



Τίτλος Εργασίας: Ανάλυση Τιμών Υγρών Καυσίμων

Ονοματεπώνυμο: Κώστας Μπότσι

Αρ. Μητρώου: 6180068

Τύπος απασχόλησης: Εξάμηνο μερικής απασχόλησης

Φορέας πρακτικής άσκησης: iMEdD

Μήνας συγγραφής: Σεπτέμβριος 2022

Επιβλέπουσα: Αλεξάνδρα Λιβάδα

Περιεχόμενα

Κεφάλαιο 1: Εισαγωγή

Κεφάλαιο 2: Μεθοδολογία και Πηγές Δεδομένων

Κεφάλαιο 3: Περιγραφική Στατιστική

Κεφάλαιο 4: Έλεγχοι Υποθέσεων

Κεφάλαιο 5: Προβλεπτικά και Ερμηνευτικά Μοντέλα

Κεφάλαιο 6: Συμπεράσματα

ΚΕΦΑΛΑΙΟ 1. ΕΙΣΑΓΩΓΗ

Είναι γνωστό ότι στις μέρες μας οι επιπτώσεις του πολέμου μεταξύ Ρωσίας και Ουκρανίας έχουν δημιουργήσει ραγδαίες αλλαγές, ειδικά στον ενεργειακό τομέα. Αυτό που θα διερευνηθεί με την παρακάτω ανάλυση είναι η μεταβολή που έχει υποστεί η τιμή της βενζίνης, ένα αγαθό πολύτιμο από το οποίο εξαρτώνται πολλά άλλα αγαθά. Για παράδειγμα, το κόστος μεταφοράς αυξάνεται, με αποτέλεσμα το κόστος για τα προϊόντα πρώτης ανάγκης να αυξάνεται και αυτό. Ομοίως, αυξάνεται το κόστος μεταφοράς με αμάξι, με αποτέλεσμα να δημιουργείται μεγάλο πρόβλημα στο κόστος διαβίωσης του πολίτη.

Σύμφωνα με διάφορες πηγές (OT.gr Newsroom, 2022)¹, στη σημερινή εποχή διανύουμε μία από τις μεγαλύτερες ενεργειακές κρίσεις στην ιστορία της Ευρώπης, με ραγδαίες μεταβολές στις τιμές του πετρελαίου. Σκοπός της παρακάτω εργασίας είναι να μετρηθεί η μεταβολή στην τιμή βενζίνης και να αναλυθεί.

ΚΕΦΑΛΑΙΟ 2. ΜΕΘΟΔΟΛΟΓΙΑ ΚΑΙ ΠΗΓΕΣ ΔΕΔΟΜΕΝΩΝ

Αρχικά να αναφερθεί ότι στο iMedD Lab έχει γίνει [μια περιγραφική ανάλυση](#) (Κέλλυ Κικί και Κωνσταντίνος Μποτσι, 2022)² που αποτυπώνει με γραφήματα την άνοδο των τιμών υγρών καυσίμων. Σε αυτήν την εργασία θα γίνει μια περαιτέρω ανάλυση για μοντελοποίηση της διακύμανσης της τιμής των υγρών καυσίμων και την πρόβλεψη της τιμής της αμόλυβδης βενζίνης 100 οκτανίων με χρήση μακροοικονομικών μεταβλητών, διότι έχει αποδειχθεί από προηγούμενες έρευνες (Boitumelo Nnoi Yolanda Sekati, 2022)³ ότι τα μακροοικονομικά μεγέθη είναι σημαντικοί παράγοντες για την πρόβλεψη της τιμής των καυσίμων.

Η συλλογή των δεδομένων έγινε από το Παρατηρητήριο Τιμών Υγρών Καυσίμων του Υπουργείου Ανάπτυξης (<http://www.fuelprices.gr/>). Στην εν λόγω ιστοσελίδα, υπάρχουν σε μορφή .pdf τα δεδομένα που αφορούν τη μέση τιμή των υγρών καυσίμων στη χώρα, σε ημερήσια και εβδομαδιαία βάση.

Τα εργαλεία και τα πακέτα που χρησιμοποιήθηκαν ήταν τα εξής:

- Η συλλογή των δεδομένων και η ανάλυσή τους έγινε με τη χρήση της γλώσσας προγραμματισμού [Python](#). Συγκεκριμένα, για τη συλλογή των δεδομένων χρησιμοποιήθηκε η βιβλιοθήκη [Requests](#) για να προσπελάσουμε προγραμματιστικά την ιστοσελίδα, η βιβλιοθήκη BeautifulSoup για να «διαβάσουμε» προγραμματιστικά τον πηγαίο html κώδικα της σελίδας και για

¹ OT.gr Newsroom 06/09/2022 Ενεργειακή κρίση. Πηγή in.gr

² Κέλλυ Κικί, Κωνσταντίνος Μποτσι 13/07/22. Η τιμή της βενζίνης τα τελευταία 5 χρόνια, πηγή lab.imedd.org

³ Boitumelo Nnoi Yolanda sekati, Johannes Tshepiso Tsoku, Lebosa Daniel Metsileg 13 jul 2022, Modeling the oil price volatility and macroeconomic variables in south Africa using the symmetric and asymmetric Garch models

να ανακτήσουμε προγραμματιστικά κάθε αναρτημένο δελτίο τιμών υγρών καυσίμων (αρχεία μορφής .pdf). Ανακτήθηκαν συνολικά περίπου 1.900 ημερήσια δελτία τιμών καυσίμων στο σύνολο της Επικράτειας.

- Η ανάγνωση των .pdf αρχείων έγινε προγραμματιστικά με τη χρήση της βιβλιοθήκης [Apache Tika](#) και ο καθαρισμός των δεδομένων έγινε με τη χρήση της βιβλιοθήκης [Pandas](#) και με τη χρήση «Regular Expressions» σε Python.
- Η ανάλυση των δεδομένων έγινε με τη με χρήση των βιβλιοθηκών Pandas, [Numpy](#) (για αριθμητικές πράξεις, π.χ. log), [statsmodels.formula.api](#) (για στατιστικά μοντέλα), [sklearn.model_selection](#) (για το «train, test set split» και το «cross-validation»), [statsmodels.tsa.stattools](#) (για ελέγχους υποθέσεων). Για στατιστικές μετρήσεις, χρησιμοποιήθηκε το πακέτο `adfuller,sklearn.metrics`. Στην προκειμένη περίπτωση, χρησιμοποιήθηκε το `mean_squared_error`. Η οπτικοποίηση των δεδομένων έγινε με τις βιβλιοθήκες [Matplotlib](#) και [Seaborn](#).

Τα δεδομένα που συγκροτήθηκαν σε dataset και στα οποία βασίζεται η ανάλυση διατίθενται ανοιχτά από το iMedD Lab και είναι δημοσιευμένα εδώ: <https://lab.imedd.org/ta-dedomena-gia-tis-times-ygron-kafsimon-stin-ellada/> (Κέλλυ Κική, 2022)

ΚΕΦΑΛΑΙΟ 3. ΠΕΡΙΓΡΑΦΙΚΗ ΣΤΑΤΙΣΤΙΚΗ

Τα δεδομένα που έχουν συλλεχθεί καλύπτουν τη χρονική περίοδο από 2017-03-14 μέχρι 2022-06-15. Πρωτογενώς υπάρχουν 7 μεταβλητές και δημιουργήθηκαν και άλλες 3: η ημερομηνία (date) διασπάστηκε σε μήνα (month), year (έτος), day (ημέρα) και weekday (ημέρα βδομάδας). Οι μεταβλητές παρουσιάζονται παρακάτω.

Ποσοτικές

Χρόνος αναφοράς

Συνεχείς

Diesel κίνησης: υγρό καύσιμο

date: χρόνος/μήνας/ήμερα

Diesel θέρμανσης: υγρό καύσιμο

year: χρόνος

Super: υγρό καύσιμο

month: μήνας

Αμόλυβδη 100: υγρό καύσιμο

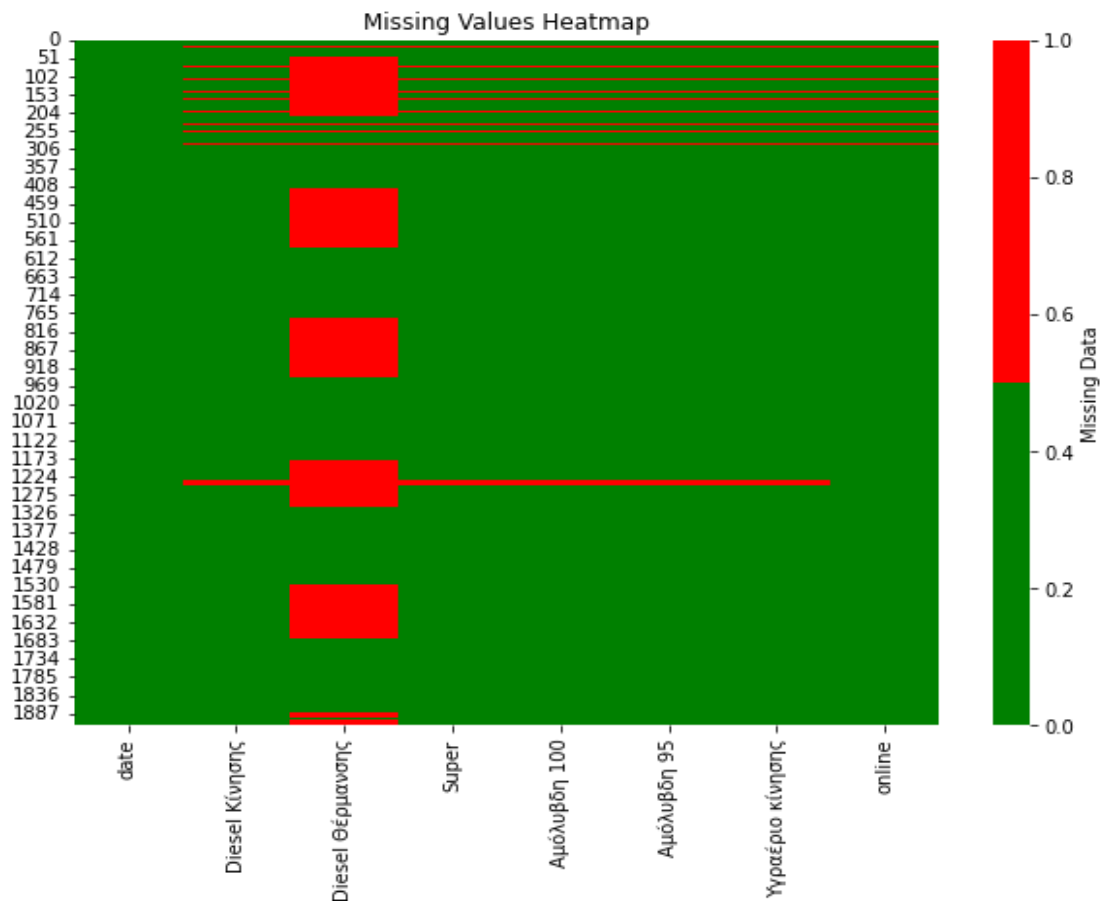
day: ημέρα

Αμόλυβδη 95: υγρό καύσιμο

weekday: εβδομάδα

Υγραερίο κίνησης: υγρό καύσιμο

Στις τιμές των παρατηρήσεων, σε κάποιες περιπτώσεις υπήρχαν ελλείπουσες τιμές (missing values), ενώ στην επόμενη με την προηγούμενη σειρά δεδομένων διατίθεντο παρατηρήσεις. Οπότε, εφαρμόστηκε η μέθοδος `interpolate` Παίρνοντας την τιμή της προηγούμενης ημέρας και της επομένης, η μέθοδος βρίσκει τον μέσο όρο της συγκεκριμένης μεταβλητής και με τον μέσο όρο αυτό αντικαθίσταται η αγνοούμενη τιμή. Παρακάτω παρουσιάζονται τα περιγραφικά μέτρα των μεταβλητών και ένα heatmap με τα missing values.



Διάγραμμα 0: Heatmap για τις αγνοούμενες τιμές

Με κόκκινο χρώμα αναπαρίστανται οι αγνοούμενες τιμές σε κάθε γραμμή, ενώ με πράσινο χρώμα τα υφιστάμενα δεδομένα. Αυτό που αξίζει να αναφερθεί είναι ότι στη μεταβλητή diesel θέρμανσης δεν μπορεί να εφαρμοστεί η μέθοδος interpolate, διότι δεν υπάρχουν δεδομένα στην προηγούμενη και στην επόμενη σειρά. Άρα, δεν μπορούμε να πάρουμε τον μέσο όρο της προηγούμενης και της επόμενης ημέρας. Άρα, αυτό που μπορούμε να κάνουμε είναι να διώξουμε τα nas.

Αμόλυβδη 95

year	Μέση τιμή	T.A	Ελάχιστη τιμή	Q1	Διάμεσος	Q3	Μέγιστη τιμή
2017	1.51	0.02	1.46	1.49	1.51	1.53	1.55
2018	1.59	0.05	1.48	1.55	1.60	1.64	1.67
2019	1.59	0.05	1.48	1.58	1.60	1.62	1.65
2020	1.44	0.09	1.31	1.41	1.42	1.43	1.62
2021	1.64	0.09	1.44	1.60	1.66	1.71	1.78
2022	2.03	0.19	1.75	1.86	2.05	2.14	2.42

Υγραέριο κίνησης

Year	Μέση τιμή	T.A	Ελάχιστη τιμή	Q1	Διάμεσος	Q3	Μέγιστη τιμή
2017	0.81	0.03	0.76	0.78	0.81	0.83	0.86
2018	0.83	0.02	0.78	0.80	0.83	0.84	0.89
2019	0.79	0.02	0.76	0.78	0.79	0.81	0.82
2020	0.74	0.04	0.70	0.72	0.73	0.76	0.82
2021	0.87	0.07	0.76	0.82	0.84	0.90	1.00
2022	1.09	0.05	0.99	1.06	1.12	1.13	1.15

Super

Year	Μέση τιμή	T.A	Ελάχιστη τιμή	Q1	Διάμεσος	Q3	Μέγιστη τιμή
2017	1.64	0.02	1.60	1.64	1.64	1.65	1.68
2018	1.70	0.04	1.62	1.65	1.72	1.74	1.76
2019	1.74	0.03	1.68	1.72	1.75	1.77	1.81
2020	1.65	0.06	1.56	1.62	1.63	1.70	1.78
2021	1.73	0.09	1.49	1.67	1.73	1.81	1.89
2022	1.86	0.17	1.49	1.71	1.87	2.02	2.20

Αμόλυβδη 100

Year	Μέση τιμή	T.A	Ελάχιστη τιμή	Q1	Διάμεσος	Q3	Μέγιστη τιμή
2017	1.71	0.02	1.68	1.70	1.70	1.72	1.74
2018	1.79	0.05	1.72	1.74	1.80	1.84	1.85
2019	1.80	0.04	1.70	1.79	1.82	1.82	1.85
2020	1.67	0.08	1.58	1.63	1.64	1.68	1.83
2021	1.82	0.08	1.63	1.78	1.82	1.88	1.93
2022	2.18	0.19	1.91	2.01	2.22	2.30	2.58

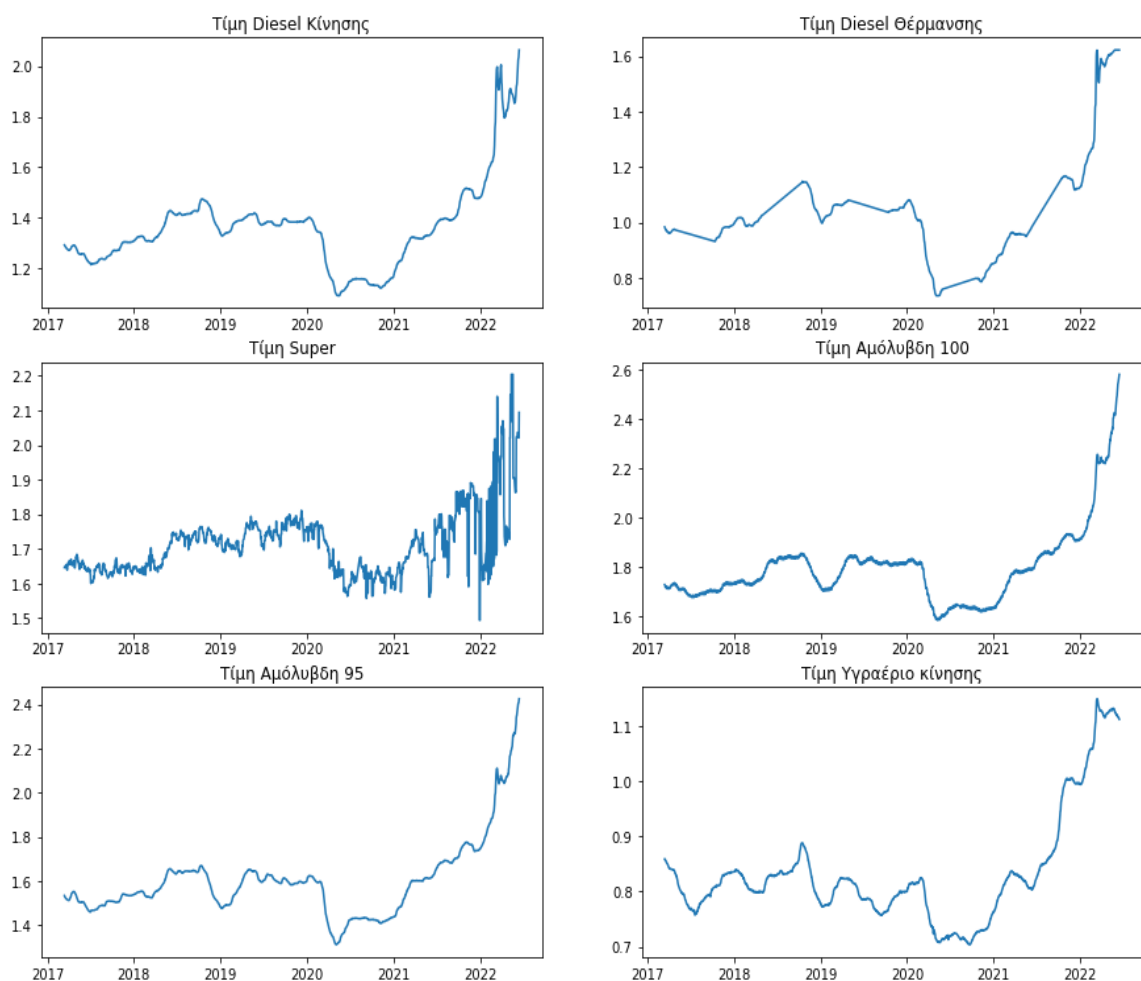
Diesel κίνησης

Year	Μέση τιμή	T.A	Ελάχιστη τιμή	Q1	Διάμεσος	Q3	Μέγιστη τιμή
2017	1.26	0.03	1.21	1.24	1.26	1.29	1.31
2018	1.39	0.05	1.30	1.33	1.41	1.42	1.47
2019	1.38	0.02	1.33	1.37	1.38	1.39	1.42
2020	1.19	0.09	1.09	1.13	1.15	1.18	1.40
2021	1.37	0.09	1.17	1.32	1.38	1.43	1.52
2022	1.77	0.17	1.48	1.60	1.83	1.91	2.06

Diesel θέρμανσης

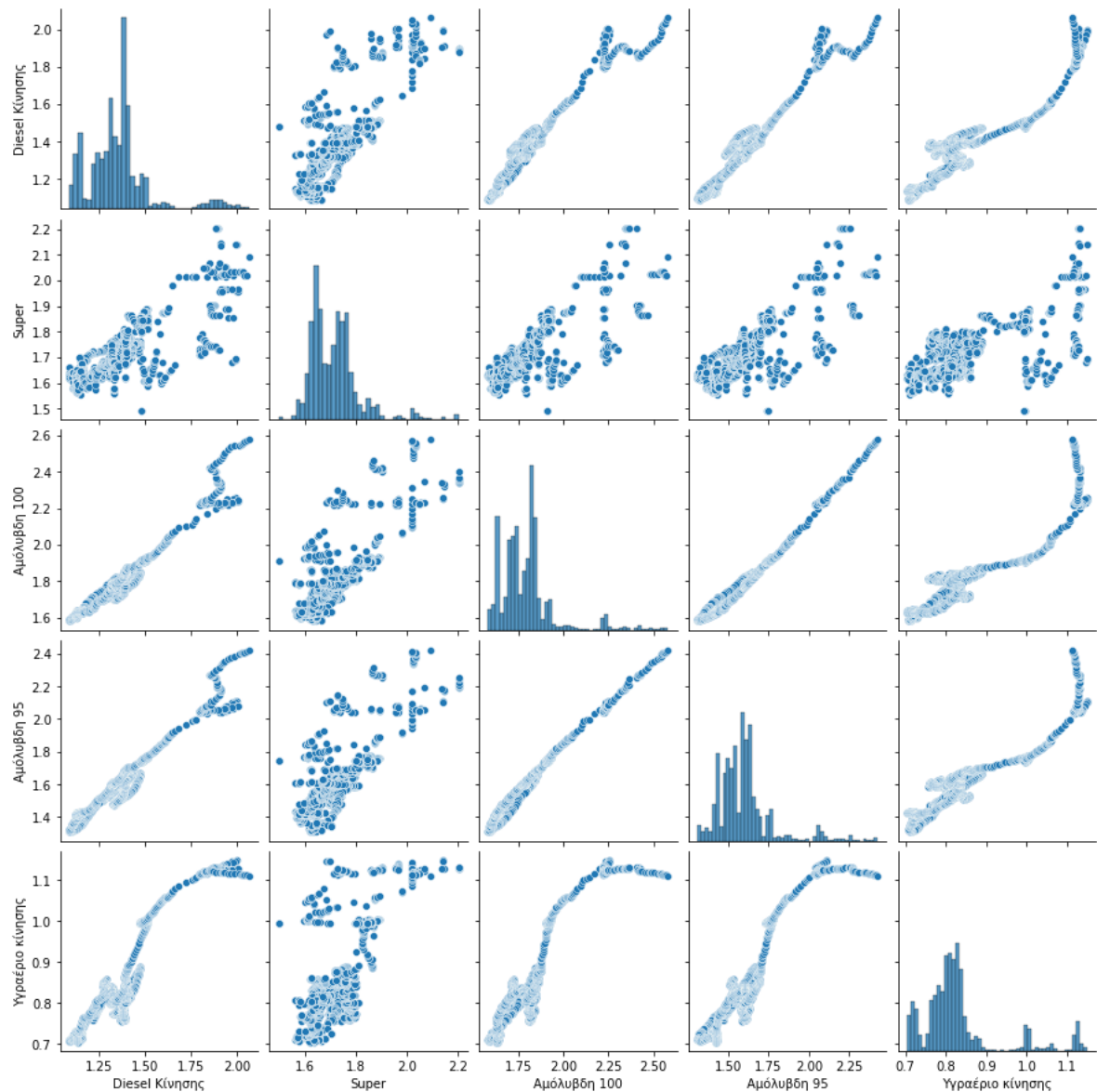
Year	Μέση τιμή	T.A	Ελάχιστη τιμή	Q1	Διάμεσος	Q3	Μέγιστη τιμή
2017	0.96	0.02	0.93	0.95	0.96	0.97	1.00
2018	1.06	0.05	0.99	1.02	1.05	1.11	1.15
2019	1.05	0.02	1.00	1.04	1.05	1.07	1.08
2020	0.84	0.10	0.74	0.77	0.80	0.85	1.08
2021	1.03	0.10	0.85	0.96	1.01	1.12	1.17
2022	1.45	0.18	1.13	1.25	1.57	1.61	1.62

Όπως παρατηρούμε από τα περιγραφικά μέτρα των ποσοτικών μεταβλητών, το 2022, για όλες της μεταβλητές, η μέση τιμή είναι έντονα πιο υψηλή σε σχέση με της προηγούμενες χρονιές. Η τυπική απόκλιση είναι μεγαλύτερη, γεγονός που μας δείχνει την επίπτωση του πόλεμου στις τιμές των καυσίμων. Άλλη διαπίστωση είναι ότι η ελάχιστη μέση τιμή που παρατηρείται το 2022, σε συγκεκριμένες μεταβλητές, είναι μεγαλύτερη σε σχέση με τη μέγιστη τιμή τις προηγούμενες χρονιές.



Διάγραμμα 1: Χρονολογικό διάγραμμα

Στα παραπάνω διαγράμματα, βλέπουμε την πορεία της μέσης τιμής των μεταβλητών στον χρόνο. Τα συμπεράσματα που προκύπτουν είναι τα εξής: Στις αρχές του 2022 η μέση τιμή για κάθε κατηγορία καυσίμου εκτινάχθηκε, παρουσιάζοντας τη μεγαλύτερη αύξηση των τελευταίων ετών. Η μέση τιμή της super φαίνεται να έχει την πιο έντονη μεταβλητότητα στην τιμή σε σχέση με τις υπόλοιπες που ακολουθούν μια ανοδική πορεία.



Διάγραμμα 2: Ιστογράμματα και συσχέτιση μεταβλητών

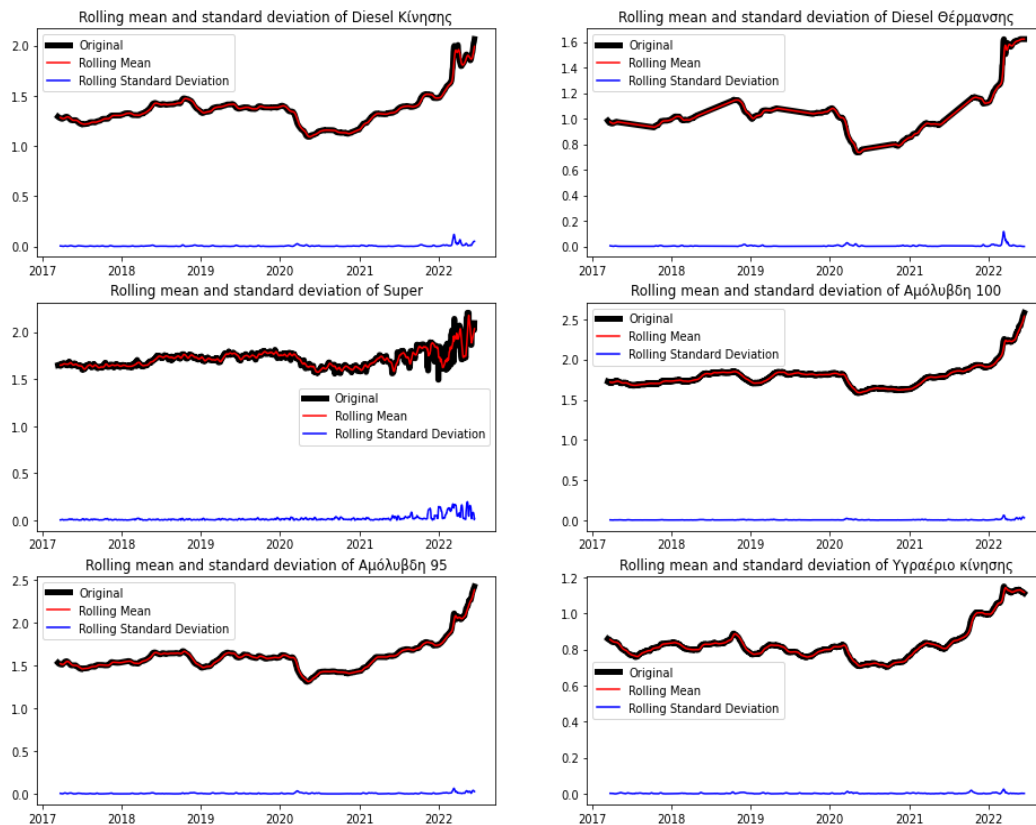
Παρατηρούμε, από τους συντελεστές κυρτώσεις, ασυμμετρίας και των ιστογραμμάτων, ότι όλες οι μεταβλητές έχουν θετική ασυμμετρία, γεγονός που σημαίνει ότι η συχνότητα των τιμών είναι μεγαλύτερη στη δεξιά πλευρά της κατανομής. Δηλαδή, οι περισσότερες τιμές βρίσκονται προς τα δεξιά της κατανομής. Απέχουν όλες οι μεταβλητές από την κανονική κατανομή.

ΚΕΦΑΛΑΙΟ 4. ΕΛΕΓΧΟΙ ΥΠΟΘΕΣΕΩΝ ΜΕΤΑΞΥ ΤΩΝ ΜΕΤΑΒΛΗΤΩΝ

Σε αυτό το κεφάλαιο, θα εξεταστούν κάποιες προϋποθέσεις που πρέπει να ισχύουν για τα μοντέλα στο επόμενο κεφάλαιο. Αρχικά, θα εξεταστεί αν υπάρχει στασιμότητα στη χρονολογική σειρά των μεταβλητών, για να μπορέσουμε να αναλύσουμε τη χρονολογική σειρά. Με τον όρο «στασιμότητα», εννοούμε ότι η χρονολογική σειρά μέσα στον χρόνο πρέπει να έχει σταθερή μέση τιμή και σταθερή διακύμανση. Ο έλεγχος για τη στασιμότητα των σειρών πραγματοποιείται με τον Dickey-Fuller και τα αποτελέσματα παρουσιάζονται παρακάτω.

	p-value
adf_Diesel_κίνησης	0.997
adf_Super	0.335
adf_Αμόλυβδη_100	1.000
adf_Αμόλυβδη_95	0.998
adf_Υγραέριο_κίνησης	0.9013

Ο έλεγχος Dickey-fuller έχει ως μηδενική υπόθεση τη μη-στασιμότητα και ως εναλλακτική υπόθεση την ύπαρξη στασιμότητας. Παρατηρούμε, σε επίπεδο στατιστικής σημαντικότητας $\alpha=5\%$, ότι το $p\text{-value} > \alpha$. Δεν απορρίπτουμε την H_0 . Άρα, δεν υπάρχει στασιμότητα των μεταβλητών στη χρονολογική σειρά. Παρουσιάζουμε και το παρακάτω γράφημα που το επιβεβαιώνει.



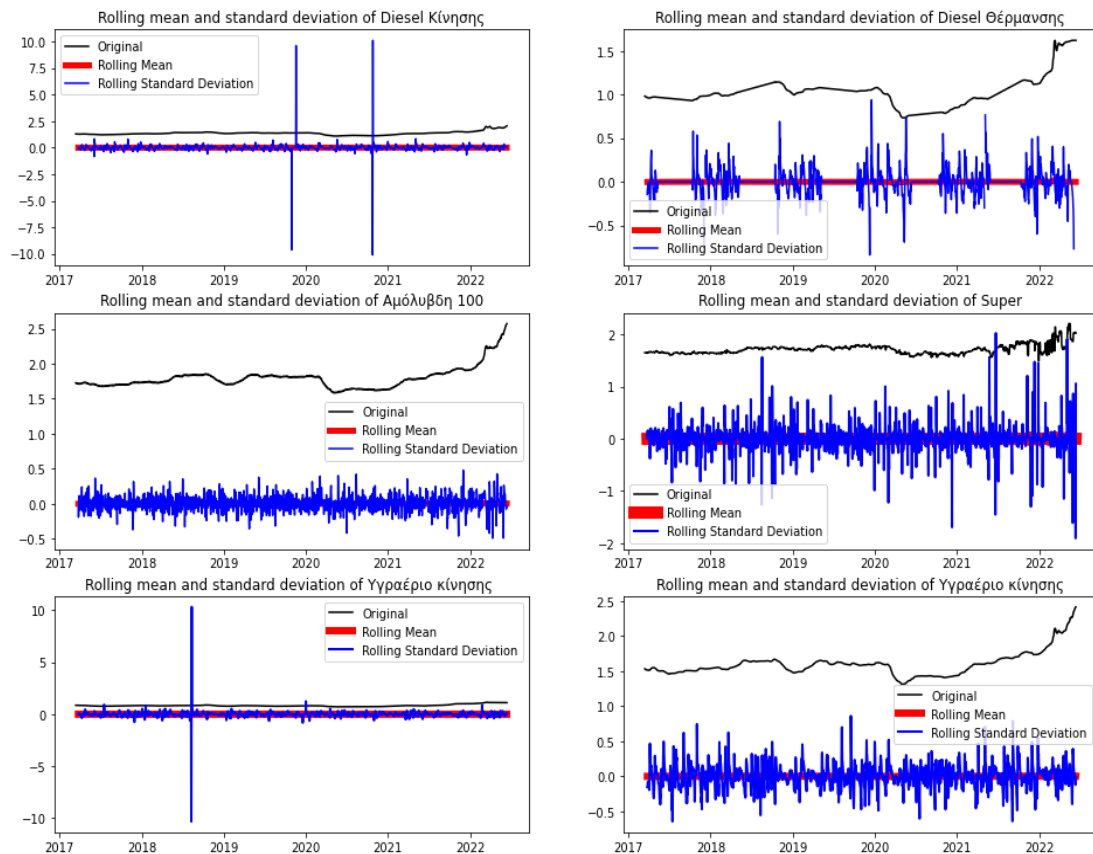
Διαγραμμα 3: Κυλιόμενος μέσος όρος και διακύμανση των τελευταίων 6 ετών.

Παρατηρούμε ότι ο κινητός μέσος όρος συμβαδίζει με τον πραγματικό μέσο όρο. Αυτό συμβαίνει διότι στα δεδομένα μας έχουμε μέσες τιμές των μεταβλητών.

Στη συνέχεια, για να διορθώσουμε το πρόβλημα, θα προχωρήσουμε σε μετασχηματισμό των μεταβλητών, ώστε να μετατραπεί η σειρά σε στάσιμη. Γι' αυτό, θα πάρουμε τη διαφορά των λογαρίθμων και θα ξανακάνουμε τον έλεγχο.

	p-value
diff(log(adf_Diesel_κίνησης))	4.677299e-09
diff(log(adf_Super))	2.005182e-20
diff(log(adf_Αμόλυβδη_100))	2.350028e-05
diff(log(adf_Αμόλυβδη_95))	2.562097e-05
diff(log(adf_Υγραέριο_κίνησης))	2.991424e-10

Από τον μετασχηματισμό βλέπουμε ότι το $p\text{-value} < \alpha = 5\%$ και, συνεπώς, απορρίπτουμε την H_0 . Η σειρά γίνεται στάσιμη, για να μπορεί να μελετηθεί εν συνεχεία.



Διάγραμμα 4: Στάσιμη σειρά των μεταβλητών.

Έπειτα, εξετάζουμε την κανονικότητα στη λογαριθμησμένη διαφορά των μεταβλητών με τον έλεγχο jarque_bera.

Diesel κίνησης: Jarque_beraResult(statistic=196575.97649467996, pvalue=0.0)

Diesel θέρμανσης: Jarque_beraResult(statistic=698056.7442444135, pvalue=0.0)

Super: Jarque_beraResult(statistic=246031.24359056941, pvalue=0.0)

Αμόλυβδη 100: Jarque_beraResult(statistic=746.3690287092458, pvalue=0.0)

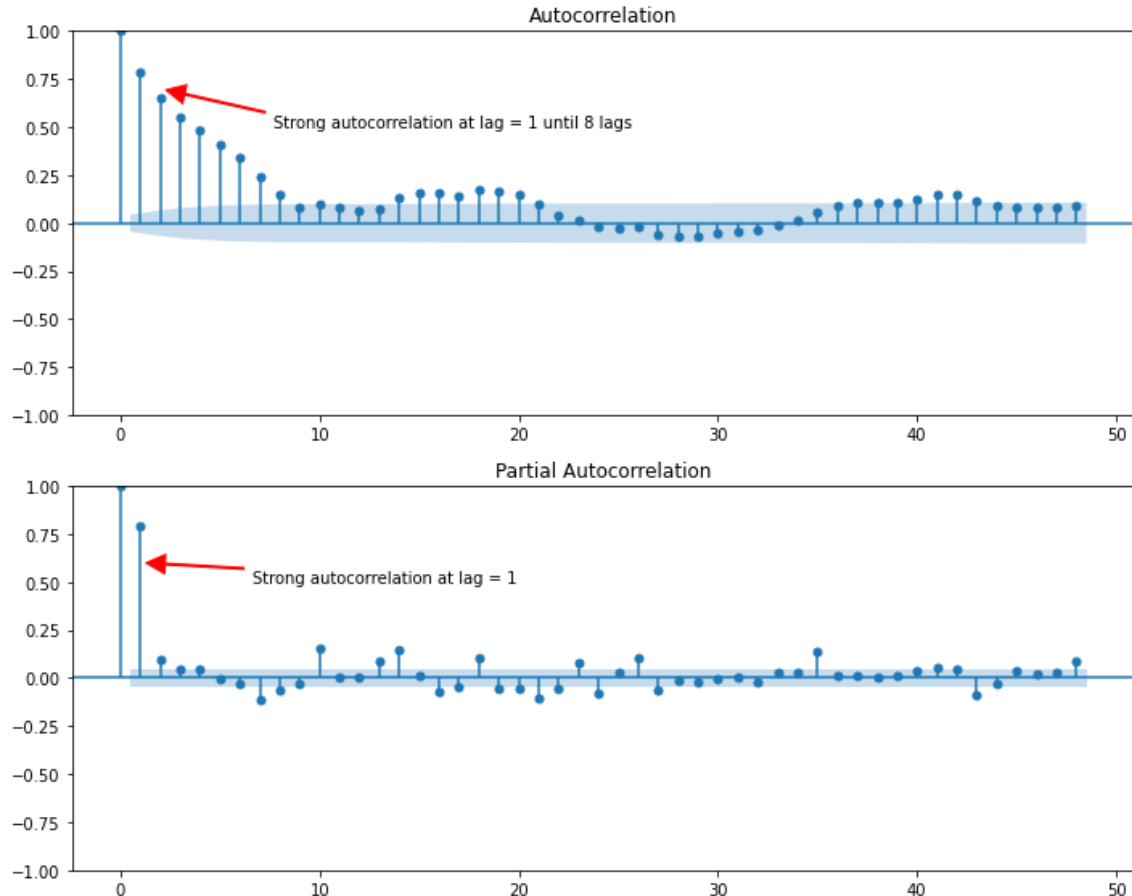
Αμόλυβδη 95: Jarque_beraResult(statistic=19960.80865876706, pvalue=0.0)

Απορρίπτουμε την H_0 που υποθέσει την ύπαρξη κανονικότητας των μεταβλητών, διότι $p\text{-value} < \alpha = 5\%$. Άρα, δεν ακολουθούν την κανονική κατανομή οι παραπάνω μεταβλητές.

ΚΕΦΑΛΑΙΟ 5. ΠΡΟΒΛΕΠΤΙΚΑ ΚΑΙ ΕΡΜΗΝΕΥΣΗ ΜΟΝΤΕΛΟΥ

Σε αυτό το κεφάλαιο, θα χρησιμοποιήσουμε χρονολογικά μοντέλα για να διορθώσουμε την χρονική αυτοσυσχέτιση των τιμών και να κάνουμε πρόβλεψη, αναλύοντας την λογαριθμησμένη διαφορά της μεταβλητής diesel κίνησης (επιλέχθηκε τυχαία η μεταβλητή).

Το πρώτο βήμα είναι να δούμε σε ποιες χρονικές υστερήσεις υπάρχει το πρόβλημα αυτοσυσχέτισης και μερικής αυτοσυσχέτισης των τιμών με τα παρακάτω διαγράμματα.



Διάγραμμα 5. Διάγραμμα αυτοσυσχέτισης και μερικής αυτοσυσχέτισης της διαφοράς των λογαριθμησμένων διαφορών.

Όπως παρατηρούμε στα διαγράμματα και τα σχόλια που έχουν αποτυπωθεί, υπάρχει ισχυρή αυτοσυσχέτιση των τιμών και μερική αυτοσυσχέτιση. Το πρόβλημα θα πρέπει να διορθωθεί με χρονολογικά μοντέλα. Μετά από κάποιες δοκιμές, εντοπίστηκε πως το κατάλληλο μοντέλο για να διορθωθεί η αυτοσυσχέτιση και η μερική αυτοσυσχέτιση είναι το ARMA(1,7), ένας συνδυασμός μοντέλων AR(1) και MA(7).

```

SARIMAX Results
=====
Dep. Variable:          y      No. Observations:      1920
Model:                ARIMA(1, 0, 7)  Log Likelihood      9875.815
Date:                 Mon, 05 Sep 2022  AIC             -19731.629
Time:                 14:40:25      BIC             -19676.029
Sample:               0      HQIC             -19711.172
                   - 1920
Covariance Type:      opg
=====
              coef      std err          z      P>|z|      [0.025      0.975]
-----
const          0.0002          0.000         1.229      0.219      -0.000         0.001
ar.L1          0.6592          0.017        39.185      0.000         0.626         0.692
ma.L1          0.0258          0.015         1.710      0.087      -0.004         0.055
ma.L2          0.1083          0.014         7.893      0.000         0.081         0.135
ma.L3          0.0946          0.009         9.964      0.000         0.076         0.113
ma.L4          0.0933          0.010         9.796      0.000         0.075         0.112
ma.L5          0.1259          0.009        14.301      0.000         0.109         0.143
ma.L6          0.1915          0.010        19.049      0.000         0.172         0.211
ma.L7          0.1012          0.008        13.289      0.000         0.086         0.116
sigma2       1.985e-06      1.33e-08      149.214      0.000      1.96e-06      2.01e-06
=====
Ljung-Box (L1) (Q):          0.02  Jarque-Bera (JB):        618624.17
Prob(Q):          0.89  Prob(JB):          0.00
Heteroskedasticity (H):      9.10  Skew:          1.96
Prob(H) (two-sided):      0.00  Kurtosis:        90.85
=====

Warnings:
[1] Covariance matrix calculated using the outer product of gradients (complex-step).

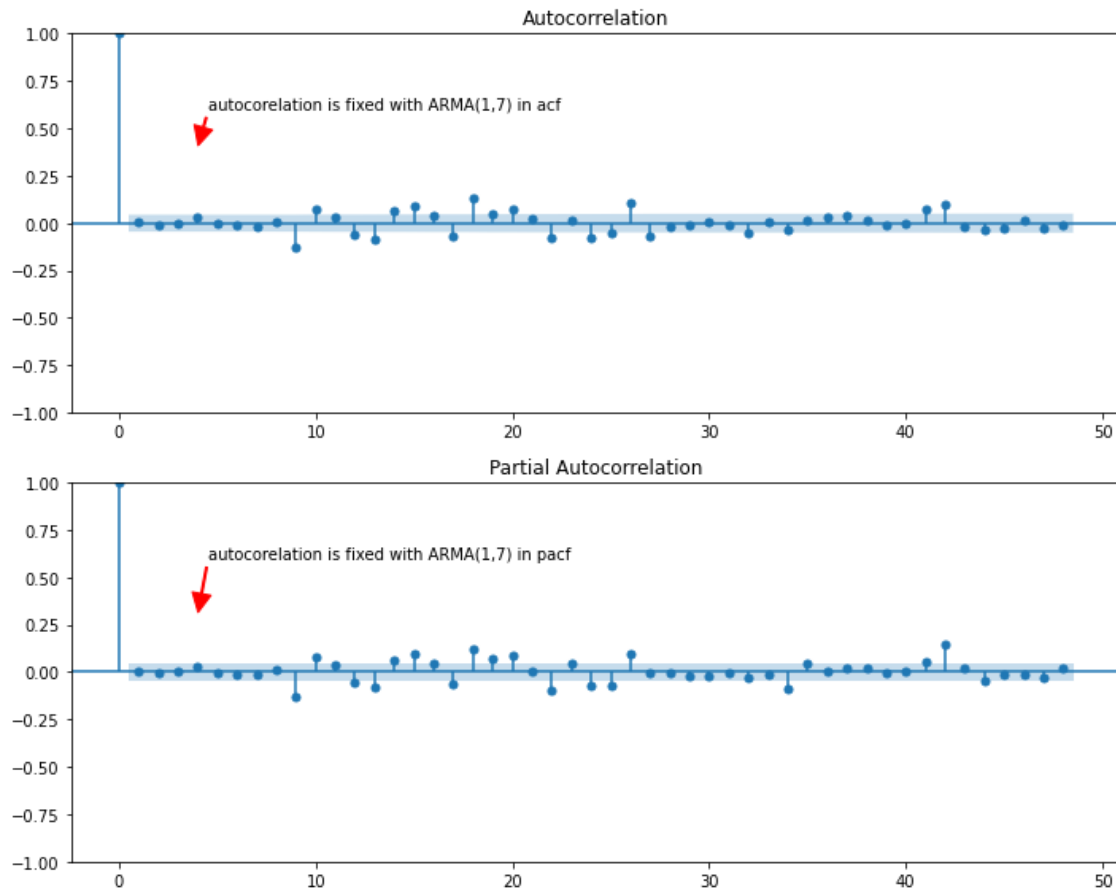
```

Πίνακας 1: Αποτελέσματα μοντέλου ARMA(1,7)

$$\begin{aligned}
 Y_t = & 0.0002 + 0.6592Y(t-1) + 0.025E(t-1) + 0.10E(t-2) + 0.094E(t-3) \\
 & + 0.0933E(t-4) \\
 & + 0.1259E(t-5) + 0.1915E(t-6) + 0.1012E(t-7)
 \end{aligned}$$

Από τον παραπάνω πίνακα, διαπιστώνουμε ότι, εκτός από τη σταθερά όλες οι υπόλοιπες μεταβλητές είναι στατιστικά σημαντικές, υπάρχει απόκλιση από την κανονικότητα του μοντέλου, και ετεροσκεδαστικότητα.

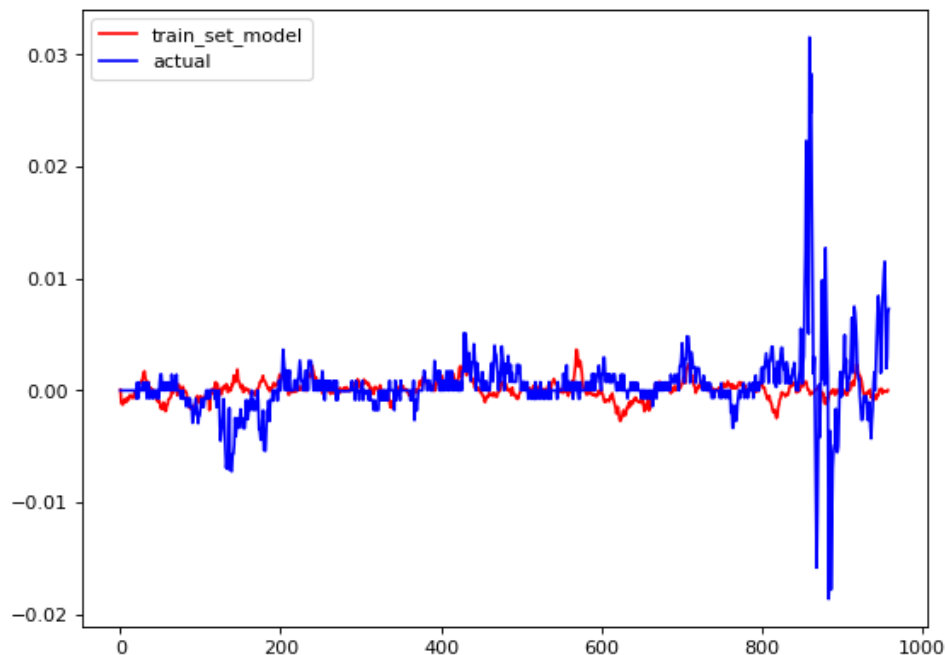
Εν συνεχεία, θα δούμε τα κατάλοιπα του μοντέλου, αν έχει διορθωθεί η αυτοσυσχέτιση και η μερική αυτοσυσχέτιση.



Διάγραμμα 6: Αυτοσυσχέτιση και μερική αυτοσυσχέτιση

Όπως βλέπουμε και στο παραπάνω γράφημα, φαίνεται ότι το μοντέλο ARMA(1,7) κατάφερε να διορθώσει την αυτοσυσχέτιση και τη μερική αυτοσυσχέτιση.

Τέλος, θα τεστάρουμε ποσό ικανό είναι το μοντέλο μας να κάνει προβλέψεις, χωρίζοντας το σετ δεδομένων, σε `train_set` που περιέχει το 70% του δείγματος και σε `test_set` που περιέχει το 30% του δείγματος. Αυτός ο διαχωρισμός του δείγματος δεν είναι τυχαίος, αλλά έχει γίνει έτσι, διότι ξέρουμε, βάσει βιβλιογραφίας⁴ (Anon., n.d.), ότι με αυτόν τον διαχωρισμό πετυχαίνουμε τη μικρότερη μεροληψία και ελαχιστοποιούμε τη διακύμανση.



Διάγραμμα 7: Προβλεπόμενες τιμές μοντέλου με τις πραγματικές

Όπως παρατηρούμε, το μοντέλο χρειάζεται διορθώσεις. Ενδεχομένως να χρειάζεται να διορθωθεί και η ετεροσκεδαστικότητα για να προβλέπει σωστά. Χρειάζεται να χρησιμοποιηθούν και garch μοντέλα.

Παρακάτω, θα δοκιμάσουμε ένα μοντέλο πολλαπλής γραμμικής παλινδρόμησης με κάποιες μακροοικονομικές μεταβλητές. Τα εν λόγω δεδομένα είναι διαθέσιμα από την ΕΛΣΤΑΤ (Anon., n.d.)⁵, σε μορφή .csv, και τα επεξεργαστήκαμε με τη γλώσσα προγραμματισμού Python.

Οι μεταβλητές αυτές είναι οι εξής,

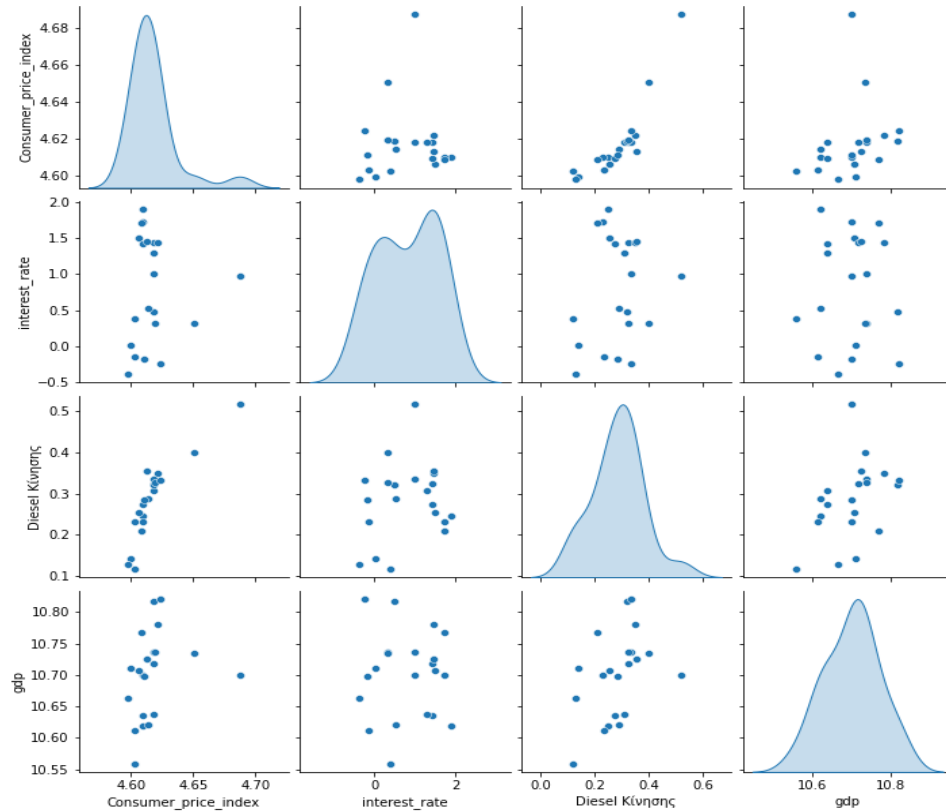
- ❖ Ακαθάριστο εγχώριο προϊόν
- ❖ Επιτόκια Ελλάδος (Anon., n.d.)⁶
- ❖ Δείκτης τιμών καταναλωτή

⁴ Cross validation for time series data, πηγή scikit-learn.org

⁵ Ακαθάριστο εγχώριο Προϊόν, δείκτης τιμών καταναλωτή, πηγή Ελληνική στατιστική, Αρχή statistics.gr

⁶ Economic research interest rate of Greece. Πηγή [Fred.stouisfed.org](https://fred.stouisfed.org)

Σκοπός είναι να προβλέψουμε τη μέση τιμή της αμόλυβδη_100 και τα δεδομένα είναι σε 3μηνιαία βάση



Διάγραμμα 8: Συσχέτιση και κατανομές των μεταβλητών

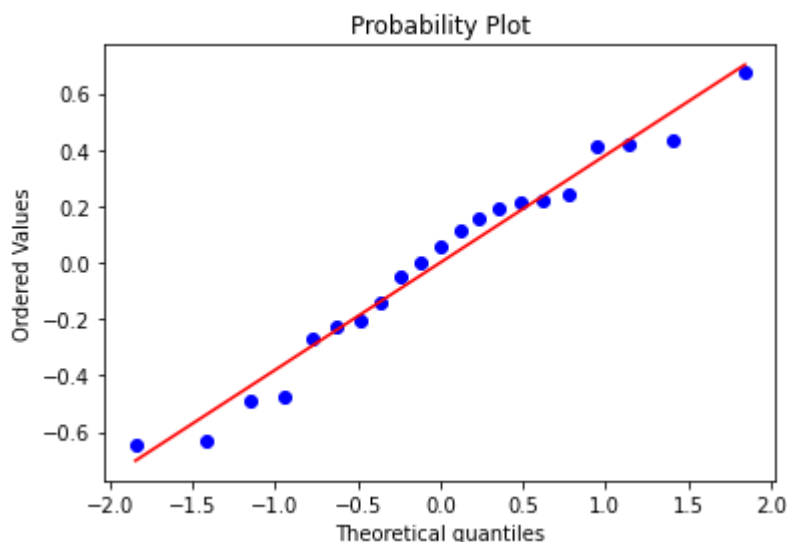


Διάγραμμα 9: Ποσοτικοποιημένες συσχετίσεις μεταβλητών

Έπειτα, εξετάζεται ποιο μοντέλο είναι το κατάλληλο που προσαρμόζεται στα δεδομένα, ώστε να μπορεί να υπολογίσει την αμολυβδη_100 και να την εκτιμήσει με τη μεγαλύτερη ακρίβεια. Αρχικά, η εξαρτημένη μεταβλητή είναι η αμολυβδη_100 και οι επεξηγηματικές οι υπόλοιπες. Ωστόσο, για να είναι έγκυρο κάποιο μοντέλο, πρέπει να ικανοποιούνται κάποιες προϋποθέσεις, να μην έχει πρόβλημα πολυσυγγραμματοκότητας, να ισχύει η κανονικότητα καταλοίπων, η γραμμικότητα, η ομοσκεδαστικότητα και η αυτοσυσχέτιση, διότι είναι χρονολογικά δεδομένα. Το μοντέλο είναι το παρακάτω.

Dep. Variable:	Αμόλυβδη_100			R-squared:	0.872	
Model:	OLS			Adj. R-squared:	0.850	
Method:	Least Squares			F-statistic:	38.68	
Date:	Wed, 14 Sep 2022			Prob (F-statistic):	8.21e-08	
Time:	12:53:34			Log-Likelihood:	-8.1956	
No. Observations:	21			AIC:	24.39	
Df Residuals:	17			BIC:	28.57	
Df Model:	3					
Covariance Type:	nonrobust					
	coef	std err	t	P> t 	[0.025	0.975]
Intercept	7.355e-16	0.087	8.48e-15	1.000	-0.183	0.183
Consumer_price_index	0.8383	0.090	9.364	0.000	0.649	1.027
interest_rate	-0.0347	0.087	-0.399	0.695	-0.218	0.149
gdp	0.2492	0.089	2.786	0.013	0.061	0.438
Omnibus:	0.665			Durbin-Watson:	1.567	
Prob(Omnibus):	0.717			Jarque-Bera (JB):	0.697	
Skew:	-0.220			Prob(JB):	0.706	
Kurtosis:	2.223			Cond. No.	1.29	

Οι προϋποθέσεις του μοντέλου είναι οι εξής: παρατηρούμε ότι δεν παραβιάζεται η υπόθεση της κανονικότητας με το τεστ jarque-bera ($jb=0.69$ $p\text{-value}=0.706>5\%$), ούτε της ομοσκεδαστικότητας ($F\text{-statistic}=1.92$, $p\text{-value}=0.20>5\%$), υπάρχουν ακραίες τιμές, ενώ δεν υπάρχει πολυσυγγραμματοκότητα, διότι οι συσχετίσεις στην πλειοψηφία είναι μικρές μεταξύ των επεξηγηματικών μεταβλητών.



Διάγραμμα 9: Κανονικότητα καταλοίπων

$N=21$, $Adjusted-R-squared=85\%$, $AIC=24.39$

Το παραπάνω μοντέλο είναι στατιστικά σημαντικό ($F-statistic= 38.68$,

$Value=0.0008<5\%$), το ποσοστό μεταβλητότητας του μοντέλου που ερμηνεύει την αμολυβδη_100 είναι 85%. Στατιστικά σημαντικές μεταβλητές είναι ο δείκτης τιμών καταναλωτή και το gdp, ενώ η σταθερά και τα επιτόκια όχι. Για παράδειγμα, αν αυξηθεί ο δείκτης τιμών καταναλωτή κατά 1%, τότε αναμένουμε μια αύξηση στη μέση τιμή της αμολύβδης_100 κατά 0.0083, έχοντας τις υπόλοιπες μεταβλητές σταθερές.

ΚΕΦΑΛΑΙΟ 6. ΣΥΜΠΕΡΑΣΜΑ

Ολοκληρώνοντας την ανάλυση από τα 1920 δεδομένα που ανακτήθηκαν, προκύπτουν κάποια συμπεράσματα. Το 2022 οι τιμές των καυσίμων φαίνονται να έχουν εκτιναχθεί σε σχέση με τις προηγούμενες χρονιές, γεγονός που αποτυπώνει τις επιπτώσεις του πολέμου μεταξύ Ουκρανίας και Ρωσίας. Η σχέση των μεταβλητών είναι θετικές. Δηλαδή, όσο αυξάνεται η μέση τιμή της μιας μεταβλητής, αυξάνεται και η άλλη.

Το μοντέλο που υπολογίστηκε για να προβλέψει τη μεταβλητότητα της διακύμανσης της τιμής diesel κίνησης, για να δούμε πως θα κυμανθεί η τιμή της στο μέλλον, φαίνεται να έχει όλες τις μεταβλητές στατιστικά σημαντικές εκτός από τη σταθερά, όπου διορθώνει την αυτοσυσχέτιση των χρονικών υστερήσεων. Όμως, το μοντέλο, αφού δοκιμάστηκε, δεν προβλέπει με μεγάλη ακρίβεια τη μεταβλητότητα της διακύμανσης και αυτό διότι παραβιάζεται η υπόθεση της ομοσκεδαστικότητας. Θα έπρεπε να προσαρμοστούν και μοντέλα garch για να μοντελοποιήσουν τη διάκυμανση. Το να προβλέψουμε τη μεταβλητότητα της διακύμανσης δείχνει ότι, αν υπάρχει στο μέλλον μεγάλη μεταβλητότητα, σημαίνει πως η τιμή βρίσκεται σε αβεβαιότητα και να νέα θα είναι κακά, ενώ αν έχει μικρή θα σήμαινε πως τα νέα είναι καλά στην τιμή της.

Στο δεύτερο μοντέλο, το γεγονός ότι έχουμε έναν πολύ υψηλό προσαρμοσμένο συντελεστή προσδιορισμού, σημαίνει ότι ένα μεγάλο ποσοστό συνολικής μεταβλητότητας της εξαρτημένης μεταβλητής ερμηνεύεται από το μοντέλο. Αυτό σημαίνει ότι τα δεδομένα επαρκούν για να δώσουν τις κατάλληλες πληροφορίες που χρειάζεται για την εκτίμηση και τον υπολογισμό της αμόλυβδης ρ_{100} .

Βιβλιογραφία

Ανόν., χ.χ. *Economic research interest rate greece*. [Ηλεκτρονικό]

Available at: <https://fred.stlouisfed.org/tags/series?t=greece%3Binterbank%3Binterest+rate>

Ανόν., χ.χ. *scikit.learn.org*. [Ηλεκτρονικό]

Available at: https://scikit-learn.org/stable/modules/cross_validation.html

Ανόν., χ.χ. *Statistics.gr*. [Ηλεκτρονικό]

Available at: <https://www.statistics.gr/el/statistics/-/publication/SEL84/->

Boitumelo Nnoi Yolanda Sekati, J. T. T. & L. D., 2022. Modelling the oil price volatility and macroeconomic variables in south africa using the symmetric and assymetric garch models. jule.

OT.gr Newsroom, 2022. Ενεργειακή κρίση: Η Ευρώπη μπορεί να αντέξει έναν κακό χειμώνα.

Κέλλυ Κική, 2022. *IMeD*. [Ηλεκτρονικό]

Available at: <https://lab.imedd.org/ta-dedomena-gia-tis-times-ygron-kafsimon-stin-ellada/>

Κέλλυ κική, Κ. Μ., 2022. Η τιμή της βενζίνης τα τελευταία πέντε χρόνια. *IMeD LAB*, 13 Ιουλίου.

Κέλλυ κικι και Κωνσταντίνος Μποτσι, 2022. Η τιμή της βενζίνης τα τελευταία 5 χρόνια. *IMeD*, 13 Ιουλίου .p. 3.

Φορέας πρακτικής άσκησης

Το [iMedD](#) (incubator for Media Education and Development) είναι μη κερδοσκοπικός δημοσιογραφικός οργανισμός, που ιδρύθηκε το 2018 με αποκλειστικό δωρητή το Ίδρυμα Σταύρος Νιάρχος, με σκοπό την ενίσχυση της δημοσιογραφίας. Η εργασία έγινε στο πλαίσιο της πρακτικής άσκησης στην επεξεργασία και ανάλυση δεδομένων, στο iMedD Lab, τον πυλώνα του οργανισμού που ειδικεύεται στη δημοσιογραφία δεδομένων.