο της τυχειότητας σε ένα σύνολο δεδομένων. Αυτή η τυχειότητα εξακριβώνεται με τον υπολογισμό των αυτοσυσχετίσεων για τις τιμές δεδομένων σε διαφορετικές χρονικές καθυστερήσεις. Εάν είναι τυχαίες, οι εν λόγω αυτοσυσχετίσεις θα πρέπει να είναι σχεδόν μηδενικές για όλες τις τιμές χρονικής καθυστέρησης. Εάν δεν είναι τυχαία, τότε μία ή περισσότερες από τις αυτοσυσχετίσεις θα είναι σημαντικά μη μηδενικές. Επιπλέον, τα διαγράμματα αυτοσυσχέτισης χρησιμοποιούνται στο στάδιο αναγνώρισης μοντέλου για Box-Jenkins αυτοπαλίνδρομο μοντέλα κινούμενου μέσου όρου (Box & Jenkins, 1970). Σημειώνεται ότι η μη συσχέτιση των δεδομένων δεν σημαίνει απαραίτητα τυχαιότητα. Τα δεδομένα που έχουν σημαντική αυτοσυσχέτιση δεν είναι τυχαία. Ωστόσο, τα δεδομένα που δεν δείχνουν σημαντική αυτοσυσχέτιση μπορούν ακόμα να παρουσιάσουν μη τυχαιότητα με άλλους τρόπους. Η αυτοσυσχέτιση είναι μόνο ένα μέτρο τυχειότητας. Στο πλαίσιο της επικύρωσης μοντέλου, ο έλεγχος για αυτοσυσχέτιση είναι συνήθως μια επαρκής δοκιμή τυχειότητας. Τα διαγράμματα αυτοσυσχέτισης σχηματίζονται από:



Κάθετος Άξονας: Συντελεστής Αυτοσυσχέτισης

$$Rh = \frac{Ch}{C_0},$$

όπου Ch η συνάρτηση αυτοδιακύμανσης

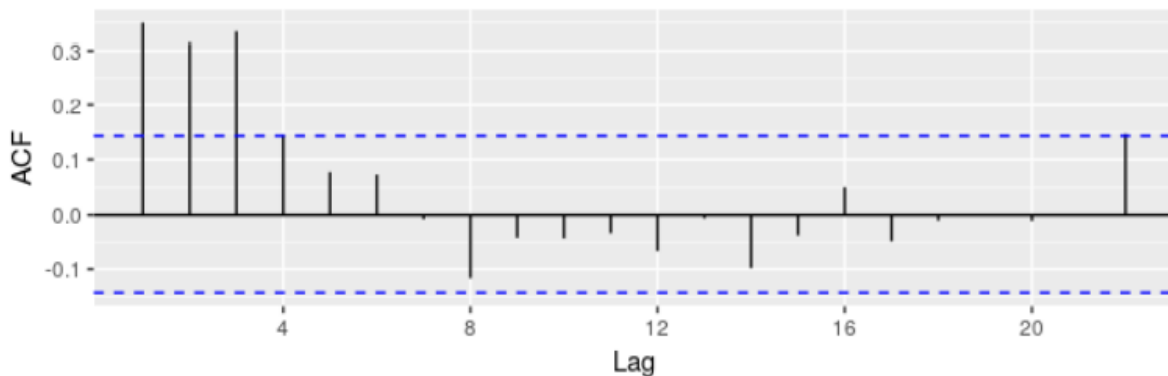
$$Ch = \frac{1}{N} \sum_{t=1}^{N-h} (Y_t - \bar{Y})(Y_{t+h} - \bar{Y}),$$

όπου Y η μεταβλητή, \bar{Y} η μέση τιμή του δείγματος, h η χρονική υστέρηση, N ο αριθμός των παρατηρήσεων του δείγματος και C_0 η συνάρτηση διακύμανσης

$$C_0 = \sum_{t=1}^N \frac{(Y_t - \bar{Y})^2}{N},$$

όπου Y η μεταβλητή, \bar{Y} η μέση τιμή του δείγματος, N ο αριθμός των παρατηρήσεων του δείγματος. Σημειώνεται ότι το Rh είναι μεταξύ -1 και $+1$.

Οριζόντιος Άξονας: Χρονική υστέρηση h ($h = 1, 2, 3, \dots$)



Εικόνα 10. Διάγραμμα Αυτοσυσχέτισης τριμηνιαίου ποσοστού αλλαγής κατανάλωσης στις ΗΠΑ

Το διάγραμμα αυτοσυσχέτισης μπορεί να δώσει απαντήσεις στις ακόλουθες ερωτήσεις:

- Υπάρχει αυτοσυσχέτιση μεταξύ των παρατηρήσεων των δεδομένων;
- Είναι η παρατηρούμενη χρονοσειρά «λευκός θόρυβος»;
- Υπάρχει κάποια σαφής σχέση μεταξύ των αυτοσυσχετίσεων της χρονοσειράς για κάθε υστέρηση;
- Ποιο είναι το κατάλληλο μοντέλο για τη συγκεκριμένη χρονοσειρά;

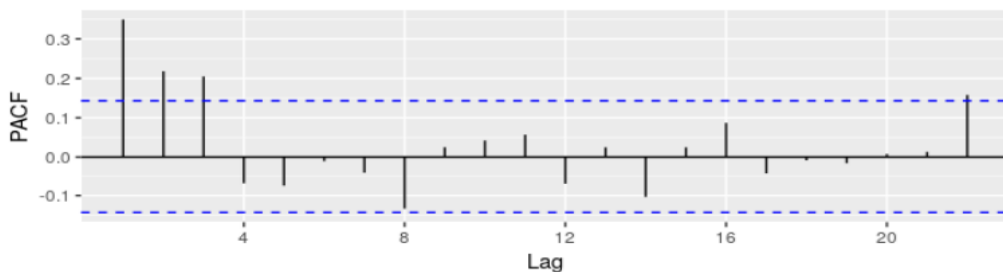
Το διάγραμμα ACF δείχνει τις αυτοσυσχετίσεις που μετρούν τη σχέση μεταξύ y_t και y_{t-k} για διαφορετικές τιμές του k . Αν το y_t και το y_{t-1} συσχετίζονται τότε και το y_t με το y_{t-2} θα πρέπει να συσχετίζονται, απλά επειδή συνδέονται με το y_{t-1} και όχι επειδή κάποια νέα πληροφορία μπορεί να περιλαμβάνεται στο y_{t-2} , η οποία θα μπορούσε να χρησιμοποιηθεί για να προβλέψει το y_t . Δεδομένου ότι στις χρονοσειρές με αυτοσυσχέτιση οι μελλοντικές τιμές επηρεάζονται σε κάποιο βαθμό από τις προηγούμενες τιμές και πιο συγκεκριμένα όσο πιο κοντά στο παρελθόν είναι μια τιμή τόσο μεγαλύτερη είναι και η επίδρασή της στη τιμή που θέλουμε να προβλέψουμε. Για παράδειγμα, η ποσότητα κατανάλωσης δημητριακών από τους καταναλωτές τον Δεκέμβριο και τον Ιανουάριο επηρεάζει τη ποσότητα κατανάλωσης δημητριακών τον Φεβρουάριο. Ο Ιανουάριος στατιστικά θα επηρεάζει περισσότερο τον



Φεβρουάριο, ωστόσο και ο Δεκέμβριος επηρεάζει έμμεσα τον Φεβρουάριο μέσω του Ιανουαρίου. Αυτό ορίζεται ως αυτοσυσχέτιση της χρονοσειράς. Ωστόσο, μπορεί κάποιες φορές να μας ενδιαφέρει να συσχετίσουμε την ποσότητα κατανάλωσης δημητριακών του Φεβρουαρίου μόνο με τον Δεκέμβριο για διάφορους λόγους και να παραλείψουμε την επίδραση του Ιανουαρίου.

Για να αντιμετωπίσουμε αυτό το πρόβλημα χρησιμοποιούμε τη μερική αυτοσυσχέτιση (PACF), η οποία μετράει τη σχέση μεταξύ y_t και y_{t-k} , αφού αφαιρεθεί η επίδραση των υστερήσεων $1, 2, 3, \dots, k - 1$. Οπότε η πρώτη μερική αυτοσυσχέτιση είναι ίδια με τη πρώτη αυτοσυσχέτιση, επειδή δεν υπάρχει τίποτα να αφαιρεθεί. Το διάγραμμα μερικής αυτοσυσχέτισης είναι ένα εργαλείο που χρησιμοποιείται συνήθως για την αναγνώριση μοντέλων σε μοντέλα Box-Jenkins. Σε επόμενο κεφάλαιο θα δούμε συγκεκριμένα πως γίνεται αναγνώριση μοντέλων βάσει ACF και PACF.

Συγκεκριμένα, οι μερικές αυτοσυσχετίσεις είναι χρήσιμες για τον προσδιορισμό της τάξης ενός αυτοπαλίνδρομου μοντέλου. Η μερική αυτοσυσχέτιση μιας διαδικασίας $AR(p)$ είναι μηδέν σε υστέρηση $p + 1$ και μεγαλύτερη. Εάν το διάγραμμα αυτοσυσχέτισης υποδεικνύει ότι ένα μοντέλο AR μπορεί να είναι κατάλληλο, τότε εξετάζεται το διάγραμμα μερικής αυτοσυσχέτισης για να βοηθήσει στον προσδιορισμό της τάξης του μοντέλου. Ψάχνουμε για το σημείο στο διάγραμμα όπου οι μερικές αυτοσυσχετίσεις ουσιαστικά γίνονται μηδενικές. Η τοποθέτηση διαστήματος εμπιστοσύνης 95 % για στατιστική σημασία είναι χρήσιμη για το σκοπό αυτό. Βλέποντας το παραπάνω διάγραμμα αυτό σημαίνει πως για κάθε υστέρηση k , όπου έχουμε PACF εκτός των ορίων σφάλματος, τότε αυτός θα ήταν ένας στατιστικά καλός παράγοντας, ώστε να συμπεριληφθεί στο μοντέλο αυτοπαλινδρόμησης, εφόσον περιέχει στατιστικά σημαντική πληροφορία για να βοηθήσει στην πρόβλεψη της τιμής.



Εικόνα 11. Διάγραμμα Μερικής Αυτοσυσχέτισης τριμηνιαίου ποσοστού αλλαγής κατανάλωσης στις ΗΠΑ



Κεφάλαιο 3 : Εισαγωγή στην Ανάλυση Χρονοσειρών

3.1 Στόχοι της Ανάλυσης Χρονοσειρών

Ο στόχος της ανάλυσης χρονοσειρών είναι μέσω παρελθοντικών τιμών της χρονοσειράς να εξαγάγει σημαντικές πληροφορίες που θα χρησιμοποιηθούν για την παραγωγή εκτιμήσεων των μελλοντικών τιμών των υπό μελέτη μεγεθών . Για την πραγματοποίηση της ανάλυσης, είναι αναγκαίο να υπάρχει αυτοσυσχέτιση των τιμών της χρονοσειράς, δηλαδή να μην παρατηρείται ο λεγόμενος «λευκός θόρυβος» (Αϊβαζίδου, 2015).

Η ανάλυση χρονοσειρών έχει νόημα μόνο όταν οι παρατηρήσεις της χρονοσειράς δεν είναι ανεξάρτητες, δηλαδή η αλληλουχία των τιμών ακολουθεί κάποια προβλέψιμη ή επαναλαμβανόμενη μορφή.

Τα βασικά στάδια κατά την ανάλυση χρονοσειρών είναι η περιγραφή με τη χρήση γραφημάτων, η επεξήγηση με τη χρήση μοντέλων για την εξεύρεση των μηχανισμών δημιουργίας της χρονοσειράς και η πρόβλεψη των δεδομένων που περιλαμβάνει τη χρήση ενός μοντέλου για την εκτίμηση μελλοντικών τιμών.

3.2 Η στοχαστική διαδικασία

Αρχικά στην ανάλυση χρονοσειρών γίνεται περιγραφή της υπό μελέτη χρονοσειράς με τη χρήση του γραφήματος των τιμών της κατά τη διάρκεια του χρόνου. Τα γραφήματα αποκαλύπτουν τα κυριότερα στοιχεία της χρονοσειράς όπως την τάση και τη εποχικότητα.

Ένα μεγάλο κομμάτι στην ανάλυση χρονοσειρών περιλαμβάνει τη διαδικασία του «φιλτραρίσματος», για παράδειγμα την αλλαγή χαρακτηριστικών της χρονοσειράς, την αποικοδόμηση στα βασικά της μέρη. Η διαδικασία αυτή ακολουθείται πάντα προτού δημιουργηθεί ένα μοντέλο που την προσομοιάζει. Η διαδικασία αυτή περιλαμβάνει την αφαίρεση της τάσης από τη χρονοσειρά όταν αυτή είναι μη στάσιμη. Ένα άλλο βήμα είναι η αποεποχικοποίηση, δηλαδή η αφαίρεση της περιοδικότητας και της κυκλικότητας της χρονοσειράς. Επίσης, μέσω της συνάρτησης αυτοσυσχέτισης εντοπίζεται η μη τυχαιότητα στα δεδομένα, δηλαδή ποσοτικοποιείται η «μνήμη» στη χρονοσειρά. Εάν αφαιρέσουμε όλα αυτά τα στοιχεία από μια χρονοσειρά αυτό που μένει είναι «λευκός θόρυβος». Ο λευκός θόρυβος είναι μια τυχαία διαδικασία, της οποίας οι τιμές είναι σειριακά ανεξάρτητες μεταξύ τους με μηδενικό μέσο όρο και άπειρη διακύμανση (Montgomery et al, 2008).

3.3 Ανάλυση Τάσης

Η τάση μιας χρονοσειράς δείχνει τη κατεύθυνση της χρονοσειράς μακροπρόθεσμα. Μπορεί να αυξάνεται, να μειώνεται ή να παραμένει σταθερή για μεγάλο χρονικό διάστημα.

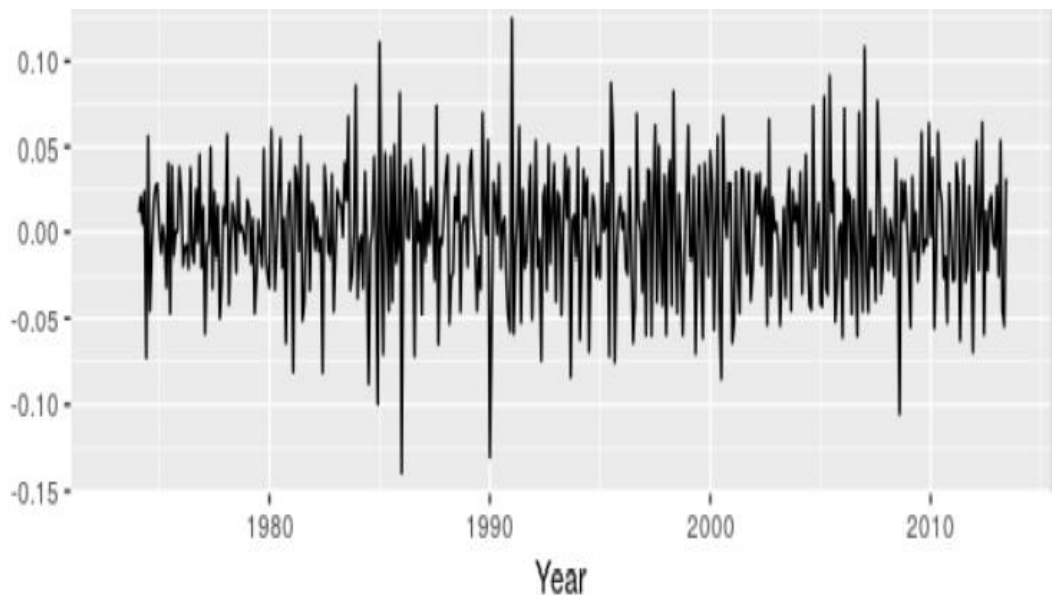
Πριν προχωρήσουμε στην ανάλυση χρονοσειρών γενικότερα είναι βασικό να ορίσουμε τη στάσιμη χρονοσειρά. Οι στάσιμες χρονοσειρές είναι αυτές των οποίων οι ιδιότητες δεν εξαρτώνται από τη χρονική στιγμή.

Μια χρονοσειρά είναι στάσιμη όταν ο μέσος όρος της είναι σταθερός, η διακύμανσή της είναι σταθερή και δεν υπάρχει εποχικότητα. Αποτελεί βασική προϋπόθεση για τη χρήση των περισσότερων μοντέλων χρονοσειρών, η χρονοσειρά να είναι στάσιμη. Αν δεν είναι, τότε θα πρέπει να μετατραπεί σε

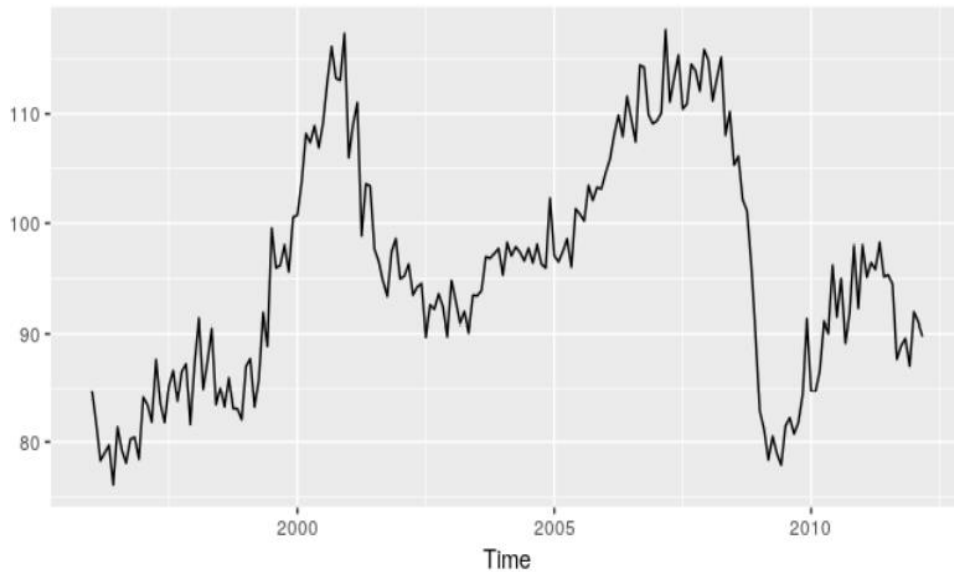


στάσιμη πριν προχωρήσουμε στην προσαρμογή κάποιου μοντέλου στη χρονοσειρά (Adhikari & Agrawal, 2013).

Έτσι, μια χρονοσειρά με τάση, ή με εποχικότητα, δεν είναι στάσιμη, που σημαίνει ότι η τάση και η εποχικότητα επηρεάζουν την τιμή της σε διαφορετικές χρονικές στιγμές. Η στασιμότητα μιας χρονοσειράς σχετίζεται με τις στατιστικές της ιδιότητες στο χρόνο. Δηλαδή, με την πιο αυστηρή έννοια, μια σταθερή χρονοσειρά παρουσιάζει παρόμοια "στατιστική συμπεριφορά" στο χρόνο. Γενικά, η στασιμότητα μιας χρονοσειράς μπορεί να προσδιοριστεί με τη λήψη αυθαίρετων «στιγμιότυπων» της διαδικασίας σε διαφορετικά χρονικά σημεία και την παρατήρηση της γενικής συμπεριφοράς της. Εάν παρουσιάζει "παρόμοια" συμπεριφορά, μπορεί κανείς στη συνέχεια να προχωρήσει με τις προσπάθειες μοντελοποίησης υπό την υπόθεση της στασιμότητας. Περαιτέρω προκαταρκτικές δοκιμές περιλαμβάνουν επίσης την παρατήρηση της συμπεριφοράς της συνάρτησης αυτοσυσχέτισης. Ένα ισχυρό και σταδιακά μειούμενο ACF θα υποδεικνύει αποκλίσεις από τη στασιμότητα (Montgomery et al, 2008).



Εικόνα 12. Παράδειγμα διαγράμματος στάσιμης χρονοσειράς.



Εικόνα 13. Παράδειγμα διαγράμματος μη στάσιμης χρονοσειράς (Hyndman & Athanasopoulos, 2018).

Η τάση εξηγείται σε μεγάλο βαθμό μέσω μιας γραμμικής συνάρτησης μ . Η προσαρμογή ενός γραμμικού μοντέλου παλινδρόμησης (AR) μ_t ως προς το t με τις παρατηρήσεις της χρονοσειράς ως τιμές του μ_t δίνει:

$$\mu_t = y + x_t,$$

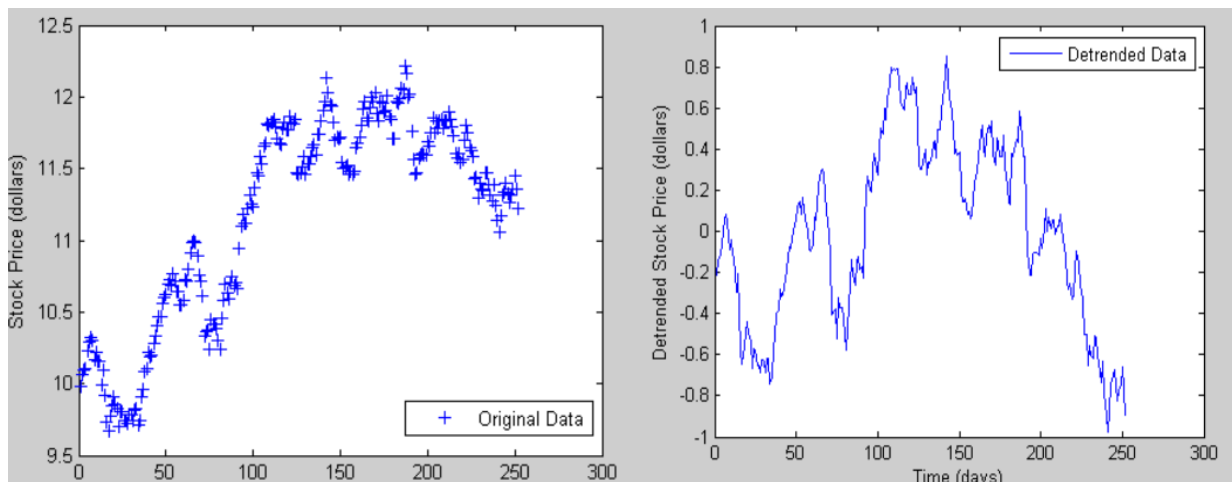
όπου y ο σταθερός όρος και x_t η συνάρτηση του x ως προς t .

Απαλείφοντας τη γραμμική τάση της χρονοσειράς έχουμε την παρακάτω χρονοσειρά:

$$y'_t = y_t - \mu,$$

όπου y_t η αρχική χρονοσειρά και μ η γραμμική συνάρτηση της τάσης

Ως επεξήγηση της διαφοράς αυτής, παρατίθεται ένα δείγμα χρονοσειράς που αποτελείται από τις ημερήσιες τιμές μιας μετοχής και η ίδια χρονοσειρά αφού αφαιρέσουμε τη τάση.



Εικόνα 14. (Αριστερά) Αρχική χρονοσειρά ημερησίων τιμών μετοχής. (Δεξιά) Η νέα χρονοσειρά μετά την αφαίρεση της τάσης (The MathWorks, 2007).



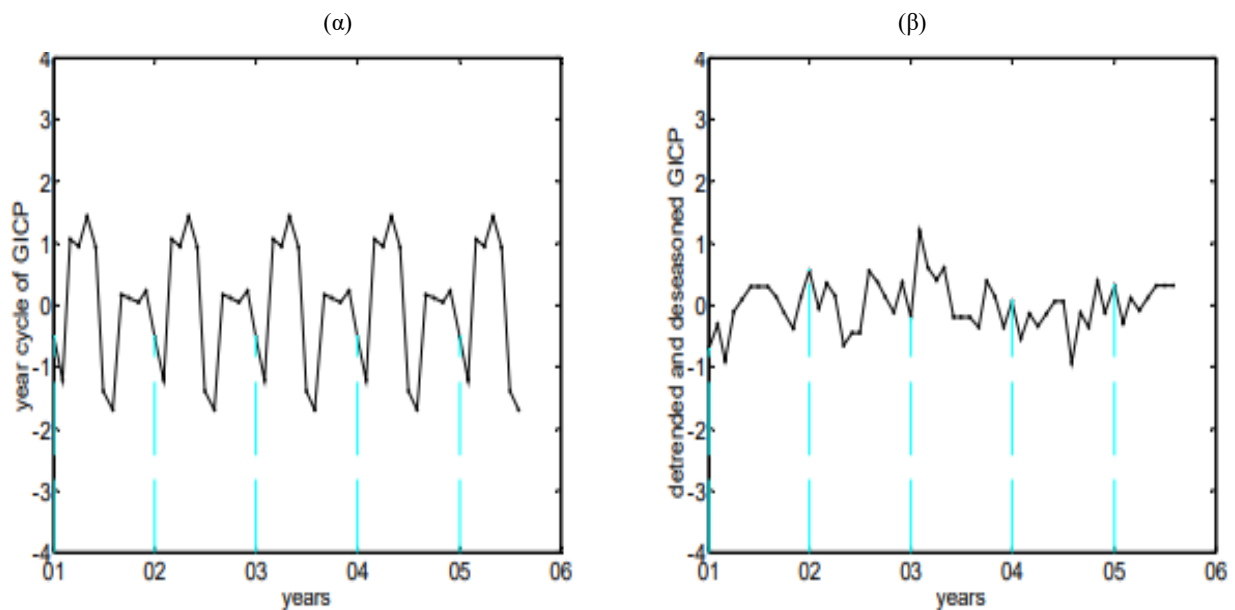
Παρατηρούμε ότι ενώ στην αριστερή εικόνα είχαμε μια συστηματική αύξηση των τιμών της μετοχής, στη δεξιά τα δεδομένα έχουν κέντρο το 0 και η γραμμική τάση έχει αφαιρεθεί.

Μια άλλη προσέγγιση για την εξάλειψη της τάσης είναι ο υπολογισμός διαφορών της χρονοσειράς (differencing) για την απόκτηση μιας νέας στάσιμης χρονοσειράς (Montgomery et al, 2008).

Το differencing μπορεί να εφαρμοστεί σαν διαδικασία παραπάνω από μια φορές διαδοχικά για την εξάλειψη του στοιχείου τάσης της χρονοσειράς και έτσι ορίζουμε και την τάξη του. Δύο είναι τα πλεονεκτήματα της διαδικασίας αυτής σε σχέση με την τοποθέτηση ενός μοντέλου τάσης στα δεδομένα. Πρώτον, δεν απαιτεί εκτίμηση των παραμέτρων, οπότε είναι μια απλούστερη προσέγγιση. Δεύτερον, η τοποθέτηση μοντέλου προϋποθέτει ότι η τάση είναι σταθερή και θα παραμείνει έτσι στο (τουλάχιστον άμεσο) μέλλον. Με άλλα λόγια, το στοιχείο της τάσης, μόλις εκτιμηθεί υποτίθεται ότι είναι ντετερμινιστικό. Ο υπολογισμός διαφορών μπορεί να επιτρέψει στο στοιχείο τάσης να αλλάξει στο χρόνο. Το πρώτο differencing αντιπροσωπεύει μια τάση που επηρεάζει την αλλαγή του μέσου όρου της χρονοσειράς. Το δεύτερο αντιστοιχεί στις μεταβολές στη κλίση της τάσης και ούτω καθεξής. Συνήθως, αν πάρουμε διαφορές της χρονοσειράς μία ή δύο φορές είναι αρκετό στη πράξη για τη κατάργηση της τάσης στα δεδομένα.

3.4 Ανάλυση Εποχικότητας

Η εποχιακή μεταβλητότητα εμφανίζεται όταν μια χρονοσειρά επηρεάζεται από εποχικούς παράγοντες πχ. μια εποχή του χρόνου ή μία μέρα της εβδομάδας. Είναι σταθερή με γνωστή συχνότητα και υπάρχει σε κάθε ημερολογιακή περίοδο (Κουγιουμτζής, 2011).



Εικόνα 15. (α) Ο εκτιμημένος ετήσιος κύκλος για την GICP. (β) Η χρονοσειρά που προκύπτει από την αφαίρεση του εκτιμημένου ετήσιου κύκλου από τη χρονοσειρά στο (α).

Εποχικότητα, ή εποχικότητα σε συνδυασμό με τάση, υπάρχει σε πολλές χρονοσειρές. Για την εξάλειψη της εποχικότητας, επίσης μπορούμε να πάρουμε διαφορές της χρονοσειράς. Όταν τόσο η τάση όσο και τα εποχικά στοιχεία είναι ταυτόχρονα παρόντα, μπορούμε διαδοχικά να πάρουμε διαφορές για να



Σύντρης Σταύρος,
“Ανάλυση και Πρόβλεψη Χρονοσειρών
με Εφαρμογές σε Θέματα Ναυτιλίας”

εξαλείψουμε αυτά τα στοιχεία. Δηλαδή, πρώτα για την αφαίρεση του εποχικού παράγοντα και στη συνέχεια για την εξάλειψη της τάσης.



Κεφάλαιο 4 : Πρόβλεψη, πλάνο και στόχοι

4.1 Αποφασίζοντας τι θα προβλέψουμε

Με τον όρο πρόβλεψη (forecast) εννοούμε την πρόβλεψη ενός μελλοντικού γεγονότος ή μιας σειράς από γεγονότα. Ωστόσο, δεν είναι πάντα εύκολο να κάνεις μια καλή πρόβλεψη. Μερικές διάσημες μη αξιόπιστες προβλέψεις ήταν οι εξής:

- «Ο πληθυσμός της Γης είναι σταθερός σε μέγεθος και θα παραμείνει έτσι έως το τέλος της ανθρωπότητας», *Εγκυκλοπαίδεια, 1756*.
- «Το 1930 θα είναι μια εξαιρετική χρονιά για την εργασία», *Υπουργείο Εργασίας ΗΠΑ, New Year's Forecast 1929*, λίγο πριν το οικονομικό κραχ της 29^{ης} Οκτωβρίου.

Οι προβλέψεις είναι ένα σημαντικό ζήτημα, το οποίο απασχολεί πολλούς κλάδους της σύγχρονης κοινωνίας όπως την οικονομία, τη διαχείριση ρίσκου, τη πολιτική, το μάρκετινγκ, τη διαχείριση περιβάλλοντος και πολλές άλλες. Κατηγοριοποιούνται σε βραχυπρόθεσμες, μεσοπρόθεσμες και μακροπρόθεσμες. Οι βραχυπρόθεσμες προβλέψεις αφορούν γεγονότα σε χρονικό ορίζοντα ημερών, εβδομάδων έως και μηνών. Οι μεσοπρόθεσμες προβλέψεις έχουν ορίζοντα ενός ή δύο χρόνων, ενώ οι μακροπρόθεσμες μπορεί να αφορούν προβλέψεις γεγονότων που επεκτείνονται πολλά χρόνια αργότερα. Οι βραχυπρόθεσμες και μεσοπρόθεσμες προβλέψεις απαιτούνται για δραστηριότητες όπως ο προϋπολογισμός ή η επιλογή προγραμμάτων έρευνας και ανάπτυξης. Οι μακροπρόθεσμες προβλέψεις αφορούν θέματα όπως η χάραξη στρατηγικής. Οι βραχυπρόθεσμες και μεσοπρόθεσμες προβλέψεις βασίζονται στην ταυτοποίηση, μοντελοποίηση και εκμετάλλευση μοτίβων που υπάρχουν σε ιστορικά δεδομένα. Επειδή τα ιστορικά δεδομένα συνήθως παρουσιάζουν στασιμότητα και δεν αλλάζουν πολύ γρήγορα, οι στατιστικές μέθοδοι είναι πολύ χρήσιμες για αυτές τις προβλέψεις.

Μια πρόβλεψη μπορεί να είναι ευκολότερη από την άλλη, ωστόσο κάποια γεγονότα δεν μπορούν να προβλεφθούν με ακρίβεια. Η πιθανότητα να πραγματοποιηθεί ένα γεγονός καθορίζεται από αρκετούς παράγοντες όπως:

- Η κατανόηση των παραγόντων που συμβάλλουν σε αυτό.
- Η ποσότητα των διαθέσιμων δεδομένων.
- Αν οι ίδιες οι προβλέψεις επηρεάζουν το αποτέλεσμα της πρόβλεψης.

Τα παραπάνω αποτελούν σημαντικά βήματα που χρήζουν ανάλυσης πριν προχωρήσει κανείς οποιαδήποτε πρόβλεψη, καθώς αποτελούν απαραίτητη προϋπόθεση για την ταυτοποίηση του προβλήματος και των μέσων που έχει στη διάθεση του, ώστε να κάνει μια ακριβή πρόβλεψη.

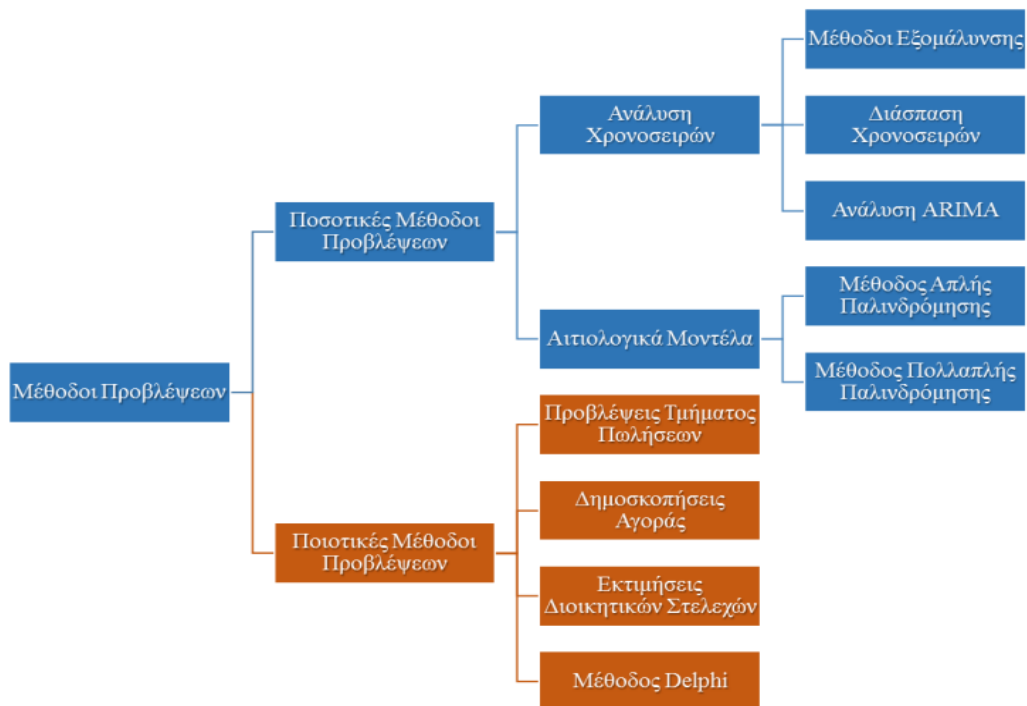
4.2 Δεδομένα και μέθοδοι πρόβλεψης

Τα περισσότερα προβλήματα προβλέψεων περιλαμβάνουν τη χρήση δεδομένων χρονοσειρών. Τα δεδομένα που χρησιμοποιούνται μπορεί να είναι καθημερινά, εβδομαδιαία, μηνιαία ή χρόνια, δεδομένα που λαμβάνονται την ίδια στιγμή, αυξανόμενα όπως οι πωλήσεις ενός προϊόντος μέσα στο μήνα ή ακόμα μπορεί να είναι στατιστικά όπως η καθημερινή τιμή κλεισίματος μια μετοχής στο Χρηματιστήριο.



Παρά το μεγάλο εύρος των λόγων για τους οποίους μπορεί να απαιτηθεί μια πρόβλεψη, οι τεχνικές χωρίζονται σε δύο βασικές κατηγορίες, τις ποιοτικές και τις ποσοτικές. Η επιλογή βασίζεται κυρίως στην ύπαρξη δεδομένων. Αν δεν υπάρχουν δεδομένα ή αν τα δεδομένα δεν είναι σχετικά με τις προβλέψεις, τότε οι ποιοτικές μέθοδοι αποτελούν μονόδρομο. Οι ποιοτικές μέθοδοι είναι συνήθως υποκειμενικές και απαιτούν τη κρίση κάποιου ειδικού. Δεν αποτελούν τυχαίες προβλέψεις, αλλά δομημένες προσεγγίσεις χωρίς τη χρήση ιστορικών δεδομένων. Μια κατάσταση στην οποία θα μπορούσε να χρησιμοποιηθεί ποιοτική μέθοδος πρόβλεψης είναι η εισαγωγή ενός νέου προϊόντος για το οποίο δεν υπάρχει σχετική προϊστορία. Πιθανές ποιοτικοί μέθοδοι στην κατάσταση αυτή θα μπορούσαν να είναι οι έρευνες πιθανών καταναλωτών ή η σύγκριση των πωλήσεων κάποιου αντίστοιχου ή παρεμφερούς προϊόντος.

Οι ποσοτικές μέθοδοι πρέπει να πληρούν δύο κριτήρια. Αρχικά, θα πρέπει να υπάρχουν αριθμητικά δεδομένα για το παρελθόν ενός μεγέθους και έπειτα να θεωρήσουμε ότι αυτό το μοτίβο που υπήρχε στο παρελθόν κατά κάποιο τρόπο θα συνεχιστεί και στο μέλλον. Για τις ποσοτικές μεθόδους υπάρχει μια γκάμα επιλογών, οι οποίες δημιουργούνται στο πλαίσιο συγκεκριμένων απαιτήσεων. Κάθε μοντέλο αναλύει στοιχεία στα δεδομένα και εκφράζει τη στατιστική σχέση μεταξύ προηγούμενων και τρεχουσών τιμών μιας μεταβλητής. Βάσει αυτών των χαρακτηριστικών δημιουργούνται προβλέψεις για τις μελλοντικές τιμές των μεγεθών. Οι πιο διαδεδομένοι τύποι μοντέλων είναι τα μοντέλα αυτοπαλινδρόμησης (autoregressive) και εξομάλυνσης (exponential smoothing) (Montgomery et al, 2008).



Εικόνα 16. Πίνακας μεθόδων πρόβλεψης (Αϊβαζίδου,2015).



4.3 Τα βασικά βήματα σε μια διαδικασία πρόβλεψης

Οποιαδήποτε διαδικασία πρόβλεψης συνήθως περιλαμβάνει κάποια βασικά βήματα:

1. Καθορισμός του προβλήματος.

Αποτελεί το δυσκολότερο μέρος της πρόβλεψης, καθώς ο ορισμός του προβλήματος απαιτεί κατανόηση του σκοπού για τον οποίο θα χρησιμοποιηθούν οι προβλέψεις, ποιος θα τις χρησιμοποιήσει και πώς θα λειτουργήσουν αυτές οι προβλέψεις μέσα στα όρια που θέτει αυτός που τις χρησιμοποιεί.

2. Συλλογή πληροφοριών.

Υπάρχουν τουλάχιστον δύο είδη πληροφοριών που απαιτούνται, τα στατιστικά δεδομένα, καθώς και η εξειδίκευση αυτών που θα συλλέξουν τα δεδομένα και θα χρησιμοποιήσουν τις προβλέψεις. Συνήθως, είναι δύσκολο να συλλέξει κανείς αρκετά ιστορικά δεδομένα για να μπορέσει δημιουργήσει ένα καλό στατιστικό μοντέλο. Σε αυτές τις περιπτώσεις μπορούν να χρησιμοποιηθούν ποιοτικοί μέθοδοι πρόβλεψης. Επίσης, μερικές φορές τα παλιά δεδομένα μπορεί να είναι λιγότερο χρήσιμα από τα πρόσφατα λόγω δομικών αλλαγών στο σύστημα, το οποίο προσπαθούμε να προβλέψουμε.

3. Διερευνητική ανάλυση

Η απεικόνιση των δεδομένων μπορεί να φανεί ιδιαίτερα χρήσιμη, καθώς μπορεί κανείς να διακρίνει επαναλαμβανόμενα και σταθερά πρότυπα στα δεδομένα, όπως και αν υπάρχει τάση, εποχικότητα, ή αν υπάρχουν ακραίες τιμές που μπορούν να εξηγηθούν από κάποιο φαινόμενο. Υπάρχουν πολλά εργαλεία που μπορούν να βοηθήσουν σε αυτό. Ένα από αυτά είναι και το MATLAB, το οποίο θα χρησιμοποιήσουμε για την ανάλυση στο 7^ο κεφάλαιο.

4. Επιλογή μοντέλου πρόβλεψης

Η επιλογή του καταλληλότερου μοντέλου πρόβλεψης βασίζεται στη διαθεσιμότητα δεδομένων, στη σχέση μεταξύ της μεταβλητής πρόβλεψης και των εξηγηματικών μεταβλητών και στο τρόπο που θα χρησιμοποιηθούν οι προβλέψεις. Κάθε μοντέλο αποτελεί ένα σύστημα το οποίο βασίζεται σε συγκεκριμένες παραδοχές και συνήθως περιλαμβάνει μια ή περισσότερες παραμέτρους που πρέπει να εκτιμηθούν βάσει των διαθέσιμων δεδομένων.

5. Χρήση και αξιολόγηση του μοντέλου πρόβλεψης

Όταν επιλεγεί το μοντέλο και εκτιμηθούν οι παράμετροί του, μένει μόνο να χρησιμοποιηθεί. Η απόδοση του μπορεί να αξιολογηθεί, αφού αποκτηθεί εικόνα για το πώς πήγαν τα δεδομένα για τα οποία έγινε η πρόβλεψη μέσω συγκεκριμένων μεθόδων υπολογισμού σφαλμάτων (Hyndman & Athanasopoulos, 2018).



Κεφάλαιο 5 : Τεχνικές Πρόβλεψης

5.1 Εισαγωγή στις βασικές αρχές των μοντέλων χρονοσειρών

Τα βασικά βήματα στη μοντελοποίηση και πρόβλεψη μιας χρονοσειράς είναι τα εξής:

1. Σχεδιάζουμε τη χρονοσειρά και προσδιορίζουμε τα βασικά χαρακτηριστικά της, όπως αν υπάρχουν τάσεις ή εποχιακή συμπεριφορά ή και τα δύο. Αναζητούμε πιθανές ακραίες τιμές ή οποιαδήποτε ένδειξη ότι η χρονοσειρά έχει αλλάξει σε σχέση με τα βασικά χαρακτηριστικά της, όπως τάσεις ή εποχικότητα, κατά τη διάρκεια της χρονικής διάρκειας μελέτης.

2. Εξαλείφουμε οποιαδήποτε τάση ή εποχιακή εξάρτηση, είτε με differencing είτε με την τοποθέτηση κατάλληλου μοντέλου στα δεδομένα. Εξετάζουμε επίσης το ενδεχόμενο να χρησιμοποιήσουμε μετασχηματισμούς δεδομένων, ιδιαίτερα εάν η μεταβλητότητα της χρονοσειράς φαίνεται να είναι ανάλογη με το μέσο επίπεδο της σειράς. Στόχος των πράξεων αυτών είναι η παραγωγή ενός συνόλου σταθερών υπολειμμάτων.

3. Ανάπτυξη μοντέλου πρόβλεψης για τα κατάλοιπα. Δεν είναι ασυνήθιστο να διαπιστώνουμε ότι υπάρχουν πολλά πιθανά μοντέλα και θα πρέπει να γίνει πρόσθετη ανάλυση για να προσδιοριστεί το καλύτερο που θα αναπτυχθεί. Μερικές φορές τα πιθανά μοντέλα μπορούν να εξαλειφθούν με βάση την προσαρμογή τους στα ιστορικά δεδομένα. Είναι απίθανο ότι ένα μοντέλο που ταιριάζει ελάχιστα θα παράγει καλές προβλέψεις

5.2 Αυτοπαλίνδρομο Μοντέλο (AR ή Autoregressive)

Μία συχνή προσέγγιση στη μοντελοποίηση χρονοσειρών μίας μεταβλητής είναι τα αυτοπαλίνδρομα μοντέλα AR:

$$x_t = \varphi_0 + \varphi_1 x_{t-1} + \dots + \varphi_p x_{t-p} + \varepsilon_t,$$

όπου οι ανεξάρτητες μεταβλητές $x_{t-1}, x_{t-2}, \dots, x_{t-p}$ είναι οι τιμές της υπό πρόβλεψη μεταβλητής x_t σε προηγούμενες χρονικές περιόδους, φ (0,1,2,...) οι συντελεστές του μοντέλου, p η τάξη και ε_t είναι «λευκός θόρυβος».

Ένα αυτοπαλίνδρομο μοντέλο είναι απλώς μια γραμμική παλινδρόμηση της τρέχουσας τιμής της σειράς έναντι μιας ή περισσότερων προηγούμενων τιμών της σειράς. Ο αριθμός των υστερήσεων που συμπεριλαμβάνουμε λέγεται η τάξη ή order του μοντέλου AR και συμβολίζεται με τον αριθμό p . Η μεταβλητή τη χρονική στιγμή t συμφώνως του μοντέλου AR(p) εξηγείται από έναν συνδυασμό των τελευταίων p τιμών της χρονοσειράς. Το κομμάτι το οποίο δεν μπορεί να εξηγηθεί από το συνδυασμό των τιμών της χρονοσειράς είναι στοχαστικό και οφείλεται σε επιδράσεις που θεωρούνται εξωτερικές τη χρονική στιγμή t και για αυτό υπάρχει η μεταβλητή ε_t , η οποία περιγράφει αυτή την τυχαιότητα.

Τα μοντέλα AR μπορούν να αναλυθούν με διάφορες μεθόδους συμπεριλαμβανομένων τεχνικών τυπικών γραμμικών ελάχιστων τετραγώνων.

Στους παρακάτω πίνακες υπάρχουν κάποια βασικά χαρακτηριστικά που καταδεικνύουν την καταλληλότητα βασικών υποδειγμάτων αυτοπαλινδρόμησης για την ανάλυση της εκάστοτε χρονοσειράς (Κατρακυλίδης κ.ά, 2015).



Υποδείγματα AR

Διαδικασία	Συνάρτηση Αυτοσυσχέτισης r_k	Συνάρτηση Μερικής Αυτοσυσχέτισης π_k
AR(1)	Φθίνει προς το μηδέν. Με θετικό πρόσημο εάν $b_1 > 0$. Με αρνητικό πρόσημο και εναλλαγή προσήμων εάν $b_1 < 0$.	Μηδενίζεται μετά το π_1 - Εμφανίζει τοπικό υψηλό στην πρώτη υστέρηση. Θετικό πρόσημο για $b_1 > 0$ Αρνητικό πρόσημο για $b_1 < 0$
AR(2)	Φθίνει προς το μηδέν όσο αυξάνεται το k.	Μηδενίζεται μετά τη 2 ^η υστέρηση π_2

5.3 Μοντέλο Κινητού Μέσου Όρου (MA ή Moving Average)

Μια άλλη κοινή προσέγγιση για τη μοντελοποίηση μοντέλων μονομεταβλητών χρονικών σειρών είναι το μοντέλο κινούμενου μέσου όρου MA:

$$X_t = \mu + A_t - \theta_1 A_{t-1} - \theta_2 A_{t-2} - \dots - \theta_q A_{t-q}$$

όπου X_t είναι η χρονική σειρά, μ είναι ο μέσος όρος της σειράς, A_{t-i} είναι όροι λευκού θορύβου και θ, \dots, θ_q είναι οι παράμετροι του μοντέλου. Η τιμή του q ονομάζεται τάξη του μοντέλου MA.

Δηλαδή, ένα μοντέλο κινητού μέσου όρου αποτελεί μια γραμμική παλινδρόμηση της παρούσας τιμής της χρονοσειράς έναντι του λευκού θορύβου ή τυχαίων σοκ μιας ή περισσότερων προηγούμενων τιμών της σειράς. Τα τυχαία σοκ σε κάθε σημείο υποτίθεται ότι προέρχονται από την ίδια κατανομή, τυπικά μια κανονική κατανομή, με μηδενική θέση και σταθερή κλίμακα. Η διάκριση σε αυτό το μοντέλο είναι ότι αυτά τα τυχαία σοκ προωθούνται σε μελλοντικές τιμές της χρονικής σειράς. Η προσαρμογή των μοντέλων MA είναι πιο περίπλοκη από ότι τα μοντέλα AR. Τα μοντέλα MA έχουν επίσης λιγότερο εμφανή ερμηνεία από τα μοντέλα AR.

Στους παρακάτω πίνακες υπάρχουν κάποια βασικά χαρακτηριστικά που καταδεικνύουν την καταλληλότητα βασικών υποδειγμάτων κινητών μέσων όρων για την ανάλυση της εκάστοτε χρονοσειράς (Κατρακυλίδης κ.ά, 2015).



Υποδείγματα MA

Διαδικασία	Συνάρτηση Αυτοσυσχέτισης ρ	Συνάρτηση Μερικής Αυτοσυσχέτισης π
MA(1)	Μηδενίζεται μετά την 1 ^η υστέρηση, δηλαδή μετά το ρ_1 . Θετικό πρόσημο για $\theta_1 < 0$ Αρνητικό πρόσημο για $\theta_1 > 0$	Φθίνει γεωμετρικά προς το 0 μετά την πρώτη μερική αυτοσυσχέτιση π_1 με αρνητικό πρόσημο για $\theta_1 > 0$ Εναλλαγή προσήμων αρχίζοντας με θετικό πρόσημο $\theta_1 < 0$
MA(2)	Μηδενίζεται μετά το ρ_2 .	Φθίνει προς το 0 σχεδόν γεωμετρικά μετά το π_2
Διαδικασία Κινητού Μέσου q τάξεως MA(q)	Μηδενίζεται μετά από q υστερήσεις, δηλαδή μετά την ρ_q	Φθίνει προς το 0 σχεδόν γεωμετρικά ή με ημιτονοειδή συμπεριφορά μετά το π_q

5.4 Αυτοπαλίνδρομο Μοντέλο Κινητού Μέσου Όρου (ARMA ή Autoregressive Moving Average)

Οι Box και Jenkins (1976) δημιούργησαν μια προσέγγιση που συνδυάζει τον κινητό μέσο όρο και τις αυτοπαλίνδρομες προσεγγίσεις. Αν και τόσο η προσέγγιση του αυτοπαλίνδρομου μοντέλου όσο και η προσέγγιση του κινητού μέσου όρου ήταν ήδη γνωστές, η συμβολή των Box και Jenkins ήταν στην ανάπτυξη μιας συστηματικής μεθοδολογίας για τον εντοπισμό και την εκτίμηση μοντέλων που θα μπορούσαν να ενσωματώσουν και τις δύο προσεγγίσεις. Αυτό καθιστά τα μοντέλα αυτά μια ισχυρή κατηγορία μοντέλων.

Η δύναμη των μοντέλων ARMA είναι ότι μπορούν να ενσωματώσουν τόσο AR όρους όσο και MA όρους. Παρόλο που τα μοντέλα AR και MA ήταν γνωστά και χρησιμοποιούνταν, οι Box και Jenkins παρείχαν μια συστηματική προσέγγιση για τη μοντελοποίηση τόσο των όρων AR όσο και MA στο μοντέλο. Τα μοντέλα ARMA είναι επίσης κοινώς γνωστά ως μοντέλα ARIMA.

Τα μοντέλα ARMA υποθέτουν ότι τα δεδομένα είναι στατικά, δηλαδή τα δεδομένα έχουν σταθερή θέση και κλίμακα. Η τάση μπορεί συχνά να αφαιρεθεί από μια μη στάσιμη σειρά για να επιτευχθεί η στασιμότητα. Το differencing είναι μια κοινή προσέγγιση για την εξάλειψη της τάσης. Το πρώτο differencing ορίζεται ως $X_t - X_{t-1}$. Στις περισσότερες περιπτώσεις, αρκεί μόνο μια φορά. Ωστόσο, μπορεί να εφαρμοστεί η διαδικασία περισσότερες από μία φορές εάν είναι απαραίτητο. Μπορεί, επίσης, να προσαρμοστεί ένα γραμμικό ή μη γραμμικό μοντέλο για να αφαιρεθεί η τάση.



Το γενικό γραμμικό μοντέλο για την πρόβλεψη στάσιμης χρονοσειράς είναι το αυτοπαλίνδρομο μοντέλο κινούμενου μέσου (Autoregressive Moving Average, ARMA) που δίνεται ως:

$$\varphi(L)y_t = \theta(L)\varepsilon_t,$$

$$\text{όπου } \varphi(L)=1 - \sum_{i=1}^p \varphi_i L^i \text{ και } \theta(L)=1 + \sum_{j=1}^q \theta_j L^j,$$

όπου L ο τελεστής υστέρησης, ο οποίος επιδρά στη χρονοσειρά y_t ως εξής:

$L^i y_t = y_{t-i}$, $\varphi(L)$ ένα AR πολυώνυμο τελεστή υστέρησης σταθερού βαθμού p (με όλες τις ρίζες εκτός του μοναδιαίου κύκλου) και αντίστοιχα $\theta(L)$ είναι ένα MA πολυώνυμο τελεστή υστέρησης αναστρέψιμου βαθμού q (με όλες τις ρίζες εκτός του μοναδιαίου κύκλου) και ε_t ο λευκός θόρυβος (Adhikari & Agrawal, 2013).

Το αυτοπαλίνδρομο μέρος AR είναι τάξης p και το μέρος του κινούμενου μέσου MA είναι τάξης q και το μοντέλο συμβολίζεται $ARMA(p, q)$. Ένα $ARMA(p, q)$ μοντέλο ισοδυναμεί με AR μοντέλο τάξης p , με σφάλματα που ακολουθούν σειρά $MA(q)$

5.5 Μοντέλο ARIMA ή Autoregressive Integrated Moving Average

Συχνά συμβαίνει ότι ενώ οι χρονοσειρές μπορεί να μην έχουν σταθερό επίπεδο, παρουσιάζουν ομοιογενή συμπεριφορά με την πάροδο του χρόνου. Μπορεί να δει κανείς ότι διαφορετικά στιγμιότυπα που λαμβάνονται παρουσιάζουν παρόμοια συμπεριφορά εκτός από το κύριο επίπεδο της χρονοσειράς. Ομοίως, οι διαδικασίες ενδέχεται να παρουσιάζουν μη στασιμότητα και στην κλίση. Μια χρονοσειρά y_t λέγεται ομοιογενής, αν είναι μη στάσιμη, αλλά η πρώτη διαφορά της που είναι

$$w_t = y_t - y_{t-1} = (1 - L)y_t$$

ή υψηλότερης τάξης διαφορές

$$w_t = (1 - L)^d y_t$$

παράγουν μια στάσιμη χρονοσειρά. Επιπρόσθετα, η y_t καλείται ως $ARIMA$ διαδικασία τάξεων p, d, q ή $ARIMA(p, d, q)$ αν η d -οστή διαφορά παράγει μια στάσιμη $ARMA(p, q)$ διαδικασία. Ο όρος Integrated χρησιμοποιείται όταν για $d = 1$ μπορούμε να γράψουμε την y_t ως άθροισμα της διαδικασίας w_t , δηλαδή:

$$\begin{aligned} y_t &= w_t + y_{t-1} \\ &= w_t + w_{t-1} + y_{t-2} \\ &= w_t + w_{t-1} + \dots + w_1 + y_0 \end{aligned}$$

Ως εκ τούτου, ένα μοντέλο $ARIMA(p, d, q)$ μπορεί να γραφτεί ως:

$$\varphi(L)(1 - L)^d y_t = \theta(L)\varepsilon_t$$

ή αναλυτικότερα:

$$\left(1 - \sum_{i=1}^p \varphi_i L^i\right) (1 - L)^d y_t = \left(1 + \sum_{j=1}^q \theta_j L^j\right) \varepsilon_t$$



όπου L ο τελεστής υστέρησης, ο οποίος επιδρά στη χρονοσειρά y_t ως εξής: $L^i y_t = y_{t-i}$, $\varphi(L)$ ένα AR πολυώνυμο τελεστή υστέρησης σταθερού βαθμού p (με όλες τις ρίζες εκτός του μοναδιαίου κύκλου) και αντίστοιχα $\theta(L)$ είναι ένα MA πολυώνυμο τελεστή υστέρησης αναστρέψιμου βαθμού q (με όλες τις ρίζες εκτός του μοναδιαίου κύκλου) και ε_t ο λευκός θόρυβος.

Έτσι, μόλις πραγματοποιηθεί το differencing και ληφθεί μια στάσιμη χρονοσειρά:

$$w_t = (1 - L)^d y_t$$

όπου w_t η νέα χρονοσειρά, L ο τελεστής υστέρησης που αναφέρθηκε παραπάνω, d ο αριθμός των διαφορών και y_t η αρχική χρονοσειρά.

Οι μέθοδοι που παρέχονται στα προηγούμενα τμήματα μπορούν να χρησιμοποιηθούν για την απόκτηση του πλήρους μοντέλου. Στις περισσότερες εφαρμογές η πρώτη διαφορά $d = 1$ και περιστασιακά η δεύτερη διαφορά $d = 2$ είναι αρκετή για την επίτευξη της στασιμότητας. Ωστόσο, μερικές φορές οι μετασχηματισμοί εκτός από το differencing είναι χρήσιμοι για τη μετατροπή μιας μη στάσιμης χρονοσειράς σε στάσιμη. Για παράδειγμα, σε πολλές οικονομικές χρονοσειρές η μεταβλητότητα των παρατηρήσεων αυξάνεται, καθώς αυξάνεται το μέσο επίπεδο της χρονοσειράς. Ωστόσο, το ποσοστό μεταβολής στις παρατηρήσεις είναι σχετικά ανεξάρτητο από το επίπεδο. Ως εκ τούτου, η λογαρίθμηση της αρχικής σειράς θα είναι χρήσιμη για την επίτευξη της στασιμότητας (Montgomery et al, 2008).

Αν συνδυάσουμε το differencing με την αυτοπαλινδρόμηση και ένα μοντέλο κινητών μέσων, τότε έχουμε ένα μη εποχικό μοντέλο *ARIMA*. Το *ARIMA* είναι ακρωνύμιο των Αυτοπαλινδρομικών Μοντέλων Κινητού Μέσου Όρου. Πολλά από τα μοντέλα που συζητήσαμε προηγουμένως αποτελούν ειδική κατηγορία του μοντέλου *ARIMA*.



Κεφάλαιο 6 : Ανάλυση και Πρόβλεψη Χρονοσειρών με χρήση MATLAB

6.1 Εισαγωγή – Στόχοι των προβλέψεων

Στο κεφάλαιο αυτό θα αναλυθεί μια χρονοσειρά με πραγματικά δεδομένα. Η ανάλυσή της έχει σκοπό να δώσει απαντήσεις σε δύο βασικά ερωτήματα. Πρώτον, ποιο είναι το πρότυπο (ή αλλιώς διαδικασία) παραγωγής των δεδομένων και δεύτερον ποια θα είναι η μελλοντική εξέλιξη του φαινομένου (πρόβλεψη). Η χρονοσειρά που θα χρησιμοποιηθεί είναι ένα μονομεταβλητό στατιστικό υπόδειγμα, δηλαδή ένα οικονομετρικό υπόδειγμα στοχαστικής διαδικασίας μίας μεταβλητής, με βάση τις πληροφορίες των παρελθόντων τιμών, καθώς και των τρεχουσών και παρελθοντικών τιμών του διαταρακτικού όρου.

Η ανάλυση των δεδομένων αποτελείται από τα εξής βήματα:

1. Αρχικά θα χωρίσουμε τα δεδομένα σε 2 μέρη. Το πρώτο μέρος αποτελείται από 55 παρατηρήσεις, οι οποίες θα χρησιμοποιηθούν για τη μοντελοποίηση της χρονοσειράς. Το δεύτερο μέρος αποτελείται από 12 παρατηρήσεις, οι οποίες θα συγκριθούν με την πρόβλεψη 12 παρατηρήσεων που θα εξάγουμε με το μοντέλο που θα δημιουργήσουμε από το πρώτο μέρος των δεδομένων. Ο λόγος που θα γίνει αυτό είναι για να έχουμε μια καλή εικόνα των σφαλμάτων του μοντέλου εκτός του δείγματος (out of sample), ώστε να δούμε την απόδοση των προβλέψεων.
2. Εφαρμογή στατιστικού ελέγχου Ljung-Box Test για έλεγχο ύπαρξης αυτοσυσχέτισης μεταξύ των παρατηρήσεων, δηλαδή ότι η χρονοσειρά δεν αποτελεί μια τυχαία διαδικασία (λευκό θόρυβο), αλλά υπάρχει κάποιο πρότυπο που μπορεί να την ερμηνεύσει. Το Ljung-Box Test αξιολογεί την μηδενική υπόθεση ότι μια σειρά σφαλμάτων δεν παρουσιάζει αυτοσυσχέτιση για έναν αριθμό υστερήσεων L , έναντι της εναλλακτικής λύσης ότι κάποιος συντελεστής αυτοσυσχέτισης $\rho(k)$, $k = 1, \dots, L$, δεν είναι μηδενικός. Το test statistic του είναι:

$$Q = (T + 2) \sum_{k=1}^L \left(\frac{\rho(k)^2}{(T-k)} \right),$$

όπου T το μέγεθος του δείγματος, L ο αριθμός των υστερήσεων της αυτοσυσχέτισης και $\rho(k)$ η αυτοσυσχέτιση σε υστέρηση k .

3. Εφαρμογή στατιστικού ελέγχου KPSS (Kwiatkowski, Phillips, Schmidt, Shin Test) για έλεγχο στασιμότητας της χρονοσειράς. Ο έλεγχος αυτός αξιολογεί την μηδενική υπόθεση ότι μια χρονοσειρά είναι στάσιμη σε αντίθεση με την εναλλακτική λύση ότι είναι μη στάσιμη (Kwiatkowski et al, 1992).

Το test statistic είναι:

$$\frac{\sum_{t=1}^T S_t^2}{s^2 T^2},$$

όπου T το μέγεθος του δείγματος, s^2 είναι η εκτίμηση της μακροπρόθεσμης διακύμανσης,

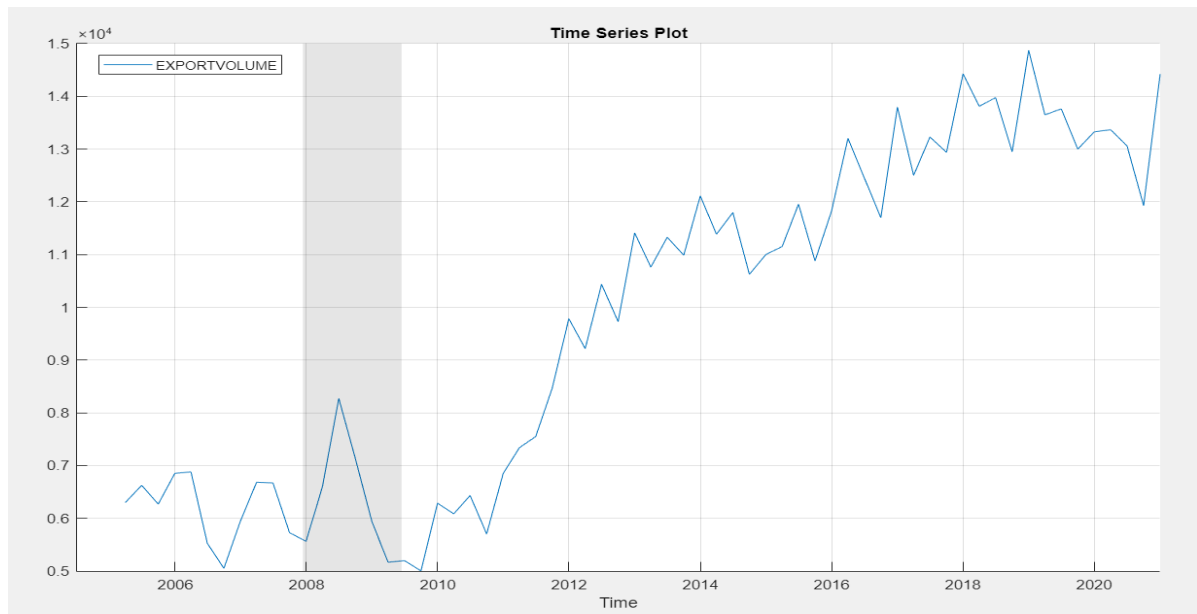
$$S_t = e_1 + e_2 + \dots + e_t$$



4. Αν η χρονοσειρά δεν είναι στάσιμη, θα κάνουμε differencing στην 1^η υστέρηση (lag) και θα επαναληφθεί το KPSS Test. Η διαδικασία ακολουθείται έως ότου η χρονοσειρά γίνει στάσιμη.
5. Μόλις έχουμε μια στάσιμη χρονοσειρά θα απεικονίσουμε την συνάρτηση αυτοσυσχέτισης (ACF) και μερικής αυτοσυσχέτισης (PACF) σε κορελλογράμματα συναρτήσεως της υστέρησης (lag), ώστε να βρούμε τις στατιστικά σημαντικές τιμές (με επίπεδο στατιστικής σημαντικότητας 5%) και μέσω αυτών θα προσδιορίσουμε τον τύπο και την τάξη του υποδείγματος που θα χρησιμοποιηθεί για την ερμηνεία της χρονοσειράς.
6. Μόλις εκτιμηθούν οι συντελεστές του μοντέλου, θα γίνει γραφική απεικόνιση του και των σφαλμάτων του ως προς την πραγματική χρονοσειρά.
7. Μέτρηση συμβατότητας μοντέλου με το AIC (Akaike Information Criterion) (Akaike, 1973) σε περίπτωση που παραπάνω από ένα μοντέλο πληροί τις προϋποθέσεις για να ερμηνεύσει τη χρονοσειρά, ώστε να αποφασίσουμε ποιο είναι το καταλληλότερο. Το συγκεκριμένο κριτήριο εκτιμάει τις πληροφορίες που χάνονται από ένα μοντέλο, όσες λιγότερες χάνονται, τόσο πιο ποιοτικό είναι το μοντέλο. Άρα το μοντέλο με το μικρότερο AIC είναι και το καταλληλότερο. Το AIC γράφεται ως εξής:
$$AIC = -2 \log(L) + 2(p + q + k + 1)$$
όπου L ο τελεστής υστέρησης, p ο βαθμός AR του μοντέλου, q ο βαθμός MA του μοντέλου και k ο αριθμός των υστερήσεων.
8. Τέλος, εφαρμογή εκ νέου Ljung-Box Test στα σφάλματα για μέτρηση αυτοσυσχέτισης μεταξύ τους. Αν δεν υπάρχει αυτοσυσχέτιση, δηλαδή τα σφάλματα είναι λευκός θόρυβος, τότε το μοντέλο έχει προβλέψει σε καλό βαθμό το ντετερμινιστικό κομμάτι της χρονοσειράς.
9. Έπειτα, αφού έχει επιλεγεί το καλύτερο μοντέλο θα δημιουργήσουμε μια πρόβλεψη για 12 παρατηρήσεις και θα παρουσιάσουμε την πρόβλεψη σε σχέση με τα αντίστοιχα πραγματικά δεδομένα.

Σε όλα τα test θα χρησιμοποιήσουμε ως default τις εξής τιμές: $L = 20$ για τις υστερήσεις και $DOF = T - K$ για τους βαθμούς ελευθερίας, καθώς η τιμή των υστερήσεων επηρεάζει τη δύναμη της δοκιμής. Εάν το L είναι πολύ μικρό, τότε η δοκιμή δεν ανιχνεύει αυτόματες συσχετίσεις υψηλής τάξης. Εάν το L είναι πολύ μεγάλο, τότε η δοκιμή χάνει ισχύ όταν μια σημαντική συσχέτιση σε μια υστέρηση ‘ξεπλένεται’ από ασήμαντες συσχετίσεις σε άλλες υστερήσεις (Box et al, 1994).

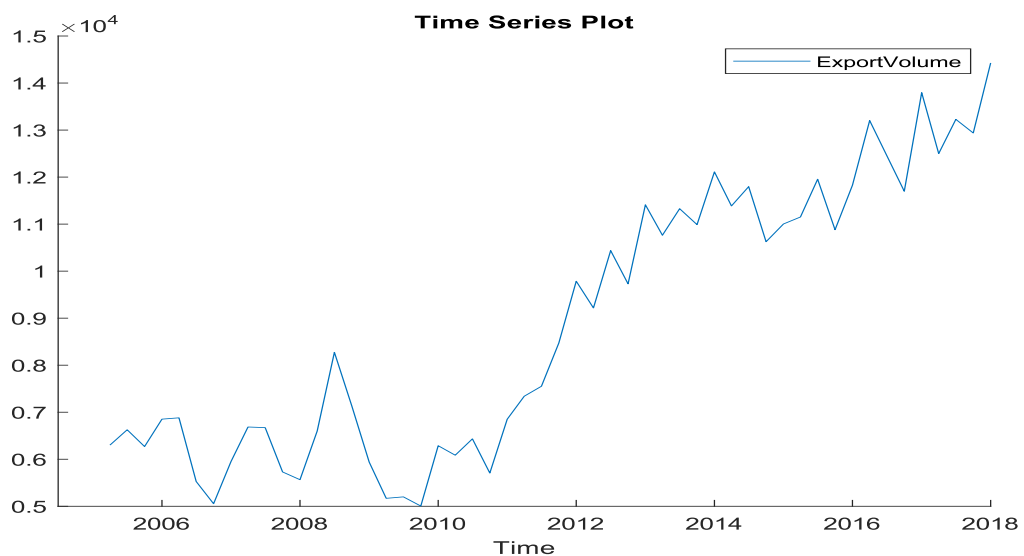
Η ανάλυση πραγματοποιήθηκε με την εφαρμογή Econometric Modeler Version 5.7 του Computational Finance Toolbox και εντολών από το Econometrics Toolbox στο MATLAB (R2021b).



Εικόνα 18. Απεικόνιση της χρονοσειράς EXPORTVOLUME (ποσότητα φορτωθέντων εμπορευμάτων εξωτερικού στους κυριότερους λιμένες Ελλάδος). Στον άξονα Y, η ποσότητα των εμπορευμάτων σε χιλιάδες τόνους. Στον άξονα X, ο αύξων αριθμός παρατήρησης (δηλαδή το κάθε τρίμηνο με έναρξη το Γ' τρίμηνο 2004)

Από την παραπάνω γραφική παράσταση συμπεραίνουμε ότι υπάρχει μια ξαφνική πτώση μεταξύ 2008 – 2010, όπως φαίνεται στη γραμμοσκιασμένη περιοχή, η οποία γνωρίζουμε ότι συνέβη λόγω της παγκόσμιας χρηματοπιστωτικής κρίσης του 2009 και επηρέασε και τη χώρα μας μεταξύ άλλων. Ομοίως παρατηρούμε μια όμοια πτώση από το 2019 έως το 2021, η οποία οφείλεται στην έναρξη της πανδημίας του COVID-19. Ωστόσο, ακολουθείται από μια μεγάλη ανάκαμψη το 2021 λόγω της αυξημένης ζήτησης αγαθών και τάσης καταναλωτισμού λόγω του ανοίγματος των αγορών.

Για την ανάλυσή μας θα χρησιμοποιήσουμε τα πρώτα 55 στοιχεία, έως το πρώτο τρίμηνο του 2018. Η χρονοσειρά που δημιουργείται από αυτά εφεξής θα λέγεται «ExportVolume».

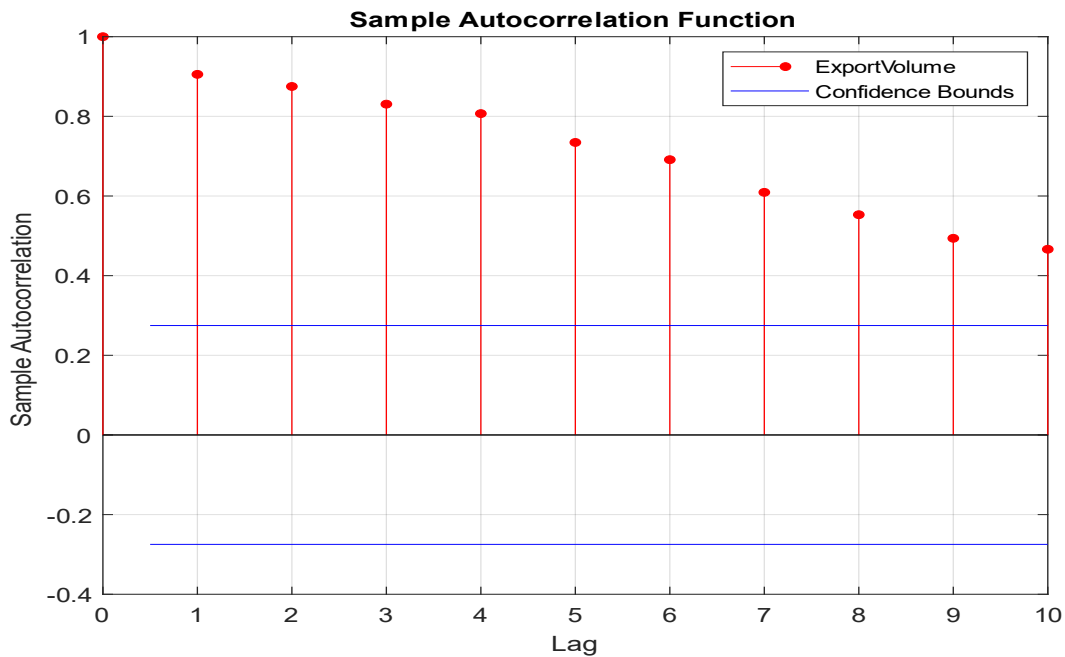


Εικόνα 19. Γραφική παράσταση της ExportVolume



6.3 Ανάλυση χρονοσειρών με Γραμμικά Μοντέλα

Αρχικά όπως επισημάνθηκε η χρονοσειρά έχει μια ανοδική τάση. Αυτό επιβεβαιώνεται και από τη συνάρτηση αυτοσυσχέτισης, η οποία φθίνει σταδιακά προς το μηδέν. Αυτό δείχνει ότι η χρονοσειρά είναι μη στάσιμη, έχει δηλαδή τάση.



Εικόνα 20. Συνάρτηση αυτοσυσχέτισης της χρονοσειράς ExportVolume.

Ωστόσο είναι σημαντικό να εφαρμόσουμε τον στατιστικό έλεγχο KPSS Test για έλεγχο στασιμότητας της χρονοσειράς, παρ’ όλο που στη συγκεκριμένη περίπτωση είναι ολοφάνερο ότι δεν είναι στάσιμη, σε άλλες ενδεχομένως να μην είναι.

KPSS Test

Μηδενική υπόθεση: Η ExportVolume είναι στάσιμη

$$H_0 : \sigma^2 = 0$$

$$H_a : \sigma^2 > 0$$

όπου H_0 η μηδενική υπόθεση, H_a η εναλλακτική υπόθεση και σ^2 η τυπική απόκλιση της ExportVolume.

Παράμετροι Test

	Lags	Include Trend	Significance Level
1	0	true	0.05



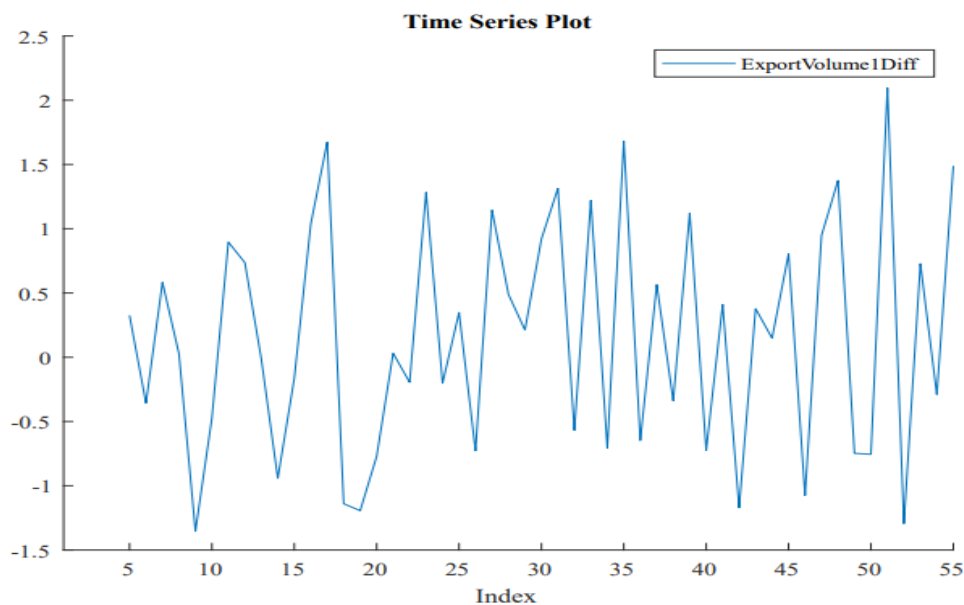
Αποτελέσματα Test

	Null Rejected	P-Value	Test Statistic	Critical Value
1	true	0.01	0.3406	0.146

Η τιμή Test Statistic είναι στατιστικά σημαντική, άρα απορρίπτεται η μηδενική υπόθεση, δηλαδή η χρονοσειρά είναι μη στάσιμη.

Οι έλεγχοι στασιμότητας είναι αναγκαίοι, διότι στις μη στάσιμες μεταβλητές τα στατιστικά μέτρα μπορεί να φαίνονται σημαντικά, ωστόσο δεν είναι έγκυρα. Επίσης, τα t στατιστικά δεν ακολουθούν τις αντίστοιχες κατανομές και συνεπώς οι στατιστικοί έλεγχοι δεν είναι έγκυροι. Το φαινόμενο αυτό ονομάζεται φαινομενική παλινδρόμηση (Granger & Newbold, 1974).

Παίρνοντας την πρώτη διαφορά της ExportVolume προκύπτει η χρονοσειρά ExportVolume1stDiff.



Εικόνα 21. Γραφική παράσταση χρονοσειράς ExportVolume1stDiff (σε εκατ. τόνους)

Για την χρονοσειρά ExportVolume1stDiff εφαρμόζουμε τον στατιστικό έλεγχο Ljung-Box, γνωστό και ως Ljung-Box Q-Test για έλεγχο ύπαρξης αυτοσυσχέτισης των 10 πρώτων υστερήσεων.

Ljung-Box Q-Test

Μηδενική υπόθεση: Οι πρώτες m αυτοσυσχετίσεις της ExportVolume1stDiff είναι συνολικά 0.

$$H_0 : \rho_1 = \rho_2 = \dots = \rho_m = 0,$$

όπου H_0 η μηδενική υπόθεση ρ_1, \dots, ρ_m οι αυτοσυσχετίσεις της ExportVolume1stDiff

$$H_a : \rho_j \neq 0, j \in 1, \dots, m$$



όπου H_a η εναλλακτική υπόθεση με ρ_j τις αυτοσυσχετίσεις της $ExportVolume1Diff$ από 1 έως m

Παράμετροι Test

	Lags	DOF	Significance Level
1	10	10	0.05

Αποτελέσματα Test

	Null Rejected	P-Value	Test Statistic	Critical Value
1	true	5.3041e-06	42.8434	18.307

Η τιμή Test Statistic είναι στατιστικά σημαντική, άρα απορρίπτεται η μηδενική υπόθεση, δηλαδή υπάρχει αυτοσυσχέτιση μεταξύ των παρατηρήσεων, που σημαίνει ότι η παρατήρηση για χρόνο t συσχετίζεται με τις προηγούμενες 10 παρατηρήσεις.

Για την χρονοσειρά $ExportVolume1Diff$ εφαρμόζουμε το KPSS Test για έλεγχο στασιμότητας.

KPSS Test

Μηδενική υπόθεση: Η $ExportVolume1Diff$ είναι στάσιμη

$$H_0 : \sigma^2 = 0$$

$$H_a : \sigma^2 > 0$$

όπου H_0 η μηδενική υπόθεση, H_a η εναλλακτική υπόθεση και σ^2 η τυπική απόκλιση της $ExportVolume1Diff$.

Παράμετροι Test

	Lags	Include Trend	Significance Level
1	0	true	0.05

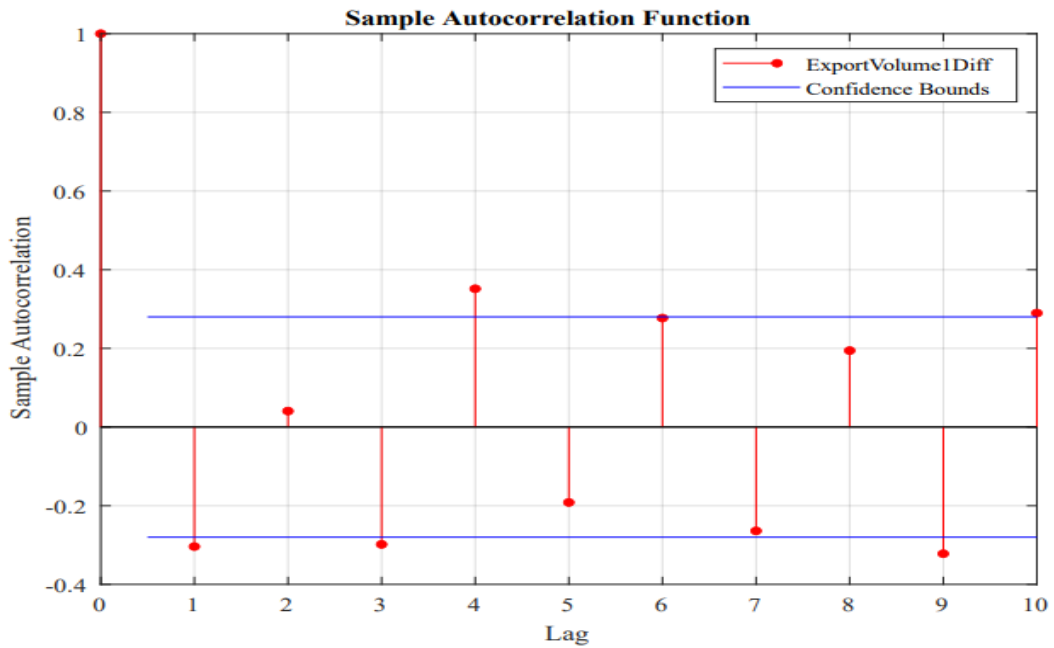
Αποτελέσματα Test

	Null Rejected	P-Value	Test Statistic	Critical Value
1	false	0.1	0.029	0.146

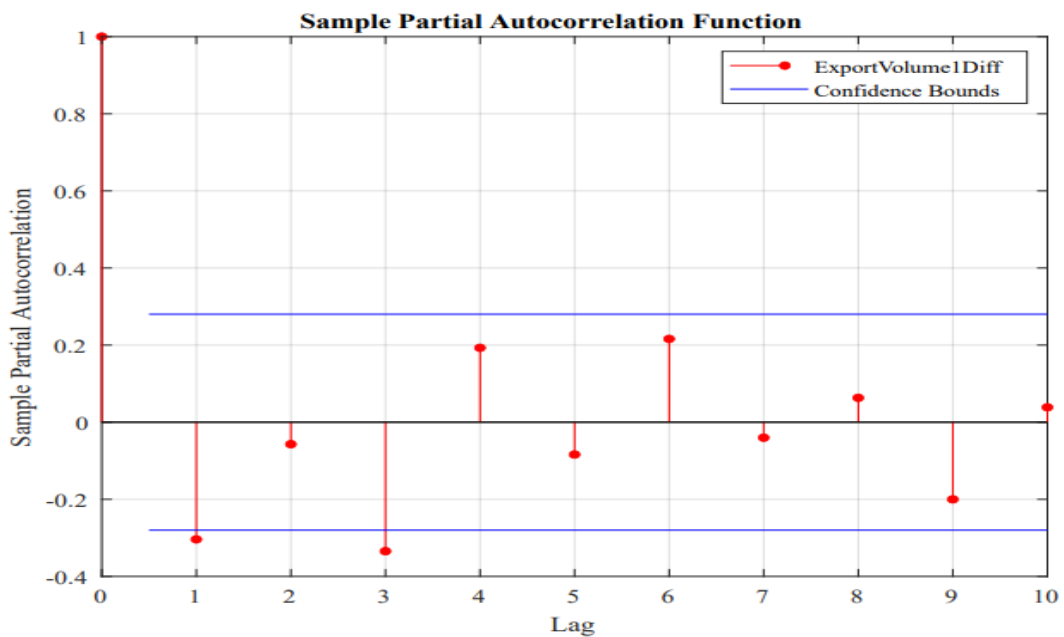


Η τιμή Test Statistic είναι στατιστικά μη σημαντική, άρα δεν απορρίπτεται η μηδενική υπόθεση, δηλαδή η χρονοσειρά είναι στάσιμη. Αυτό σημαίνει ότι μπορούμε να προχωρήσουμε στην ανάλυσή της με κάποιο γραμμικό υπόδειγμα της γενικότερης κατηγορίας ARIMA.

Παρακάτω απεικονίζονται οι συναρτήσεις αυτοσυσχέτισης και μερικής αυτοσυσχέτισης της χρονοσειράς *ExportVolume1Diff*.



Εικόνα 22. Συνάρτηση αυτοσυσχέτισης της χρονοσειράς *ExportVolume1Diff*



Εικόνα 23. Συνάρτηση μερικής αυτοσυσχέτισης της χρονοσειράς *ExportVolume1Diff*



Τα διαγράμματα ACF και PACF καταδεικνύουν ένα μοντέλο ARIMA με συντελεστή $p=1$ ή $p=3$, καθώς παρατηρείται μια ημιτονοειδής σχέση μεταξύ των παρατηρήσεων στο διάγραμμα ACF φθίνοντας σταδιακά προς το μηδέν, ενώ στο διάγραμμα PACF παρατηρούμε στατιστικά σημαντικές τιμές για $k=1$ και $k=3$. Η χρονοσειρά γίνεται στάσιμη παίρνοντας τη 1^η διαφορά της οπότε θα έχουμε $d=1$, ενώ το $q=0$, καθώς βάσει διαγραμμάτων δεν παρατηρείται κάποια σχέση που να υποδεικνύει ότι θα πρέπει να χρησιμοποιήσουμε MA όρους στο μοντέλο. Οπότε τα μοντέλα *ARIMA* (3,1,0) και *ARIMA* (1,1,0) είναι τα κατάλληλα για την ανάλυση (Hyndman & Athanasopoulos, 2018).

Υπολογίζουμε αρχικά ένα μοντέλο *ARIMA* (1,1,0) με την ακόλουθη εξίσωση:

$$(1 - \phi_1 L)(1 - L)y_t = c + \varepsilon_t$$

όπου ϕ_1 ο παράγοντας AR, ο L τελεστής υστέρησης, c μια σταθερά και ε_t ο λευκός θόρυβος.

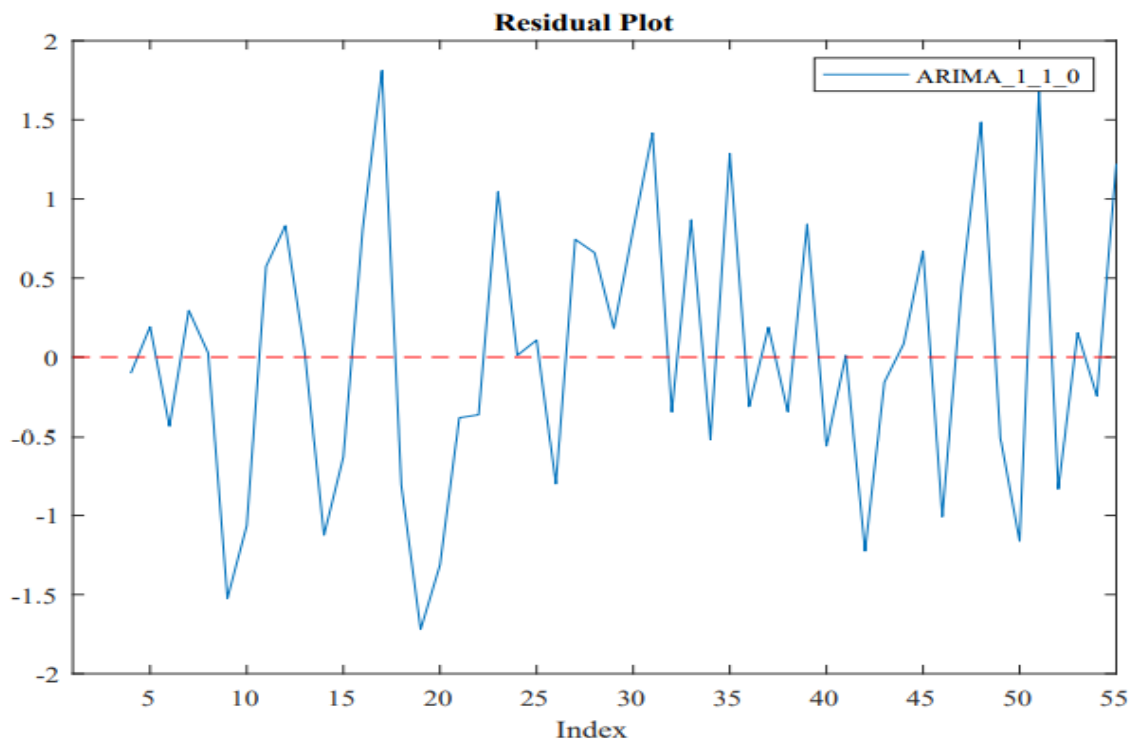
Εκτίμηση Μοντέλου *ARIMA* (1,1,0)

Αποτελέσματα εκτίμησης

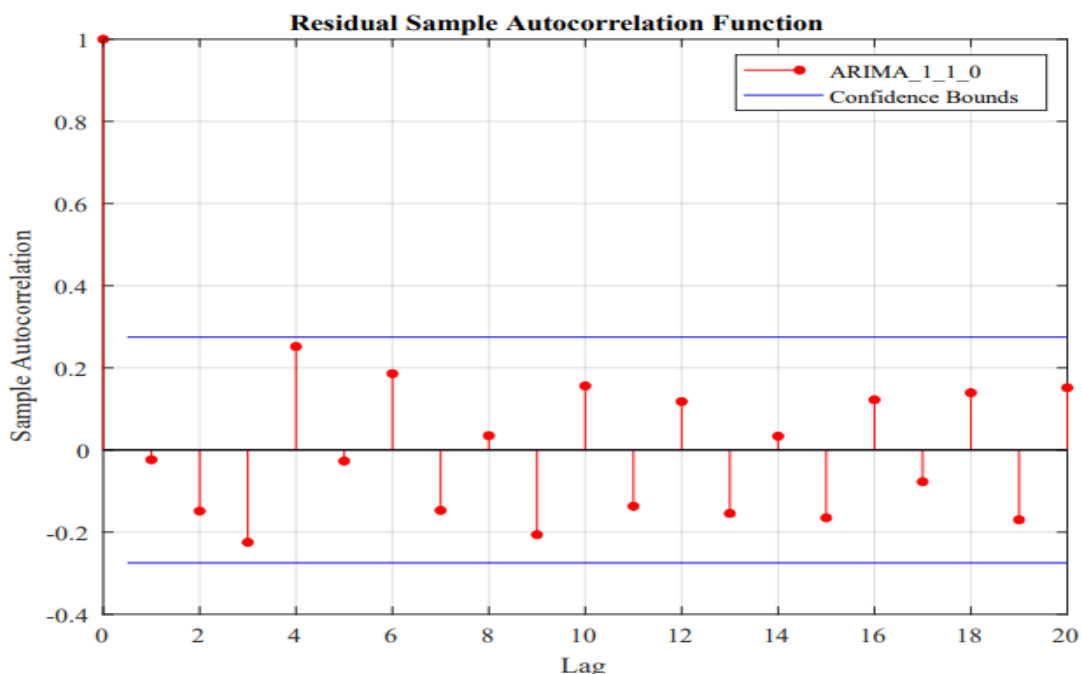
Parameter	Value	Standard Error	t Statistic	P-Value
Constant	0.1785	0.1162	1.5354	0.1247
AR{1}	-0.3053	0.1431	-2.1330	0.0329
Variance	0.7136	0.1833	3.8934	0.0001

Goodness of Fit

AIC	138.521
BIC	144.317



Εικόνα 24. Αναπαράσταση σφαλμάτων μοντέλου $ARIMA(1,1,0)$



Εικόνα 25. Συνάρτηση αυτοσυσχέτισης των σφαλμάτων του μοντέλου $ARIMA(1,1,0)$

Από την συνάρτηση αυτοσυσχέτισης των σφαλμάτων του μοντέλου συμπεραίνουμε ότι δεν υπάρχει αυτοσυσχέτιση στα σφάλματα, δηλαδή τα σφάλματα αποτελούν μια τυχαία διαδικασία, άρα το μοντέλο περιγράφει σε καλό βαθμό το ντετερμινιστικό κομμάτι της χρονοσειράς.



Εφαρμόζοντας το Ljung-Box για τα σφάλματα του μοντέλου συμπεραίνουμε ακριβώς το ίδιο.

Ljung-Box Q-Test

Μηδενική υπόθεση: Οι πρώτες m αυτοσυσχετίσεις των σφαλμάτων του $ARIMA(1,1,0)$ είναι συνολικά 0.

$$H_0 : \rho_1 = \rho_2 = \dots = \rho_m = 0,$$

όπου H_0 η μηδενική υπόθεση, ρ_1, \dots, ρ_m οι αυτοσυσχετίσεις των σφαλμάτων

$$H_a : \rho_j \neq 0, j \in 1, \dots, m$$

όπου H_a η εναλλακτική υπόθεση με ρ_j τις αυτοσυσχετίσεις των σφαλμάτων από 1 έως m

Παράμετροι Test

	Lags	DOF	Significance Level
1	10	9	0.05

Αποτελέσματα Test

	Null Rejected	P-Value	Test Statistic	Critical Value
1	false	0.0646	16.1124	16.9190

Η τιμή Test Statistic είναι στατιστικά μη σημαντική, άρα δεν απορρίπτεται η μηδενική υπόθεση, δηλαδή δεν υπάρχει αυτοσυσχέτιση μεταξύ των σφαλμάτων, τα σφάλματα είναι μια τυχαία διαδικασία.

Ακολουθούμε την ίδια διαδικασία για το μοντέλο $ARIMA(3,1,0)$. Υπολογίζουμε ένα μοντέλο $ARIMA(3,1,0)$ με την ακόλουθη εξίσωση:

$$(1 - \phi_1 L - \phi_2 L^2 - \phi_3 L^3)(1 - L)y_t = c + \varepsilon_t$$

όπου ϕ_1, ϕ_2, ϕ_3 οι παράγοντες AR, L ο τελεστής υστέρησης για $k = 1, 2, 3$, c μια σταθερά και ε_t ο λευκός θόρυβος.

Εκτίμηση Μοντέλου $ARIMA(3,1,0)$

Αποτελέσματα εκτίμησης

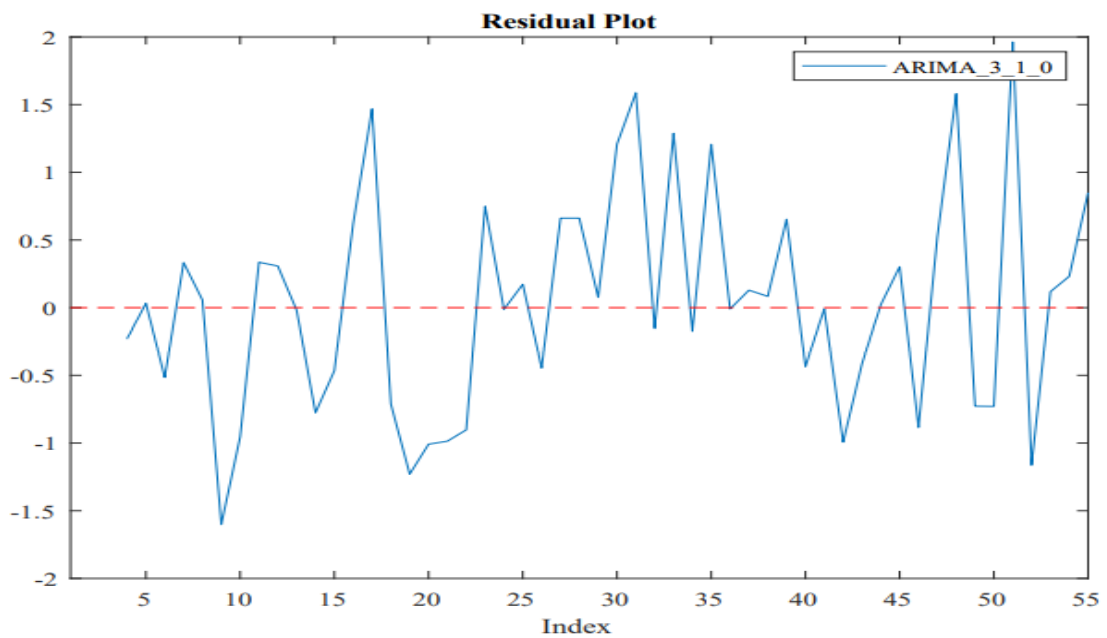
Parameter	Value	Standard Error	t Statistic	P-Value
Constant	0.2227	0.1167	1.9088	0.0563
AR{1}	-0.33486	0.1364	-2.4544	0.0141
AR{2}	-0.13912	0.1367	-1.0180	0.3087



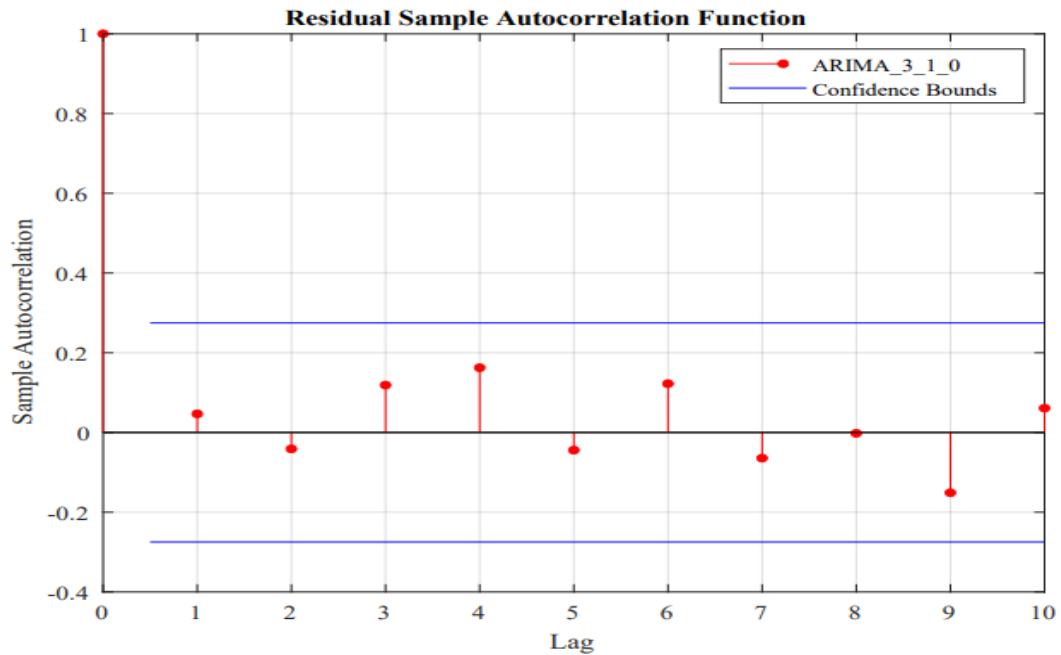
AR{3}	-0.3248	0.1421	-2.2857	0.0223
Variance	0.6407	0.1571	4.0789	0.000098

Goodness of Fit

AIC	136.8088
BIC	146.2679



Εικόνα 26. Αναπαράσταση σφαλμάτων μοντέλου ARIMA (3,1,0)



Εικόνα 27. Συνάρτηση αυτοσυσχέτισης των σφαλμάτων του μοντέλου ARIMA (3,1,0).

Από την συνάρτηση αυτοσυσχέτισης των σφαλμάτων του μοντέλου συμπεραίνουμε ότι δεν υπάρχει αυτοσυσχέτιση στα σφάλματα, δηλαδή τα σφάλματα αποτελούν μια τυχαία διαδικασία, άρα το μοντέλο περιγράφει σε καλό βαθμό το ντετερμινιστικό κομμάτι της χρονοσειράς.

Ljung-Box Q-Test

Μηδενική υπόθεση: Οι πρώτες m αυτοσυσχετίσεις των σφαλμάτων του ARIMA (3,1,0) είναι συνολικά 0.

$$H_0 : \rho_1 = \rho_2 = \dots = \rho_m = 0$$

όπου H_0 η μηδενική υπόθεση, ρ_1, \dots, ρ_m οι αυτοσυσχετίσεις των σφαλμάτων

$$H_a : \rho_j \neq 0, j \in 1, \dots, m$$

όπου H_a η εναλλακτική υπόθεση με ρ_j τις αυτοσυσχετίσεις των σφαλμάτων από 1 έως m .

Παράμετροι Test

	Lags	DOF	Significance Level
1	10	9	0.05

Αποτελέσματα Test

	Null Rejected	P-Value	Test Statistic	Critical Value
	false	0.7703	5.692	16.9190



Ο έλεγχος οδήγησε σε στατιστικά μη σημαντικό αποτέλεσμα, άρα δεν απορρίπτεται η μηδενική υπόθεση, δηλαδή δεν υπάρχει αυτοσυσχέτιση μεταξύ των σφαλμάτων, τα σφάλματα είναι μια τυχαία διαδικασία.

Από τα παραπάνω συμπεραίνουμε ότι και τα δύο μοντέλα περιγράφουν καλά τη χρονοσειρά, οπότε θα πρέπει να χρησιμοποιήσουμε κάποιο κριτήριο για να επιλέξουμε το καταλληλότερο από τα δύο και να προχωρήσουμε στη πρόβλεψη. Το κριτήριο είναι το AIC (Akaike Information Criterion), το οποίο αναφέρθηκε παραπάνω.

Πραγματοποιήθηκαν δοκιμές με διάφορα μοντέλα πάνω στα δεδομένα, όπου διαπιστώθηκε ότι τα μοντέλα ARIMA (3,1,0) ωστόσο για λόγους σύντμησης της εργασίας, παρουσιάστηκε η ανάλυση των 2 καταλληλότερων μοντέλων.

<u>Μοντέλο</u>	<u>AIC</u>	<u>BIC</u>
ARIMA (1,0,1)	139.2749	147.0799
ARIMA (1,1,0)	138.5215	144.3170
ARIMA (2,1,0)	140.4164	148.0644
ARIMA (3,1,0)	136.8088	146.2679

Επομένως θα χρησιμοποιήσουμε τα μοντέλα ARIMA (1,1,0) και ARIMA (3,1,0), καθώς έχει το ένα έχει μικρότερο AIC, ενώ το άλλο μικρότερο BIC. Οπότε δεν υπάρχει ένα σαφώς καλύτερο μοντέλο βάσει των κριτηρίων αυτών.

6.4 Παρουσίαση αποτελεσμάτων και διαγραμμάτων προβλέψεων

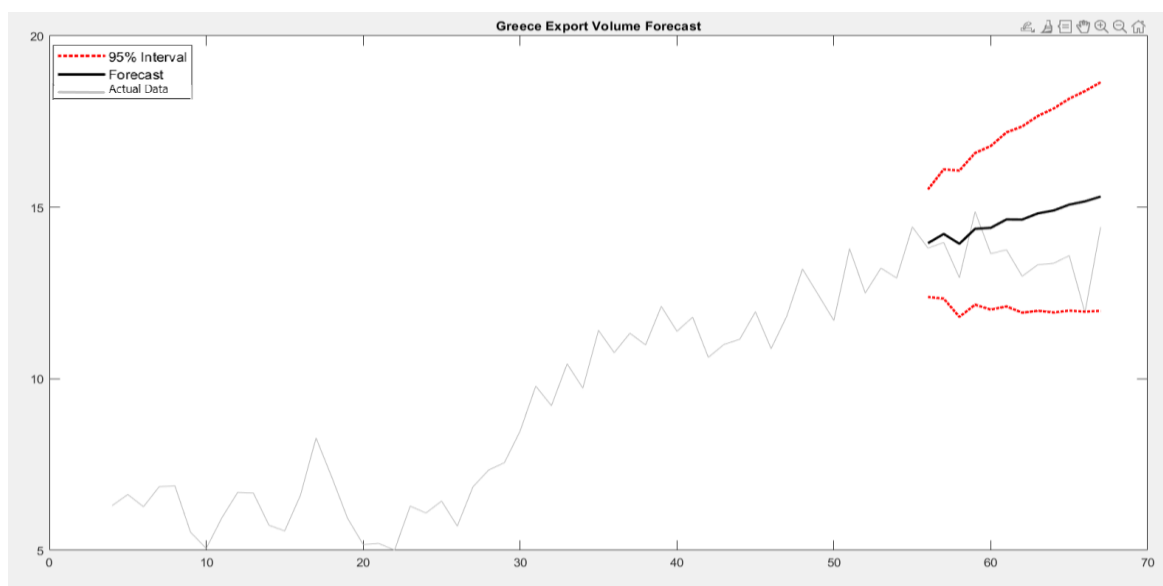
Χρησιμοποιώντας τα μοντέλα ARIMA παράγουμε προβλέψεις για τις 12 παρατηρήσεις των δεδομένων, τις οποίες δεν χρησιμοποιήσαμε στην ανάλυση, ώστε να τις έχουμε σαν μέτρο σύγκρισης σε σχέση με τις παραγόμενες προβλέψεις από το μοντέλο.

Από το μοντέλο ARIMA (3,1,0) παράγονται οι παρακάτω 12 προβλέψεις. Η μονάδα μέτρησης είναι σε εκατομμύρια τόνους.

	Ημερομηνία	Προβλέψεις	Πραγματικά Δεδομένα
1	'Q2 2018'	13.955	14.427
2	'Q3 2018'	14.222	13.811
3	'Q4 2018'	13.932	13.975



4	'Q1 2019'	14.372	12.947
5	'Q2 2019'	14.402	14.872
6	'Q3 2019'	14.646	13.648
7	'Q4 2019'	14.642	13.762
8	'Q1 2020'	14.823	12.999
9	'Q2 2020'	14.906	13.327
10	'Q3 2020'	15.077	13.366
11	'Q4 2020'	15.173	13.059
12	'Q1 2021'	15.313	11.928



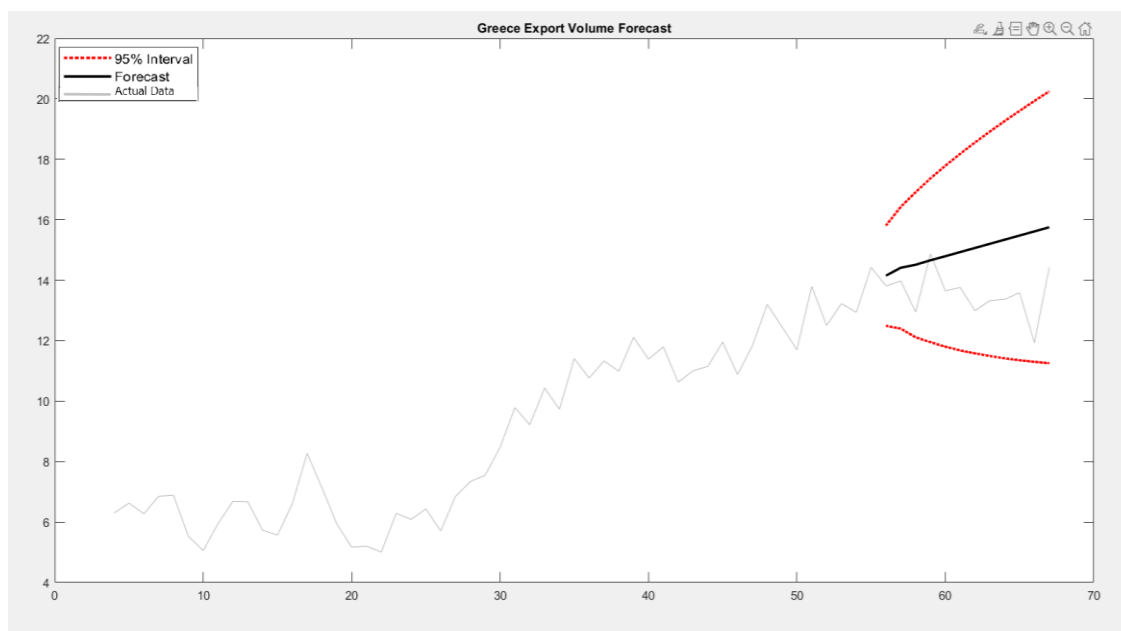
Εικόνα 28. Απεικόνιση πρόβλεψης μοντέλου $ARIMA(3,1,0)$ σε διάστημα εμπιστοσύνης 95%.

Ομοίως από το μοντέλο $ARIMA(1,1,0)$ παράγονται οι παρακάτω 12 προβλέψεις. Η μονάδα μέτρησης είναι σε εκατομμύρια τόνους.

	Ημερομηνία	Προβλέψεις	Πραγματικά Δεδομένα
1	'Q2 2018'	14.150	14.427
2	'Q3 2018'	14.413	13.811
3	'Q4 2018'	14.511	13.975



4	'Q1 2019'	14.660	12.947
5	'Q2 2019'	14.793	14.872
6	'Q3 2019'	14.931	13.648
7	'Q4 2019'	15.067	13.762
8	'Q1 2020'	15.204	12.999
9	'Q2 2020'	15.341	13.327
10	'Q3 2020'	15.478	13.366
11	'Q4 2020'	15.614	13.059
12	'Q1 2021'	15.751	11.928



Εικόνα 29. Απεικόνιση πρόβλεψης μοντέλου ARIMA(1,1,0) σε διάστημα εμπιστοσύνης 95%.

6.5 Μέτρηση Απόδοσης Μοντέλων Πρόβλεψης

Βασικό κομμάτι της ανάλυσης χρονοσειρών αποτελεί η αξιολόγηση των σφαλμάτων του μοντέλου που χρησιμοποιείται για την εξαγωγή των προβλέψεων. Για αυτό το λόγο είναι σημαντικό να ορίσουμε τα μέτρα που χρησιμοποιούνται για την εύρεση των σφαλμάτων.

Το MBE (Mean Bias Error) ορίζεται ως:

$$MBE = \frac{1}{n} \sum_{i=1}^n (\hat{y}_i - y_i) ,$$



όπου n το σύνολο των παρατηρήσεων, \hat{y}_i οι προβλέψεις και y_i τα πραγματικά δεδομένα.

Σαν στατιστικό μέτρο χρησιμοποιείται, γιατί αναδεικνύει αν η πρόβλεψη τείνει προς υπερεκτίμηση ή υποεκτίμηση των πραγματικών τιμών. Αν η τιμή του είναι θετική, τότε το μοντέλο τείνει να υπερεκτιμά τις πραγματικές τιμές, ενώ αν είναι αρνητική γίνεται υποεκτίμηση των πραγματικών τιμών

Το MAE (Mean Absolute Error) ορίζεται ως:

$$MAE = \frac{1}{n} \sum_{i=1}^n |\hat{y}_i - y_i|,$$

όπου n το σύνολο των παρατηρήσεων, \hat{y}_i οι προβλέψεις και y_i τα πραγματικά δεδομένα.

Το MAE είναι ένα στατιστικό μέτρο, το οποίο περιγράφει καλύτερα σφάλματα των οποίων η κατανομή είναι ενιαία, ωστόσο επειδή τα σφάλματα του μοντέλου είναι σύνηθες να έχουν κανονική κατανομή, το RMSE είναι καλύτερο στατιστικό μέτρο για τη παρουσίαση τέτοιου τύπου δεδομένων (Chai & Draxler, 2014).

Το RMSE (Root Mean Square Error) ορίζεται ως:

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (\hat{y}_i - y_i)^2},$$

όπου n το σύνολο των παρατηρήσεων, \hat{y}_i οι προβλέψεις και y_i τα πραγματικά δεδομένα.

Το RMSE παρέχει πληροφορίες για τη βραχυπρόθεσμη απόδοση του μοντέλου επιτρέποντας μια σύγκριση μεταξύ κάθε παρατήρησης και πρόβλεψης. Όσο μικρότερη η τιμή του, τόσο καλύτερη η απόδοση του μοντέλου. Ένα μειονέκτημά του είναι ότι μερικά μεγάλα σφάλματα αυξάνουν δραστικά τη τιμή του, καθώς τετραγωνίζονται. Επίσης δεν παρέχει πληροφορία για το αν έχει γίνει υποεκτίμηση ή υπερεκτίμηση στη πρόβλεψη (Stone, 1993).

Το MAPE (Mean Absolute Percentage Error) ορίζεται ως:

$$MAPE = \frac{1}{n} \sum_{i=1}^n \left| \frac{\hat{y}_i - y_i}{y_i} \right| \times 100\%,$$

όπου n το σύνολο των παρατηρήσεων, \hat{y}_i οι προβλέψεις και y_i τα πραγματικά δεδομένα. Το MAPE εκφράζει το MAE ως ποσοστό. Η χρησιμότητά του είναι ότι ο καθένας μπορεί να κατανοήσει το μέγεθος ενός σφάλματος ως ποσοστό καλύτερα απ' ότι ως αριθμό.

Τα στατιστικά μέτρα των σφαλμάτων των δύο μοντέλων παρουσιάζονται στον παρακάτω πίνακα.

	ARIMA (1,1,0)	ARIMA (3,1,0)
MAE	1.542	1.2753
MAPE	11,854%	9,801%
MSE	3.4615	2.4229
RMSE	1.8605	1.5565
MBE	1.482	1.112



Βλέπουμε ότι το μοντέλο ARIMA (3,1,0) αποδίδει καλύτερα από το ARIMA (1,1,0) στην πρόβλεψη, καθώς όλα τα σφάλματα είναι μικρότερα.

Ερμηνεύοντας τους παραπάνω δείκτες σφαλμάτων βλέπουμε ότι το MBE είναι 1.482, οπότε συμπεραίνουμε ότι οι προβλέψεις του μοντέλου οδήγησαν σε υπερεκτίμηση των τιμών. Σε αυτό συνέβαλλε η δραματική μείωση των εξαγωγών κατά την περίοδο έναρξης της πανδημίας του COVID-19 κατά το 3^ο τρίμηνο του 2019, που θα μπορούσε να χαρακτηριστεί ως random shock, το οποίο γίνεται προφανές αν παρατηρήσουμε το διάγραμμα της χρονοσειράς. Η χρονοσειρά από εκείνη τη χρονική περίοδο και μετά έχει μια σημαντική τάση μείωσης. Τέτοιου είδους επιδράσεις από εξωτερικούς παράγοντες είναι εξαιρετικά δύσκολο να προβλεφθούν από οποιοδήποτε μοντέλο και λόγω αυτού, το MBE είναι θετικό.

Το RMSE είναι μεγαλύτερο από το MAE, καθώς το συγκεκριμένο μέτρο είναι πιο ευαίσθητο στην ύπαρξη μεγάλου σφάλματος μιας μόνο περιόδου σε σχέση με το MAE (Αϊβαζίδου, 2015).

Από το 2020 και μετά τα επιμέρους σφάλματα είναι μεγαλύτερα με αποτέλεσμα να αυξάνουν τετραγωνικά την τιμή του συγκεκριμένου μέτρου.

Το MAPE είναι 9.8%, δηλαδή οι τιμές των προβλέψεων κατά μέσο όρο διαφέρουν 9,8% από την τιμή των πραγματικών δεδομένων κατ' απόλυτη τιμή, ενώ το MAE είναι 1.2753, δηλαδή οι προβλέψεις διαφέρουν κατά μέσο όρο 1.2753 εκατομμύρια τόνους από τα πραγματικά δεδομένα κατ' απόλυτη τιμή.



Κεφάλαιο 7 : Συμπεράσματα – Επίλογος

Στα πλαίσια της συγκεκριμένης διπλωματικής εργασίας πραγματοποιήθηκε μια βιβλιογραφική έρευνα με σκοπό την ανάλυση των χρονοσειρών και των βασικών τους στοιχείων, όπως και των μεθοδολογιών Ανάλυσης και Πρόβλεψης Χρονοσειρών. Από τη συγκεκριμένη έρευνα προέκυψε ότι υπάρχει πληθώρα μεθόδων που αποσκοπούν στην ανάλυση και παραγωγή προβλέψεων. Η παρούσα διπλωματική εστίασε κατά κύριο λόγο στη μεθοδολογία που ακολουθείται για την ανάπτυξη μοντέλων ARIMA και υποπεριπτώσεών τους, γνωστή και ως μεθοδολογία Box-Jenkins, η οποία χρησιμοποιείται ευρέως για τη παραγωγή προβλέψεων σε πολλούς τομείς της σύγχρονης κοινωνίας.

Όσον αφορά το πρακτικό στατιστικό κομμάτι εφαρμογής της διπλωματικής εργασίας χρησιμοποιήθηκαν στατιστικά δεδομένα των εξαγωγών στην Ελλάδα από το 2004 έως το 2021, έγινε στατιστική ανάλυση βάσει των μεθοδολογιών που αναφέρονται στην εργασία, καθώς επίσης ανάπτυξη και επιλογή του καταλληλότερου μοντέλου για την παραγωγή προβλέψεων. Κάποια βασικά συμπεράσματα από τη συγκεκριμένη μελέτη είναι ότι το μέγεθος του διαθέσιμου δείγματος για τον υπολογισμό των κατάλληλων παραμέτρων του μοντέλου που θα χρησιμοποιηθεί για παραγωγή προβλέψεων διαδραματίζει πολύ σημαντικό ρόλο στην απόδοση του. Όση περισσότερη η διαθέσιμη πληροφορία, τόσο πιο ισχυρό είναι και το μοντέλο που θα αναπτυχθεί. Στη στατιστική ανάλυση της συγκεκριμένης διπλωματικής εργασίας υπήρξαν κάποιοι περιορισμοί που αφορούν κυρίως στην εύρεση δεδομένων, το οποίο ίσως να αποτέλεσε και το απαιτητικότερο κομμάτι της σε συνδυασμό βέβαια με τον περιορισμένο χρόνο για την εκπόνησή της.

Ο Ναυτιλιακός κλάδος αποτελεί έναν από τους πιο δυναμικούς κλάδους της οικονομίας, για αυτό και πολλές φορές η παραγωγή αξιόπιστων προβλέψεων παίζει σημαντικό ρόλο στη λήψη αποφάσεων. Παρ' όλα αυτά, η ευαισθησία του στις μεταβολές που επιφέρει το παγκόσμιο εμπόριο είναι μεγάλη. Η πανδημία του COVID-19 μείωσε σημαντικά την κίνηση της εφοδιαστικής αλυσίδας και κατ' επέκταση και την κίνηση εμπορευμάτων δια θαλάσσης από την Ελλάδα προς το εξωτερικό. Ερμηνεύοντας τα αποτελέσματα των προβλέψεων στο προηγούμενο κεφάλαιο το μοντέλο οδήγησε σε υπερεκτίμηση των τιμών ίσως λόγω της επίδρασης της πανδημίας που οδήγησε σε ύφεση του εμπορίου μειώνοντας την εμπορευματική κίνηση από την Ελλάδα προς το εξωτερικό. Έχοντας υπ' όψει τα παραπάνω πρέπει να τονιστεί πως οι προβλέψεις όταν χρησιμοποιούνται με σύνεση αναλογιζομένου του ρίσκου και της αβεβαιότητας μπορούν να γίνουν χρήσιμα εργαλεία, δίνοντας συγκριτικό πλεονέκτημα έναντι ανταγωνιστών και βοηθώντας στην ορθότερη λήψη αποφάσεων και χάραξης στρατηγικής. Ωστόσο, πάντα υπάρχει το περιθώριο σφάλματος, πόσο μάλλον όταν επιδρούν τυχαίοι παράγοντες (random shocks), των οποίων η επίδραση είναι σχεδόν αδύνατο να υπολογιστεί εκ των προτέρων.

Η παρούσα διπλωματική εργασία ανέλυσε τις χρονοσειρές με τη μεθοδολογία ARIMA για την παραγωγή προβλέψεων. Υπάρχουν πολλές αξιολογικές μέθοδοι προβλέψεων όπως οι αιτιακές μέθοδοι (Armstrong et al , 2005), η μηχανική μάθηση (Athey & Imbens, 2019), τα νευρωνικά δίκτυα (Livingston, 2008) και οι ποιοτικές μέθοδοι (Sanders & Ritzman, 2004). Όλες οι μέθοδοι παρουσιάζουν θετικά και αρνητικά σημεία. Η σε βάθος μελέτη τους για την εύρεση του καταλληλότερου τρόπου προσέγγισης, μοντελοποίησης και ανάλυσης της εκάστοτε χρονοσειράς αποτελεί σίγουρα ένα θέμα που αξίζει την προσοχή της επιστημονικής κοινότητας.



Κεφάλαιο 8: Βιβλιογραφία

- Adhikari, R. & Agrawal, R. K. (2013). *An Introductory Study on Time Series Modeling and Forecasting*.
- Akaike, H. (1973). *Information theory and an extension of the maximum likelihood principle*. In B. N. Petrov & F. Csaki (Eds.), 2nd International Symposium on Information Theory (pp. 267-281). Budapest: Akademiai Kiado.
- Armstrong, J. S., Collopy, F. & Yokum, J. T. (2005). *Decomposition by causal forces: a procedure for forecasting complex time series*. *International Journal of forecasting*, 21(1), 25-36.
- Athey, S. & Imbens, G. W. (2019). *Machine learning methods that economists should know about*. *Annual Review of Economics*, 11, pp. 685-725.
- Box, G.E.P. & Jenkins, G. (1970). *Time Series Analysis, Forecasting and Control*. Holden-Day, San Francisco, CA.
- Box, G. E. P., Jenkins, G. M. & Reinsel, G. C. (1994). *Time Series Analysis: Forecasting and Control*. 3rd ed. Englewood Cliffs, NJ: Prentice Hall.
- Burger, C. E. (2000, June). *A Multivariate Time Series Analysis of U.S. Army Recruiting*. Monterey, CA 93943-5000: Naval Postgraduate School.
- Chai, T. & Draxler, R. R. (2014, June 30). *Root mean square error (RMSE) or mean absolute error (MAE)? Arguments against avoiding RMSE in the literature*. USA: Geoscientific model development. doi:10.5194/gmd-7-1247-2014
- Gomez, V. (2019). *Linear Time Series with MATLAB and OCTAVE*. Switzerland: Springer Nature.
- Granger, C. W., & Newbold, P. (1974). *Spurious regressions in econometrics*. *Journal of econometrics*, 2(2), 111-120.
- Hyndman, R. J. (2014, January 24). *Thoughts on the Ljung-Box Test*. *Hyndsight*, published online. Available: <https://robjhyndman.com/hyndsight/ljung-box-test/>
- Hyndman, R. J. & Athanasopoulos, G. (2018). *Forecasting: Principles and Practice, 2nd Edition*. Australia.
- Kwiatkowski, D., Phillips, P. C. B., Schmidt, P. & Shin, Y. (1992). *Testing the Null Hypothesis of Stationarity against the Alternative of a Unit Root*. *Journal of Econometrics*. Vol. 54, σσ. pp. 159–178.
- Livingstone, D. J. (2008). *Artificial neural networks: methods and applications* (pp. 185-202). Totowa, NJ, USA: Humana Press.
- Makridakis, S., Wheelwright, S. & Hyndman, R. (1998). *Forecasting*. John Wiley and Sons.
- Montgomery, D. C., Jennings C. L. & Kulahci M. (2008). *Introduction to Time Series Analysis and Forecasting*. USA: WILEY-INTERSCIENCE.
- NIST/SEMATECH. *e-Handbook of Statistical Methods*, Available: <http://www.itl.nist.gov/div898/handbook/>
- Pedregal, D. J. (2019). *Time series analysis and forecasting with ECOTOOL*. Ciudad Real, Spain: PLoS ONE. Ανάκτηση από <https://doi.org/10.1371/journal.pone.0221238>



Σούντρης Σταύρος,
“Ανάλυση και Πρόβλεψη Χρονοσειρών
με Εφαρμογές σε Θέματα Ναυτιλίας”

Sanders, N. R. & Ritzman, L. P. (2004). *Integrating judgmental and quantitative forecasts: methodologies for pooling marketing and operations information*. International Journal of Operations & Production Management.

Stone, R. (1993). *Improved statistical procedure for the evaluation of solar radiation estimation models*, Solar Energy (Issue 4, Τόμ. 51). Available:
<https://www.sciencedirect.com/science/article/pii/0038092X93901247>

The MathWorks. (2007). *MATLAB Data Analysis*. Natick, MA, USA.

Αθηνιώτης, Ν. (2019, Φεβρουάριος). *Ομαδοποίηση Χρονοσειρών Βάσει Ποιοτικών Χαρακτηριστικών και Μοτίβων και Αξιοποίηση Ευρημάτων για τη Βελτίωση της Ακρίβειας Πρόβλεψης Κλασικών Μεθόδων*. Αθήνα: Ε.Μ.Π.

Αϊβαζίδου, Ε. (2015, Νοέμβριος). *Ανάπτυξη μοντέλων χρονοσειρών και παλινδρόμησης για την αξιολόγηση των επιπτώσεων της οικονομικής κρίσης στη θαλάσσια επιβατική και εμπορευματική κίνηση στην Ελλάδα*. Θεσσαλονίκη: Α.Π.Θ.

Ελληνική Στατιστική Αρχή. Available: [//www.statistics.gr/el/statistics/itr](http://www.statistics.gr/el/statistics/itr)

Κατρακυλίδης, Κ., Κοντέος, Γ. & Σαριαννίδης, Ν. (2015). *Εισαγωγή στη Σύγχρονη Οικονομετρική Ανάλυση*. Θεσσαλονίκη.

Κουγιουμτζής, Δ. (2011, Μάιος 10). *Χρονοσειρές*. Ελλάδα.

Νταβίντ, Κ. (2018, Φεβρουάριος). *Συλλογή Μακροχρόνιων Χρονοσειρών για Παραγωγή και Αξιολόγηση Μακροπρόθεσμων Προβλέψεων*. Αθήνα: Ε.Μ.Π., ΠΑ.ΠΕΙ.