



ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΕΙΡΑΙΩΣ

Τμήμα Ψηφιακών Συστημάτων

Πρόγραμμα Μεταπτυχιακών Σπουδών

**“Πληροφοριακά Συστήματα και Υπηρεσίες”
με κατεύθυνση στα “Μεγάλα Δεδομένα και
Αναλυτική”**

Διπλωματική Εργασία

**ΑΝΑΛΥΣΗ ΤΗΣ ΤΙΜΗΣ ΤΟΥ ΑΡΓΟΥ
ΠΕΤΡΕΛΑΙΟΥ ΚΑΙ ΕΦΑΡΜΟΓΗ ΣΥΓΧΡΟΝΩΝ
ΜΕΘΟΔΩΝ ΓΙΑ ΤΗΝ ΠΡΟΒΛΕΨΗ ΤΗΣ**

του

Ιωάννη Λαζαρίδη

Επιβλέπων Καθηγητής: Μιχαήλ Φιλιππάκης

ΑΘΗΝΑ

2021

Ευχαριστίες

Με την ολοκλήρωση της διπλωματικής μου εργασίας θα ήθελα να εκφράσω τις ευχαριστίες μου σε όλους όσους βοήθησαν και συνέβαλλαν καθοριστικά στην εκπόνησή της.

Ευχαριστώ θερμά τον επιβλέπων καθηγητή μου κύριο Φιλιππάκη Μιχαήλ που από την πρώτη στιγμή με εμπιστεύτηκε και δέχτηκε να αναλάβει το συγκεκριμένο θέμα. Η επιστημονική του καθοδήγηση, το αμείωτο ενδιαφέρον του και η συμπαράστασή του αποτέλεσαν αρωγές καθοριστικής σημασίας στην εκπόνηση της διπλωματικής μου εργασίας. Επίσης θα ήθελα να ευχαριστήσω τη Δρ. Μαρία Ελένη Πούλου για την πολύτιμη βοήθειά της στην επίβλεψη της διπλωματικής.

Επιπλέον, ιδιαίτερες ευχαριστίες θα ήθελα να απευθύνω στην καθηγήτρια κυρία Πρέντζα Ανδριάνα και στην αναπληρώτρια καθηγήτρια κυρία Χαλκίδη Μαρία για τις πολύτιμες υποδείξεις και συμβουλές τους ως μέλη της τριμελούς επιτροπής καθώς και σε όλους τους καθηγητές του μεταπτυχιακού για τις γνώσεις που αποκόμισα.

Τέλος, θα ήθελα να εκφράσω τη βαθιά ευγνωμοσύνη μου σε όλα τα μέλη της οικογένειάς μου για την πολύτιμη στήριξη, συμπαράσταση και κατανόηση τους.

Περίληψη

Στην παρούσα εργασία μελετάται το αργό πετρέλαιο ως παράγοντας επιρροής της παγκόσμιας οικονομίας αλλά και ως επένδυση και γίνεται μία εμπειρική ανάλυση της τιμής του και των παραγόντων που συσχετίζονται με αυτήν.

Η διαμόρφωση της τιμής του πετρελαίου είναι ένα πολύ σημαντικό ζήτημα που αφορά την παγκόσμια κοινότητα καθώς σχετίζεται αμφίδρομα με τους περισσότερους χρηματοοικονομικούς δείκτες της οικονομίας, τους οποίους επηρεάζει σε πολύ μεγάλο βαθμό αλλά και επηρεάζεται από αυτούς. Κατά πολλούς αποτελεί τον βασικότερο παράγοντα ρύθμισης της λειτουργίας της παγκόσμιας οικονομίας, μιας και παραμένει η κυριότερη πηγή ενέργειας του πλανήτη. Οι διακυμάνσεις στην τιμή του πετρελαίου έχουν άμεσο αντίκτυπο σε όλα τα μακροοικονομικά μεγέθη όπως ο πληθωρισμός, τα διεθνή επιτόκια και οι ισοτιμίες και μία ανεξέλεγκτη ανοδική πορεία της τιμής του μπορεί να οδηγήσει σε επιβράδυνση της παγκόσμιας οικονομικής δραστηριότητας.

Στις μέρες μας αποτελεί κατά πολλούς επενδυτική επιλογή ως παράγοντας διαφοροποίησης χαρτοφυλακίου προσελκύνοντας ολοένα και περισσότερους επενδυτές και κάνοντας πλέον επιτακτική την ανάγκη κατανόησης των χαρακτηριστικών της αγοράς του. Για το λόγο αυτό, πληθώρα ερευνητών εξετάζει τους παράγοντες που μεταβάλλουν την τιμή του αργού πετρελαίου αναπτύσσοντας προβλεπτικά μοντέλα που έχουν ως βάση ιστορικές καταγραφές των τιμών άλλων χρηματοοικονομικών στοιχείων και εμπορευμάτων αλλά κυρίως της ίδιας της τιμής του αργού πετρελαίου. Ωστόσο, η πρόβλεψη της τιμής του πετρελαίου παραμένει ένα από τα δυσκολότερα προβλήματα στον κόσμο των αναλύσεων λόγω της μεγάλης μεταβλητότητας που παρουσιάζει.

Σκοπός της παρούσας εργασίας είναι να παράσχει χρήσιμες πληροφορίες στις ομάδες ενδιαφερομένων της αγοράς του πετρελαίου μελετώντας τις κινήσεις της παγκόσμιας αγοράς και δημιουργώντας κατάλληλο προβλεπτικό μοντέλο συγκρίνοντας ως προς την ακρίβεια ενδεδειγμένες εφαρμογές μηχανικής μάθησης. Στην προσπάθεια εξαγωγής χρήσιμων συμπερασμάτων για τη μεταβλητότητα της τιμής του πετρελαίου θα αναλυθούν οι συσχετίσεις των τιμών του χρυσού, του ασημιού, του 10ετούς συμβολαίου μελλοντικής εκπλήρωσης (10 Year T-Note), του πληθωρισμού των Ηνωμένων Πολιτειών και του χρηματιστηριακού δείκτη NASDAQ 100 σε μηνιαία βάση από τον Ιανουάριο του 2000 μέχρι το Δεκέμβριο του 2020. Θα πραγματοποιηθεί καταρχάς περιγραφική ανάλυση των παραπάνω δεικτών για να βρεθούν πιθανές συσχετίσεις μεταξύ τους και στη συνέχεια δημιουργία κατάλληλων προβλεπτικών μοντέλων μηχανικής μάθησης με βάση τις ιστορικές καταγραφές της τιμής του πετρελαίου.

Abstract

In the present work, crude oil is studied as a factor of influence of the world economy and as an investment and an empirical analysis is made of its price and the factors related to it.

The oil's price initiation is a very important issue for the world community as it is bilaterally related to most of the financial indicators of the economy, which it affects to a great extent and is also influenced by them. According to many people, it is the main factor regulating the functioning of the world economy since it remains the main source of energy on the planet. Oil's price fluctuations have a direct impact on all macroeconomic variables such as inflation, international interest rates and exchange rates, and an uncontrolled upward trend in the price of oil can lead to a slowdown in global economic activity.

Nowadays crude oil is an investment choice as a factor of portfolio diversification attracting more and more investors and making it imperative to understand the characteristics of its market. For this reason, a multitude of researchers are examining the factors that change the price of crude oil by developing predictive models based on historical records of the prices of other financial assets and commodities but mainly the price of crude oil itself. However, predicting the price of oil remains one of the most difficult problems in the world of analysis due to its high volatility.

The purpose of this work is to provide useful information to oil market stakeholders by studying global market movements and creating an appropriate predictive model comparing in terms of accuracy machine learning applications. In order to draw useful conclusions about oil price volatility, the correlations of gold, silver, 10-year T-Note, US inflation and the NASDAQ 100 monthly stock index will be analysed on a monthly basis from January 2000 until December 2020. A descriptive analysis of the above indicators will be carried out first in order to detect possible correlation between them, followed by the creation of appropriate predictive machine learning models based on the historical records of the oil price.

Περιεχόμενα

1. Εισαγωγή	8
2. Επισκόπηση Βιβλιογραφίας.....	10
3. Η Αγορά του Πετρελαίου	13
3.1. Εύρεση, εξόρυξη και επεξεργασία	13
3.2. Οι χρήσεις του πετρελαίου και η ζήτηση ανά κατηγορία	14
3.2.1. Μεταφορές	14
3.2.2. Βιομηχανικά Προϊόντα Μη Ενεργειακών Σκοπών και λοιπές χρήσεις	15
3.3. Παράγοντες επιρροής της τιμής του πετρελαίου και συσχετιζόμενα χρηματοοικονομικά προϊόντα	15
4. Η τιμή του αργού πετρελαίου την περίοδο Ιανουάριος 2000 – Δεκέμβριος 2020: Εμπειρική διερεύνηση των συσχετιζόμενων δεικτών και δημιουργία προβλεπτικού μοντέλου	17
4.1. Περιγραφικά στοιχεία των υπό μελέτη δεικτών και Ανάλυση των Συσχετίσεων μεταξύ τους.....	17
4.2. Ανάλυση Χρονοσειρών – Επισκόπηση Θεωρίας	19
4.2.1 Χαρακτηριστικά Χρονοσειρών	19
4.2.2 Ολοκληρωμένα Αυτοπαλίνδρομα Μοντέλα Κινητού Μέσου Όρου (ARIMA)	22
4.2.3 Η μεθοδολογία Box – Jenkins	23
4.2.4 Τεχνητά Νευρωνικά Δίκτυα	25
4.2.5 Το μοντέλο LSTM.....	27
4.3. Προσδιορισμός κατάλληλου μοντέλου ARIMA στη χρονοσειρά Ιανουάριος 2000 – Δεκέμβριος 2020 και δημιουργία πρόβλεψης της τιμής του πετρελαίου κατά την περίοδο Ιανουάριος 2021 – Οκτώβριος 2021.	30
4.4. Προσδιορισμός κατάλληλου μοντέλου LSTM στη χρονοσειρά Ιανουάριος 2000 – Δεκέμβριος 2020 και δημιουργία πρόβλεψης της τιμής του πετρελαίου κατά την περίοδο Ιανουάριος 2021 – Οκτώβριος 2020.	37
5. Συμπεράσματα	42
Βιβλιογραφία	44
Παράρτημα 1. Διαγράμματα καταλοίπων των υπό αξιολόγηση μοντέλων ARIMA	46

Λίστα γραφημάτων

Γράφημα 1 Κατανάλωση Ενέργειας ανά τύπο καυσίμου (EU)	14
Γράφημα 2 Κατανάλωση Ενέργειας ανά κατηγορία (EU)	15
Γράφημα 3 Χρονοσειρές Συσχετιζόμενων Παραγόντων.....	16
Γράφημα 4 Χρονοσειρά Λευκού Θορύβου	19
Γράφημα 5 Μορφή Στάσιμης και μη Στάσιμης Χρονοσειράς.....	20
Γράφημα 6 Διαγράμματα Αυτοσυσχέτισης μη στάσιμης (αριστερά) και στάσιμης (δεξιά) χρονοσειράς.....	21
Εικόνα 7 Artificial Neural Network	26
Εικόνα 8 RNN Network (Olah)	27
Εικόνα 9 LSTM Network (Yan)	28
Γράφημα 10 Περιγραφικά Διαγράμματα της χρονοσειράς Ιανουάριος 2000 – Δεκέμβριος 2020	30
Γράφημα 11 Διάγραμμα αυτοσυσχέτισης της χρονοσειράς Ιανουάριος 2000 – Δεκέμβριος 2020	31
Γράφημα 12 Διάγραμμα μερικής αυτοσυσχέτισης της χρονοσειράς Ιανουάριος 2000 – Δεκέμβριος 2020.....	31
Γράφημα 13 Διάγραμμα της χρονοσειράς Ιανουάριος 2000 – Δεκέμβριος 2020 με 1ης τάξης διαφορές στις τιμές	32
Γράφημα 14 Διάγραμμα της χρονοσειράς Ιανουάριος 2000 – Δεκέμβριος 2020 με λογαρίθμους και 1 ^{ης} τάξης διαφορές στις τιμές	32
Γράφημα 15 Διάγραμμα αυτοσυσχέτισης της χρονοσειράς Ιανουάριος 2000 – Δεκέμβριος 2020 με λογαρίθμους και 1 ^{ης} τάξης διαφορές στις τιμές.....	33
Γράφημα 16 Διάγραμμα μερικής αυτοσυσχέτισης της χρονοσειράς Ιανουάριος 2000 – Δεκέμβριος 2020 με λογαρίθμους και 1 ^{ης} τάξης διαφορές στις τιμές	33
Γράφημα 17 Διαγράμματα καταλοίπων μοντέλου ARIMA (1, 1, 2)(1, 0, 0)[12]	36
Γράφημα 18 Σφάλμα Πρόβλεψης μοντέλου ARIMA (1, 1, 2)(1, 0, 0)[12]	37
Γράφημα 19 Train/Validation Losses of Implemented LSTMs	41
Γράφημα 20 Actual – Forecasted Prices of LSTM & ARIMA Models	43
Γράφημα 21 Διαγράμματα καταλοίπων μοντέλου ARIMA (1,1,1).....	46
Γράφημα 22 Διαγράμματα καταλοίπων μοντέλου ARIMA (2,1,1).....	47
Γράφημα 23 Διαγράμματα καταλοίπων μοντέλου ARIMA (1,1,2).....	48
Γράφημα 24 Διαγράμματα καταλοίπων μοντέλου ARIMA (1,1,3).....	49
Γράφημα 25 Διαγράμματα καταλοίπων μοντέλου ARIMA (2,1,3).....	50
Γράφημα 26 Διαγράμματα καταλοίπων μοντέλου ARIMA (3,1,1).....	51
Γράφημα 27 Διαγράμματα καταλοίπων μοντέλου ARIMA (3,1,2).....	52
Γράφημα 28 Διαγράμματα καταλοίπων μοντέλου ARIMA (3,1,3).....	53
Γράφημα 29 Διαγράμματα καταλοίπων μοντέλου ARIMA (1,1,1)(1,0,0)[12]	54

Λίστα Πινάκων

Πίνακας 1 Περιγραφικά στοιχεία των υπό μελέτη δεικτών για την περίοδο Ιανουάριος 2000 – Δεκέμβριος 2020.....	17
Πίνακας 2 Βαθμός συσχέτισης και επίπεδο σημαντικότητας της συσχέτισης των υπό μελέτη μεταβλητών	18
Πίνακας 3 Κριτήρια αξιολόγησης των μοντέλων ARIMA στη χρονοσειρά Ιανουάριος 2000 – Δεκέμβριος 2020.....	34
Πίνακας 4 Εκτιμημένοι συντελεστές ARIMA (1, 1, 2)(1, 0, 0)[12].....	35
Πίνακας 5 Αποτελέσματα ελέγχου Ljung – Box στο ARIMA (1, 1, 2)(1, 0, 0)[12]	35
Πίνακας 6 Προβλεπόμενες τιμές χρυσού για τις μηνιαίες τιμές της περιόδου Ιανουάριος 2021 – Οκτώβριος 2021.....	36
Πίνακας 7 LSTM(4), Dropout(0.2), Dense(1) Results.....	40
Πίνακας 8 LSTM(4), Dropout(0.2), Dense(1) Forecasted Prices & MAPE.....	41
Πίνακας 9 biLSTM(4), Dropout(0.2), Dense(1) & 4 Layer LSTM Results	41

1. Εισαγωγή

Ο κόσμος των επενδύσεων προσελκύει ένα μεγάλο μέρος ανθρώπων στις σύγχρονες κοινωνίες καθορίζοντας παράλληλα σε σημαντικό βαθμό τον οικονομικό βίο και των μη συμμετεχόντων. Το παγκόσμιο οικονομικό σύστημα εν γένει για να λειτουργήσει απαιτεί την πραγματοποίηση επενδύσεων σε μεγάλη κλίμακα ούτως ώστε να πραγματοποιηθούν έργα και να επιτευχθεί κέρδος ως απόδοση των επενδυτικών και των δανειακών κεφαλαίων.

Οι επενδύσεις στην Κεϋνσιανή ανάλυση κατέχουν τον κεντρικό ρόλο στη δυναμική του οικονομικού συστήματος, η έλλειψη των οποίων οδηγεί σε μείωση του εθνικού εισοδήματος και κατ' επέκταση της ζήτησης και της αποταμίευσης. Στις σύγχρονες οικονομίες, καταλυτικό ρόλο στον κόσμο των επενδύσεων παίζει η παγκοσμιοποίηση η οποία έχει οδηγήσει σε μεγαλύτερη διασύνδεση μεταξύ των αγορών και στην αύξηση της επικοινωνίας και των επιχειρηματικών ευκαιριών ακόμα και στις απομακρυσμένες περιοχές του πλανήτη. Έτσι περισσότεροι επενδυτές μπορούν να έχουν πρόσβαση σε νέες επενδυτικές ευκαιρίες και οι πιθανοί κίνδυνοι γίνονται ευκολότερα προσπελάσιμοι χάρη στη βελτιωμένη τεχνολογία και τις νέες μεθόδους ανάλυσης. Οι χώρες με θετικές σχέσεις μεταξύ τους είναι πλέον σε θέση να ενοποιήσουν όλο και περισσότερο τις οικονομίες τους μέσω της αύξησης των επενδύσεων και του εμπορίου. Τα προϊόντα και οι υπηρεσίες που υπήρχαν στο παρελθόν καθίστανται ευκολότερα διαθέσιμα σε νέες αγορές με αποτέλεσμα τη βελτίωση των οικονομικών ευκαιριών για τους εργαζομένους στις οικονομίες αυτές και τη βελτίωση των εισοδημάτων των νοικοκυριών.

Για τους επενδυτές, οι ευκαιρίες αυτές παρουσιάζουν ένα ευρύ φάσμα επενδυτικών επιλογών και νέους τρόπους επίτευξης κέρδους. Προκειμένου όμως να επιτύχουν κέρδος καλούνται να περιορίσουν όσο το δυνατόν περισσότερο την πιθανότητα λάθους ή αλλιώς τον κίνδυνο της επένδυσης.

Ο κίνδυνος αυτός προκύπτει από την έλλειψη γνώσης ή πληροφόρησης, από την πολυπλοκότητα της διαδικασίας που καλούνται να ακολουθήσουν καθώς και από τη μεταβλητότητα των αποδόσεων του στοιχείου στο οποίο επενδύουν. Από άποψη οικονομικών παραγόντων η μεταβλητότητα των τιμών και των αποδόσεων των επενδύσεων ταξινομείται συνήθως σε επενδυτικό και επιχειρηματικό κίνδυνο και σχετίζεται άμεσα με τη γενικότερη κατάσταση και τις προοπτικές της εθνικής και της διεθνούς οικονομίας. Η σημασία της παρακολούθησης της εθνικής και της διεθνούς οικονομίας αυξάνει, όσο αναπτύσσεται και διεθνοποιείται η χρηματαγορά και η κεφαλαιαγορά. Ιδιαίτερα στην περίπτωση της κεφαλαιαγοράς, οι συνεχείς αλλαγές και η δυναμική της φύση έχουν οδηγήσει τους επενδυτές να αναζητούν εμπειριστατωμένες μεθόδους παρακολούθησης και ανάλυσης των τεκταινόμενων ούτως ώστε να είναι σε θέση να επιτύχουν μεγαλύτερη ακρίβεια στις προβλέψεις τους που θα τους οδηγήσουν στην κατάλληλη λήψη αποφάσεων.

Το πετρέλαιο συγκεντρώνει μεγάλο ενδιαφέρον στις αγορές είτε ως σταθερή κερδοσκοπική επιλογή είτε ως παράγοντας αντιστάθμισης κινδύνου. Παρόλα αυτά πληθώρα ερευνητών επισημαίνουν πως η αγορά πετρελαίου κάθε άλλο παρά ασφαλής και σταθερή μπορεί να χαρακτηριστεί τονίζοντας πως κάθε πρόβλεψη για την πορεία της τιμής του κρίνεται εξαιρετικά επισφαλής κυρίως λόγω της ισχυρής μεταβλητότητας που παρουσιάζει. Η συγκεκριμένη θέση κάνει ακόμα επιτακτικότερη την ανάγκη στους επενδυτές για αναζήτηση νέων μεθόδων

πρόβλεψης των διακυμάνσεων της τιμής του και μοντελοποίησης των παραγόντων που επιδρούν σε αυτήν.

Στην παρούσα εργασία θα μελετηθεί το αργό πετρέλαιο ως αντικείμενο επένδυσης, θα πραγματοποιηθεί εμπειρική ανάλυση των παραγόντων που συσχετίζονται με την τιμή του και θα προσδιοριστεί κατάλληλο προβλεπτικό μοντέλο. Σκοπός της εργασίας είναι να παράσχει χρήσιμες πληροφορίες στις ομάδες ενδιαφερομένων της αγοράς του πετρελαίου μελετώντας την πορεία της τιμής του καθώς και τις κινήσεις της παγκόσμιας αγοράς. Ειδικότερα θα αναλυθούν οι συσχετίσεις των τιμών του χρυσού, του ασημιού, του 10ετούς συμβολαίου μελλοντικής εκπλήρωσης (10 Year T-Note), του πληθωρισμού των Ηνωμένων Πολιτειών και του χρηματιστηριακού δείκτη NASDAQ 100 σε μηνιαία βάση από τον Ιανουάριο του 2000 μέχρι και το Δεκέμβριο του 2020. Στη συνέχεια, με βάση τις μηνιαίες καταγραφές της τιμής του πετρελαίου και ακολουθώντας τη μεθοδολογία Box – Jenkins θα επιλεγεί το καταλληλότερο μοντέλο ARIMA το οποίο θα αξιολογηθεί ως προς την προσαρμογή και την προβλεπτική του ικανότητα. Τέλος θα εφαρμοστεί στη χρονοσειρά των τιμών του πετρελαίου ένα νευρωνικό δίκτυο LSTM και θα συγκριθεί σε όρους προβλεπτικής ακρίβειας με το μοντέλο ARIMA.

Αναφορικά με τη δομή της εργασίας, στο δεύτερο κεφάλαιο θα γίνει η επισκόπηση της βιβλιογραφίας και των τρόπων προσέγγισης της διεθνούς επιστημονικής κοινότητας ως προς τη διαδικασία πρόβλεψης της τιμής του πετρελαίου. Στο τρίτο κεφάλαιο θα παρουσιαστεί η αγορά του πετρελαίου αποτυπώνοντας την πορεία του σε βάθος εικοσαετίας αλλά και την υφιστάμενη κατάσταση των σχετικών επενδύσεων σε παγκόσμια κλίμακα. Κλείνοντας στα δύο τελευταία κεφάλαια θα γίνει η στατιστική μελέτη των ιστορικών δεδομένων της τιμής του και των επιλεγθέντων εξωγενών δεικτών, θα δημιουργηθούν κατάλληλα προβλεπτικά μοντέλα και θα εξαχθούν συμπεράσματα.

2. Επισκόπηση Βιβλιογραφίας

Για την πρόβλεψη της τιμής του αργού πετρελαίου έχει χρησιμοποιηθεί πληθώρα προβλεπτικών μοντέλων τα οποία είναι σε θέση να διαχειριστούν τα φαινόμενα έλλειψης στασιμότητας και μεταβαλλόμενης διασποράς που συχνά εμφανίζονται στη μοντελοποίηση χρονοσειρών. Ιδιαίτερα στην ανάλυση και πρόβλεψη χρηματοοικονομικών δεικτών, νομισμάτων και λοιπών οικονομικών στοιχείων έρευνες τα τελευταία χρόνια έχουν δείξει πως τα μοντέλα τύπου ARIMA μπορούν να πετύχουν καλύτερες προβλέψεις σε χρονοσειρές οι οποίες είναι μη στάσιμες, δηλαδή παρουσιάζουν τάσεις, εποχικότητα και περιοδικότητα.

Στο μοντέλο ARIMA (αυτοπαλίνδρομο ολοκληρωμένο υπόδειγμα κινητού μέσου) η εκτιμώμενη αξία μιας μεταβλητής είναι ένας γραμμικός συνδυασμός παλαιότερων τιμών και σφαλμάτων του παρελθόντος. Αποτελεί εξέλιξη του μοντέλου ARMA και μπορεί να επεκταθεί και σε μη στάσιμες χρονοσειρές με τη χρήση της μεθόδου της διαφορίσης. Έχει μελετηθεί εκτεταμένα από τους Box και Jenkins (Box and Jenkins) των οποίων η μεθοδολογία περιγράφει την κατασκευή μοντέλων χρονοσειρών μέσω τριών σταδίων: της ταυτοποίησης (identification), της εκτίμησης (estimation) και του διαγνωστικού ελέγχου (diagnostic checking).

Ως εναλλακτική των προσεγγίσεων πρόβλεψης της τιμής του αργού πετρελαίου μέσω γραμμικών μοντέλων συναντάμε όλο και συχνότερα τα τελευταία έτη τη χρήση νευρωνικών δικτύων (ANN). Το κυριότερο πλεονέκτημα των νευρωνικών δικτύων στην ανάλυση χρονοσειρών είναι το ότι αποτελούν εν γένει μία μέθοδο η οποία καθοδηγείται από τα δεδομένα και προσαρμόζεται σε αυτά χωρίς την αυστηρή δέσμευση αρχικών υποθέσεων. Ως μη γραμμικά δίκτυα, δεν πραγματοποιούν προβλέψεις έχοντας ως βασική υπόθεση ότι η υπό μελέτη χρονοσειρά παράγεται από γραμμικές συναρτήσεις αλλά αντίθετα έχουν τη δυνατότητα χάρη στην ευελιξία τους να εντοπίζουν τη σχέση μεταξύ των δεδομένων καθώς και οποιαδήποτε συνάρτηση τα περιγράφει. Στην απόδοση των νευρωνικών δικτύων επιδρούν τρεις πολύ σημαντικοί παράγοντες: ο αριθμός των κόμβων εισόδου, ο αριθμός των κρυφών νευρώνων και το μέγεθος του σετ δεδομένων για την εκπαίδευση του δικτύου.

Οι Lubna A Gabralla και Ajith Abraham (Lubna A Gabralla) πραγματοποίησαν μία επισκόπηση των μεθόδων που εφαρμόζονται στην πρόβλεψη της τιμής του αργού πετρελαίου και των αποτελεσμάτων τους, δίνοντας έμφαση στα γενικά βήματα που ακολουθούνται και ερευνώντας παράλληλα τις οικονομικές μεταβλητές που παρουσιάζουν ισχυρή συσχέτιση με την τιμή του αργού πετρελαίου και οδηγούν τις διακυμάνσεις της. Πιο συγκεκριμένα, παρουσίασαν τις ισχυρές συσχετίσεις της τιμής του αργού πετρελαίου με τις ισοτιμίες του αμερικανικού δολαρίου, τα διεθνή επιτόκια και τους χρηματιστηριακούς δείκτες, τον δείκτη ανεργίας και τις τιμές άλλων πολύτιμων μετάλλων όπως ο χρυσός και το ασήμι. Στη συνέχεια παρουσίασαν τους κύριους παράγοντες επιρροής όπως η ζήτηση και η προσφορά του πετρελαίου, οι διεθνείς πολιτικές και οι οδηγίες του Οργανισμού Εξαγωγών Πετρελαιοπαραγωγών Χωρών (OPEC) καθώς και τους παράγοντες επιρροής του φυσικού περιβάλλοντος. Τέλος, με είσοδο της τιμής των χρονοσειρών των παραπάνω οικονομικών μεταβλητών, παρουσίασαν συνδυασμούς προβλεπτικών μοντέλων μηχανικής μάθησης όπως τα Τεχνητά Νευρωνικά Δίκτυα (ANN), τα μοντέλα ARIMA και τα Support Vector Machines (SVMs).

Οι A. C. Akpanta et al. (A.C. Akpanta) χρησιμοποίησαν τη μεθοδολογία Box-Jenkins για να προβλέψουν τις μηνιαίες τιμές του αργού πετρελαίου της Νιγηρίας λαμβάνοντας μία σχετική

χρονοσειρά τιμών από το 1982 έως το 2013. Στη μελέτη τους έλαβαν υπόψη τις τιμές του ελέγχου Augmented Dickey -Fuller όπου εντοπίστηκε η μη στασιμότητα της χρονοσειράς και στη συνέχεια μετέτρεψαν τη χρονοσειρά σε στάσιμη με τη διαδικασία της διαφορίσης. Τα διαγράμματα αυτοσυσχέτισης και μερικής αυτοσυσχέτισης ACF και PACF αντίστοιχα τους υπέδειξαν ένα μοντέλο ARIMA(6,1,7) του οποίου οι περισσότεροι συντελεστές όμως δεν εμφάνισαν στατιστική σημαντικότητα. Έτσι κατέληξαν σε ένα μοντέλο ARIMA(2,1,2) το οποίο εμφάνισε την καλύτερη προσαρμογή στα δεδομένα τους καθώς είχε τις χαμηλότερες τιμές στα στατιστικά AIC και BIC.

Οι J.Jaya Selvi et al. (J.Jaya Selvi) μελέτησαν τις τεχνικές ARIMA στην πραγματοποίηση προβλέψεων για την τιμή του αργού πετρελαίου. Χρησιμοποιώντας τη μεθοδολογία Box-Jenkins προσάρμοσαν ένα μοντέλο ARIMA(2,1,1) στις ετήσιες παρατηρήσεις της τιμής του αργού πετρελαίου από το 1946 έως το 2016 πραγματοποιώντας προβλέψεις για την περίοδο 2017-2021.

Οι Jessin Shah P A et al. (Jessin Shah P A) χρησιμοποίησαν ένα σύνολο δεδομένων με μηνιαίες τιμές αργού πετρελαίου σύμφωνα με το ορόσημο West Texas Intermediate (WTI) από τον Ιούλιο του 1987 έως και το Μάρτιο του 2020. Λόγω της μεγάλης διακύμανσης των τιμών της χρονοσειράς του πετρελαίου μετέτρεψαν τις αρχικές τιμές λαμβάνοντας τους λογαρίθμους αυτών και χώρισαν τα δεδομένα τους σε σύνολα εκπαίδευσης και ελέγχου με αναλογία 80/20. Στη συνέχεια ελέγχοντας τις τιμές του ελέγχου Augmented Dickey -Fuller διαπίστωσαν τη μη στασιμότητα της χρονοσειράς την οποία μετέτρεψαν σε στάσιμη εφαρμόζοντας πρώτης τάξης διαφορές στους λογαρίθμους των τιμών του αργού πετρελαίου. Τέλος, κατέληξαν σε ένα μοντέλο ARIMA(0,1,4) το οποίο εμφάνισε την καλύτερη προσαρμογή στα δεδομένα τους.

Οι Nidhi Moitra et al. (Nidhi Moitra) ανέπτυξαν ένα Νευρωνικό Δίκτυο LSTM για την πρόβλεψη της τιμής του αργού πετρελαίου δοκιμάζοντας διαφορετικές αρχιτεκτονικές και ρυθμίσεις παραμέτρων στο μοντέλο τους. Πιο συγκεκριμένα, έχοντας εκπαιδεύσει διάφορα LSTM μοντέλα κατέληξαν σε ένα LSTM Νευρωνικό Δίκτυο με συνολικά τέσσερα layers εκ των οποίων ένα εισαγωγής, δύο hidden layers και ένα εξαγωγής, 10 lookbacks και 50 εποχές. Το παραπάνω μοντέλο εμφάνισε την καλύτερη επίδοση σε όρους ακρίβειας με 283.34 RMSE στο train υποσύνολο και 532.13 RMSE στο test υποσύνολο.

Οι Anh H. Vo et al. (Anh H. Vo1) εκπαιδύσαν ένα Bi – LSTM (Bidirectional LSTM) για την πρόβλεψη της τιμής του πετρελαίου. Η δομή του μοντέλου τους περιελάμβανε τρία Bi – LSTM στρώματα νευρώνων (Layers) τα οποία βοήθησαν στο να συνυπολογισθούν χρήσιμες πληροφορίες τόσο μέσω εμπρόσθιας κατεύθυνσης όσο και μέσω οπίσθιας. Το τελευταίο πλήρως συνδεδεμένο στρώμα νευρώνων χρησιμοποιήθηκε για να πραγματοποιήσει τις τελικές προβλέψεις στην τιμή του πετρελαίου χρησιμοποιώντας χρήσιμες πληροφορίες από τα προηγούμενα στρώματα. Από τη μελέτη τους και μέσω συγκρίσεων στους βασικούς δείκτες ακρίβειας MSE, RMSE, MAE και MAPE, κατέληξαν πως τα sequence-based Bi-LSTM και LSTM μοντέλα είναι περισσότερο αποτελεσματικά από τα σύνθετα CNN-LSTM.

Οι Yingrui Zhou et al. (Yingrui Zhou) εφάρμοσαν τις τεχνικές CEEMDAN (Complete Ensemble Empirical Mode Decomposition with Adaptive Noise) και XGBOOST (Extreme Gradient Boosting) σε ένα συνδυαστικό μοντέλο CEEMDAN – XGBOOST για την πρόβλεψη της τιμής του αργού πετρελαίου. Αρχικά εφάρμοσαν την τεχνική CEEMDAN για να μετατρέψουν τις μη γραμμικές και μη στάσιμες ακολουθίες της χρονοσειράς σε διάφορες συναρτήσεις IMF (intrinsic mode functions) και ένα κατάλοιπο. Στη συνέχεια, μέσω της τεχνικής XGBOOST πραγματοποίησαν προβλέψεις σε κάθε συνάρτηση IMF και σε κάθε κατάλοιπο. Τα

αποτελέσματά τους έδειξαν πως το μοντέλο CEEMDAN – XGBOOST που εκπαίδευσαν ξεπερνούσε σε ακρίβεια μοντέλα ARIMA, SVR και FNN συγκρινόμενο ως προς τους δείκτες RMSE και MAE.

Οι Li Shu-rong et al. (Li Shu-rong) στη μελέτη που εκπόνησαν για τις προβλέψεις της τιμής του αργού πετρελαίου, προτείνουν ένα μοντέλο ϵ -SVR (Support Vector Regression) με συντελεστή δυναμικής διόρθωσης στα σφάλματα της πρόβλεψης. Τα αποτελέσματά τους έδειξαν πως το μοντέλο ϵ -SVR μπορεί να προβλέψει με ακρίβεια της διακυμάνσεις της τιμής του αργού πετρελαίου καθώς και να επεκτείνει τη χρησιμότητά του και σε άλλες χρονοσειρές χρηματοοικονομικών στοιχείων.

3. Η Αγορά του Πετρελαίου

3.1. Εύρεση, εξόρυξη και επεξεργασία

Το αργό (ορυκτό - ακατέργαστο) πετρέλαιο είναι ένα μείγμα υδρογονανθράκων, αλκαλίων και άλλων οργανικών ενώσεων που βρίσκεται μέσα σε πετρώματα των στρωμάτων του φλοιού της Γης. Αποτελεί την κυριότερη πηγή ενέργειας του πλανήτη αλλά και βάση για την κατασκευή πολλών προϊόντων της καθημερινότητάς μας.

Το αργό πετρέλαιο μπορεί να ποικίλει σε εμφάνιση, σύνθεση και καθαρότητα. Οι κύριες κατηγορίες σύνθεσης αργού πετρελαίου είναι τα παραφινικά πετρέλαια, τα ασφαλτικά πετρέλαια και τα ασφαλτοπαραφινικά πετρέλαια. Η δεύτερη κατηγορία (ασφαλτικά πετρέλαια) είναι και αυτή που δίνει τα βαρέα κλάσματα μαζούτ και ορυκτέλαια. Αναφορικά με τον εντοπισμό πετρελαϊκών κοιτασμάτων χρησιμοποιούνται μέθοδοι όπως η σεισμική, η ηλεκτρική, η ηλεκτρομαγνητική, η σταθμική ή βαρυτομετρική και η ραδιενεργή μέθοδος. Στη σεισμική μέθοδο, προκαλείται μέσω εκρηκτικών μηχανισμών ένας τεχνητός σεισμός και από τα σεισμικά κύματα που παράγονται δύναται να εντοπιστούν μέσω των φαινομένων της διάθλασης ή ανάκλασης πιθανά υπόγεια κοιτάσματα. Το μειονέκτημα της μεθόδου αυτής είναι πως αντί για πετρελαϊκά κοιτάσματα δύναται να υποδείξει μεγάλες ποσότητες υπόγειων υδάτων. Η ηλεκτρική μέθοδος βασίζεται στην ιδιότητα του φλοιού της Γης να παρουσιάζει διαφορετικά επίπεδα αντίστασης στη διέλευση ηλεκτρικού ρεύματος και δεδομένου ότι το πετρέλαιο δεν είναι καλός αγωγός του ηλεκτρισμού, μεγάλη αντίσταση στη διέλευση μπορεί να υποδεικνύει την ύπαρξη κοιτάσματος. Η ηλεκτρομαγνητική μέθοδος μπορεί μέσω ευαίσθητων οργάνων που μετρούν την ένταση μαγνητικών πεδίων από τόπο σε τόπο να υποδείξει πιθανά κοιτάσματα πετρελαίου και η σταθμική ή βαρυτομετρική μέθοδος βασίζεται στην ένταση του πεδίου βαρύτητας ανά σημείο της Γης. (Wikipedia, Πετρέλαιο)

Η άντληση του πετρελαίου γίνεται από ειδικές εγκαταστάσεις που βρίσκονται πάνω από το σημείο του κοιτάσματος. Οι εγκαταστάσεις αυτές έχουν το σχήμα πύργου και μέσω της γεώτρησης αρτεσιανού φρέατος που πραγματοποιούν, το πετρέλαιο αναβλύζει υπό τη μορφή πίδακα λόγω των υφιστάμενων πιέσεων. Στις μέρες μας, οι συνηθέστερες μορφές άντλησης είναι μέσω προκαλούμενης πίεσης νερού μέσα από το οποίο επιπλέει το πετρέλαιο, με εξακόντιση νιτρογλυκερίνης και με εισαγωγή υδροχλωρικού οξέος υπό πίεση.

Το πετρέλαιο αναβλύζει από τις πετρελαιοπηγές αναμεμιγμένο με αέρια, νερό και ποσότητες άμμου τα οποία πρέπει να διαχωριστούν από αυτό για να προκύψει το καθαρό ακατέργαστο πετρέλαιο. Ο διαχωρισμός των αερίων γίνεται μέσω ενός ειδικού διαχωριστή και χρησιμοποιούνται είτε προς επανεισαγωγή εντός των πηγών για καινούριες εξορύξεις είτε οδηγούνται προς το εμπόριο ως φυσικά αέρια, ενώ το νερό και η άμμος αποχωρίζονται από το πετρέλαιο με παραμονή τους σε δεξαμενές (καθίζηση). Το καθαρό ακατέργαστο πετρέλαιο που δημιουργείται οδηγείται πλέον στα διυλιστήρια όπου γίνεται η κλασματική απόσταξη και αποθείωσή του και εν συνεχεία παράγονται μία σειρά προϊόντων με συγκεκριμένα ποιοτικά χαρακτηριστικά. (Δ. Καρώνης)

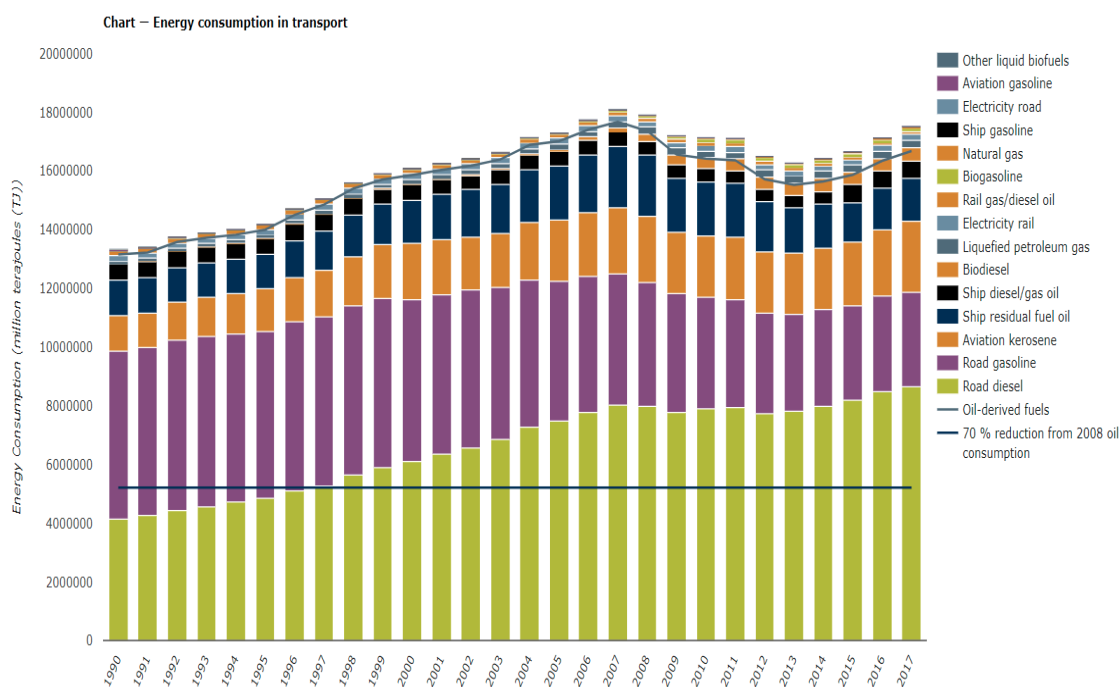
3.2. Οι χρήσεις του πετρελαίου και η ζήτηση ανά κατηγορία

3.2.1. Μεταφορές

Το πετρέλαιο και τα παράγωγα προϊόντα του, όπως έχει αναφερθεί και προγενέστερα, αποτελεί την κυριότερη πηγή ενέργειας του πλανήτη. Στον τομέα των μεταφορών, ο οποίος αποτελεί το 66,1% των συνολικών χρήσεων του πετρελαίου σύμφωνα με στοιχεία της Eurostat (Eurostat), ο πιο συνηθισμένος τύπος είναι το καύσιμο ντίζελ και ιδιαίτερα η μορφή πετροντίζελ, μία κλασματική απόσταξη του αργού πετρελαίου μεταξύ 200 και 350 βαθμών κελσίου σε ατμοσφαιρική πίεση. Άλλες διαδεδομένες μορφές είναι το συνθετικό ντίζελ και το βιοντίζελ οι οποίες βρίσκουν εφαρμογή σε τομείς όπως οι μεταφορές, η παραγωγή και ο μετασχηματισμός, η θέρμανση κλπ.

Σύμφωνα με τα στοιχεία του Ευρωπαϊκού Οργανισμού Περιβάλλοντος (European Environment Agency), η ετήσια κατανάλωση ενέργειας στις μεταφορές αυξήθηκε κατά 38% μεταξύ των ετών 1990 και 2007 ενώ εξαιτίας της οικονομικής κρίσης παρατηρήθηκε μείωση της τάξεως του 3% κατά την περίοδο 2007 – 2017. Ο τομέας των θαλάσσιων μεταφορών παρουσίασε τη μεγαλύτερη πτώση, 10% σε μία χρονιά (2008 – 2009) και συνολικά 19% την περίοδο 2007 – 2017 ενώ οι αερομεταφορές παρουσίασαν αύξηση 7% κατά την ίδια περίοδο. Οι χερσαίες μεταφορές, οι οποίες αποτελούν το μεγαλύτερο μέρος της ζήτησης (73% το 2017), παρουσιάζουν κατά 34% υψηλότερη ζήτηση ενέργειας το 2017 από το 1990 (Agency, Final energy consumption in Europe by mode of transport). Στο παρακάτω γράφημα παρουσιάζεται η ετήσια κατανάλωση ενέργειας ανά τύπο καυσίμου.

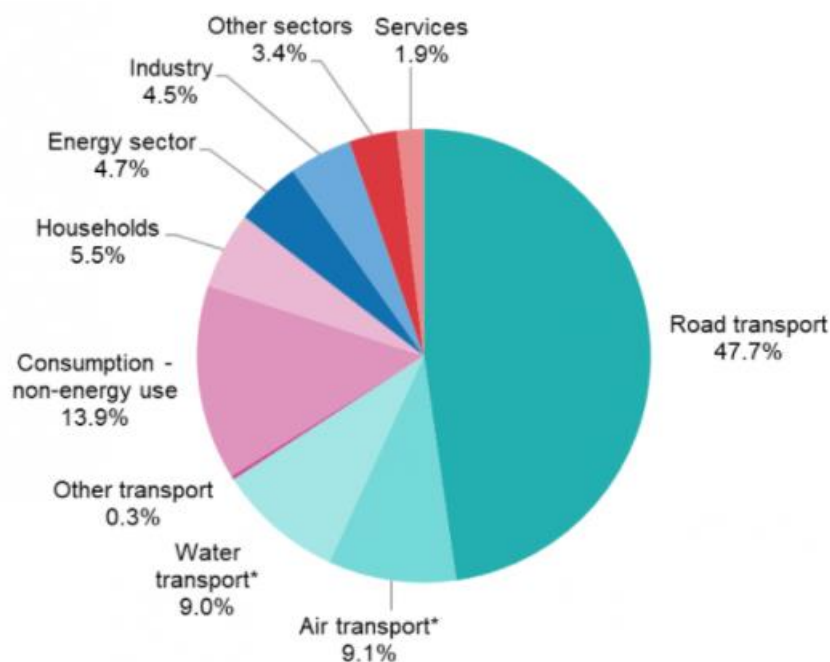
Γράφημα 1 Κατανάλωση Ενέργειας ανά τύπο καυσίμου (EU)



3.2.2. Βιομηχανικά Προϊόντα Μη Ενεργειακών Σκοπών και λοιπές χρήσεις

Τα πετρελαϊκά προϊόντα που χρησιμοποιούνται για μη ενεργειακούς σκοπούς είναι η δεύτερη μεγαλύτερη κατηγορία χρήσης πετρελαίου, αντιπροσωπεύοντας το 13,9 % της κατανάλωσης για το 2019 σύμφωνα με στοιχεία της Eurostat. Η κατηγορία αυτή περιλαμβάνει για παράδειγμα πίσσα για οδικές επιφάνειες, λιπαντικά για τη μείωση της τριβής καθώς και τη χρήση προϊόντων πετρελαίου στη βιομηχανία πετροχημικών για τις χημικές τους ιδιότητες και όχι για το ενεργειακό τους περιεχόμενο. Σε αυτή την περίπτωση, τα πρωτογενή προϊόντα πετρελαίου μετατρέπονται σε άλλα προϊόντα όπως πλαστικά, απορρυπαντικά, ελαστικά κλπ. Όσον αφορά στις υπόλοιπες κατηγορίες, τα νοικοκυριά αποτελούν το 5,5 %, ο ενεργειακός τομέας το 4,7 % και οι βιομηχανίες το 4,5 % των χρήσεων του πετρελαίου. Ο τομέας των υπηρεσιών (εμπορικών και δημόσιων) χρησιμοποίησαν περίπου 1,9 % του συνόλου ενώ οι υπόλοιποι τομείς, συμπεριλαμβανομένης της γεωργίας, της δασοκομίας και της αλιείας, χρησιμοποίησαν περίπου το 3,4 %.

Γράφημα 2 Κατανάλωση Ενέργειας ανά κατηγορία (EU)



3.3. Παράγοντες επιρροής της τιμής του πετρελαίου και συσχετιζόμενα χρηματοοικονομικά προϊόντα

Το αργό πετρέλαιο, όντας η κύρια πηγή ενέργειας του πλανήτη, επηρεάζει και επηρεάζεται από πολλούς μακροοικονομικούς δείκτες και χρηματοοικονομικά προϊόντα. Η συνθήκη αυτή οφείλεται στο γεγονός ότι το πετρέλαιο είναι ένα από τα απαραίτητα συστατικά κίνησης της

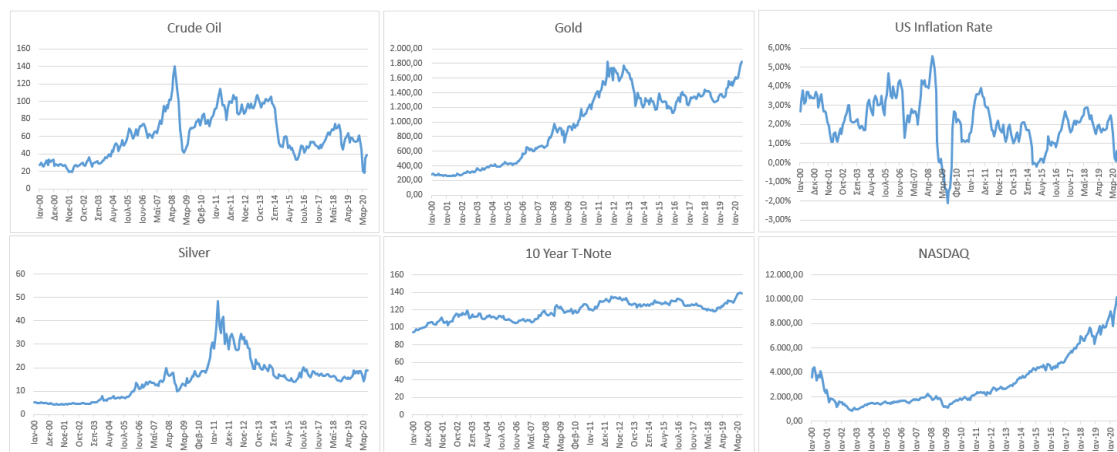
παγκόσμιας οικονομίας μιας και αποτελεί τον ακρογωνιαίο λίθο της παραγωγής, της μεταποίησης και των μεταφορών. Επομένως μία ενδεχόμενη κρίση στην τιμή του θα μπορούσε να επιφέρει δραματικές αλλαγές στα τεκταινόμενα της παγκόσμιας οικονομίας, ακόμα και παγκόσμια οικονομική κρίση. Αντιστρόφως, αλλαγές σε παγκόσμιους οικονομικούς δείκτες και προϊόντα δύναται να μεταβάλλουν την τιμή του αργού πετρελαίου. Εξαιρετικά σημαντικές ακόμα αιτίες μεταβολής της τιμής του πετρελαίου είναι το εκάστοτε κοινωνικό-πολιτικό πλαίσιο και σύστημα αποφάσεων, τα κυτασμικά αποθέματα, η τεχνολογία καθώς και η κλιματική αλλαγή.

Στην παρούσα εργασία θα εξετάσουμε τους παράγοντες που επηρεάζουν την τιμή του πετρελαίου από την πλευρά των επενδύσεων και δη στη χρηματιστηριακή αγορά, θα δούμε άρα συσχετιζόμενους χρηματοοικονομικούς δείκτες και προϊόντα όπως Συμβόλαια Μελλοντικής Εκπλήρωσης και Ομόλογα.

Η τιμή του αργού πετρελαίου παρουσιάζει παραδοσιακά μεγάλη και σταθερή συσχέτιση με την τιμή του χρυσού, δηλαδή όταν αυξάνεται η τιμή του πετρελαίου αυξάνεται και του χρυσού και αντιστρόφως. Σύμφωνα με την εταιρία χρηματοοικονομικών ερευνών Market Realist (Shawn), η σχέση αυτή παραμένει ενεργή και άμεση το 60% του χρόνου. Ωστόσο στις μέρες μας παρατηρείται αποκοπή της σχέσης πετρελαίου – χρυσού, κάτι το οποίο φαντάζει λογικό ως αποτέλεσμα της παγκόσμιας πανδημίας COVID – 19 μιας και η εν μέρει παύση της οικονομικής δραστηριότητας (παραγωγής, μεταφορών) έφερε πτώση στην τιμή του πετρελαίου, ενώ παράλληλα η γενικότερη αβεβαιότητα έφερε στροφή των επενδυτών στο χρυσό ο οποίος παραδοσιακά αποτελεί επενδυτικό καταφύγιο σε περιόδους κρίσεων. (Ng)

Ένα ακόμα μέταλλο που εμφανίζει ισχυρή θετική συσχέτιση με την τιμή του αργού πετρελαίου είναι το ασήμι. Οι Imtiaz Arif et al. (Imtiaz Arif) έδειξαν πως υπάρχει στατιστικά σημαντική θετική συσχέτιση μεταξύ των τιμών του πετρελαίου και του ασημιού, αυξάνονται και μειώνονται παράλληλα δηλαδή. Στο Γράφημα 3 παρουσιάζονται οι χρονοσειρές των εμπορευμάτων που συσχετίζονται ισχυρά με το πετρέλαιο, του χρυσού και του ασημιού, αλλά και άλλων πιθανώς συσχετιζόμενων παραγόντων που θα μελετήσουμε εκτενέστερα στο επόμενο κεφάλαιο όπως το Δεκαετές Αμερικανικό Ομόλογο US 10 Year T-Note, ο δείκτης NASDAQ και ο δείκτης Αμερικανικού Πληθωρισμού.

Γράφημα 3 Χρονοσειρές Συσχετιζόμενων Παραγόντων



4. Η τιμή του αργού πετρελαίου την περίοδο Ιανουάριος 2000 – Δεκέμβριος 2020: Εμπειρική διερεύνηση των συσχετιζόμενων δεικτών και δημιουργία προβλεπτικού μοντέλου

4.1. Περιγραφικά στοιχεία των υπό μελέτη δεικτών και Ανάλυση των Συσχετίσεων μεταξύ τους

Στον παρακάτω πίνακα παρουσιάζονται τα περιγραφικά στοιχεία των μηνιαίων τιμών των υπό μελέτη δεικτών για την περίοδο Ιανουάριος 2000 – Δεκέμβριος 2020. Τα δεδομένα της τιμής του αργού πετρελαίου, του δεκαετούς Αμερικανικού Ομολόγου US 10 Year T-Note, του χρυσού, του δείκτη NASDAQ 100 και της τιμής του ασημιού συλλέχθηκαν από την ιστοσελίδα Investing.com ενώ η τιμή του Αμερικανικού Πληθωρισμού από το US Bureau of Labor Statistics. Όπως παρατηρούμε η μέση μηνιαία τιμή του πετρελαίου είναι ίση με 61 δολάρια (μέση απόκλιση τετραγώνου 26,2) και με εύρος τιμής από 18,8 έως 140. Αντίστοιχα αναφορικά με τις τιμές των συσχετιζόμενων μεταβλητών, η μέση μηνιαία τιμή του χρυσού είναι ίση με 996 (μέση απόκλιση τετραγώνου 501,5 και εύρος τιμών από 257,9 έως 2017,1), η μέση μηνιαία τιμή του δείκτη NASDAQ είναι ίση με 3392,2 (μέση απόκλιση τετραγώνου 2475,5 και εύρος τιμών από 832,5 έως 12.888,3), η μέση μηνιαία τιμή του δεκαετούς αμερικανικού ομολόγου είναι ίση με 119,4 (μέση απόκλιση τετραγώνου 10,3 και εύρος τιμών από 94,8 έως 140,1) και η μέση μηνιαία τιμή του ποσοστού πληθωρισμού είναι ίση με 0,02 (μέση απόκλιση τετραγώνου 0,01 και εύρος τιμών από -0,02 έως 0,06). Τέλος η μέση μηνιαία τιμή του ασημιού είναι ίση με 15,3 με μέση απόκλιση τετραγώνου 8,6 και εύρος τιμών από 4,1 έως 48,6. Από τις τιμές των τυπικών αποκλίσεων παρατηρούμε πως στην υπό μελέτη χρονική περίοδο τη μεγαλύτερη μεταβλητότητα τιμής εμφάνισαν ο χρηματιστηριακός δείκτης NASDAQ (συντελεστής μεταβλητότητας ίσος με 72,97%) και η τιμή του ποσοστού πληθωρισμού (συντελεστής μεταβλητότητας ίσος με 58,36%) ενώ φαίνεται πως η τιμή του δεκαετούς αμερικανικού ομολόγου μεταβλήθηκε συγκριτικά λιγότερο (συντελεστής μεταβλητότητας ίσος με 8,6%).

Πίνακας 1 Περιγραφικά στοιχεία των υπό μελέτη δεικτών για την περίοδο Ιανουάριος 2000 – Δεκέμβριος 2020

	Μέσος	Μέση απόκλιση τετραγώνου	Ελάχιστο	Μέγιστο
Crude Oil	61.0	26.2	18.8	140.0
Gold	995.9	501.5	257.9	2017.1
Silver	15.3	8.6	4.1	48.6
10 Year T-Note	119.4	10.3	94.8	140.1
US Inflation Rate	0.02	0.01	-0.02	0.06
NASDAQ	3392.2	2475.2	832.5	12888.3

Στην παρούσα ενότητα θα μελετήσουμε επιπλέον και τη σχέση μεταξύ των επιλεχθέντων μεταβλητών (παράγοντες επιρροής) και της υπό εξέταση μεταβλητής (τιμή του πετρελαίου) ερμηνεύοντάς τες με το δείκτη r (συντελεστής συσχέτισης Pearson) και το επίπεδο σημαντικότητας p της συσχέτισης μεταξύ των μεταβλητών.

Πίνακας 2 Βαθμός συσχέτισης και επίπεδο σημαντικότητας της συσχέτισης των υπό μελέτη μεταβλητών

	Crude Oil	Gold	Silver	10 Year T-Note	US Infl. Rate
Συντελεστής συσχέτισης r μεταξύ των μεταβλητών					
Gold	0.50				
Silver	0.70	0.87			
10 Year T-Note	0.31	0.90	0.71		
US Infl. Rate	0.19	-0.33	-0.12	-0.48	
NASDAQ 100	-0.10	0.64	0.28	0.57	-0.22
Επίπεδο σημαντικότητας p της συσχέτισης μεταξύ των μεταβλητών					
Gold	0.000				
Silver	0.000	0.000			
10 Year T-Note	0.000	0.000	0.000		
US Infl. Rate	0.003	0.000	0.059	0.000	
NASDAQ 100	0.107	0.000	0.000	0.000	0.000

Όπως παρατηρούμε στον Πίνακα 2 το αργό πετρέλαιο εμφανίζει στατιστικά σημαντική συσχέτιση (επίπεδο σημαντικότητας 5%) με όλες τις μεταβλητές εκτός από την τιμή του χρηματιστηριακού δείκτη NASDAQ 100. Αναλυτικότερα τις μεγαλύτερες θετικές συσχετίσεις τις εμφανίζει με το ασημί (70%) και το χρυσό (50%). Ακολουθούν οι θετικές συσχετίσεις με το δεκαετές αμερικανικό ομόλογο (31%) και την τιμή του αμερικανικού πληθωρισμού (19%) ενώ παρατηρείται αρνητική συσχέτιση μεταξύ του αργού πετρελαίου και του χρηματιστηριακού δείκτη NASDAQ 100 (-10%).

Μεταξύ των συσχετιζόμενων μεταβλητών στατιστικά σημαντικές ισχυρές συσχετίσεις είναι αυτές μεταξύ του χρυσού και του δεκαετούς αμερικανικού ομολόγου (90%) καθώς και μεταξύ του χρυσού και του ασημιού (87%). Δεν παρατηρείται στατιστικά σημαντική συσχέτιση σε επίπεδο σημαντικότητας 5% μεταξύ της τιμής του δείκτη NASDAQ 100 και της τιμής του αργού πετρελαίου καθώς και του αμερικανικού πληθωρισμού με την τιμή του ασημιού.

4.2. Ανάλυση Χρονοσειρών – Επισκόπηση Θεωρίας

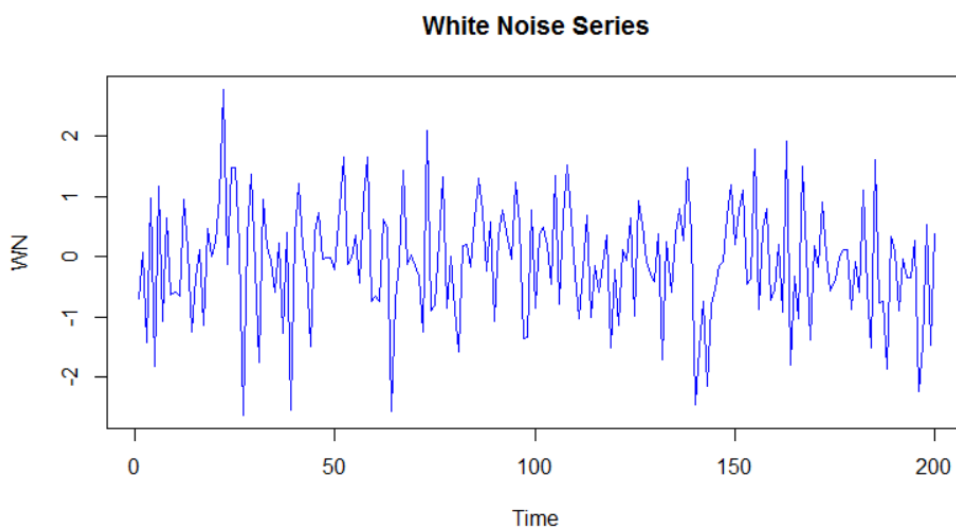
4.2.1 Χαρακτηριστικά Χρονοσειρών

Χρονοσειρά ή χρονολογική σειρά είναι ένα σύνολο παρατηρήσεων οι οποίες λαμβάνονται σε ίσα χρονικά διαστήματα ή περιόδους (μέρα, εβδομάδα, μήνας, έτος κ.α.) και αποτυπώνουν την εξέλιξη της τιμής μίας μεταβλητής στο χρόνο. Οι τιμές της μεταβλητής αποτυπώνονται ως παρατηρήσεις παρελθοντικών περιόδων και αναλύοντάς τες μπορούμε να προβούμε σε προβλέψεις για το μέλλον εξάγοντας χρήσιμα συμπεράσματα για τη συμπεριφορά της και διευκολύνοντας τη διαδικασία λήψης αποφάσεων.

Η ανάλυση χρονοσειρών μπορεί να εφαρμοστεί πρακτικά οπουδήποτε λαμβάνουν χώρα επαναλαμβανόμενα γεγονότα σε χρονικά διαστήματα τα οποία παίρνουν τη μορφή ποσοτικών μεταβλητών. Πολλά επιστημονικά πεδία τα οποία εξετάζουν ποσοτικές μεταβολές, όπως τα Οικονομικά, η Ιατρική, οι Πολιτικές Επιστήμες και η Μηχανολογία εφαρμόζουν ανάλυση χρονοσειρών για να ερμηνεύσουν φαινόμενα και να αποκτήσουν περαιτέρω γνώσεις στους τομείς που δραστηριοποιούνται.

Η διαδικασία της ανάλυσης μίας χρονοσειράς περιλαμβάνει γραφικές απεικονίσεις και στατιστικά μεγέθη – μοντέλα ανάλυσης. Για την επίτευξη της εξαγωγής συμπερασμάτων παρατηρώντας μία χρονοσειρά, θα πρέπει καταρχήν να διασφαλιστεί πως η χρονοσειρά που αναλύεται δεν αποτελείται από εντελώς τυχαίες μεταβολές τιμών (λευκός θόρυβος). Όταν δεν παρατηρείται λευκός θόρυβος, δηλαδή η μεταβολές των τιμών της χρονοσειράς δεν είναι πλήρως ανεξάρτητες, τότε μπορούμε να εξάγουμε χρήσιμες πληροφορίες και να εντοπίσουμε πιθανά μοτίβα στα δεδομένα που αναλύουμε.

Γράφημα 4 Χρονοσειρά Λευκού Θορύβου



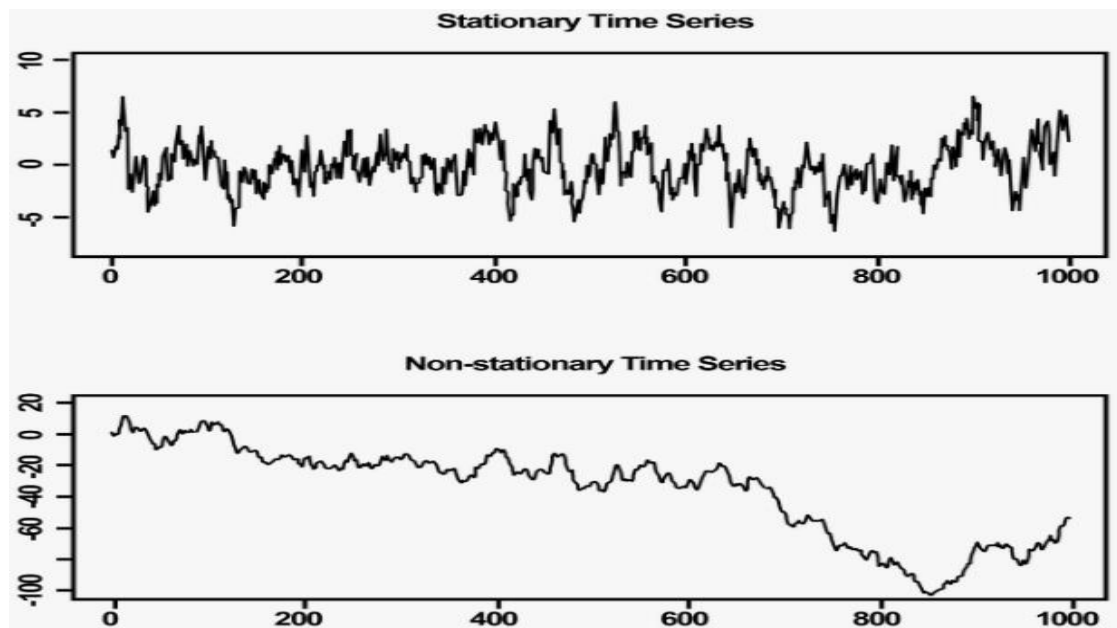
Σε χρονοσειρές όπου οι παρατηρήσεις παρουσιάζουν κάποια μορφή εξάρτησης, θα πρέπει να εξετάζονται χαρακτηριστικά όπως η στασιμότητα, η αιτιοκρατία και στοχαστικότητα καθώς και η γραμμικότητα και η μη γραμμικότητα.

Στασιμότητα

Μία στάσιμη χρονοσειρά περιλαμβάνει τιμές των οποίων οι διακυμάνσεις δεν διαφοροποιούνται με το χρόνο και η μέση τιμή της παραμένει σταθερή. Η στασιμότητα αποτελεί πλέον κοινή παραδοχή σε πολλές πρακτικές και μεθόδους στην ανάλυση χρονοσειρών, όπως η εκτίμηση της τάσης μίας χρονοσειράς και οι προβλέψεις, καθώς απλοποιεί τη μελέτη τους.

Σε μία μη στάσιμη χρονοσειρά παρατηρείται τάση, περιοδικότητα και εποχικότητα. Με τον όρο τάση εννοούμε τη μακροχρόνια γενική κίνηση μίας χρονοσειράς, αυξητική ή μειωτική. Για παράδειγμα μία στάσιμη χρονοσειρά δεν μπορεί να παρουσιάζει τάση καθώς η μακροχρόνια μεταβολή των τιμών της αποτελεί διαγραμματικά ευθεία γραμμή παράλληλη με τον οριζόντιο άξονα του χρόνου.

Γράφημα 5 Μορφή Στάσιμης και μη Στάσιμης Χρονοσειράς



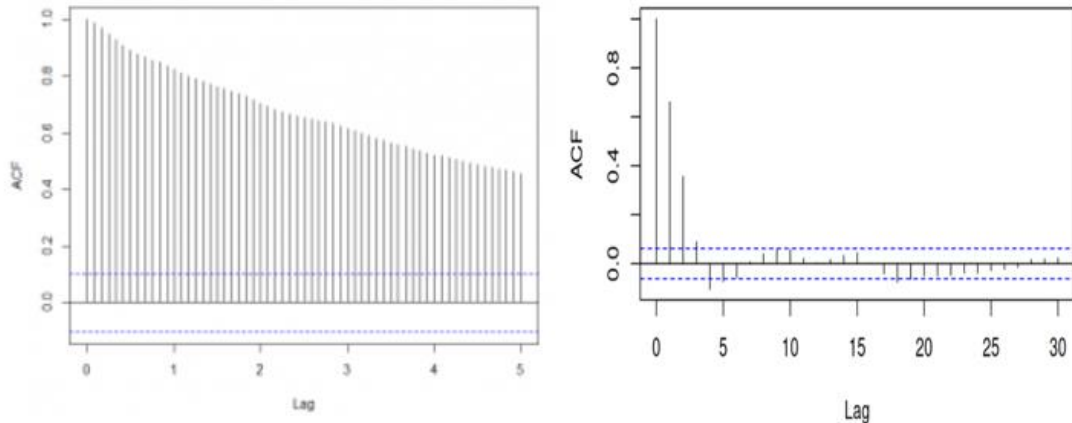
Περιοδικότητα ή κυκλικότητα είναι επαναλαμβανόμενες κυκλικές κυμάνσεις γύρω από την τάση της χρονοσειράς με διάρκεια μεγαλύτερη του ενός έτους. Μία ολοκληρωμένη κυκλική κίνηση μπορεί να είναι ανοδική ή καθοδική, ενδεχομένως με διαφορετική διάρκεια κάθε φορά κατά μήκος της χρονοσειράς. Περιοδικό φαινόμενο μπορεί να αποτελέσει για παράδειγμα ένας οικονομικός κύκλος.

Με τον όρο εποχικότητα αναφέρεται μία κυκλική κίνηση με περίοδο ίση με ένα έτος, μέσα στο οποίο παρουσιάζει ανοδικές και καθοδικές κινήσεις με σταθερό σχετικά ρυθμό. Παράδειγμα

χρονοσειράς με εποχικότητα αποτελεί η ζήτηση παγωτού της οποίας η τιμή αυξομειώνεται μέσα σε ένα ημερολογιακό έτος με αρκετά προβλέψιμο ρυθμό.

Η στασιμότητα ή μη μίας χρονοσειράς μπορεί να ελεγχθεί διαγραμματικά, π.χ. με απλό διάγραμμα της εξέλιξης των τιμών της στο χρόνο (Γράφημα 12) και με διαγράμματα αυτοσυσχέτισης (Γράφημα 13). Οι αυτοσυσχετίσεις μίας στάσιμης χρονοσειράς φθίνουν στο μηδέν με γρήγορο ρυθμό όσο αυξάνονται οι χρονικές υστερήσεις (lags) ενώ οι για τις μη στάσιμες φθίνουν με αργό ρυθμό.

Γράφημα 6 Διαγράμματα Αυτοσυσχέτισης μη στάσιμης (αριστερά) και στάσιμης (δεξιά) χρονοσειράς



Ως αυτοσυσχέτιση σε μία χρονοσειρά ορίζεται η συσχέτιση μίας τιμής της μεταβλητής της με προηγούμενες τιμές της ίδιας μεταβλητής με καθυστέρηση μίας ή και περισσότερων περιόδων. Στην ανάλυση χρονοσειρών είναι πολύ σημαντικό να γνωρίζουμε εάν οι παρατηρούμενες τιμές έχουν σχέση αλληλεξάρτησης μεταξύ τους και για το σκοπό αυτό χρησιμοποιούμε το συντελεστή αυτοσυσχέτισης. Ο συντελεστής αυτοσυσχέτισης συμβολίζεται με r_k και υπολογίζεται ως εξής:

$$r_k = \frac{\sum_{t=k+1}^n (Y_t - \bar{Y})(Y_{t-k} - \bar{Y})}{\sum_{t=1}^n (Y_t - \bar{Y})^2}$$

Ο συντελεστής r_k δηλώνει πως οι παρατηρήσεις Y περιόδου $t = 1, 2, \dots, n$ με χρονοκαθυστέρηση (lag) k σχετίζονται μεταξύ τους και δείχνει αν τα δεδομένα είναι τυχαία, αν έχουν τάση, αν είναι στάσιμα και αν έχουν εποχικότητα. Σε περίπτωση που τα δεδομένα είναι τυχαία ο συντελεστής αυτοσυσχέτισης θα έχει τιμή κοντά στο μηδέν ενώ αν υπάρχει τάση ο συντελεστής αυτοσυσχέτισης της πρώτης χρονοκαθυστέρησης θα είναι κοντά στο 1. Όταν τα δεδομένα παρουσιάζουν εποχικότητα τότε ο συντελεστής αυτοσυσχέτισης θα παρουσιάζει κάποιο peak σε χρονοκαθυστέρηση k ($k=12$ για μηνιαίες παρατηρήσεις, $k=4$ για τριμηνιαίες παρατηρήσεις κ.α.).

Αιτιοκρατία και Στοχαστικότητα

Αν οι τιμές μίας χρονοσειράς μπορούν να καθοριστούν ακριβώς μέσω μίας απλής συνάρτησης που περιλαμβάνει χρονική συνιστώσα, τότε η χρονοσειρά ονομάζεται αιτιοκρατική. Όταν μία χρονοσειρά είναι αιτιοκρατική, πρακτικά απουσιάζει σε κάποιο σημαντικό βαθμό η τυχαιότητα. Στον πραγματικό κόσμο όμως, σχεδόν όλες οι χρονοσειρές περιέχουν κάποιο ποσοστό θορύβου και έτσι μπορούν όλες να θεωρηθούν στοχαστικές. Με τον όρο στοχαστικότητα λοιπόν,

περιγράφεται το φαινόμενο κατά το οποίο οι τιμές μίας χρονοσειράς μεταβάλλονται με κάποια μικρή ή μεγάλη τυχαιότητα στο χρόνο.

Γραμμικότητα και μη-Γραμμικότητα

Γραμμικότητα σε μία χρονοσειρά έχουμε όταν οι τιμές της μεταβλητής της αλληλοεπιδρούν γραμμικά στο χρόνο. Σε μία γραμμική χρονοσειρά μπορούμε να ορίσουμε την εξέλιξη των τιμών της ως γραμμικό συνδυασμό των προηγούμενων παρατηρήσεων της χρονοσειράς. Αντίθετα, σε ένα μη-γραμμικό σύστημα έχουμε μεταβαλλόμενη επίδραση μεταξύ της τιμής και των προηγούμενων παρατηρήσεων της σε διαφορετικές χρονικές στιγμές.

4.2.2 Ολοκληρωμένα Αυτοπαλίνδρομα Μοντέλα Κινητού Μέσου Όρου (ARIMA)

Τα ολοκληρωμένα αυτοπαλίνδρομα μοντέλα κινητού μέσου όρου (ARIMA) χρησιμοποιούνται σε ευρεία κλίμακα στην ανάλυση χρονοσειρών για τη διενέργεια προβλέψεων. Αποτελούν μοντέλα τα οποία χρησιμοποιούν τις παρελθοντικές τιμές μίας χρονοσειράς και πραγματοποιούν προβλέψεις με γραμμική παλινδρόμηση.

Τα μοντέλα ARIMA όπως υποδηλώνει και το όνομά τους αποτελούνται από τρία βασικά συστατικά, τους όρους AR, MA και I. Ο όρος AR, ή αλλιώς αυτοπαλινδρομούμενο μέρος, υποδηλώνει πως το μοντέλο χρησιμοποιεί την εξαρτημένη σχέση μεταξύ των πρόσφατων και των παλαιότερων παρατηρήσεων μιας μεταβλητής, παλινδρομώντας πρακτικά τα δεδομένα στις παλαιότερες τιμές τους. Έτσι τα αυτοπαλινδρομούμενα μοντέλα AR είναι μοντέλα γραμμικής παλινδρόμησης με εξαρτημένη μεταβλητή την τυχαία μεταβλητή της χρονοσειράς σε μία χρονική στιγμή t και ανεξάρτητες μεταβλητές τις τυχαίες μεταβλητές σε προηγούμενους χρόνους $t-1, \dots, t-p$. Ο αριθμός των υστερήσεων που χρησιμοποιούμε λέγεται τάξη του μοντέλου και συμβολίζεται με p και το αυτοπαλινδρομούμενο μοντέλο τάξης p συμβολίζεται AR(p) και ορίζεται ως

$$y_t = \beta_0 + \beta_1 y_{t-1} + \dots + \beta_p y_{t-p}$$

Ο όρος MA αποτελεί το μοντέλο που υπολογίζει τα κατάλοιπα των σφαλμάτων παρελθοντικής χρονοσειράς ως γραμμικό συνδυασμό των παρόντων ή μελλοντικών τιμών της χρονοσειράς. Έτσι στα MA μοντέλα η εξαρτημένη μεταβλητή είναι μία τυχαία μεταβλητή της χρονοσειράς σε μία χρονική στιγμή t και οι ανεξάρτητες μεταβλητές είναι τα σφάλματα που προέκυψαν σε χρονικές υστερήσεις $t-1, \dots, t-q$.

$$y_t = \alpha_0 + \alpha_1 \varepsilon_{t-1} + \dots + \alpha_p \varepsilon_{t-p}$$

Ο συνδυασμός των μοντέλων AR και MA μας δίνουν το μοντέλο ARMA το οποίο ορίζεται ως

$$y_t = \beta_0 + \alpha_0 + \beta_1 y_{t-1} + \alpha_1 \varepsilon_{t-1} + \dots + \beta_p y_{t-p} + \alpha_p \varepsilon_{t-p}$$

Όπου τα β και α αναπαριστούν τους συντελεστές του AR και MA μοντέλου αντίστοιχα. Τέλος, ο όρος I υποδηλώνει την διαφορίση στη χρονοσειρά έτσι ώστε να γίνει στάσιμη. Για παράδειγμα, σε ένα μοντέλο ARIMA του οποίου ο όρος I λαμβάνει την τιμή $d = 1$ σημαίνει πως χρησιμοποιήθηκαν 1^ης τάξης διαφορές έτσι ώστε να γίνει στάσιμη η χρονοσειρά.

Τα μοντέλα ARIMA είναι κατάλληλα για αναλύσεις και προβλέψεις σε δεδομένα με τάση, περιοδικότητα και εποχικότητα.

4.2.3 Η μεθοδολογία Box – Jenkins

Οι George Box και Gwilym Jenkins ήταν Άγγλοι στατιστικοί οι οποίοι έγιναν ευρέως γνωστοί από τη μεταξύ τους συνεργασία στα αυτοπαλίνδρομα μοντέλα κινητού μέσου όρου στην ανάλυση χρονοσειρών. Στο βιβλίο με τίτλο “TIME SERIES ANALYSIS forecasting and control” που εκδόθηκε το 1976 περιγράφουν τη μέθοδό τους η οποία αποτελείται από 3 στάδια: την ταυτοποίηση (identification), την εκτίμηση (estimation) και το διαγνωστικό έλεγχο (diagnostic check). Η μέθοδος αυτή είναι κατάλληλη για μεσαίες και μεγάλες χρονοσειρές οι οποίες περιλαμβάνουν το λιγότερο 50 παρατηρήσεις και είναι μη-στάσιμες.

Ταυτοποίηση (Identification)

Κατά τη διαδικασία της ταυτοποίησης προσδιορίζονται οι παράμετροι p , d και q ενός μοντέλου ARIMA(p,d,q) όπου p είναι ο βαθμός υστέρησης (lag order), d ο βαθμός διαφορίσης και q ο βαθμός κινητού μέσου όρου. Επειδή οι χρονοσειρές κατά πλειοψηφία παρουσιάζουν μη στασιμότητα και εποχικότητα, το πρώτο βήμα που πρέπει να πραγματοποιηθεί είναι η μετατροπή μίας μη στάσιμης χρονοσειράς σε στάσιμη και η εξάλειψη της εποχικότητας με μοντελοποίηση. Μία χρονοσειρά παρατηρήσεων X_t (όπου t η χρονική περίοδος) είναι μη στάσιμη όταν στις παρατηρούμενες τιμές υπάρχει τάση, μεταβολή στο μέσο ή εποχική διακύμανση (ο μέσος και η διακύμανση μεταβάλλονται για ορισμένες τιμές του t). Την πληροφορία για στασιμότητα ή μη την λαμβάνουμε παρατηρώντας τα διαγράμματα αυτοσυσχέτισης και μερικής αυτοσυσχέτισης της χρονοσειράς αλλά και μέσω του ελέγχου μοναδιαίας ρίζας.

Το διάγραμμα αυτοσυσχέτισης (ACF) δείχνει τη συσχέτιση μίας παρατήρησης με υπόλοιπες παρατηρήσεις με κάποια χρονική υστέρηση (lag). Ο άξονας x δείχνει τη χρονική υστέρηση και ο άξονας y δείχνει το συντελεστή συσχέτισης ο οποίος διακυμαίνεται από -1 έως 1 για αρνητική και θετική συσχέτιση αντίστοιχα. Το διάγραμμα μερικής αυτοσυσχέτισης (PACF) δείχνει τις συσχετίσεις μίας παρατήρησης με υπόλοιπες παρατηρήσεις με κάποια χρονική υστέρηση μη λαμβάνοντας όμως υπόψιν την επίδραση που μπορούν ενδεχομένως να επιφέρουν οι ενδιάμεσες παρατηρήσεις. Και τα δύο διαγράμματα περιλαμβάνουν σε οριζόντιες γραμμές τα διαστήματα εμπιστοσύνης 95% και 99% και όταν μία μπάρα ξεπερνά τα όρια αυτά, είναι σημαντική και αξίζει προσοχής καθώς υποδεικνύει χρήσιμους όρους του μοντέλου όπως θα δούμε και στη συνέχεια.

Την πληροφορία για στασιμότητα ή μη στασιμότητα της χρονοσειράς λαμβάνουμε επίσης όπως αναφέραμε και προηγουμένως και μέσω του ελέγχου μοναδιαίας ρίζας. Όταν ο συντελεστής της αυτοπαλίνδρομης διαδικασίας ισούται με 1 (μοναδιαία ρίζα) η χρονοσειρά είναι μη στάσιμη ενώ σε περίπτωση που είναι μικρότερος της μονάδας τότε η χρονοσειρά είναι στάσιμη. Ο συνηθέστερος έλεγχος για την εξέταση μοναδιαίας ρίζας είναι ο έλεγχος των Dickey – Fuller οι οποίοι υλοποίησαν το τεστ το 1979 (Dickey και Fuller). Στην διευρυμένη μορφή του (Augmented Dickey – Fuller Test), που θα εφαρμόσουμε και στην παρούσα εργασία, η εξίσωση του αρχικού Dickey – Fuller Test διευρύνεται περιλαμβάνοντας μεγαλύτερη τάξη στη διαδικασία παλινδρόμησης. Έτσι η νέα εξίσωση θα λαμβάνει τη μορφή

$$y_t = c + \beta_t + \alpha y_{t-1} + \varphi_1 \Delta y_{t-1} + \varphi_2 \Delta y_{t-2} + \dots + \varphi_p \Delta y_{t-p} + e_t$$

Όπου y_{t-1} η υστέρηση 1 της χρονοσειράς και Δy_{t-1} οι πρώτες διαφορές της χρονοσειράς τη στιγμή $t-1$. Η μηδενική υπόθεση παραμένει ίδια με του Dickey – Fuller Test ($H_0 : \alpha=1$) και υποδηλώνει την ύπαρξη μοναδιαίας ρίζας. Η p-value του ελέγχου πρέπει να είναι μικρότερη από το επίπεδο σημαντικότητας προκειμένου να απορριφθεί η μηδενική υπόθεση και να υποθέσουμε στασιμότητα στη χρονοσειρά.

Σε περίπτωση που έχουμε ενδείξεις για την ύπαρξη μη στασιμότητας, λαμβάνουμε κατάλληλης τάξης διαφορές για τη μετατροπή της χρονοσειράς σε στάσιμη. Έτσι αν η αρχική χρονοσειρά είναι της μορφής $W = X_t - X_{t-1}$ ή αλλιώς $W_t = (1 - B)X_t$ μετά τη διαφόριση λαμβάνει τη γενική μορφή

$$\Phi_p(B)(1 - B)^d X_t = \theta_q(B)\alpha_t(I)$$

όπου d είναι ο βαθμός διαφόρισης. Αν η χρονοσειρά δεν μετατραπεί σε στάσιμη με πρώτης τάξης διαφορές τότε διαφορίζουμε ξανά και παρατηρούμε τα διαγράμματα καθώς και τον έλεγχο μοναδιαίας ρίζας. Αν για παράδειγμα η αυτοσυσχέτιση εξαλείφεται με δεύτερης τάξης διαφορές τότε ο συντελεστής d ισούται με 2.

Οι τιμές q και p προσδιορίζονται από τα διαγράμματα αυτοσυσχέτισης και μερικής αυτοσυσχέτισης αντίστοιχα ύστερα από της διαδικασία της διαφόρισης. Αν οι μερικές αυτοσυσχετίσεις εξαλείφονται ύστερα από λίγες υστερήσεις τότε η τελευταία υστέρηση με μεγάλη αυτοσυσχέτιση και πάνω από τα διαστήματα εμπιστοσύνης εκτιμά και τις τιμές των q και p .

Σε περίπτωση που οι αυτοσυσχετίσεις και οι μερικές αυτοσυσχετίσεις παραμένουν, προτείνεται η δημιουργία μεικτών μοντέλων μέσω δοκιμών στις επιλογές των p και q μέχρις ότου τα κατάλοιπα να μην αυτοσυσχετίζονται (λευκός θόρυβος).

Εκτίμηση του μοντέλου (model estimation) και διαγνωστικός έλεγχος (checking)

Αφού καθοριστούν οι τιμές των p , q , και d γίνεται ο υπολογισμός των παραμέτρων ϕ και θ της εξίσωσης 1 μέσω της διαδικασίας εκτίμησης μέγιστης πιθανοφάνειας. Η εξίσωση μέγιστη πιθανοφάνειας επιλύεται με μη γραμμική μεγιστοποίηση της συνάρτησης και μόλις το μοντέλο προσαρμοστεί στα δεδομένα γίνεται ο διαγνωστικός έλεγχος.

Κατά τον διαγνωστικό έλεγχο ελέγχεται αν το εκτιμηθέν υπόδειγμα στο προηγούμενο στάδιο είναι αποτελεσματικό, δηλαδή αν τα κατάλοιπα κατανομούνται ομοιόμορφα με μηδενικό μέσο και διακύμανση σ^2 (λευκός θόρυβος). Ο έλεγχος για την ανεξαρτησία των καταλοίπων γίνεται με το Test των Ljung-Box όπου για $p - value > 0.05$ απορρίπτεται η μηδενική υπόθεση η οποία υποδεικνύει εξάρτηση μεταξύ των καταλοίπων.

Στο στάδιο του διαγνωστικού ελέγχου εξετάζεται επίσης και η καλή προσαρμογή του υποδείγματος στα δεδομένα συγκρίνοντάς το με ένα υπόδειγμα μεγαλύτερης τάξης. Αν στο υπόδειγμα μεγαλύτερης τάξης οι επιπλέον συντελεστές είναι ίσοι με μηδέν, τότε το αρχικό εκτιμηθέν υπόδειγμα παρουσιάζει την καλύτερη προσαρμογή. Στην ανάλυση χρονοσειρών, ως δείκτης καλής προσαρμογής χρησιμοποιείται ευρέως ο δείκτης AIC (Akaike Information Criterion) ο οποίος δίνεται από τον τύπο

$$AIC = \ln(s^2) + \frac{2n}{T}$$

Όπου s^2 είναι η εκτίμηση διακύμανσης των καταλοίπων, n ο αριθμός των εκτιμωμένων παραμέτρων του υποδείγματος και T ο αριθμός των παρατηρήσεων που χρησιμοποιούνται στην παλινδρόμηση. Η χαμηλότερη τιμή του δείκτη AIC μεταξύ των εκτιμημένων υποδειγμάτων δίνει το καλύτερο υπόδειγμα σε όρους προσαρμογής στα δεδομένα.

4.2.4 Τεχνητά Νευρωνικά Δίκτυα

Τα Τεχνητά Νευρωνικά Δίκτυα είναι ένα υπολογιστικό σύστημα το οποίο επεξεργάζεται τα δεδομένα που του εισάγονται και πραγματοποιεί υπολογισμούς με τρόπο που μιμείται το ανθρώπινο νευρικό σύστημα. Στις μέρες μας είναι ευρέως διαδεδομένα στον κλάδο της επιστήμης των δεδομένων καθώς έχουν τη δυνατότητα να αντιμετωπίζουν μη γραμμικές σχέσεις και να δίνουν ακριβείς προβλέψεις μέσα από σύνολα δεδομένων διάφορων μορφών όπως αριθμοί, κείμενα, εικόνες και ήχος.

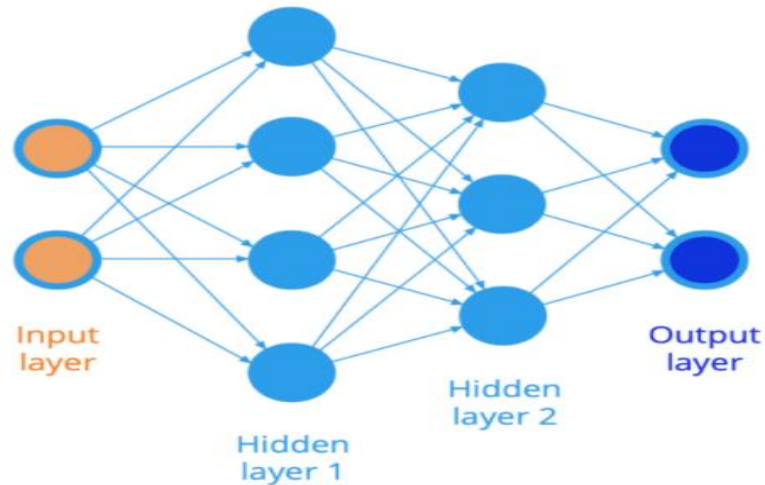
Το Νευρωνικό Δίκτυο αποτελείται από απλούς υπολογιστικούς κόμβους (νευρώνες) οι οποίοι είναι διασυνδεδεμένοι μεταξύ τους. Ένα στρώμα νευρώνων (layer) αποτελείται ουσιαστικά από πολλούς νευρώνες και κατ' επέκταση ένα Νευρωνικό Δίκτυο δύναται να αποτελείται από πολλά στρώματα νευρώνων. Κάθε νευρώνας λαμβάνει ένα σύνολο αριθμητικών εισόδων, είτε απευθείας από το εξωτερικό περιβάλλον είτε από άλλους νευρώνες, πραγματοποιεί υπολογισμούς και παράγει μία έξοδο. Η έξοδος αυτή με τη σειρά της, είτε αποτελεί είσοδο για άλλους νευρώνες είτε κατευθύνεται στο εξωτερικό περιβάλλον ως αποτέλεσμα του δικτύου.

Τα Νευρωνικά Δίκτυα αποτελούνται από τρεις τύπους νευρώνων: τους νευρώνες εισόδου, τους νευρώνες εξόδου και τους υπολογιστικούς ή κρυμμένους νευρώνες (hidden layers). Οι Νευρώνες εισόδου δεν έχουν υπολογιστική λειτουργία καθώς απλώς αποτελούν απλώς την πύλη εισόδου για τα εκάστοτε δεδομένα. Οι υπολογιστικοί ή κρυμμένοι νευρώνες λαμβάνουν κάθε είσοδο από τους νευρώνες εισόδου και την πολλαπλασιάζουν με ένα συναπτικό βάρος. Στη συνέχεια αθροίζουν το σύνολο των γινομένων το οποίο τροφοδοτείται ως όρισμα στη συνάρτηση ενεργοποίησης του κάθε κόμβου. Τέλος, το αποτέλεσμα της συνάρτησης ενεργοποίησης αποτελεί την έξοδο του νευρώνα και αντίστοιχα την είσοδο για τους νευρώνες του επόμενου στρώματος.

Η έξοδος y_k του ενός νευρώνα k δίνεται από την εξίσωση

$$y_k = \Phi\left(\sum_{i=0}^n x_{ki} * w_{ki}\right)$$

όπου x_{ki} η i -οστή είσοδος του νευρώνα k , w_{ki} το i -οστό συναπτικό βάρος του νευρώνα k και Φ η συνάρτηση ενεργοποίησης του νευρωνικού δικτύου. (Wikipedia, Neural Networks)



Εικόνα 7 Artificial Neural Network

Η συνάρτηση ενεργοποίησης υποδεικνύει ουσιαστικά τη λειτουργία ή μη κάθε νευρώνα και μπορεί να είναι η βηματική συνάρτηση ή συνάρτηση κατωφλίου, η συνάρτηση πρόσημου, η σιγμοειδής συνάρτηση ή η γραμμική συνάρτηση. Στους νευρώνες με βηματική συνάρτηση ενεργοποίησης η τιμές που μπορούν να πάρουν τα σήματα εξόδου είναι 0 ή 1, στους νευρώνες με συνάρτηση πρόσημου είναι -1 ή 1 ενώ στους νευρώνες με σιγμοειδή συνάρτηση ενεργοποίησης οι τιμές μπορούν να είναι οποιοσδήποτε πραγματικός αριθμός μεταξύ 0 και 1. Η σιγμοειδής συνάρτηση ενεργοποίησης είναι ευρέως διαδεδομένη στην κατασκευή νευρωνικών δικτύων και δύο παραδείγματα αυτής είναι η λογιστική σιγμοειδής (πεδίο τιμών $[0,1]$) και η υπερβολική εφαιπτομένη (πεδίο τιμών $[-1,1]$).

Αναφορικά με της αρχιτεκτονικές μορφές που μπορεί να λάβει ένα τεχνητό νευρωνικό δίκτυο, οι νευρώνες των στρωμάτων (layers) μπορεί να είναι πλήρως συνδεδεμένοι (fully connected) ή μερικώς συνδεδεμένοι (partially connected). Στη μεν πρώτη περίπτωση κάθε νευρώνας συνδέεται με όλους τους νευρώνες του επόμενου στρώματος (layer), ενώ στη δεύτερη περίπτωση η συνθήκη αυτή δεν είναι απαραίτητη και έτσι υπάρχουν νευρώνες που δεν συνδέονται με όλους τους νευρώνες του επόμενου στρώματος. Αναλόγως με το πως οι νευρώνες τροφοδοτούν άλλους νευρώνες, τα τεχνητά νευρωνικά δίκτυα χωρίζονται σε πρόσθιας τροφοδότησης (feed forward) και ανατροφοδότησης (feed backward or recurrent). Στα τεχνητά νευρωνικά δίκτυα πρόσθιας τροφοδότησης οι νευρώνες είναι οργανωμένοι σε διαφορετικά επίπεδα ώστε οι νευρώνες του εκάστοτε επιπέδου να τροφοδοτούν τους νευρώνες του αμέσως επόμενου επιπέδου. Στα τεχνητά νευρωνικά δίκτυα ανατροφοδότησης οι νευρώνες ενός επιπέδου μπορούν να τροφοδοτούν νευρώνες του ίδιου επιπέδου αλλά και προηγούμενων επιπέδων. Η λειτουργία αυτή δίνει στο νευρωνικό δίκτυο τη δυνατότητα της προσωρινής μνήμης η οποία διακρίνεται σε αυτοσυσχετιζόμενη μνήμη (ανατροφοδότηση νευρώνων ίδιου επιπέδου) και σε ετεροσυσχετιζόμενη μνήμη (ανατροφοδότηση νευρώνων διαφορετικών επιπέδων). Και στις δύο περιπτώσεις η διαδικασία της μάθησης δικτύου

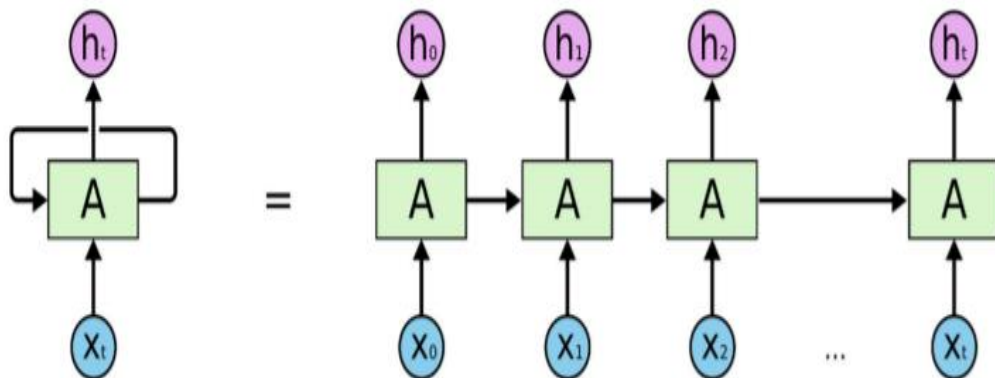
πραγματοποιείται με έναν αριθμό “περασμάτων” μεταξύ των στρωμάτων/επιπέδων του δικτύου που ονομάζονται εποχές (epochs).

Τα στοιχεία που καλείται να παραμετροποιήσει ο σχεδιαστής του νευρωνικού δικτύου αφορούν στον αριθμό των ενδιάμεσων κρυφών επιπέδων-στρωμάτων (layers), τον αριθμό των κόμβων ανά επίπεδο, τον τρόπο σύνδεσης των κόμβων μεταξύ τους, την τιμή ενεργοποίησης (τιμή κατωφλίου), τη μορφή της συνάρτησης μετάβασης, οι τιμές των αρχικών βαρών μεταξύ των κόμβων και οι αλγόριθμοι (κανόνες εκπαίδευσης) που χρησιμοποιούνται για να ενισχυθούν οι σύνδεσμοι μεταξύ των κόμβων.

Απαραίτητο βήμα για την υλοποίηση μίας ανάλυσης μέσω τεχνητού νευρωνικού δικτύου είναι η διαδικασία μετασχηματισμού (κανονικοποίησης) των δεδομένων εισαγωγής του δικτύου. Δύο τεχνικές μετασχηματισμού δεδομένων είναι η μέγιστη και ελάχιστη τιμή, όπου τα αρχικά δεδομένα μετασχηματίζονται σε εύρος τιμών από -1 έως 1, και μέσης τιμής και διασποράς όπου μετά το μετασχηματισμό ή νέα μέση τιμή του συνόλου των δεδομένων να είναι ίση με 0 και η διασπορά να είναι ίση με 1.

4.2.5 Το μοντέλο LSTM

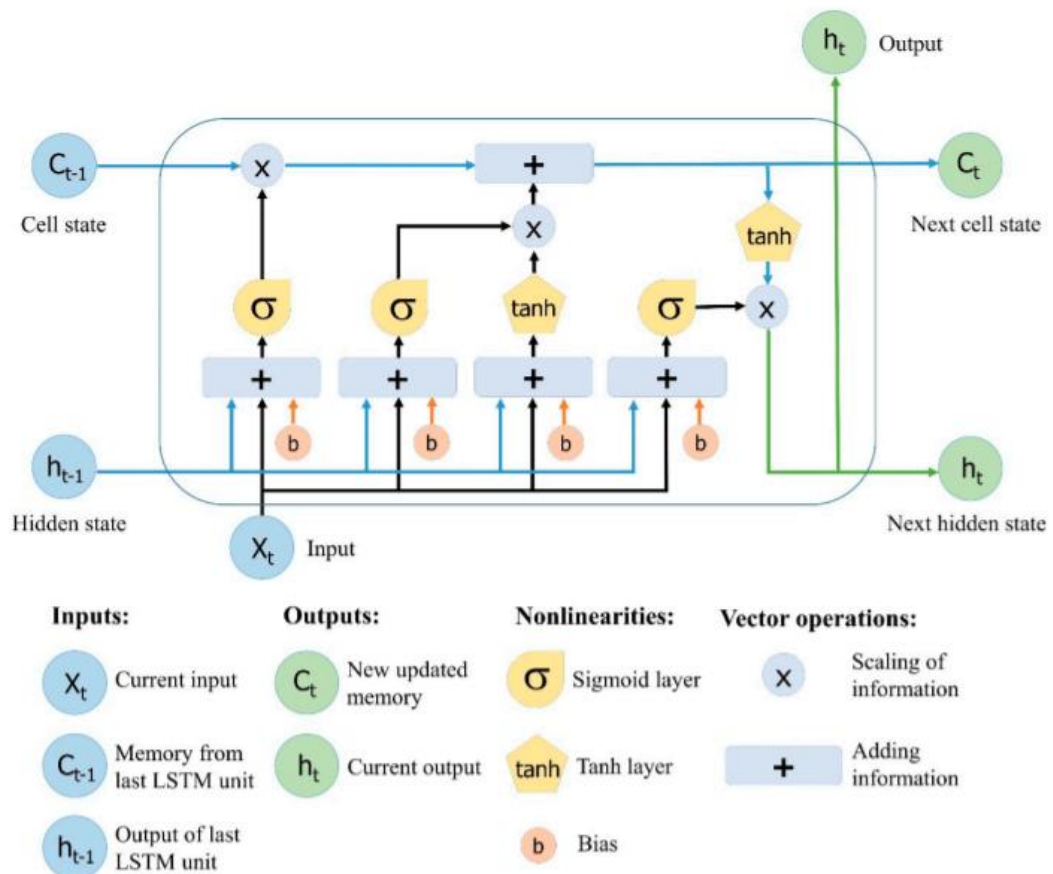
Τα μοντέλα LSTM ανήκουν στην κατηγορία RNN (Recurrent Neural Networks), είναι δηλαδή ουσιαστικά μία εξελιγμένη μορφή νευρωνικών δικτύων ανατροφοδότησης. Τα Recurrent Neural Networks αποτελούνται από ένα στρώμα/επίπεδο εισόδου, ένα ή περισσότερα κρυμμένα επίπεδα και ένα επίπεδο εξόδου. Η δομή τους αποτελείται από μία αλυσιδωτή αλληλουχία (sequential process) επαναλαμβανόμενων ενοτήτων με βασική ιδέα τη χρησιμοποίησή τους ως προσωρινή μνήμη πληροφοριών των παρελθοντικών βημάτων. Η παρακάτω εικόνα δείχνει τη μορφή ενός RNN δικτύου .



Εικόνα 8 RNN Network (Olah)

Η Εικόνα 9 παρουσιάζει ένα δίκτυο RNN με μία μονάδα εισόδου, και μία επαναλαμβανόμενη κρυφή μονάδα επεκτεινόμενη σε ένα πλήρες δίκτυο όπου X_t είναι η είσοδος στο χρονικό σημείο t και h_t η έξοδος στο χρονικό σημείο t . Κατά τη διαδικασία της εκπαίδευσης το δίκτυο χρησιμοποιεί τον αλγόριθμο backpropagation ο οποίος υπολογίζει την παράγωγο της συνάρτησης σφάλματος και προσαρμόζει τα συναπτά βάρη των νευρώνων με τη μέθοδο gradient descent. Ωστόσο, τα δίκτυα RNN δεν είναι αρκετά αποτελεσματικά στην εύρεση μοτίβων μακροχρόνιων εξαρτήσεων εξαιτίας των φαινομένων gradient vanishing και exploding gradient (Bengio Y.).

Τα δίκτυα LSTM παρουσιάστηκαν για πρώτη φορά από τους Hochreiter και Schmidhuber (Hochreiter και Schmidhuber) με σκοπό να ξεπεράσουν τα ανωτέρω μειονεκτήματα των παραδοσιακών δικτύων RNN προσθέτοντας επιπλέον δια δράσεις ανά κελί (Xuan Hien Le, Application of Long Short Term Memory (LSTM) Neural Network for Flood Forecasting). Είναι κατάλληλα για μάθηση μακροχρόνιων εξαρτήσεων συγκρατώντας στην προσωρινή μνήμη τους πληροφορίες από προγενέστερες περιόδους. Η Εικόνα 6 δείχνει τη δομή και τα συστατικά ενός LSTM δικτύου.



Εικόνα 9 LSTM Network (Yan)

Ένα τυπικό δίκτυο LSTM αποτελείται από μπλοκ μνήμης τα οποία ονομάζονται κύτταρα (cells). Σε κάθε επόμενο κελί της ακολουθίας μεταφέρονται δύο μέρη, το μέρος του κυττάρου και το

κρυφό μέρος. Το μέρος του κυττάρου είναι η κύρια αλυσίδα της ροής δεδομένων, η οποία επιτρέπει στα δεδομένα να μετακινούνται προς τα εμπρός (feedforward) ουσιαστικά αμετάβλητα εκτός ορισμένων ενδεχόμενων γραμμικών μετασχηματισμών. Τα δεδομένα μπορούν να προστεθούν ή να αφαιρεθούν από το μέρος του κυττάρου μέσω των σιγμοειδών πυλών. Μία πύλη είναι όμοια με ένα στρώμα από πράξεις πινάκων, οι οποίες περιλαμβάνουν διαφορετικά βάρη ελέγχοντας τη διαδικασία απομνημόνευσης, και δίνει τη δυνατότητα σε ένα δίκτυο LSTM να αποφεύγει μακροχρόνιες εξαρτήσεις. Η διαδικασία αναγνώρισης και εξαίρεσης δεδομένων επιτελείται από τη σιγμοειδή συνάρτηση η οποία λαμβάνει την έξοδο h_{t-1} της προγενέστερης LSTM μονάδας (κρυφό μέρος) τη χρονική στιγμή $t - 1$ και την είσοδο X_t τη χρονική στιγμή t . Επιπλέον η σιγμοειδής συνάρτηση καθορίζει πιά μέρη από τις παρελθοντικές εισόδους πρέπει να αγνοηθούν. Η πύλη αυτή ονομάζεται forget gate (ft) όπου f_t είναι ένα διάνυσμα με τιμές εύρους 0 και 1 που αντιστοιχεί σε κάθε αριθμό στο μέρος του κυττάρου C_{t-1} .

$$f_t = \sigma(W_f \cdot [h_{t-1}, X_t] + b_f),$$

$$N_t = \tanh(W_n \cdot [h_{t-1}, X_t] + b_n),$$

$$C_t = C_{t-1}f_t + N_t i_t$$

C_{t-1} και C_t είναι τα μέρη των κυττάρων τη χρονική στιγμή $t - 1$ και t αντίστοιχα και τα W και b τα σχετικά βάρη και κατώφλια ενεργοποίησης του κυτταρικού μέρους.

Στο τελευταίο βήμα, οι εξαγόμενες τιμές h_t είναι οι φιλτραρισμένες τιμές της εξόδου του μέρους του κυττάρου. Αρχικά, ένα σιγμοειδές layer αποφασίζει ποια μέρη του κυτταρικού μέρους θα αποτελέσουν τις εξόδους του. Στη συνέχεια, οι έξοδοι της σιγμοειδούς πύλης O_t πολλαπλασιάζονται από τις νέες τιμές που προέκυψαν από το επίπεδο της υπερβολικής εφαιπτομένης tanh του κυτταρικού μέρους, με εύρος τιμών από -1 έως 1. (Xuan Hien Le, Application of Long-Short Term Memory (LSTM) Neural Network for Flood Forecasting)

$$O_t = \sigma(W_o \cdot [h_{t-1}, X_t] + b_o),$$

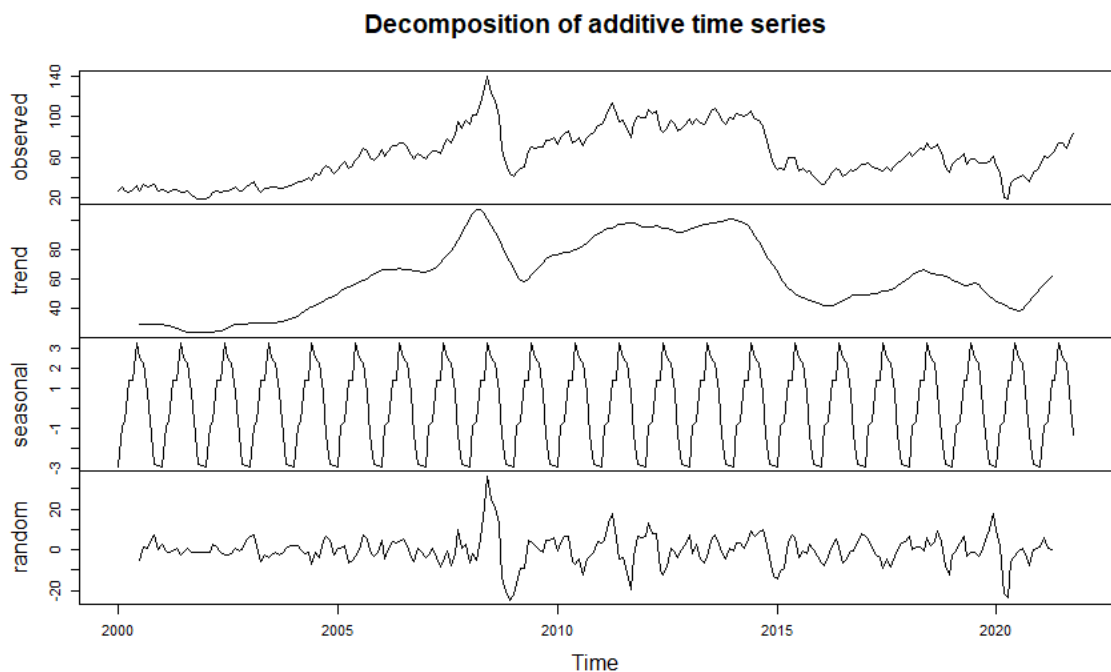
$$h_t = O_t \tanh(C_t)$$

Τέλος, η συνάρτηση ενεργοποίησης tanh χρησιμοποιείται στην κανονικοποίηση των τιμών που ρέουν μέσα στο δίκτυο LSTM συμπιέζοντάς τις για να βρίσκονται πάντα στο εύρος (-1,1) και ρυθμίζοντας έτσι την έξοδο του δικτύου.

4.3. Προσδιορισμός κατάλληλου μοντέλου ARIMA στη χρονοσειρά Ιανουάριος 2000 – Δεκέμβριος 2020 και δημιουργία πρόβλεψης της τιμής του πετρελαίου κατά την περίοδο Ιανουάριος 2021 – Οκτώβριος 2021.

Στην παρούσα ενότητα παρουσιάζονται τα αποτελέσματα της προσαρμογής ενός μοντέλου ARIMA στις μηνιαίες καταγραφές της τιμής του αργού πετρελαίου την περίοδο Ιανουάριος 2000-Δεκέμβριος 2020 όπως αυτά προέκυψαν μέσω του στατιστικού πακέτου R. Το μοντέλο προσαρμόστηκε στο σύνολο της παραπάνω χρονοσειράς με δεδομένα από τον Ιανουάριο του 2000 έως τον Δεκέμβριο του 2020 και χρησιμοποιήθηκε για να προβλέψει τις μηνιαίες τιμές του αργού πετρελαίου την περίοδο Ιανουάριος – Οκτώβριος 2021. Ο παραπάνω διαχωρισμός σε training/test υποσύνολα έγινε με τρόπο ώστε να συμπεριληφθεί στην ανάλυση και η επίδραση στην τιμή του αργού πετρελαίου που έχει η χρονική περίοδος Μάρτιος – Δεκέμβριος 2020 όπου το φαινόμενο της πανδημίας του COVID-19 έχει παγκόσμιο αντίκτυπο μεταβάλλοντας την εικόνα της τιμής του πετρελαίου αλλά και πολλών χρηματοοικονομικών δεικτών.

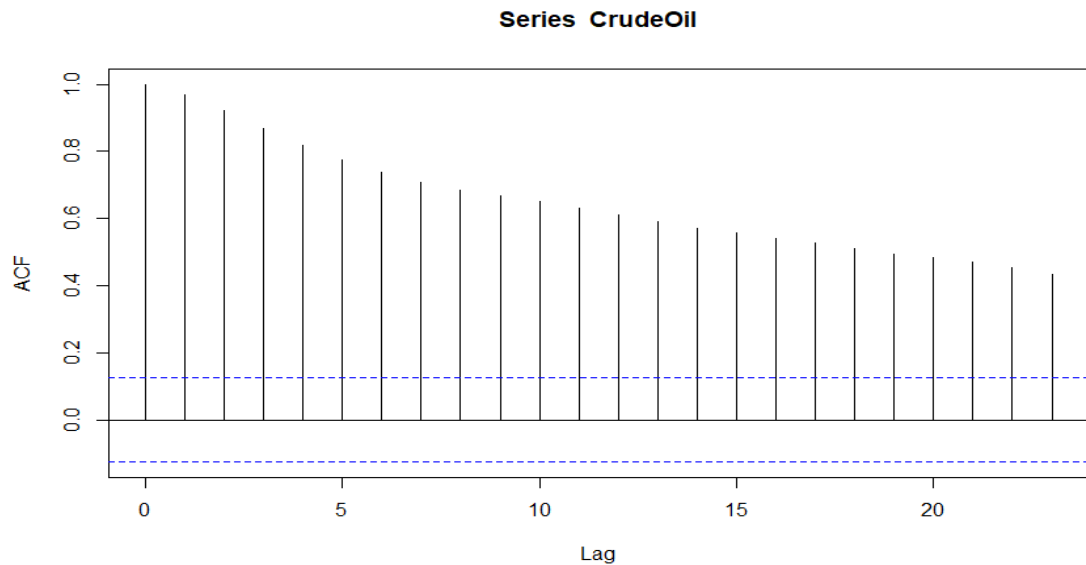
Γράφημα 10 Περιγραφικά Διαγράμματα της χρονοσειράς Ιανουάριος 2000 – Δεκέμβριος 2020



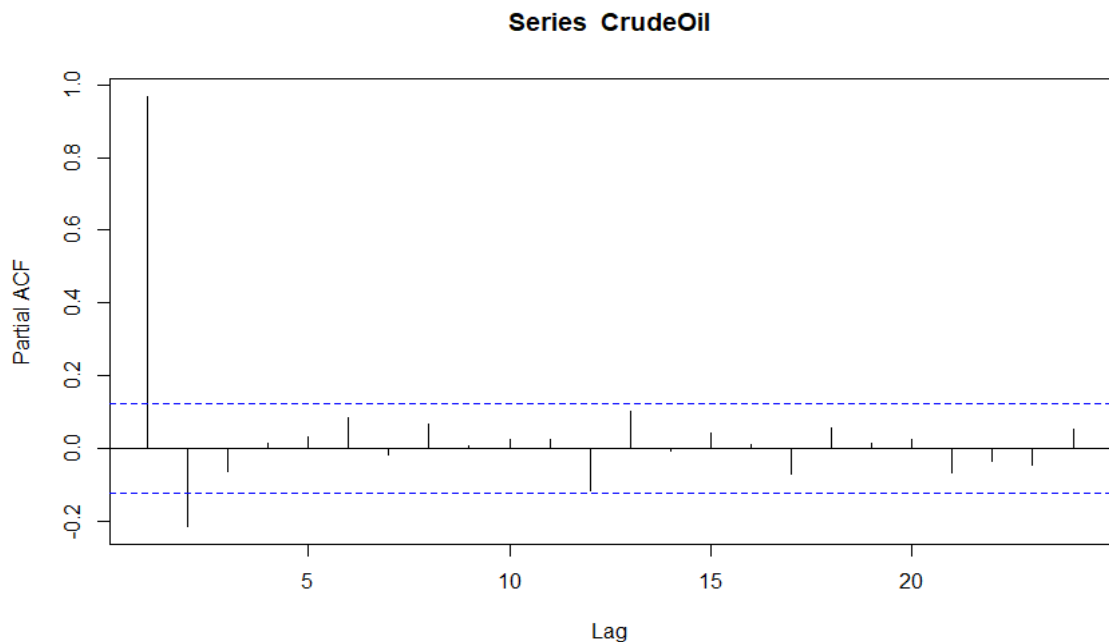
Όπως παρατηρούμε στο Γράφημα 10, η χρονοσειρά της τιμής του αργού πετρελαίου εμφανίζει μακροχρόνια ελαφρώς ανοδική τάση με σημαντικά πτωτικές περιόδους κατά τα έτη 2008 (παγκόσμια οικονομική κρίση), 2015 (πετρελαϊκή κρίση) και το Μάρτιο του 2020 (παγκόσμια πανδημία COVID-19). Ακόμα παρατηρούμε πως η χρονοσειρά εμφανίζει εποχιακές μεταβολές και σχετικά τυχαία κυμαινόμενα κατάλοιπα.

Στα γραφήματα αυτοσυσχέτισης (ACF) και μερικής αυτοσυσχέτισης (PACF) που ακολουθούν φαίνεται πως η χρονοσειρά του test δείγματος δεν είναι στάσιμη καθώς οι αυτοσυσχετίσεις δεν εξαλείφονται γρήγορα ύστερα από κάποιους βαθμούς υστέρησης.

Γράφημα 11 Διάγραμμα αυτοσυσχέτισης της χρονοσειράς Ιανουάριος 2000 – Δεκέμβριος 2020



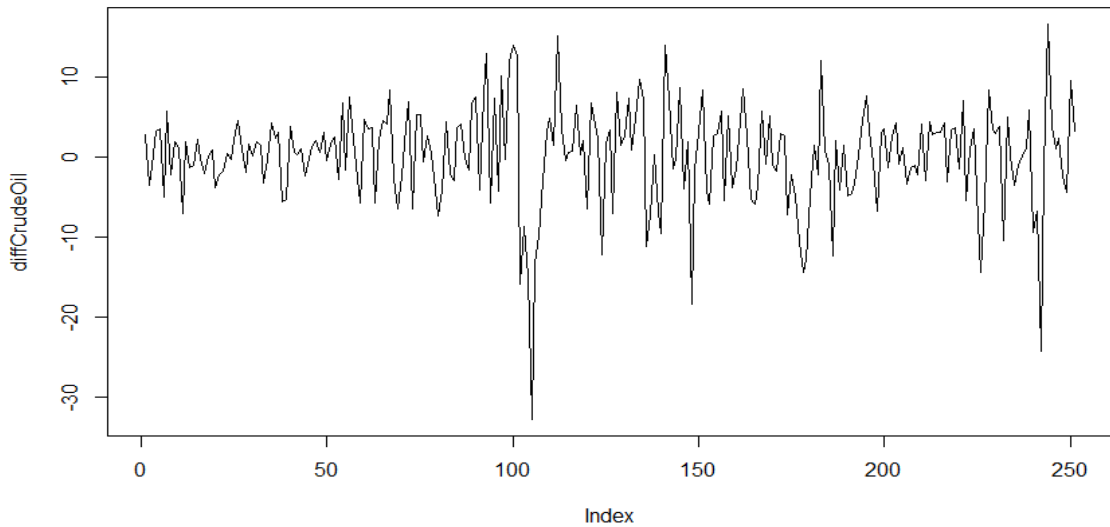
Γράφημα 12 Διάγραμμα μερικής αυτοσυσχέτισης της χρονοσειράς Ιανουάριος 2000 – Δεκέμβριος 2020



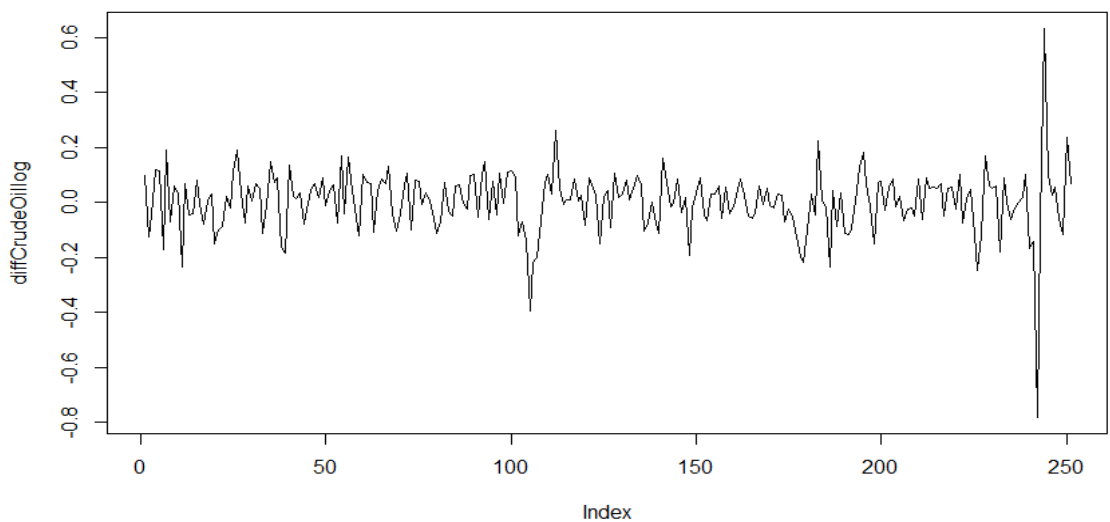
Η παραπάνω εικόνα επιβεβαιώνεται και μέσω του ελέγχου μοναδιαίας ρίζας Augmented Dickey – Fuller Test όπου με p-value 0.548 γίνεται δεκτή η μηδενική υπόθεση περί μη στασιμότητας της υπό μελέτη χρονοσειράς. Επαναλαμβάνοντας τη διαδικασία του ελέγχου Augmented Dickey – Fuller λαμβάνοντας πρώτης τάξης διαφορές στην τιμή του πετρελαίου παρατηρούμε πως η χρονοσειρά γίνεται στάσιμη σύμφωνα με τον έλεγχο με p-value 0,01 και βαθμό υστέρησης 6.

Παρατηρώντας το διάγραμμα της τιμής του αργού πετρελαίου με 1^{ης} τάξης διαφορές παρατηρούμε πως η χρονοσειρά έχει μετατραπεί σχεδόν σε λευκό θόρυβο με ακραίες τιμές (outliers) να παρατηρούνται μόνο στις περιόδους της παγκόσμιας οικονομικής κρίσης (2008) και της παγκόσμιας πανδημίας COVID-19 (2020). Λαμβάνοντας τους λογαρίθμους της τιμής του αργού πετρελαίου και διαφορίζοντας ξανά λαμβάνουμε ακόμα καλύτερη εικόνα σε όρους λευκού θορύβου μιας και σημαντικά ακραίες τιμές παρατηρούνται μόνο κατά την περίοδο της παγκόσμιας πανδημίας.

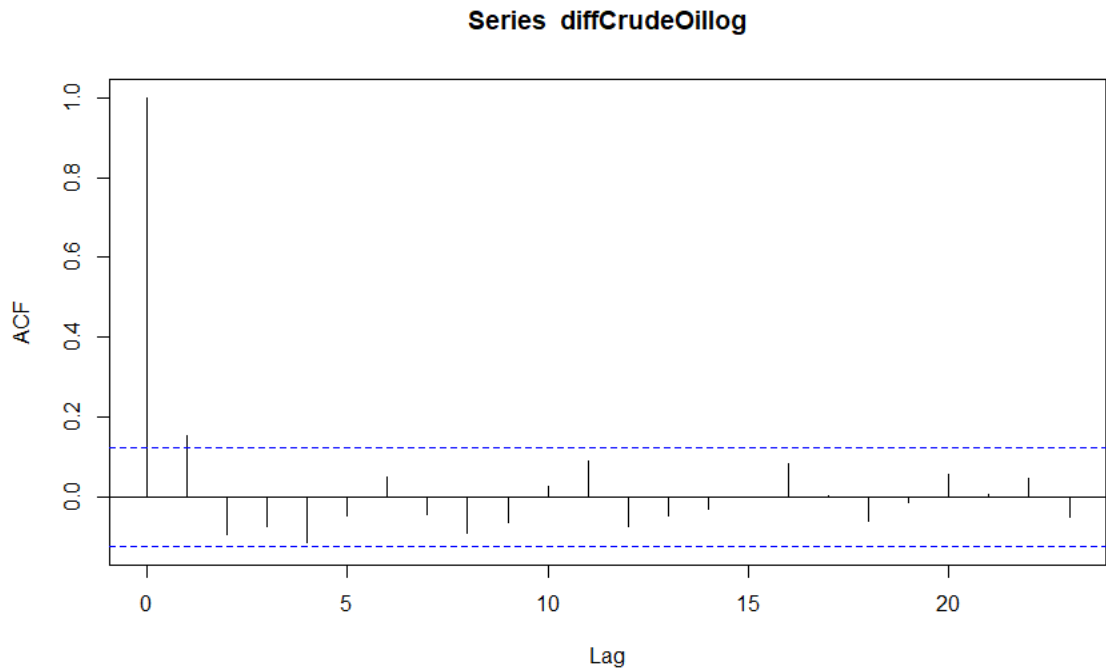
Γράφημα 13 Διάγραμμα της χρονοσειράς Ιανουάριος 2000 – Δεκέμβριος 2020 με 1ης τάξης διαφορές στις τιμές



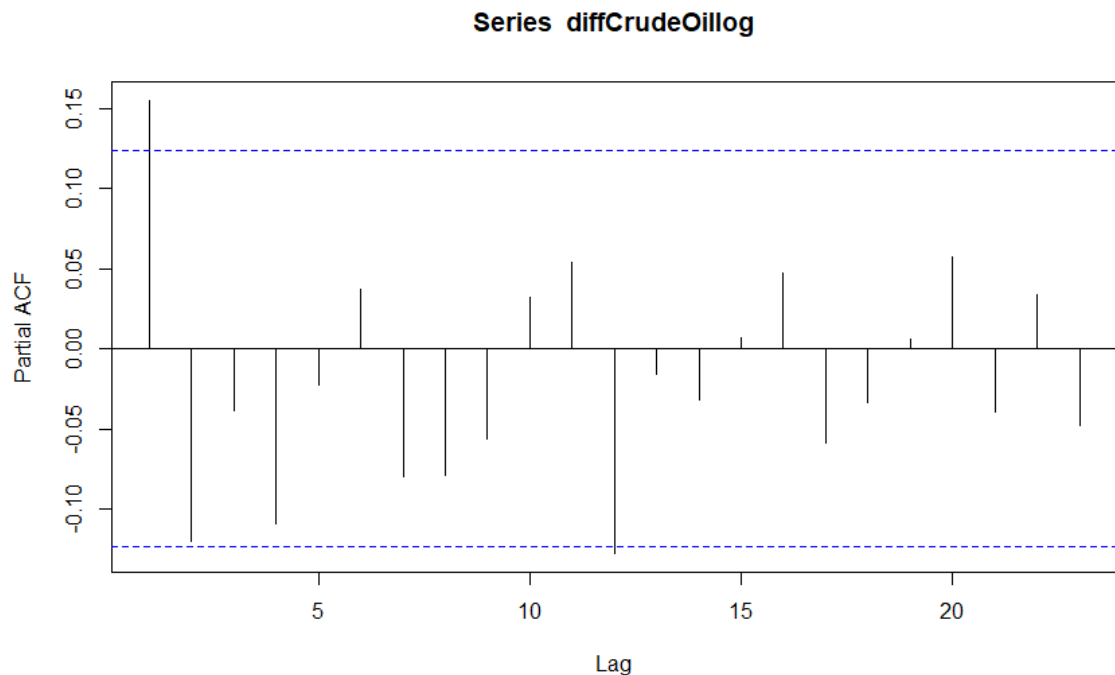
Γράφημα 14 Διάγραμμα της χρονοσειράς Ιανουάριος 2000 – Δεκέμβριος 2020 με λογαρίθμους και 1^{ης} τάξης διαφορές στις τιμές



Γράφημα 15 Διάγραμμα αυτοσυσχέτισης της χρονοσειράς Ιανουάριος 2000 – Δεκέμβριος 2020 με λογαρίθμους και 1^η τάξης διαφορές στις τιμές



Γράφημα 16 Διάγραμμα μερικής αυτοσυσχέτισης της χρονοσειράς Ιανουάριος 2000 – Δεκέμβριος 2020 με λογαρίθμους και 1^η τάξης διαφορές στις τιμές



Τα διαγράμματα αυτοσυσχέτισης και μερικής αυτοσυσχέτισης (Γράφημα 15 και Γράφημα 16) φαίνεται να υποδεικνύουν μία διαδικασία MA(1) και AR(1), ωστόσο στη δεύτερη περίπτωση φαίνεται να εξακολουθούν να υπάρχουν σχετικά ισχυρές και κοντά στα όρια εμπιστοσύνης μερικές αυτοσυσχετίσεις ακόμα και μετά από μία χρονική υστέρηση (lag). Ακόμα υπάρχει και

ένδειξη για εποχιακό όρο SAR(1) στη χρονική υστέρηση lag = 12. Για το λόγο αυτό θα εκτιμηθούν διάφορα μοντέλα ARIMA (p,1,q) στους λογαρίθμους των τιμών της χρονοσειράς Ιανουάριος 2000 – Δεκέμβριος 2020 και θα αξιολογηθούν με βάση τη στατιστική σημαντικότητα των εκτιμημένων συντελεστών, το κριτήριο AIC (Akaike Information Criterion), το κριτήριο BIC (Bayesian Information Criterion) και το μέσο απόλυτο ποσοστό σφάλματος (MAPE) στις προβλέψεις των μηνιαίων τιμών της περιόδου Ιανουάριος – Οκτώβριος 2021 καθώς και τη συμπεριφορά των καταλοίπων μέσα από τη διαγραμματική απεικόνιση και τον έλεγχο Ljung – Box για το συντελεστή αυτοσυσχέτισης τους.

Πίνακας 3 Κριτήρια αξιολόγησης των μοντέλων ARIMA στη χρονοσειρά Ιανουάριος 2000 – Δεκέμβριος 2020

ARIMA models	P-Value of coefficients	AIC	BIC	Ljung - Box Test (15 lags)
ARIMA (1,1,1)	0,22 (AR1) 0,03 (MA1)	-383,06	-372,48	0,6527
ARIMA (2,1,1)	0,00 for all coefficients	-385,44	-371,34	0,8533
ARIMA (1,1,2)	0,00 for all coefficients	-386,03	-371,93	0,9107
ARIMA (2,1,2)	0,02 (AR1), 0,92 (AR2) 0,06 (MA1), 0,39 (MA2)	-384,04	-366,41	0,9146
ARIMA (1,1,3)	0,00 for all coefficients except from 0,94 (MA3)	-384,04	-366,41	0,9131
ARIMA (2,1,3)	NaN Error	-382,06	-360,90	0,9161
ARIMA (3,1,1)	0,00 for all coefficients except from 0,53 (AR3)	-383,83	-366,19	0,8957
ARIMA (3,1,2)	0,39 (AR1), 0,62 (AR2), 0,67 (AR3) 0,55 (MA1), 0,3 (MA2)	-382,2	-361,05	0,9178
ARIMA (3,1,3)	0,92 (AR1), 0,37 (AR2), 0,99 (AR3) 0,83 (MA1), 0,18 (MA2), 0,76 (MA3)	-380,05	-355,37	0,9161
ARIMA (1,1,1),(1,00)[12]	0,04 (MA1), 0,29 (AR1), 0,11 (SAR1)	-383,64	-369,54	0,7771

ARIMA (1,1,2),(1,0,0)[12]	0,00 for all coefficients except from SAR1 (0.12)	-386,36	-368,73	0,9744
------------------------------	--	---------	---------	--------

Σύμφωνα με τα αποτελέσματα των ελέγχων Ljung – Box για ανεξαρτησία των καταλοίπων παρατηρούμε πως όλες οι p-values είναι μεγαλύτερες από 0,05 οπότε μπορούμε να υποθέσουμε πως τα δεδομένα είναι ανεξάρτητα κατανομημένα και δεν υπάρχει σειριακή αυτοσυσχέτιση σε κανένα από τα παραπάνω μοντέλα (Πίνακας 3). Επομένως επιβεβαιώνεται και η εικόνα για λευκού θορύβου κατάλοιπα όπως αυτή προκύπτει και από τα διαγράμματα των υπό αξιολόγηση μοντέλων (Παράρτημα 1). Συγκρίνοντας τις τιμές του κριτηρίου AIC για το βαθμό προσαρμογής στα δεδομένα της χρονοσειράς βλέπουμε πως τη μικρότερη τιμή του ελέγχου (AIC = -386,36) και άρα την καλύτερη προσαρμογή παρουσιάζει το seasonal ARIMA (1,1,2),(1,0,0)[12] ακολουθούμενο από το ARIMA (1,1,2) με AIC = -386,03. Και τα δύο αυτά μοντέλα έχουν στατιστικά σημαντικούς όλους τους συντελεστές τους, ωστόσο το ARIMA (1,1,2),(1,0,0)[12] παρουσιάζει μεγαλύτερη προβλεπτική ικανότητα καθώς το μέσο απόλυτο σφάλμα πρόβλεψης είναι το μικρότερο από τα υπόλοιπα μοντέλα. Συνυπολογίζοντας όλα τα ανωτέρω προκύπτει πως το καταλληλότερο μοντέλο για τα δεδομένα της χρονοσειράς της τιμής του αργού πετρελαίου είναι το ARIMA (1,1,2),(1,0,0)[12] με AIC = -386,36 και μέσο απόλυτο σφάλμα πρόβλεψης 10 βημάτων MAPE = 22,29%. Αξίζει να αναφέρουμε πως το μοντέλο αυτό παρουσίασε πολύ καλή ακρίβεια πρόβλεψης ενός βήματος (Ιανουάριος 2021) όπου προέβλεψε τιμή πετρελαίου στα 49,2\$ όταν η πραγματική ήταν 52,2 (5,78% σφάλμα). Η εξίσωση του υποδείγματος ύστερα από την εκτίμηση των συντελεστών του δίνεται από την εξίσωση:

$$\Delta y_t = 0,98\Delta y_{t-1} - 0,57\varepsilon_{t-1} - 0,24\varepsilon_{t-2} - 0,12\Delta y_{t-12}$$

Πίνακας 4 Εκτιμημένοι συντελεστές ARIMA (1, 1, 2)(1, 0, 0)[12]

Coefficients	Estimate	Std Error	P-value
AR1	0,732937	0,149416	0,00
MA1	-0,568870	0,153451	0,00
MA2	-0,238291	0,061398	0,00
SAR1	-0,123075	0,079937	0,12

Log likelihood = 198,18

AIC = -386,36

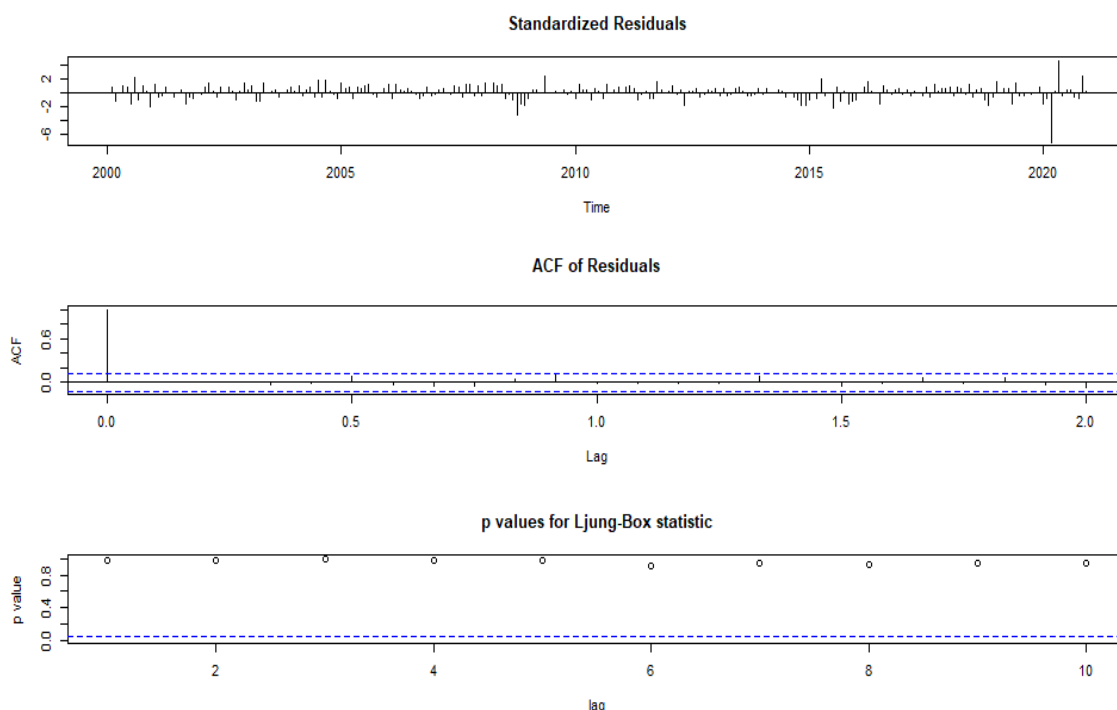
Πίνακας 5 Αποτελέσματα ελέγχου Ljung – Box στο ARIMA (1, 1, 2)(1, 0, 0)[12]

Ljung - Box Test

X - Squared	Lags	P - value
0,64551	5	0,9858
3,8537	10	0,9537
6,2928	15	0,9744
9,6909	20	0,9734
14,236	25	0,9575

Στη συνέχεια (Γράφημα 17) παρουσιάζονται τα γραφήματα των καταλοίπων του ARIMA (1,1,2),(1,0,0)[12]. Η εικόνα δείχνει πως τα σφάλματα είναι ανεξάρτητα κατανομημένα και πως δεν υπάρχει ισχυρή ένδειξη αυτοσυσχέτισης μεταξύ τους (λευκός θόρυβος).

Γράφημα 17 Διαγράμματα καταλοίπων μοντέλου ARIMA (1, 1, 2)(1, 0, 0)[12]



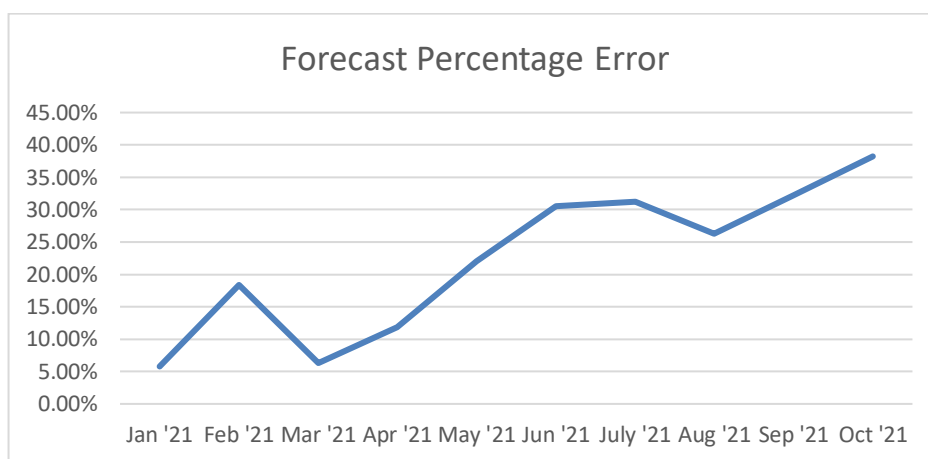
Συνοψίζοντας, μπορούμε να πούμε πως το ARIMA (1,1,2),(1,0,0)[12] παρουσιάζει καλή προσαρμογή στα δεδομένα της χρονοσειράς Ιανουάριος 2000 – Δεκέμβριος 2020 των λογαρίθμων των τιμών του αργού πετρελαίου και είμαστε σε θέση να πραγματοποιήσουμε πρόβλεψη στις μηνιαίες τιμές της περιόδου Ιανουάριος 2021 – Οκτώβριος 2021. Οι προβλέψεις που έδωσε το μοντέλο καθώς και τα σφάλματα μεταξύ προβλεπόμενης και παρατηρούμενης παρουσιάζονται στον παρακάτω πίνακα.

Πίνακας 6 Προβλεπόμενες τιμές χρυσού για τις μηνιαίες τιμές της περιόδου Ιανουάριος 2021 – Οκτώβριος 2021

Date	Actual Price	Forecasted Price	Percentage Error
Jan '21	52.2	49.2	5.78%
Feb '21	61.5	50.2	18.40%
Mar '21	59.16	55.5	6.27%
Apr '21	63.58	56.1	11.83%
May '21	66.32	51.7	22.07%
Jun '21	73.47	51.0	30.56%
July '21	73.95	50.8	31.24%
Aug '21	68.5	50.5	26.30%
Sep '21	75.03	50.9	32.22%
Oct '21	83.57	51.6	38.23%

Όπως παρατηρούμε, το μέσο απόλυτο σφάλμα πρόβλεψης δεν ξεπερνάει το 22,29%. Παρατηρούμε ωστόσο πως καθώς μεταφερόμαστε σε μελλοντικές παρατηρήσεις το σφάλμα πρόβλεψης αυξάνεται σημαντικά δίνοντάς μας έτσι μία σαφή εικόνα αβεβαιότητας στις συγκεκριμένες προβλέψεις. Τέλος αξίζει να σταθούμε στην καλή ακρίβεια που εμφάνισε η πρώτη πρόβλεψη κατά το μήνα Ιανουάριο με σφάλμα μόλις 5,78%.

Γράφημα 18 Σφάλμα Πρόβλεψης μοντέλου ARIMA (1, 1, 2)(1, 0, 0)[12]



4.4. Προσδιορισμός κατάλληλου μοντέλου LSTM στη χρονοσειρά Ιανουάριος 2000 – Δεκέμβριος 2020 και δημιουργία πρόβλεψης της τιμής του πετρελαίου κατά την περίοδο Ιανουάριος 2021 – Οκτώβριος 2020.

Στην παρούσα ενότητα παρουσιάζεται η υλοποίηση μοντέλων LSTM και η προσαρμογή τους στην χρονοσειρά της τιμής του αργού πετρελαίου την περίοδο Ιανουάριος 2000 – Δεκέμβριος 2020 για την πραγματοποίηση προβλέψεων την περίοδο Ιανουάριος 2021 – Οκτώβριος 2021. Πιο συγκεκριμένα, θα υλοποιηθούν διάφορες παραλλαγές μοντέλων δύο βασικών τεχνικών, της μονής (unidirectional) και της αμφίδρομης λειτουργίας (bidirectional) οι οποίες θα αξιολογηθούν ως προς την ακρίβεια της προσαρμογής τους στο train και test υποσύνολο των δεδομένων εξετάζοντας τους δείκτες MSE, MAE και RMSE. Κατά τη δημιουργία προβλέψεων 6 βημάτων, θα αξιολογηθεί η ακρίβεια πρόβλεψης με βάση το δείκτη MAPE και θα συγκριθεί με την αντίστοιχη τιμή του δείκτη του μοντέλου ARIMA της προηγούμενης ενότητας.

Τα μοντέλα LSTM αμφίδρομης λειτουργίας διατρέχουν τις εισόδους με δύο τρόπους, έναν από το παρελθόν στο μέλλον όπως και τα unidirectional και έναν από το μέλλον στο παρελθόν. Στη δεύτερη περίπτωση, είναι σε θέση να διατηρούν πληροφορίες από το μέλλον και συνδυάζοντάς τες με το παρελθόν να κατανοούν και να προσαρμόζονται καλύτερα στα σύνολα δεδομένων. Η χρήση των bidirectional LSTMs δεν μπορούμε να πούμε με βεβαιότητα πως προτείνεται στο

σύνολο των περιπτώσεων έναντι των απλών unidirectional LSTMs, ωστόσο κατά πλειοψηφία δίνουν καλύτερα αποτελέσματα.

Η υλοποίηση των μοντέλων LSTM και biLSTN θα γίνει με τη χρήση της rython και των βιβλιοθηκών KERAS και TensorFlow. Αρχικά, τα δεδομένα εισόδου (χρονοσειρά) θα πρέπει να κανονικοποιηθούν προκειμένου να βελτιστοποιηθεί η διαδικασία μάθησης του αλγόριθμου. Η διαδικασία αυτή είναι σημαντικό να πραγματοποιείται γιατί οι διαφορετικές κλίμακες στις τιμές των δεδομένων μπορεί να οδηγήσουν το δίκτυο στο να δώσει βάρη με άνισο τρόπο και έτσι να μπουνε σε προτεραιότητα μεταβλητές που δεν θα έπρεπε. Όσον αφορά στα δεδομένα της παρούσας εργασίας, η χρονοσειρά του πετρελαίου μετασχηματίζεται σε νέες τιμές με μέσο 0 και διακύμανση 1. (Lu Hou)

Μετά την κανονικοποίηση τα δεδομένα χωρίζονται σε train/validation υποσύνολα με την ποσότητα 60/40 αντίστοιχα και πλέον λαμβάνουν την τελική μορφή τους για να εισαχθούν στο LSTM. Τα μοντέλα LSTM είναι μία ακολουθιακή διαδικασία με συγκεκριμένη λογική υλοποίησης, επομένως και τα δεδομένα εισόδου θα πρέπει να έχουν συγκεκριμένο σχήμα. Τα δεδομένα εισόδου στο LSTM Keras είναι πάντα ένας τρισδιάστατος πίνακας διαστάσεων (w, x, y) όπου x είναι τα χρονικά βήματα (look back periods or time steps), y είναι οι μονάδες δεδομένων και w είναι το πλήθος (batch size) των ζευγαριών των προηγούμενων δύο στοιχείων.

Σε ένα δίκτυο LSTM τα δεδομένα χωρίζονται σε y και x όπου τα x χρησιμοποιούνται για να προβλέψουν τα y. Για παράδειγμα, στα δεδομένα της χρονοσειράς του αργού πετρελαίου χρησιμοποιούνται οι τιμές από τον Ιούλιο έως το Δεκέμβριο του 2020 για να προβλέψουν την τιμή του Ιανουαρίου 2021, οι τιμές από τον Αύγουστο 2020 έως τον Ιανουάριο 2021 για να προβλέψουν την τιμή του Φεβρουαρίου 2021 κλπ. Ο αριθμός των μηνών που χρησιμοποιούνται για να προβλεφθεί η τιμή ενός μήνα είναι τα Look Backs που αναφέραμε νωρίτερα όπου στο παράδειγμά μας και στην υλοποίηση της εργασίας ο αριθμός των Look Backs είναι ίσος με 6. Η τιμή αυτή προέκυψε ύστερα από πολλές επαναλήψεις με διαφορετικό αριθμό Look Backs (8, 10, 12, 24) μιας και ήταν αυτή που έδωσε τα καλύτερα αποτελέσματα σε όρους ακρίβειας.



Με παρόμοια λογική ορίσθηκαν και οι βασικές παράμετροι του μοντέλου LSTM οι οποίες παρουσιάζονται στη συνέχεια και θα παραμένουν σταθερές όσο επιχειρούνται δοκιμές στις υπόλοιπες παραμέτρους (epochs, batch size, no of neurons(units), no of LSTM Layers).

- LSTM Activation Function: tanh
- Recurrent Activation Function: sigmoid
- Loss Function: Mean Squared Error (MSE)
- Optimizer: Adam

Σε αντίθεση με τη συνάρτηση ενεργοποίησης tanh η οποία αφορά στο hidden state του κυττάρου, η σιγμοειδής συνάρτηση ενεργοποίησης (sigmoid recurrent activation function) αφορά στις πύλες εισόδου, εξόδου και απόρριψης (forget gate). Η χρήση των συναρτήσεων αυτών είναι κοινή στα LSTM δίκτυα και στην πλειοψηφία των περιπτώσεων μοντελοποιεί επιτυχώς τις χρονολογικές σειρές.

Η Mean Squared Error Loss Function αποτελεί μέρος του αλγορίθμου βελτιστοποίησης του LSTM δικτύου καθώς υπολογίζει επαναλαμβανόμενα το σφάλμα του κάθε μέρους έτσι ώστε να αναπροσαρμόζονται τα βάρη και να μειώνεται το σφάλμα σε επόμενες επαναλήψεις. Αποτελεί την default επιλογή σε regression προβλήματα και το μέσο τετραγωνικό σφάλμα υπολογίζεται ως ο μέσος όρος των τετραγωνικών διαφορών μεταξύ των προβλεπόμενων και πραγματικών τιμών. Το τετράγωνο στη συνάρτηση οδηγεί σε επιπλέον “τιμωρία” του μοντέλου για τιμές που απέχουν περισσότερο από τις προβλεπόμενες.

Ο Adam Optimizer είναι ένας αλγόριθμος βελτιστοποίησης και χρησιμοποιείται για την αναπροσαρμογή των συναπτικών βαρών του δικτύου. Διαφέρει από την κλασσική στοχαστική διαδικασία gradient descent όπου το learning rate παραμένει σταθερό και χρήση του ενδείκνυται για μη στάσιμες χρονολογικές σειρές και για δεδομένα με θόρυβο. Στην πλειοψηφία των περιπτώσεων συμπεριφέρεται καλύτερα από άλλους optimizers, όπως για παράδειγμα οι AdaGrad και RMSProp, και μαζί με τους προαναφερθέντες συντελεστές θα αποτελέσει σταθερή επιλογή για την εκπαίδευση του LSTM δικτύου της παρούσας εργασίας.

Για τη μοντελοποίηση της χρονοσειράς του αργού πετρελαίου και την πραγματοποίηση προβλέψεων 10 βημάτων (Ιανουάριος – Οκτώβριος 2021) θα εκπαιδευτούν δύο μοντέλα LSTM, ένα unidirectional και ένα bidirectional (biLSTM), με σταθερούς τους συντελεστές που αναφέρθηκαν προηγουμένως και πραγματοποιώντας δοκιμές – παραμετροποιήσεις στη δομή του δικτύου (πλήθος και είδος layers, πλήθος units per layer), στον αριθμό των εποχών (epochs) και στο batch size.

Η χρονοσειρά της τιμής του αργού πετρελαίου αποτελείται από 252 παρατηρήσεις και ύστερα από το διαχωρισμό 60/40 σε train/validation το μοντέλο λαμβάνει σαν είσοδο δύο υποσύνολα μεγέθους 151 και 101 παρατηρήσεων αντίστοιχα. Ο διαχωρισμός έγινε σε αυτό το ποσοστό καθώς το σχετικά μικρό μέγεθος της χρονοσειράς απαιτούσε ένα ικανού μεγέθους τεστ υποσύνολο για να μπορεί να γενικεύει τις προβλέψεις και να αποφευχθεί το overfitting.

Η βασική δομή του δικτύου περιλαμβάνει 3 layers, ένα εισόδου με 4 LSTM units, ένα dropout layer με ποσοστό απόρριψης 20% και ένα dense (fully connected layer) με μία έξοδο. Τέλος, πραγματοποιούνται δοκιμές με διαφορετικό αριθμό εποχών και batch size οι οποίες παρουσιάζονται στους επόμενους πίνακες.

Πίνακας 7 LSTM(4), Dropout(0.2), Dense(1) Results

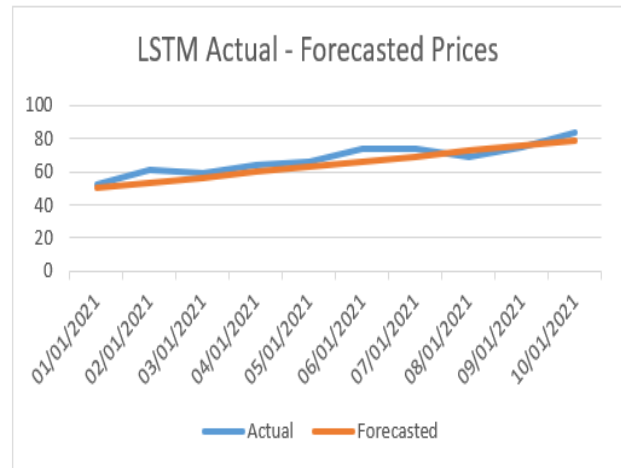
LSTM(4), Dropout(0.2), Dense(1)				
epochs	batchsize	train rmse	test rmse	mape
40	24	62.24%	63.61%	22.96%
	18	62.74%	64.06%	25.22%
	12	63.38%	64.58%	26.57%
	6	64.37%	65.40%	25.07%
	1	63.88%	64.18%	19.32%
60	24	63.50%	64.76%	26.12%
	18	63.33%	64.50%	27.06%
	12	63.96%	65.01%	26.48%
	6	63.79%	64.57%	26.15%
	1	64.55%	64.83%	16.87%
80	24	63.68%	64.81%	27.18%
	18	63.54%	64.57%	27.11%
	12	63.65%	64.49%	27.48%
	6	64.47%	64.96%	21.60%
	1	64.17%	64.41%	16.23%
100	24	64.03%	65.05%	26.83%
	18	63.38%	64.30%	28.75%
	12	63.84%	64.54%	27.17%
	6	64.79%	65.13%	19.60%
	1	66.34%	66.57%	6.14%

Όπως παρατηρούμε στον Πίνακα 7, οι τιμές των δεικτών train/test RMSE αυξάνονται καθώς μειώνεται το batch size και καθώς αυξάνονται οι εποχές χωρίς ωστόσο να παρατηρείται overfitting (train RMSE < test RMSE). Η καλύτερη προσαρμογή στα train και test υποσύνολα παρατηρείται στην εκπαίδευση του δικτύου σε 40 εποχές και με batch size 24 πετυχαίνοντας 62.24% train RMSE και 63.61 test RMSE. Ωστόσο καλύτερη προβλεπτική ικανότητα για την περίοδο Ιανουάριος – Οκτώβριος 2021 εμφάνισε το δίκτυο που εκπαιδεύτηκε με αριθμό εποχών 100 και batch size = 1 με μέσο απόλυτο ποσοστιαίο σφάλμα MAPE = 6.14%. Σε όλα τα πειράματα που πραγματοποιήθηκαν παρατηρήθηκε πως με batch size = 1 μειώνεται σημαντικά το σφάλμα πρόβλεψης.

Στον παρακάτω πίνακα παρουσιάζονται οι προβλέψεις για την περίοδο Ιανουάριος – Οκτώβριος 2021 όπως αυτές προέκυψαν από το δίκτυο με batch size = 1 και epochs = 100. Όπως παρατηρούμε, το δίκτυο προβλέπει με σφάλμα 3.64% την τιμή του Ιανουαρίου και στη συνέχεια με σχετικά μικρές αυξομειώσεις στο σφάλμα (εκτός από το Φεβρουάριο και τον Ιούνιο με 13.31% και 9.92% σφάλμα πρόβλεψης αντίστοιχα) ακολουθώντας με σχετική ακρίβεια την πορεία των πραγματικών τιμών προβλέποντας συνεχόμενη μηνιαία αύξηση στις τιμές του αργού πετρελαίου της τάξεως του 5% κατά μέσο όρο. Αυτή η συνεχόμενη αύξηση στις προβλέψεις μπορούμε να πούμε πως αποτελεί μειονέκτημα στην προβλεπτική ακρίβεια του μοντέλου, καθώς αδυνατεί να προβλέψει της πραγματικές μειώσεις στην τιμή του αργού πετρελαίου το Μάρτιο και τον Αύγουστο του 2021.

Πίνακας 8 LSTM(4), Dropout(0.2), Dense(1) Forecasted Prices & MAPE

LSTM			
Date	Actual	Forecasted	Absolute Error
01/01/2021	52.2	50.29956134	3.64%
02/01/2021	61.5	53.31498526	13.31%
03/01/2021	59.16	56.52961253	4.45%
04/01/2021	63.58	59.80448321	5.94%
05/01/2021	66.32	62.94640602	5.09%
06/01/2021	73.47	66.1835	9.92%
07/01/2021	73.95	69.4264616	6.12%
08/01/2021	68.5	72.5697096	5.94%
09/01/2021	75.03	75.56248267	0.71%
10/01/2021	83.57	78.33300796	6.27%



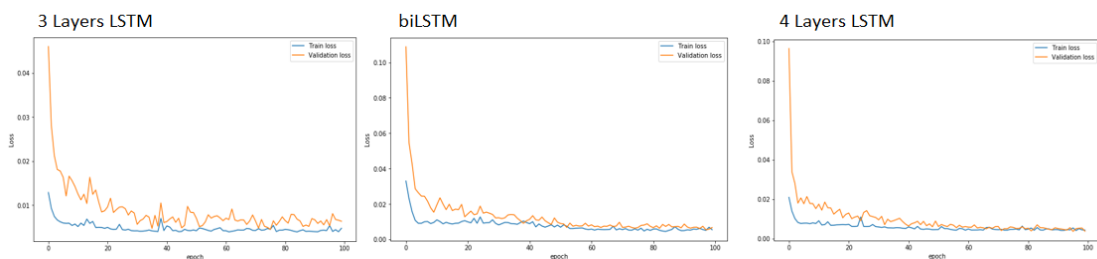
Στους επόμενους 2 πίνακες παρουσιάζονται τα αποτελέσματα ενός biLSTM παρόμοιας δομής με το αρχικό δίκτυο και ενός LSTM με την προσθήκη ενός κρυφού layer επιπλέον. Οι δοκιμές έγιναν με αριθμό εποχών 100 καθώς με αυτόν τον αριθμό επιτευχθήκαν τα καλύτερα αποτελέσματα σε όρους μέσου απόλυτου ποσοστού σφάλματος πρόβλεψης.

Πίνακας 9 biLSTM(4), Dropout(0.2), Dense(1) & 4 Layer LSTM Results

biLSTM(4), Dropout(0.2), Dense(1)					LSTM(4,4), Dropout(0.2), Dense(1)				
epochs	batchsize	train rmse	test rmse	mape	epochs	batchsize	train rmse	test rmse	mape
100	24	63.88%	65.84%	29.80%	100	24	63.20%	65.77%	30.31%
	18	64.50%	66.43%	27.75%		18	63.45%	66.24%	28.29%
	12	64.77%	66.57%	27.30%		12	61.84%	64.53%	32.07%
	6	64.20%	65.64%	26.21%		6	62.62%	65.09%	27.72%
	1	65.17%	65.36%	16.67%		1	65.66%	66.10%	14.41%

Τα αποτελέσματα δείχνουν πως και στα δύο μοντέλα επιτυγχάνεται βελτίωση στους δείκτες train/test RMSE και άρα καλύτερη προσαρμογή στα δεδομένα από το αρχικό δίκτυο, ωστόσο και τα δύο μοντέλα εμφάνισαν φτωχότερα αποτελέσματα σε όρους μέσου απόλυτου ποσοστού σφάλματος πρόβλεψης. Ακόμα παρατηρήθηκε overfitting στην προσπάθεια με batch size = 1 σε αρκετά μεγαλύτερο βαθμό από το αρχικό δίκτυο όπως δείχνουν τα επόμενα διαγράμματα.

Γράφημα 19 Train/Validation Losses of Implemented LSTMs



5. Συμπεράσματα

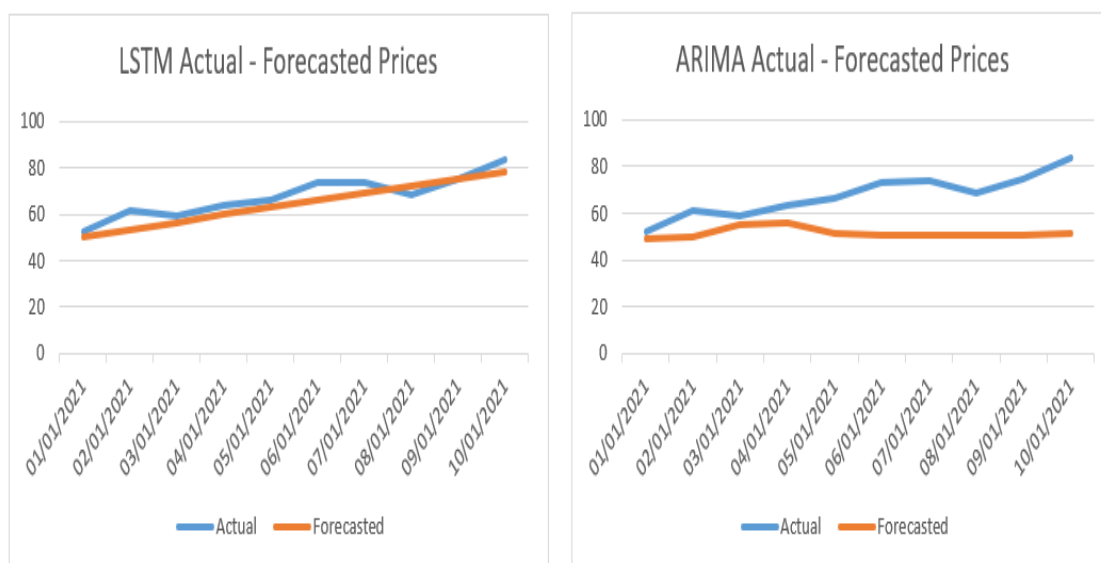
Με βάση την ανάλυση της συγκεκριμένης διπλωματικής και την μελέτη που πραγματοποιήθηκε στα τεκταινόμενα της αγοράς του πετρελαίου προκύπτει πως η τιμή του παρουσιάζει ισχυρές και στατιστικά σημαντικές συσχετίσεις με την τιμή του ασημιού (70%) και την τιμή του χρυσού (50%) με θετικό πρόσημο. Αναφορικά με τους υπόλοιπους δείκτες που παρουσιάστηκαν στην εργασία, παρατηρήθηκε ισχυρή και στατιστικά σημαντική συσχέτιση με το δεκαετές αμερικανικό ομόλογο (31%). Στην προκειμένη περίπτωση όμως δεν μπορούμε να κάνουμε λόγω για επιρροή των δεικτών αυτών στην τιμή του αργού πετρελαίου καθώς ενδέχεται να υπάρχει αντίστροφη αιτιότητα χρίζοντας επιπλέον έρευνας που ξεπερνάει το χρόνο εκπόνησης μιας διπλωματικής εργασίας.

Ακόμα, στη συγκεκριμένη διπλωματική αναπτύχθηκαν και αξιολογήθηκαν 11 αυτοπαλίνδρομα ολοκληρωμένα μοντέλα κινητού μέσου όρου με σκοπό την καλύτερη προσαρμογή στα δεδομένα των μηνιαίων τιμών του αργού πετρελαίου την περίοδο Ιανουάριος 2000 – Δεκέμβριος 2020 και την πραγματοποίηση προβλέψεων για τις μηνιαίες τιμές της περιόδου Ιανουάριος - Οκτώβριος 2021. Ο παραπάνω διαχωρισμός σε training/test υποσύνολα έγινε με τρόπο ώστε να συμπεριληφθεί στην ανάλυση και η επίδραση στην τιμή του αργού πετρελαίου που έχει η χρονική περίοδος Μάρτιος 2020 – Οκτώβριος 2021 όπου το φαινόμενο της πανδημίας του COVID-19 έχει παγκόσμιο αντίκτυπο μεταβάλλοντας την εικόνα της τιμής του πετρελαίου αλλά και πολλών χρηματοοικονομικών δεικτών. Τα αποτελέσματα έδειξαν πως το seasonal ARIMA (1, 1, 2)(1, 0, 0)[12] προσαρμόζεται καλύτερα στα δεδομένα της διπλωματικής, έχοντας στατιστικά σημαντικούς όλους τους εκτιμημένους συντελεστές, μέσο απόλυτο ποσοστό σφάλματος MAPE 22,29%, τυχαία κατανομημένα σφάλματα και τις πραγματικές τιμές του πετρελαίου να βρίσκονται εντός των ορίων της πρόβλεψης. Ωστόσο η αύξηση του μέσου απόλυτου ποσοστιαίου σφάλματος καθώς μεγαλώνει το διάστημα πρόβλεψης περιορίζει τη χρήση του μοντέλου σε πιο βραχυχρόνιες προβλέψεις, ενός έως και τριών μηνών, όπου το μέσο απόλυτο ποσοστιαίο σφάλμα είναι 5.78% και 10.15% αντίστοιχα.

Στο δεύτερο σκέλος της εμπειρικής ανάλυσης εκπαιδεύτηκαν 3 δίκτυα LSTM, δύο unidirectional με τρία και τέσσερα layers αντίστοιχα και ένα bidirectional με τρία layers. Διατηρώντας σταθερούς τους συντελεστές των δικτύων optimizer (Adam), loss function (MSE), activation function (tanh) και recurrent activation function (sigmoid), επιχειρήθηκαν πειράματα στους συντελεστές epochs και batch size. Τα αποτελέσματα έδειξαν πως το δίκτυο LSTM με τρία layer (1 εισόδου, 1 dropout(0.2), 1 fully connected (dense) με μία έξοδο) με 100 εποχές εκπαίδευσης και batch size = 1 προσαρμόστηκε καλύτερα στα δεδομένα και έδωσε προβλέψεις για την περίοδο Ιανουάριος – Οκτώβριος 2021 με μέσο απόλυτο ποσοστιαίο σφάλμα MAPE = 6,14%. Ωστόσο υπάρχουν ενδείξεις για οριακό overfitting που πιστεύουμε πως δικαιολογείται από τη φύση των δεδομένων καθώς στη μοντελοποίηση χρονοσειρών μέσω LSTM δικτύων είναι πιθανό το μοντέλο να δυσκολεύεται περισσότερο στο train παρά στο validation set. Η θέση αυτή τεκμηριώνεται από το γεγονός ότι στην τιμή του αργού πετρελαίου για παράδειγμα, όπου ο θόρυβος και ο αντίκτυπος πολιτικών και κοινωνικών παραγόντων επιφέρει αστάθεια και υψηλή διακύμανση, μία περίοδος (train set) μπορεί να είναι πολυπλοκότερη σε ανάλυση από μια άλλη περίοδο (validation set). Πράγματι η περίοδος επιλογής του train set (Ιανουάριος 2000 – Ιούλιος 2012) εμφανίζει μεγαλύτερο συντελεστή μεταβλητότητας από την περίοδο που αφορά στο validation set (Αύγουστος 2012 – Δεκέμβριος 2020) με συντελεστή μεταβλητότητας (CV) ίσο με 35,28% και 47,86% αντίστοιχα.

Λαμβάνοντας υπόψη τα παραπάνω συμπεράσματα, φαίνεται πως τα Τεχνητά Νευρωνικά δίκτυα και συγκεκριμένα τα δίκτυα LSTM υπερτερούν στη μοντελοποίηση χρονοσειρών όπως η τιμή του αργού πετρελαίου καθώς δύναται να δώσουν καλύτερες και περισσότερο μακροχρόνιες προβλέψεις από τα μοντέλα ARIMA τα οποία μπορούν να δώσουν προβλέψεις σημαντικά λιγότερων βημάτων με αποδεκτό σφάλμα. Το Διάγραμμα 18 δείχνει ακριβώς αυτή την εικόνα, όπου το δίκτυο LSTM δίνει πολύ καλή ακρίβεια στην πρόβλεψη της τιμής του αργού πετρελαίου τον Ιανουάριο του 2021 και συνεχίζει ακολουθώντας το trend των πραγματικών τιμών, ενώ το μοντέλο ARIMA περιορίζεται σε μία πολύ καλής και σχεδόν αντίστοιχης σε ακρίβεια πρόβλεψης της πρώτης τιμής, ωστόσο παρουσιάζει σημαντική αύξηση του σφάλματος στη συνέχεια.

Γράφημα 20 Actual – Forecasted Prices of LSTM & ARIMA Models



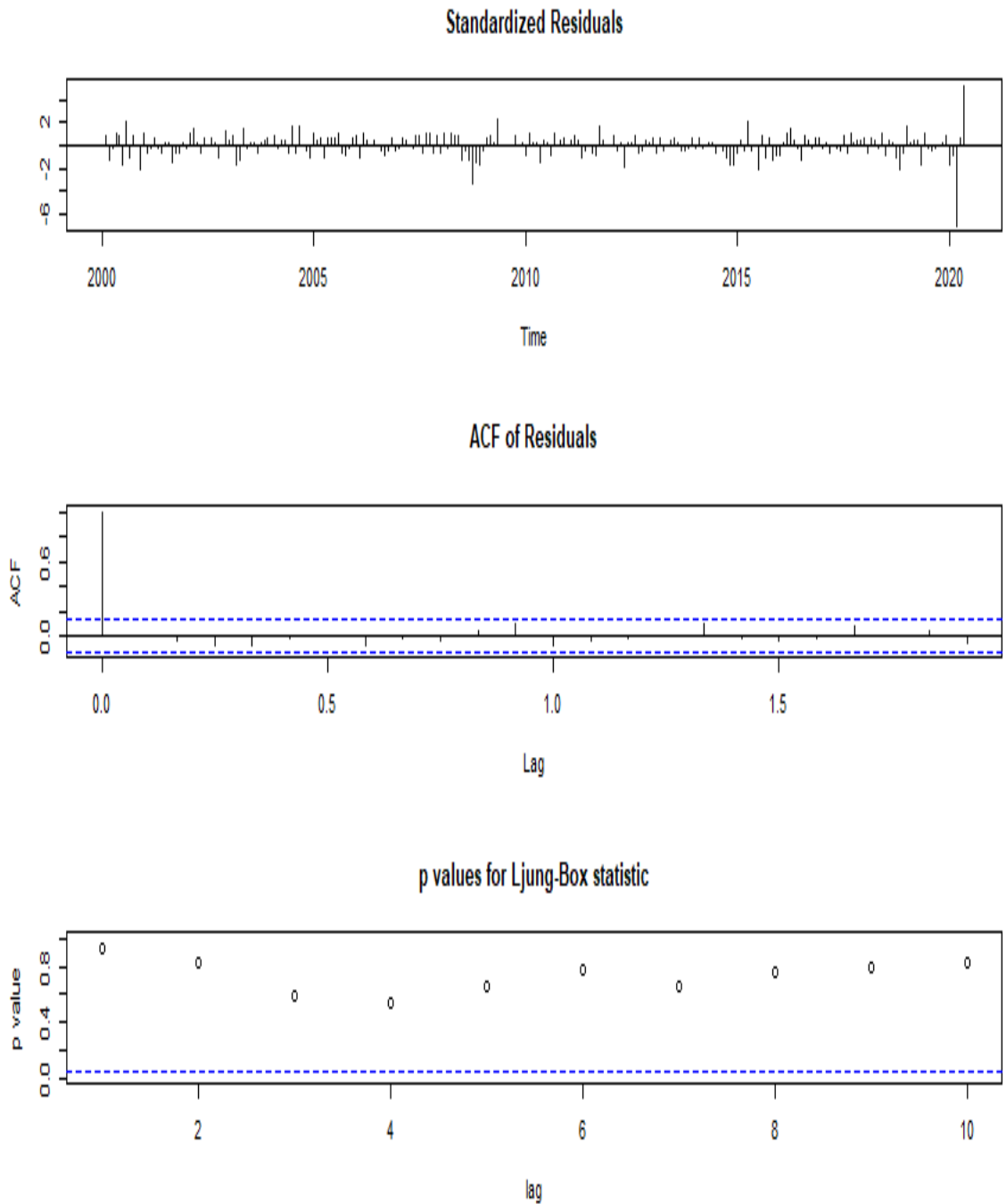
Βιβλιογραφία

- A.C. Akpanta, I. E. Okorie. «Application of Box-Jenkins Techniques in Modelling and Forecasting Nigeria Crude Oil Prices.» *International Journal of Statistics and Applications* 2014.
- Agency, European Environment. *Final energy consumption in Europe by mode of transport*. 10 2 2021. 10 2021.
- . *Final energy consumption in Europe by mode of transport*. 10 2 2021. 13 10 2021.
- Anh H. Vo¹, Trang Nguyen, Tuong Le. «Brent Oil Price Prediction Using Bi-LSTM Network.» *Tech Science Press* (2020).
- Bengio Y., Simard P., Frasconi P. «Learning long-term dependencies with gradient descent is difficult.» *IEEE* 1994: 157-166.
- Box, George και Gwilym Jenkins. *Time Series Analysis: Forecasting and Control*. San Francisco, 1970.
- Dickey, D. A. και W. A. Fuller. «Distribution of the Estimators for Autoregressive Time Series with a Unit Root.» *Journal of the American Statistical Association*, 1979. 427-431.
- Eurostat. «Oil and petroleum products - a statistical overview.» 2021.
- Hochreiter, S. και J. Schmidhuber. «Long short-term memory. *Neural Comput.*» 1997. 1735–1780.
- Imtiaz Arif, Lubna Khan, Khalid Mehmood Iraqi. «Relationship between Oil Price and White Precious Metals Return: New Evidence from Quantile-on Quantile Regression.» *Pakistan Journal of Commerce and Social Sciences* 2019.
- J.Jaya Selvi, R.Kaviya Shree, J.Krishnan. «Forecasting Crude Oil Price Using ARIMA Models.» *International Journal of Advance Research in Science and Engineering* March 2018.
- Jessin Shah P A, Dr. G Kiruthiga. «Crude Oil Price Forecasting using ARIMA model.» *International Research Journal of Engineering and Technology (IRJET)* March 2020.
- Li Shu-rong, Ge Yu-lei. «Crude Oil Price Prediction Based on a Dynamic Correcting Support Vector Regression Machine.» *Hindawi Publishing Corporation Abstract and Applied Analysis* (2013): 7.
- Lu Hou, Jinhua Zhu, James T. Kwok, Fei Gao, Tao Qin, Tie-Yan Liu. «Normalization Helps Training of Quantized LSTM.» *University of Science and Technology of China, Hefei, China* χ.χ.
- Lubna A Gabralla, Ajith Abraham. «Computational Modeling of Crude Oil Price Forecasting: A Review of Two Decades of Research.» *International Journal of Computer Information Systems and Industrial Management Applications* 24 11 2014.
- Ng, Kay. *yahoo!finance*. 28 5 2020. 14 10 2021.

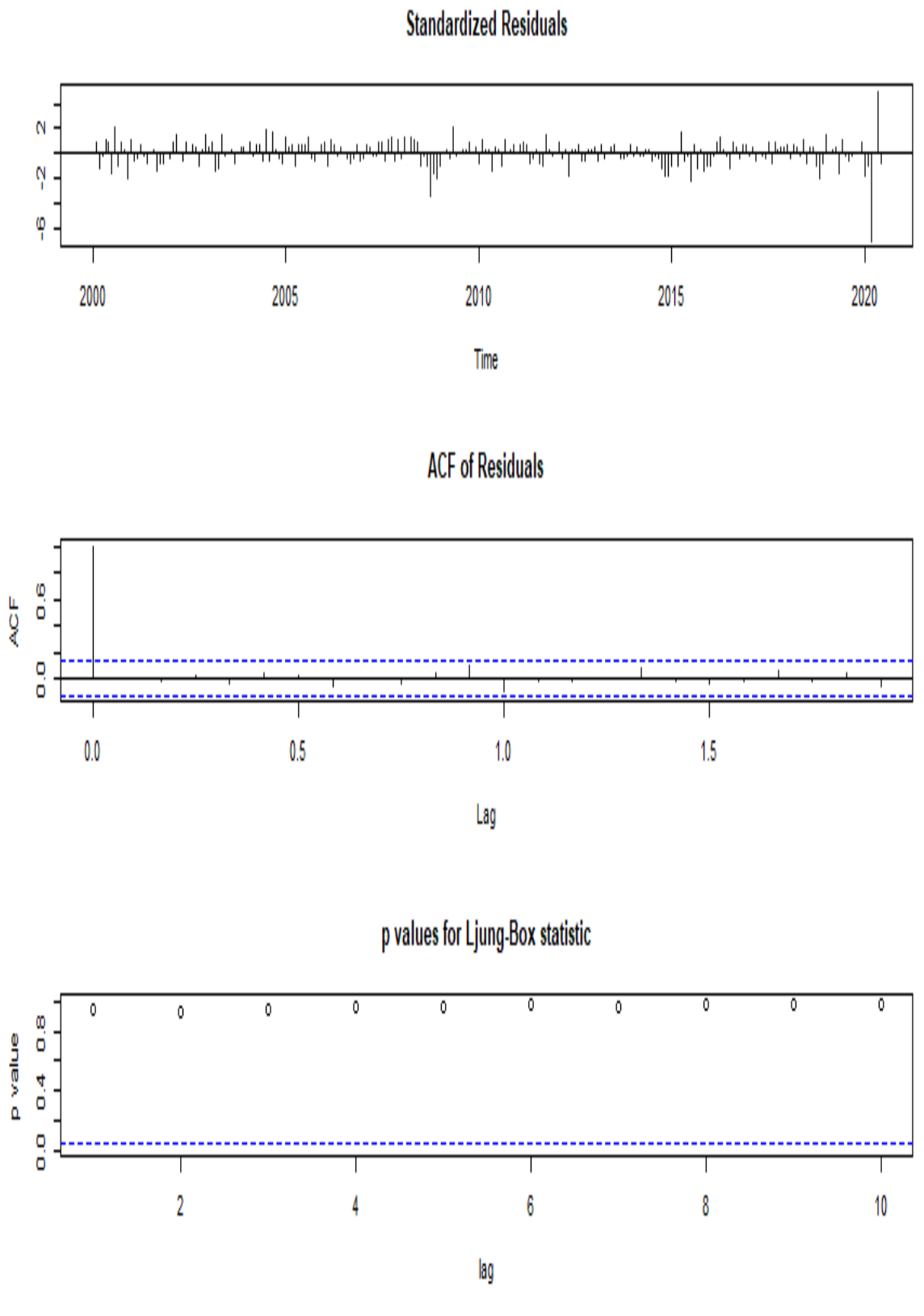
- Nidhi Moitra, Priya Raj, Sanidhya Saxena, Rohit Kumar. «Crude Oil Prediction Using LSTM.» *International Journal of Innovative Science and Research Technology* (2020).
- Olah, C. Understanding LSTM Networks. <http://colah.github.io/posts/2015-08-Understanding-LSTMs>. χ.χ.
- Shawn, Meera. «The Correlation between Gold and Oil.» *Market Realist* (2020).
- Sonnichsen, N. «Leading oil demanding sectors in the OECD 2019.» 2021. 10 2021.
- Wikipedia. *Neural Networks*. χ.χ.
- . *Πετρέλαιο*. χ.χ. 10 2021.
- Xuan Hien Le, Sungho Jung, Giha Lee. «Application of Long Short Term Memory (LSTM) Neural Network for Flood Forecasting.» *Water* 7 2019.
- . «Application of Long-Short Term Memory (LSTM) Neural Network for Flood Forecasting.» *Water* 7 2019: 7-8.
- Yan, S. «<https://medium.com/mlreview/understanding-lstm-and-its-diagrams-37e2f46f1714>.» χ.χ. *Understanding LSTM and Its Diagrams*.
- Yingrui Zhou, Taiyong Li, Jiayi Shi, Zijie Qian. «A CEEMDAN and XGBOOST-Based Approach to Forecast Crude Oil Prices.» (2019).
- Zhang, G. Peter. *Time series forecasting using a hybrid ARIMA and neural network model*. Atlanta USA: Georgia State University, 2001.
- Δ. Καρώνης, Ε. Λόης, Φ. Ζαννίκος. *Τεχνολογία Πετρελαίου και Φυσικού Αερίου*. Αθήνα: ΕΜΠ, 2011.

Παράρτημα 1. Διαγράμματα καταλοίπων των υπό αξιολόγηση μοντέλων ARIMA

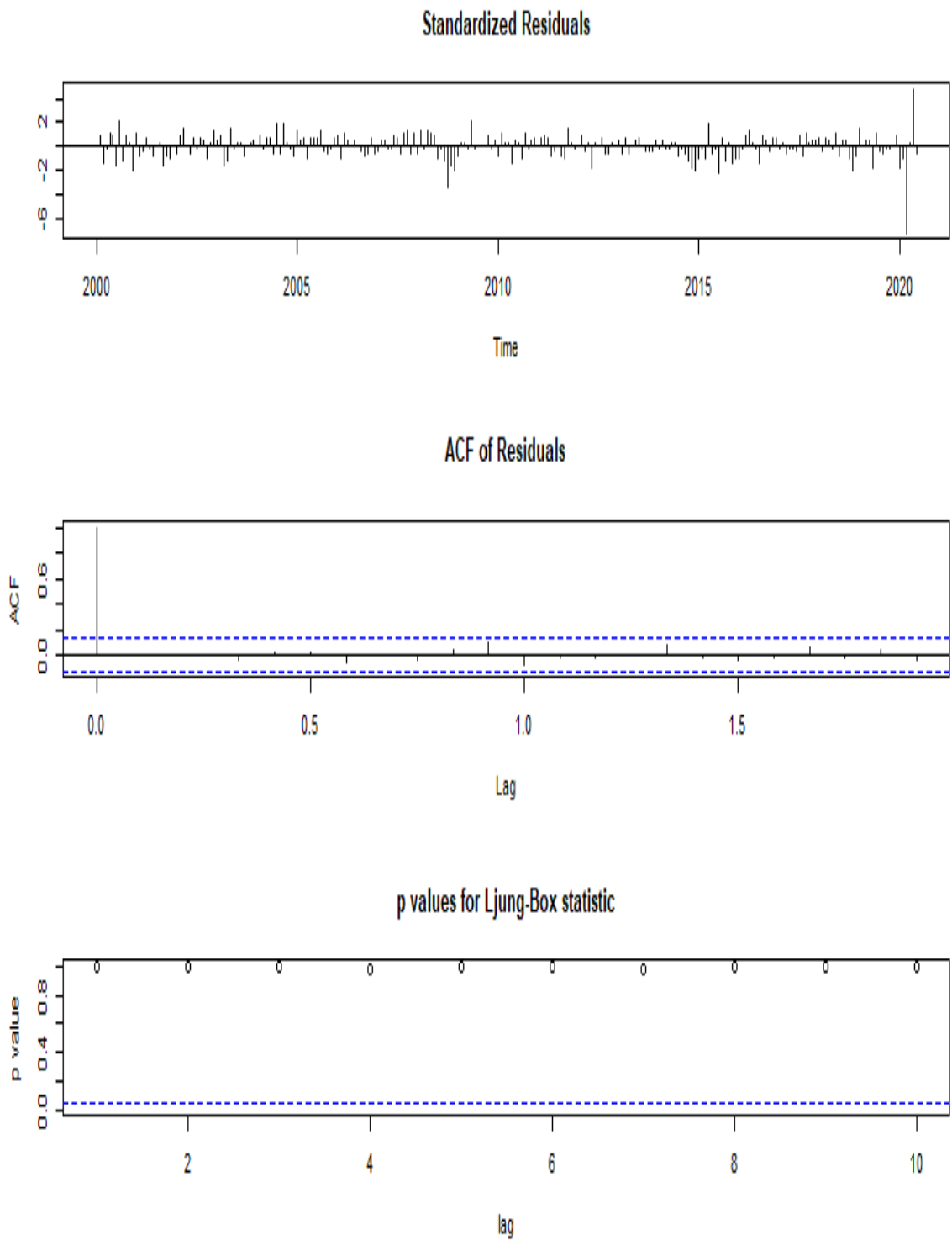
Γράφημα 21 Διαγράμματα καταλοίπων μοντέλου ARIMA (1,1,1)



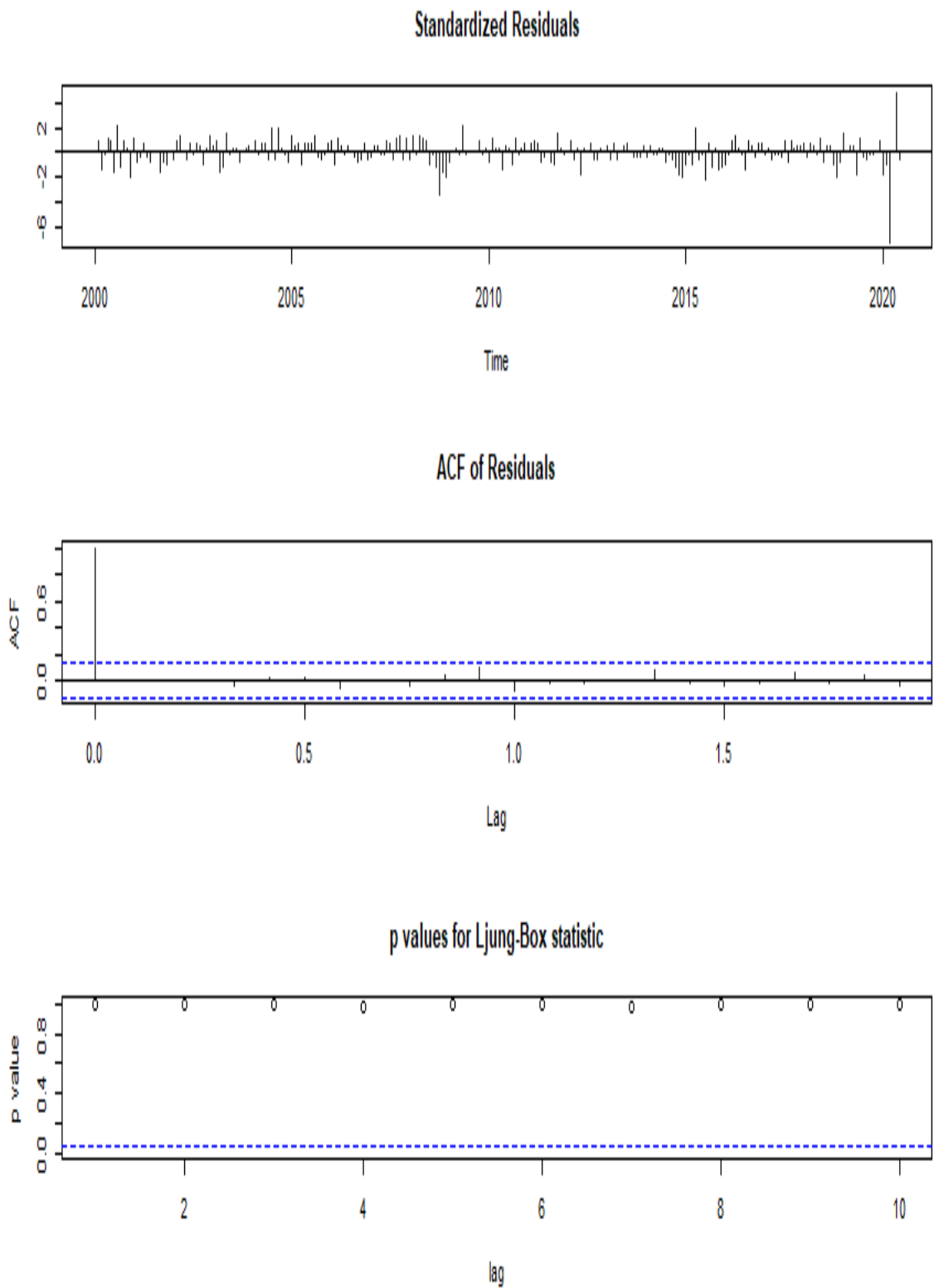
Γράφημα 22 Διαγράμματα καταλοίπων μοντέλου ARIMA (2,1,1)



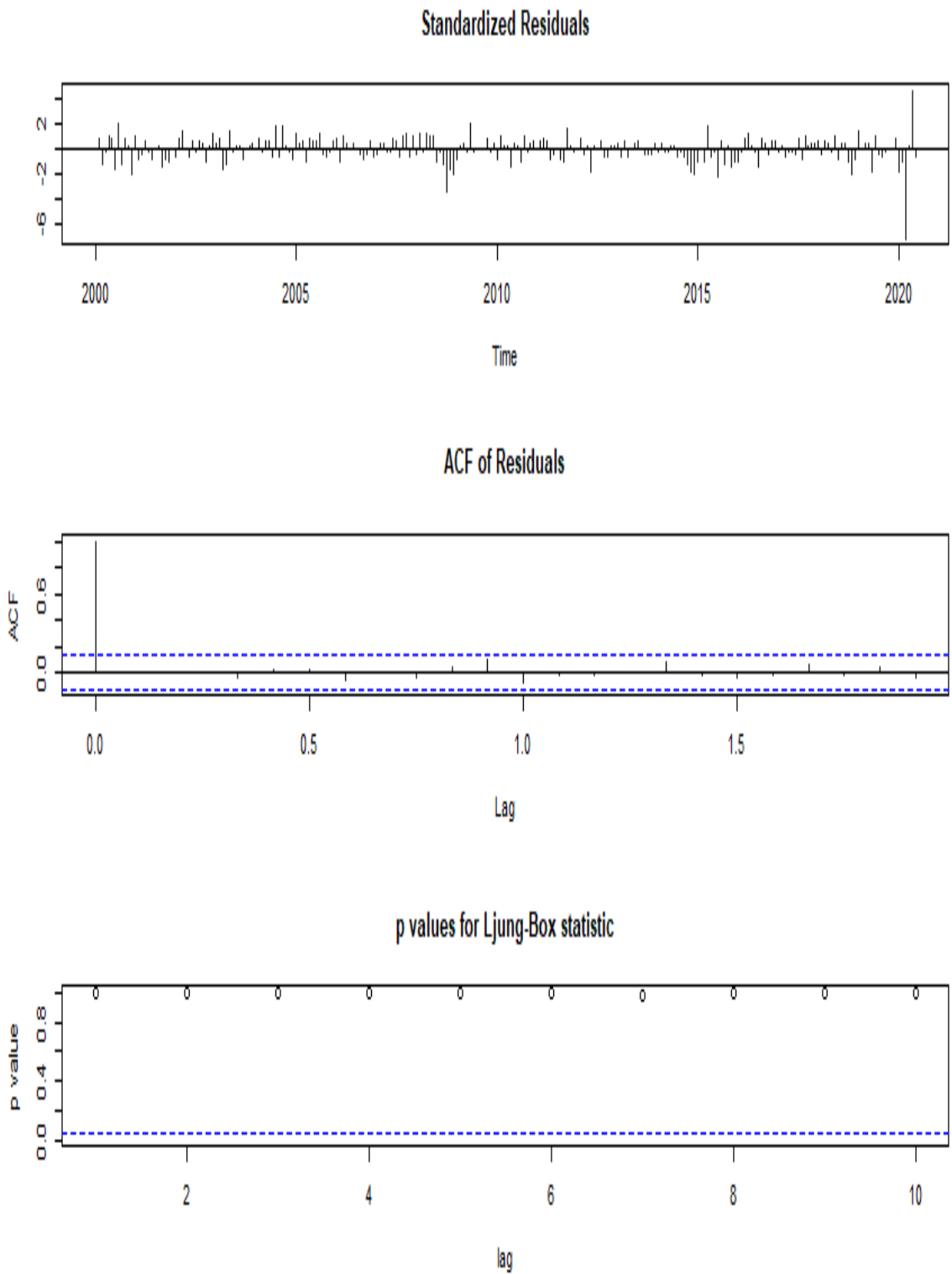
Γράφημα 23 Διαγράμματα καταλοίπων μοντέλου ARIMA (1,1,2)



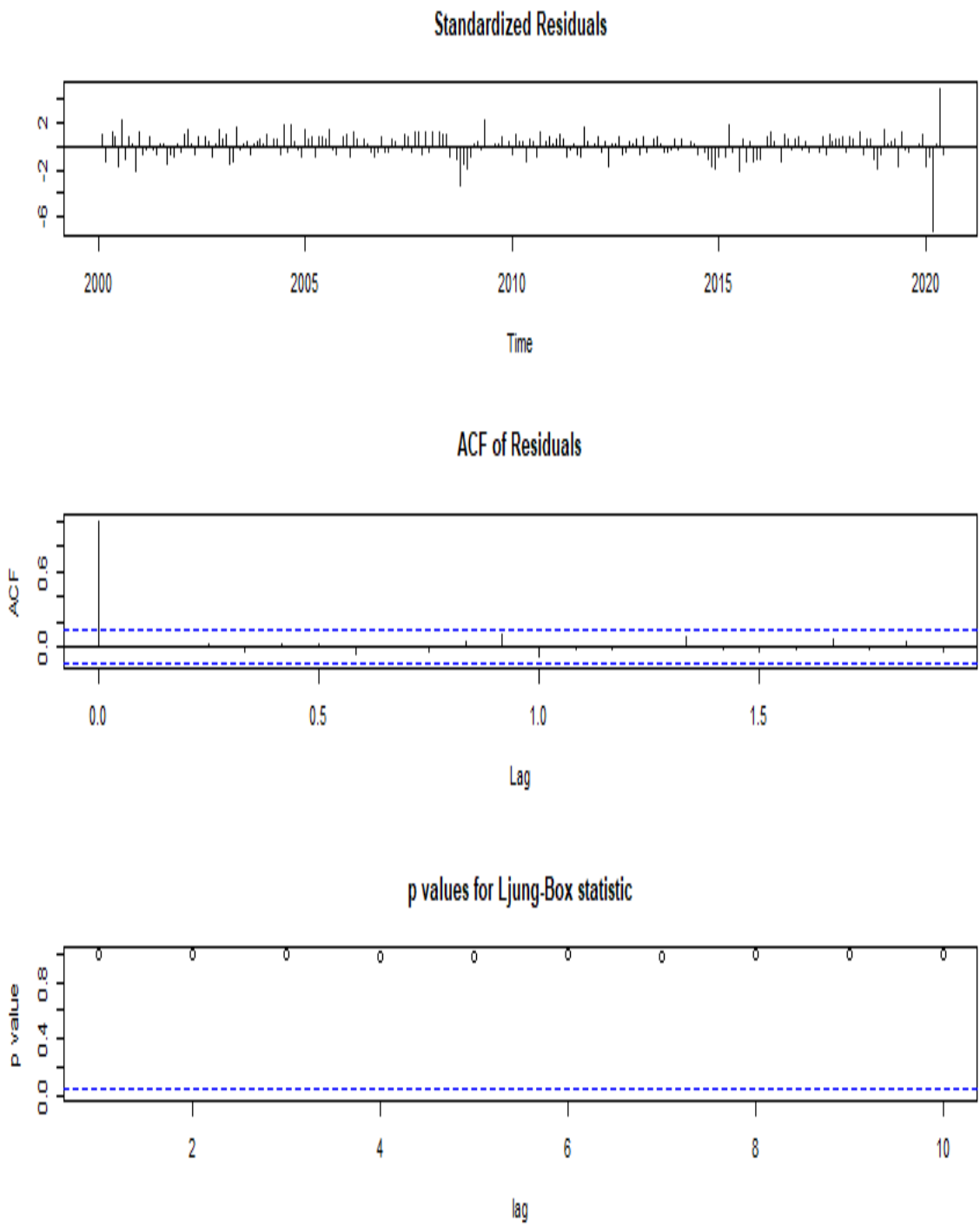
Γράφημα 24 Διαγράμματα καταλοίπων μοντέλου ARIMA (1,1,3)



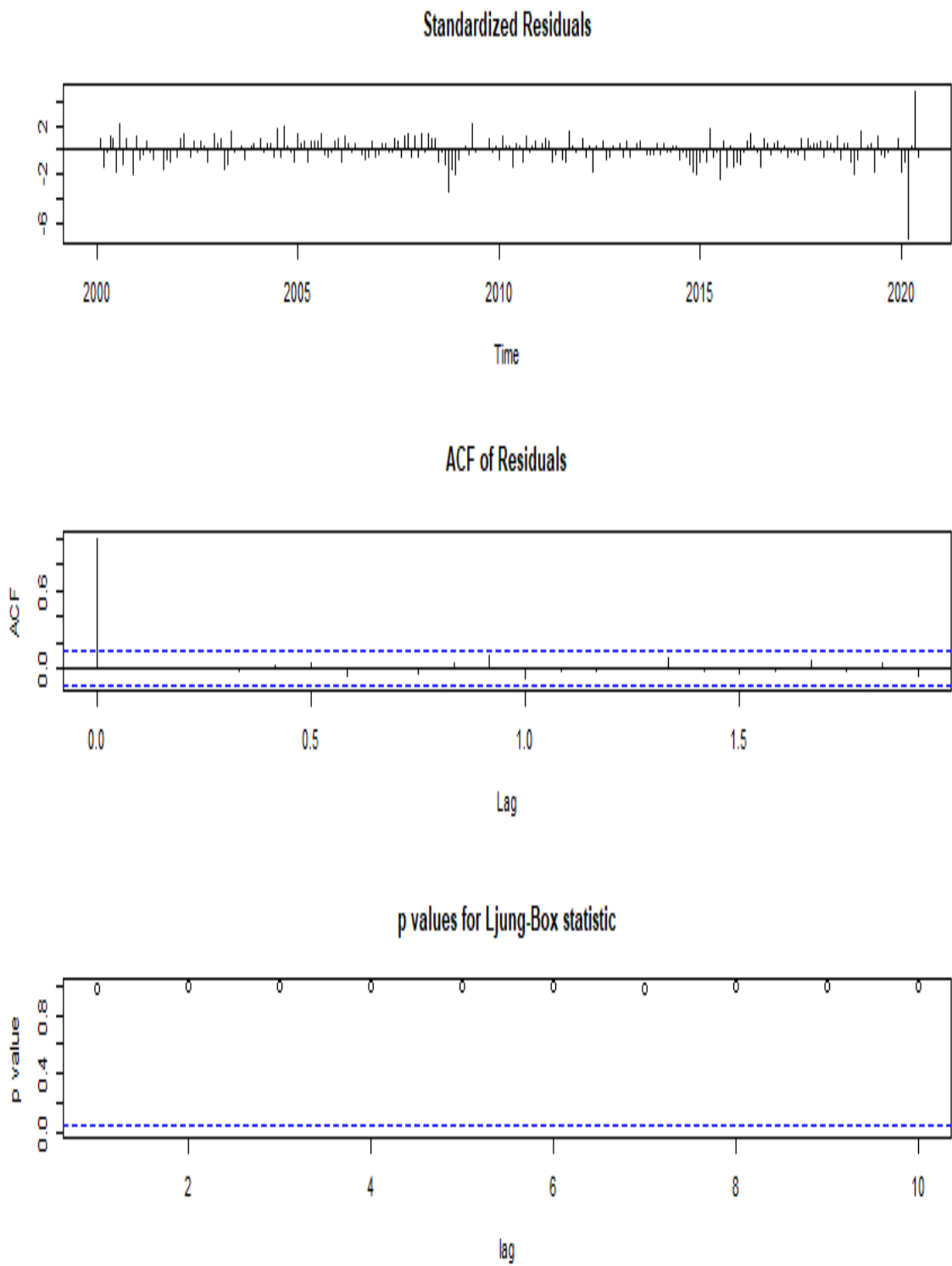
Γράφημα 25 Διαγράμματα καταλοίπων μοντέλου ARIMA (2,1,3)



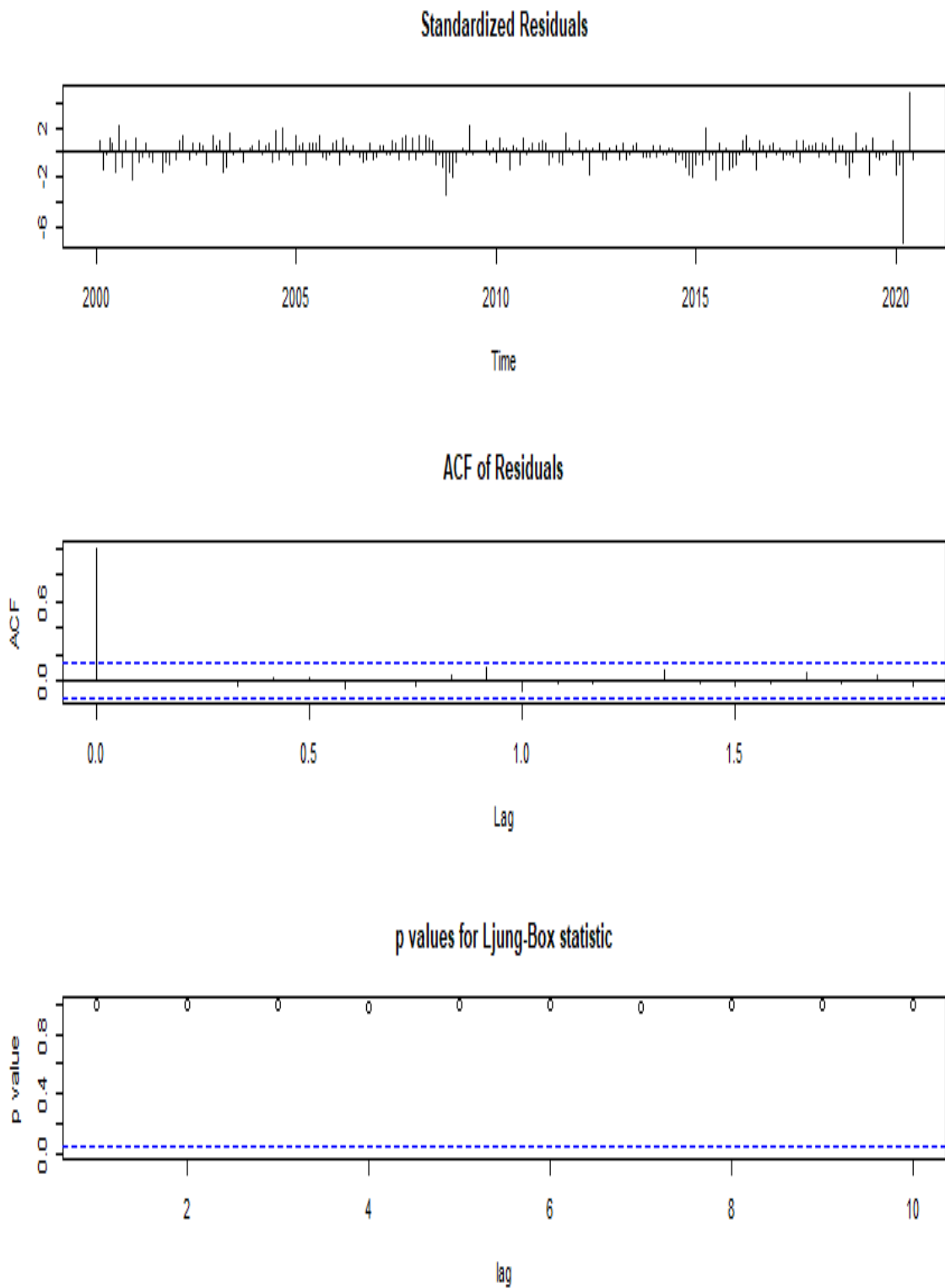
Γράφημα 26 Διαγράμματα καταλοίπων μοντέλου ARIMA (3,1,1)



Γράφημα 27 Διαγράμματα καταλοίπων μοντέλου ARIMA (3,1,2)



Γράφημα 28 Διαγράμματα καταλοίπων μοντέλου ARIMA (3,1,3)



Γράφημα 29 Διαγράμματα καταλοίπων μοντέλου ARIMA (1,1,1)(1,0,0)[12]

