



Πανεπιστήμιο Αιγαίου

Τμήμα Στατιστικής και Αναλογιστικών-Χρηματοοικονομικών Μαθηματικών

Π.Μ.Σ Στατιστικής και Ανάλυσης Δεδομένων

**Ανάπτυξη Μεθόδων Stresstesting
με Χρήση Χρονοσειρών**

Διπλωματική Εργασία

Τζιάλλας

Ελευθέριος

Σάμος, 2021

Τριμελής Επιτροπή

Καραγρηγορίου Αλέξανδρος

Καθηγητής Τμήματος Σ.Α.Χ.Μ.

Κουντζάκης Χρήστος

Επ. Καθηγητής Τμήματος Σ.Α.Χ.Μ. (Επιβλέπων)

Ξανθόπουλος Στυλιανός

Αν. Καθηγητής Τμήματος Σ.Α.Χ.Μ.

Περιεχόμενα

Περίληψη	1
1 Μονομεταβλητή Ανάλυση Χρονοσειρών	4
1.1 Εισαγωγή στις Χρονοσειρές	5
1.2 Τύποι και Χαρακτηριστικά Χρονοσειρών	6
1.3 Στασιμότητα–Μη Στασιμότητα	9
1.4 Υπόδειγμα Τυχαίου Περιπάτου–Λευκός Θόρυβος.....	10
1.5 Αυτοπαλίνδρομα Ολοκληρωμένα Κινητού Μέσου Υποδείγματα	11
1.5.1 Αυτοπαλίνδρομο υπόδειγμα πρώτης τάξεως (AR(1)).....	11
1.5.2 Αυτοπαλίνδρομο υπόδειγμα τάξεως p (AR(p))	11
1.5.3 Υπόδειγμα κινητού μέσου πρώτης τάξεως (MA(1)).....	13
1.5.4 Υπόδειγμα κινητού μέσου τάξεως q (MA(q)).....	14
1.5.5 Το υπόδειγμα ARMA(p,q).....	15
1.5.6 Η γενική περίπτωση του ARIMA (p,d,q) υποδείγματος.....	16
1.5.7 Συνάρτηση αυτοσυσχέτισης (ACF)	17
1.5.8 Συνάρτηση μερικής αυτοσυσχέτισης (PACF)	18
1.5.9 Μέθοδος ανάλυσης χρονοσειράς Box–Jenkins.....	19
1.6 Exponentially Weighted Moving Average–EWMA	19
1.7 Διαγνωστικοί Έλεγχοι.....	20
1.7.1 Έλεγχος στασιμότητας Dickey–Fuller	20
1.7.2 Έλεγχος κανονικότητας Jarque–Bera	21
1.7.3 Έλεγχος ανεξαρτησίας Ljung–Box	22
1.7.4 Έλεγχος ετεροσκεδαστικότητας Goldfeld–Quandt.....	23
1.8 Κριτήρια Επιλογής Μοντέλων.....	23
1.8.1 Akaike's Information Criterion	24
1.8.2 Bayesian Information Criterion	24
1.8.3 Hannan–Quinn Information Criterion.....	25
1.8.4 Μέτρα αξιολόγησης προβλεπτικής ικανότητας μοντέλων.....	25
1.9 Μετασχηματισμός Δεδομένων.....	27
1.9.1 Λογαριθμικός μετασχηματισμός.....	27

1.9.2 Μετασχηματισμός Box–Cox.....	28
2 Backtesting και Stresstesting	30
2.1 Μέθοδοι Υπολογισμού Αξίας σε Κίνδυνο (VaR).....	30
2.1.1 Ιστορική προσομοίωση	31
2.1.2 Monte–Carlo προσομοίωση	32
2.1.3 Η μέθοδος διακύμανσης–συνδιακύμανσης.....	33
2.2 Expected Shortfall.....	35
2.3 Backtesting και Τεχνικές	35
2.3.1 Kupiec’s Test	37
2.3.2 Christoffersen’s Test	38
2.4 Stresstesting και Τεχνικές	40
2.4.1 Ανάλυση ευαισθησίας.....	40
2.4.2 Ανάλυση σεναρίου.....	41
2.4.3 Αντίστροφη προσομοίωση ακραίων καταστάσεων	41
3 Ένα Τροποποιημένο ARIMA(1,1,0) για την Μοντελοποίηση και Πρόβλεψη Αποδόσεων Χαρτοφυλακίου.....	44
3.1 Ο Προτεινόμενος Αλγόριθμος – Το Τροποποιημένο ARIMA μοντέλο	44
3.2 Πειραματική Μελέτη	45
3.3 Συμπεράσματα.....	47
Βιβλιογραφία	49
Παράρτημα	51

Περίληψη

Η παρούσα διπλωματική εργασία εκπονήθηκε στα πλαίσια του Προγράμματος Μεταπτυχιακών Σπουδών ‘‘Στατιστική και Ανάλυση Δεδομένων’’, του Τμήματος Στατιστικής και Αναλογιστικών–Χρηματοοικονομικών Μαθηματικών, του Πανεπιστημίου Αιγαίου. Σκοπός της, είναι η μελέτη της καλής προσαρμογής και κατ’ επέκταση προβλεπτικής ικανότητας μοντέλων υπέρβασης. Για το λόγο αυτό, αφού παρουσιαστεί η απαραίτητη θεωρία γίνεται χρήση μιας παραλλαγής του ARIMA μοντέλου, το οποίο συνδυάζει EWMA βαρύτητες και δείχνουμε ότι το εν λόγω μοντέλο είναι κατάλληλο για μοντελοποίηση αλλά και πρόβλεψη τιμών αποδόσεων χαρτοφυλακίου.

Η δομή της διατριβής είναι η ακόλουθη. Στο Κεφάλαιο 1 γίνεται αναφορά σε βασικές έννοιες των χρονολογικών υποδειγμάτων και γενικότερα του κλάδου των χρονολογικών σειρών. Το υπόδειγμα ARIMA, καθώς και υποκατηγορίες του, παρουσιάζονται συνοδευόμενα, για σκοπούς πληρότητας, από κριτήρια επιλογής μοντέλου.

Στο Κεφάλαιο 2, γίνεται συζήτηση γύρω από γνωστές μεθόδους Backtesting και Stresstesting, όπου μεταξύ άλλων, αναλύονται οι μέθοδοι Backtesting του Kupiec και του Christoffersen.

Στο Κεφάλαιο 3, προτείνεται μια νέα αλγοριθμική διαδικασία 5 βημάτων, η οποία ενσωματώνει ένα τροποποιημένο ARIMA(1,1,0) μοντέλο με EWMA βάρη, για την μοντελοποίηση και πρόβλεψη χρηματιστηριακών αποδόσεων. Έτσι, εφαρμόζεται πειραματική μελέτη πάνω σε ημερήσιες τιμές αποδόσεων ενός χαρτοφυλακίου με σκοπό να ελεγχθεί η απόδοση της προτεινόμενης αλγοριθμικής διαδικασίας. Τα αποτελέσματα συγκρίνονται με αυτά ενός τυπικού ARIMA μοντέλου με ενδιαφέροντα αποτελέσματα.

Στο τέλος της διπλωματικής έχει δημιουργηθεί Παράρτημα, στο οποίο παρατίθεται ο κώδικας (σε γλώσσα R), ο οποίος δημιουργήθηκε στα πλαίσια της εφαρμογής του Κεφαλαίου 3.

Abstract

This thesis is conducted at the Department of Statistics and Actuarial–Financial Mathematics, of the University of the Aegean, in the context of the MSc Program in *Statistics and Data Analysis*. Its purpose, is to examine the goodness of fit as well as the predictive accuracy of overrun models. For this reason, after the discussion of the appropriate theory, a modified ARIMA model is presented, which utilizes EWMA weights, with the goal of showing that such a model is suitable for both modeling and predicting returns values.

The thesis is organized as follows:

In Chapter 1, we discuss the basic aspects of time–series modeling. The ARIMA model, as well as its subcategories, are presented accompanied, for the shake of completeness, by model selection criteria.

In Chapter 2, some well–known methods of Backtesting and Stresstesting are being discussed.

In Chapter 3, a new 5-step algorithmic procedure is proposed, which implements a modified ARIMA(1,1,0) model with EWMA weights, for modeling and forecasting stock returns. To that end, an experimental study is applied to the returns of a portfolio in order to test the fitting and forecasting capability of the proposed algorithmic procedure. The results are compared to the ones of a typical ARIMA model with interesting conclusions.

Finally, in the Appendix section, the R–code, that was developed within the context of the application of Chapter 3, is cited.

1 Μονομεταβλητή Ανάλυση Χρονοσειρών

Η ανάλυση χρονολογικών σειρών αποτελεί μία από τις σημαντικότερες τεχνικές της επιστήμης του Data Science. Η διαδικασία αυτή στοχεύει στην αναζήτηση μαθηματικού μοντέλου το οποίο περιγράφει το υπό εξέταση φαινόμενο και ταυτόχρονα να έχει καλή προβλεπτική ικανότητα.

Ένας από τους βασικότερους στόχους της ανάλυσης χρονοσειρών, είναι η δημιουργία εύρεσης ενός μοντέλου που εκτός από το να περιγράψει την συμπεριφορά της χρονοσειράς είναι ικανό να προβλέψει την συμπεριφορά της στο μέλλον. Η ακριβής πρόβλεψη θεωρείται αποτέλεσμα της ανάλυσης αλλά και της χρήσης των διαθέσιμων ιστορικών πληροφοριών–δεδομένων.

Τα μοντέλα πρόβλεψης χρονοσειρών διακρίνονται σε γραμμικά και μη γραμμικά ανάλογα με το αν η συνάρτηση, η οποία περιγράφει το εκάστοτε μοντέλο είναι γραμμική ή όχι. Επιπρόσθετα, οι μέθοδοι πρόβλεψης χρονολογικών σειρών μπορούν να διαχωριστούν σε **υποκειμενικές** ή **ποιοτικές** (subjective or qualitative) και σε **αντικειμενικές** ή **ποσοτικές** (objective or quantitative). Στην πρώτη κατηγορία οι προβλέψεις εφαρμόζονται από έμπειρους αναλυτές, οι οποίοι χρησιμοποιούν κατά κύριο λόγο την προσωπική τους κρίση, παρά τις μαθηματικές–στοχαστικές μεθόδους. Από την άλλη πλευρά, οι αντικειμενικές μέθοδοι στηρίζονται αποκλειστικά σε κάποιο μαθηματικό ή στατιστικό κανόνα που εφαρμόζεται σε ποσοτικά δεδομένα. Μία ακόμη ιδιαίτερα σημαντική διάκριση μεταξύ των μοντέλων πρόβλεψης χρονοσειρών είναι ο διαχωρισμός τους αυτή σε **αιτιατά** (casual) και **μη–αιτιατά** (non–casual). Αιτιατά ονομάζονται τα υποδείγματα στα οποία οι προβλέψεις μια μεταβλητής βασίζονται σε μια οικονομική ή στατιστική σχέση που συνδέει την μεταβλητή αυτή με άλλες σχετιζόμενες μεταβλητές. Χαρακτηριστικό παράδειγμα τέτοιων υποδειγμάτων είναι τα οικονομετρικά υποδείγματα μιας εξίσωσης ή συστημάτων εξισώσεων. Μη–αιτιατά καλούνται τα υποδείγματα, τα οποία στηρίζονται αποκλειστικά και μόνο στις προηγούμενες τιμές της ίδιας χρονοσειράς προκειμένου να γίνει η πρόβλεψη.

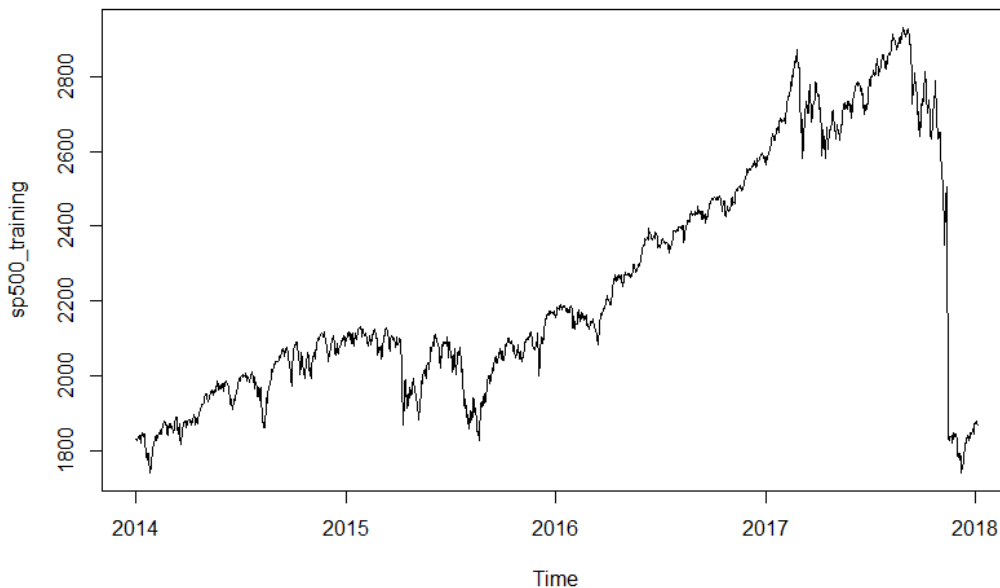
1.1 Εισαγωγή στις Χρονοσειρές

Ως **χρονοσειρά**, χαρακτηρίζεται το σύνολο των δεδομένων που συλλέγονται ανά τακτά χρονικά διαστήματα και υποδηλώνουν την αλλαγή στις τιμές μιας μεταβλητής κατά την διάρκεια χρονικών διαδοχικά ίσων περιόδων. Θα μπορούσε να θεωρηθεί λοιπόν ως μια στοχαστική διαδικασία καθώς οι τυχαίοι παράγοντες επιδρούν στις τιμές και η κάθε τιμή διαφορετικών χρονικών στιγμών αποτελεί μια τυχαία μεταβλητή.

Αν συμβολίσουμε τη χρονολογική σειρά με Y_t και με τον δείκτη t δηλώσουμε τη χρονική στιγμή που ελήφθη η παρατήρηση τότε η ακολουθία τιμών:

$$\{Y_t\}, t = 1, 2, \dots, n,$$

αποτελεί ένα δείγμα n παρατηρήσεων της χρονολογικής σειράς Y_t . (Δημέλη, 2013)



Γράφημα 1.1.1: Χρονοσειρά δείκτη S&P500

1.2 Τύποι και Χαρακτηριστικά Χρονοσειρών

Οι χρονοσειρές διακρίνονται σε:

- **Διακριτού χρόνου** όταν οι τυχαίες μεταβλητές λαμβάνονται μόνο σε συγκεκριμένες στιγμές συνήθως με την ίδια συχνότητα, ακόμα και αν η μεταβλητή είναι συνεχής.
- **Συνεχούς χρόνου** όπου οι τυχαίες μεταβλητές λαμβάνονται από έναν συνεχή χρόνο, για παράδειγμα μια δυαδική διαδικασία σε έναν συνεχή χρόνο είναι μια συνεχής χρονοσειρά.

Μια χρονοσειρά μπορεί να εκφραστεί ως το άθροισμα τεσσάρων συνιστωσών (προσθετικό υπόδειγμα):

$$Y_t = T_t + S_t + C_t + I_t,$$

όπου:

- T: Τάση
- S: Εποχικότητα
- C: Κυκλικότητα
- I: Άρρυθμη Συνιστώσα.

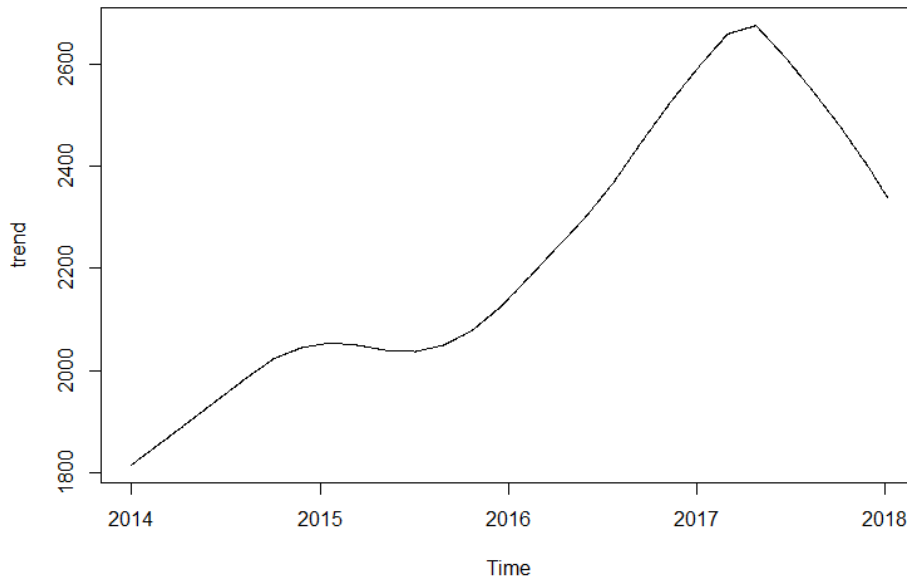
Τάση: Η μακροπρόθεσμη μεταβολή του μέσου επιπέδου των τιμών σε μια χρονοσειρά. Η τάση μπορεί να χαρακτηριστεί ως ανοδική καθοδική δεδομένου ενός συγκεκριμένου χρονικού διαστήματος, και η εκτίμηση της μπορεί να γίνει από διάφορες οικογένειες καμπυλών, (π.χ. μια ευθεία γραμμή ή μια εκθετική καμπύλη). Συνήθως ένας ικανός αριθμός παρατηρήσεων επιφέρει πιο ασφαλή και ακριβή συμπεράσματα για την ύπαρξη τάσης ή όχι στη χρονοσειρά Y_t , ενώ η τάση, η εποχικότητα και η κυκλικότητα, (T_t , S_t , C_t) χαρακτηρίζονται ως ανεξάρτητοι δείκτες από τις μονάδες μέτρησης. (Chatfield, 1984)

Η τάση μπορεί να οριστεί, με την βοήθεια της μεταβλητής του χρόνου t ως:

$$Y_t = f(t) + \varepsilon_t, \quad t = 1, 2, \dots, n,$$

όπου:

- $f(t)$ μια γραμμική ή μη συνάρτηση του χρόνου
- ε_t το τυχαίο σφάλμα με $E(\varepsilon_t) = 0$ και $var(\varepsilon_t) = E(\varepsilon_t^2) = \sigma_\varepsilon^2$.

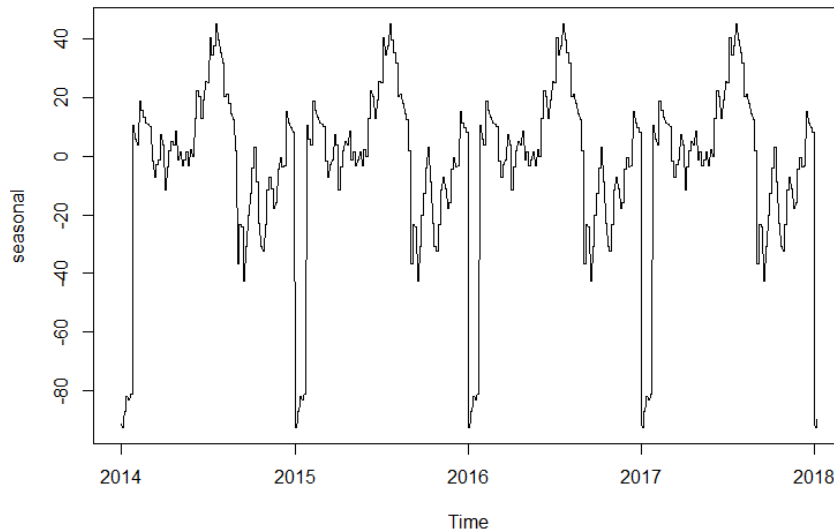


Γράφημα 1.2.1: Διάγραμμα τάσης

Κυκλικότητα: Η μεταβολή λόγω εξωγενών παραγόντων σε μεγάλες περιόδους, οι περίοδοι αυτοί είναι μεγαλύτεροι του έτους, πενταετίας, δεκαετίας δίχως όμως αυτό να σημαίνει ότι χαρακτηρίζεται σταθερού μήκους. Παρατηρώντας τις γραφικές παραστάσεις των χρονοσειρών η κυκλικότητα εμφανίζεται ως μια κυματοειδής γραμμή που τοποθετείται ανάμεσα στην υψηλότερη και χαμηλότερη στάθμη, με μεγαλύτερη εμφάνιση στις οικονομικές χρονοσειρές όπως το ακαθάριστο εθνικό προϊόν εξαιτίας των μεταβολών που παρουσιάζουν οι οικονομίες.

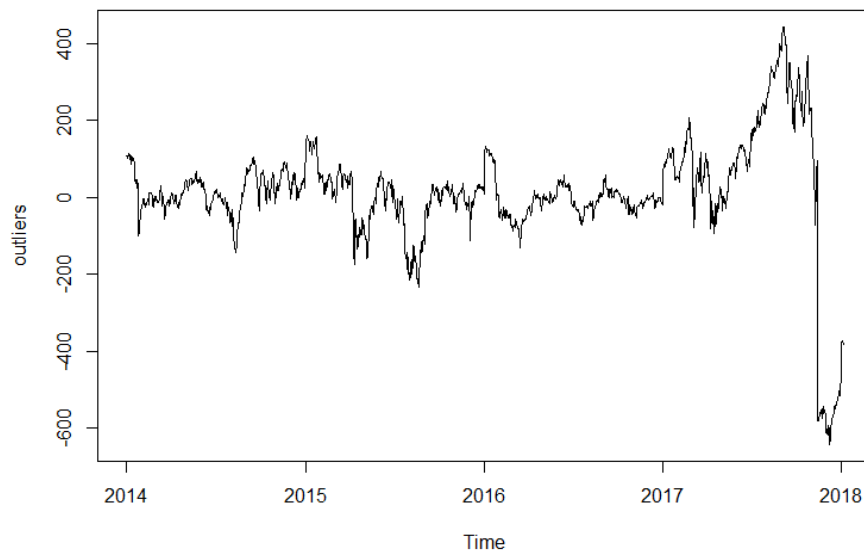
Εποχικότητα: Η περιοδική διακύμανση με σταθερό και μικρότερο, ή ακόμα και ίσος μήκος ενός έτους. Γίνεται άμεσα κατανοητή και προβλέψιμη, εφόσον τα δεδομένα σε μια χρονοσειρά παρουσιάζουν επανάληψη με το ίδιο μοτίβο σε κάθε διάστημα. Όταν η εποχική διακύμανση παρουσιάζεται με συστηματικά, είναι ένα χαρακτηριστικό το οποίο είναι εύκολα παρατηρήσιμο αλλά και μπορεί να απομονωθεί για την αποφυγή επιρροής στα

δεδομένα. Η χρονοσειρά στην οποία απουσιάζει η εποχικότητα ονομάζεται αποεποχικοποιημένη.



Γράφημα 1.2.1: Διάγραμμα εποχικότητας

Άρρυθμη Συνιστώσα: η άρρυθμη συνιστώσα ή αλλιώς θόρυβος εκφράζει το συνολικό αποτέλεσμα των μη συστηματικών παραγόντων. Οι επιπλέον και οι στιγμιαίες ακραίες τιμές που εμφανίζονται συσχετίζονται στην άρρυθμη συνιστώσα. (Μηλιώνης, 2017)



Γράφημα 1.2.2: Διάγραμμα άρρυθμης συνιστώσας

1.3 Στασιμότητα–Μη Στασιμότητα

Στάσιμη χρονοσειρά ονομάζεται εκείνη της οποίας οι στατιστικές ιδιότητες παραμένουν αναλλοίωτες στο χρόνο. Διακρίνεται σε **ασθενώς στάσιμη** και **αυστηρώς στάσιμη**, οι ορισμοί των οποίων είναι οι ακόλουθοι:

Ορισμός

Μια χρονοσειρά Y_t χαρακτηρίζεται ως **ασθενώς στάσιμη**, αν η μέση τιμή μ_t και η αυτοσυνδιακύμανση j -στης τάξης γ_{it} των τυχαίων παρατηρήσεων Y_t και Y_{t-j} δεν εξαρτώνται από τη χρονική στιγμή t .

$$M_t = E(Y_t) = \mu, \quad t = 1, 2, \dots, n,$$

$$Var(Y_t) = \sigma^2, \quad t = 1, 2, \dots, n,$$

$$\gamma_{it} = E(Y_t - \mu_t)(Y_{t+j} - \mu_{t+j}) = \gamma_j, \quad t = 1, 2, \dots, n.$$

Σύμφωνα με τον παραπάνω ορισμό, εφόσον η μέση τιμή μ_t των παρατηρήσεων της χρονοσειράς πρέπει να είναι σταθερή, φτάνουμε στο συμπέρασμα ότι για να είναι μια χρονοσειρά ασθενώς στάσιμη δεν μπορεί να έχει αυξητική (καθοδική) πορεία.

Ορισμός

Μια χρονοσειρά ονομάζεται **αυστηρώς στάσιμη** αν για οποιοσδήποτε χρονικές τιμές μετατοπίσεων j_1, j_2, \dots, j_n η από κοινού κατανομή των τυχαίων μεταβλητών $(Y_t, Y_t + j_1, Y_t + j_2, \dots, Y_t + j_n)$ εξαρτάται μόνο από τις χρονικές μετατοπίσεις και όχι από τη χρονική στιγμή t .

1.4 Υπόδειγμα Τυχαίου Περιπάτου–Λευκός Θόρυβος

Μια χρονοσειρά χαρακτηρίζεται ως τυχαίος περίπατος, αν αφαιρεθεί η τάση με την βοήθεια της διαφοροποίησης όπου το κάθε ένα από τα στοιχεία της y_t είναι αποτέλεσμα της προσθήκης (μη προβλέψιμου όρου), στο προηγούμενο στοιχείο της y_{t-1} , μιας τυχαίας μεταβλητής ε_t . Η γενική μορφή ενός υποδείγματος τυχαίου περιπάτου είναι η ακόλουθη:

$$Y_t = Y_{t-1} + \varepsilon_t, \quad t = 1, 2, \dots, n,$$

όπου ε_t είναι ο λευκός θόρυβος μηδενικού μέσου.

Το πιο απλό μοντέλο χρονολογικής σειράς είναι αυτό του τυχαίου περιπάτου. Μια σειρά είναι λευκός θόρυβος αν δεν έχει κάποιο συγκεκριμένο σχήμα ή πρότυπο. Συμβολίζοντας με ε_t την σειρά τότε θα ονομάζεται λευκός θόρυβος αν έχει μέσο σταθερό (συνήθως είναι ίσο με το μηδέν), σταθερή διακύμανση, και οι τιμές της δεν αυτοσυσχετίζονται, δηλαδή:

- $E(\varepsilon_t) = 0$, για όλα τα t
- $\gamma_0 = E(\varepsilon_t)^2 = \sigma_t^2$, για όλα τα t
- $\gamma_k = E(\varepsilon_t \varepsilon_{t-k}) = 0$, για όλα τα t και $k \neq 0$.

Με βάση τα παραπάνω, η διαδικασία του λευκού θορύβου χαρακτηρίζεται εξ ορισμού στάσιμη. Θα μπορούσε να θεωρηθεί και ως μια στοχαστική διαταραχή της οποίας οι τιμές δεν επηρεάζονται από τις παρελθούσες τιμές της σειράς, και ταυτόχρονα δεν κάποια επίδραση στις μελλοντικές τιμές. Ένα χαρακτηριστικό παράδειγμα λευκού θορύβου αποτελούν τα κατάλοιπα μιας παλινδρόμησης. (Δημέλη, 2013)

1.5 Αυτοπαλίνδρομα Ολοκληρωμένα Κινητού Μέσου Υποδείγματα

1.5.1 Αυτοπαλίνδρομο υπόδειγμα πρώτης τάξεως (AR(1))

Το αυτοπαλίνδρομο υπόδειγμα πρώτης τάξης AR(1) ορίζεται ως εξής:

$$Y_t = \delta + \alpha Y_{t-1} + \varepsilon_t, \quad t = 1, 2, \dots, n,$$

όπου:

- δ η μέση τιμή της AR(1) διαδικασίας
- α η παράμετρος της AR(1) διαδικασίας
- ε_t ο λευκός θόρυβος ανεξάρτητος της Y_{t-1} .

Η λύση της AR(1) επιτυγχάνεται χάρη στη μέθοδο διαδοχικών αντικαταστάσεων. Θεωρώντας το Y_0 ως αρχική τιμή η τελική μορφή καταλήγει χρησιμοποιώντας την ακολουθία τιμών Y_1, Y_2, \dots, Y_t αν γίνει αντικατάσταση στην παραπάνω σχέση, δηλαδή:

$$\begin{aligned} Y_1 &= \delta + \alpha Y_0 + \varepsilon_1 \\ Y_2 &= \delta + \alpha(Y_1 = \delta + \alpha Y_0 + \varepsilon_1) + \varepsilon_2 = \delta(1 - \alpha) + \alpha^2 Y_0 + \alpha \varepsilon_1 + \varepsilon_2 \\ &\quad \cdot \\ &\quad \cdot \\ &\quad \cdot \\ Y_t &= \delta \sum_{j=0}^{t-1} \alpha^j + \alpha^t Y_0 + \sum_{j=0}^{t-1} \alpha^j \varepsilon_{t-j}. \end{aligned}$$

Η μορφή αυτή χαρακτηρίζεται ως αναδρομική και συμβάλλει στην μελέτη των συνθήκων στασιμότητας. (Δημέλη, 2013)

1.5.2 Αυτοπαλίνδρομο υπόδειγμα τάξεως p (AR(p))

Η γενική μορφή ενός αυτοπαλίνδρομου υποδείγματος τάξεως p AR(p), σε αποκλίσεις από το μέσο, έχει τη μορφή:

$$y_t = a_1 y_{t-1} + a_2 y_{t-2} + \dots + a_p y_{t-p} + \varepsilon_t, \quad t = 1, 2, \dots, n, \quad (1)$$

όπου:

- ε_t είναι ο λευκός θόρυβος και τα
- $\alpha_1, \dots, \alpha_p$ χαρακτηρίζονται ως παράμετροι του μοντέλου με $\alpha_p \neq 0$ για μια διαδικασία τάξεως p .

Γράφοντας το υπόδειγμα AR(p) σε όρους L έχουμε:

$$(1 - \alpha_1 L - \dots - \alpha_p L^p)y_t = \varepsilon_t \quad \text{ή} \quad A(L)y_t = \varepsilon_t,$$

όπου το y_t έχει μέση τιμή ίση με το μηδέν.

Αναλύοντας το χαρακτηριστικό πολυώνυμο σε παράγοντες έχουμε:

$$A(L) = (1 - \lambda_1 L) \dots (1 - \lambda_p L) = \prod_{j=1}^p (1 - \lambda_j L), \quad (2)$$

που $\lambda_1^{-1}, \dots, \lambda_p^{-1}$ είναι οι ρίζες του πολυωνύμου $A(L) = 0$, το υπόδειγμα γράφεται ως:

$$y_t = A^{-1}(L)\varepsilon_t. \quad (3)$$

Με αντικατάσταση της (2) στην (3) και αναπτύσσοντας το πολυώνυμο σε επιμέρους κλάσματα έχουμε:

$$y_t = \frac{1}{(1-\lambda_1 L)\dots(1-\lambda_p L^p)} \varepsilon_t = \sum_{j=1}^p \frac{\xi_j}{(1-\lambda_j L)} \varepsilon_t.$$

Η παραπάνω διαδικασία συγκλίνει μόνο αν $|\lambda_i| \leq 1$ για $i = 1, 2, \dots, p$. Οι συνθήκες αυτές ισοδυναμούν με τις συνθήκες $|\lambda_i^{-1}| > 1$ για $i = 1, 2, \dots, p$. Θα πρέπει δηλαδή οι ρίζες της χαρακτηριστικής εξίσωσης του AR(p) να είναι εντός του μοναδιαίου κύκλου ή εναλλακτικά οι ρίζες του πολυωνύμου $A(L)$ να είναι εκτός του μοναδιαίου κύκλου. (Δημέλη, 2013)

Η μέση τιμή σε μια στάσιμη AR(p) ορίζεται ως εξής:

$$\mu = E(Y_t) = \frac{\delta}{(1 - \alpha_1 - \alpha_2 - \dots - \alpha_p)}.$$

Αν πολλαπλασιαστεί η σχέση (1) με το y_t κατά μέλη και κάνοντας χρήση των αναμενόμενων τιμών, προκύπτει η εξίσωση διακύμανσης:

$$\gamma_0 = \alpha_1\gamma_1 + \alpha_2\gamma_2 + \dots + \alpha_p\gamma_p + \sigma_\varepsilon^2.$$

Θέτοντας $\gamma_k = \rho_k\gamma_0$ και $k=1,2, \dots$ προκύπτει η διακύμανση της Y_t σε συνάρτηση των α_j , $j = 1,2, \dots, p$,

$$\gamma_0 = \frac{\sigma_\varepsilon^2}{(1-\alpha_1\rho_1-\alpha_2\rho_2-\dots-\alpha_p\rho_p)}.$$

1.5.3 Υπόδειγμα κινητού μέσου πρώτης τάξεως (MA(1))

Η χρονοσειρά MA(1) πρώτης τάξης ορίζεται από την ακόλουθη σχέση:

$$Y_t = \delta + \varepsilon_t - \theta_1\varepsilon_{t-1}, t=1,2, \dots, n.$$

Η Y_t παρατήρηση είναι εξαρτημένη από τον όρο του σφάλματος ε_t καθώς και από το προηγούμενο σφάλμα ε_{t-1} και ο συντελεστής θ_1 δέχεται τιμές από -1 έως $+1$.

Για την χρονοσειρά πρώτης τάξης κινητού μέσου MA(1) ισχύουν τα εξής:

- $\mu = E(Y_t) = E(\mu + \varepsilon_t + \theta_1\varepsilon_{t-1}) = \mu + E(\varepsilon_t) + \theta_1E(\varepsilon_{t-1}) = \mu$
- $\gamma_0 = E(Y_t - \mu)^2 = E(\varepsilon_t + \theta_1\varepsilon_{t-1})^2 = (1 + \theta_1^2)\sigma^2$
- $\gamma_1 = E(Y_t - \mu)(Y_{t-1} - \mu) = E(\varepsilon_t + \theta_1\varepsilon_{t-1})(\varepsilon_{t-1} + \theta_1\varepsilon_{t-2}) = \dots = \theta_1\sigma^2, s = 1.$

Οι αυτοδιακυμάνσεις δεύτερης τάξης και πάνω είναι όλες μηδενικές:

- $\gamma_s = E(Y_t - \mu)(Y_{t-s} - \mu) = E(\varepsilon_t + \theta_1\varepsilon_{t-1})(\varepsilon_{t-s} + \theta_1\varepsilon_{t-s-1}) = \dots = 0, s > 1.$

Η μέση τιμή και οι αυτοδιακυμάνσεις φαίνεται πως δεν έχουν εξάρτηση από το χρόνο t , συνεπώς ανεξάρτητα της τιμής της παραμέτρου θ , η χρονοσειρά MA(1) είναι στάσιμη.

1.5.4 Υπόδειγμα κινητού μέσου τάξεως q (MA(q))

Τα υποδείγματα κινητού μέσου είναι χρήσιμα στις περιπτώσεις, όπου τα γεγονότα δημιουργούν αποτελέσματα τα οποία έχουν διάρκεια στο χρόνο παρόλο που το γεγονός αυτό καθ' αυτό έχει πάψει να υφίσταται.

Μια χρονοσειρά Y_t ονομάζεται χρονοσειρά κινητού μέσου τάξεως q MA(q), όταν κάθε παρατήρηση Y_t εκφράζεται ως ένα σταθμισμένο άθροισμα μια σταθερά μ , μιας χρονοσειράς λευκού θορύβου και q χρονικές υστερήσεις της χρονοσειράς λευκού θορύβου. Γενικά, ένα υπόδειγμα MA(q) χρησιμοποιείται όταν οι συντελεστές συσχέτισης πηγαίνουν εκθετικά προς στο μηδέν και εμφανίζονται στατιστικά q συντελεστές αυτοσυσχέτισης.

Η γενική μορφή του υποδείγματος MA(q) είναι η ακόλουθη:

$$Y_t = \delta + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \dots - \theta_q \varepsilon_{t-q}, t = 1, 2, \dots, n.$$

Η παραπάνω σχέση είναι μια χρονοσειρά λευκού θορύβου, ενώ οι παράμετροι μ και $\theta_1, \theta_2, \dots, \theta_q$ μπορούν να είναι οποιοδήποτε πραγματικοί αριθμοί. Τα $\theta_1, \theta_2, \dots, \theta_q$ χρησιμεύουν ως βαρύτητες έτσι ώστε μεγάλες χρονικές υστερήσεις του ε_t να έχουν μικρότερη συνεισφορά στην επεξήγηση της συμπεριφοράς της υπό-εξέτασης χρονοσειράς. (Δημέλη, 2013)

Κάνοντας χρήση του τελεστή υστέρησης L η παραπάνω σχέση μπορεί να γραφτεί ως εξής:

$$Y_t = \delta + (1 - \theta_1 B - \dots - \theta_q B^q) \varepsilon_t = \delta + \theta(B) \varepsilon_t,$$

όπου το μέρος $\theta(B) = 1 - \theta_1 B - \dots - \theta_q B^q$ είναι το χαρακτηριστικό πολυώνυμο της MA(q).

Για την χρονοσειρά κινητού μέσου τάξης q MA(q) ισχύουν τα εξής:

- $\delta = E(Y_t)$
- $\gamma_0 = (1 + \theta_1^2 + \dots + \theta_q^2) \sigma_\varepsilon^2$

$$\bullet \gamma_k = \begin{cases} (-\theta_k + \theta_1\theta_{k+1} + \dots + \theta_{q-k}\theta_q)\sigma_\varepsilon^2, & k = 1, \dots, q \\ 0, & k > q \end{cases} .$$

1.5.5 Το υπόδειγμα ARMA(p,q)

Η μορφή ενός ARMA υποδείγματος είναι η ακόλουθη:

$$Y_t = \delta + \alpha_1 Y_{t-1} + \alpha_2 Y_{t-2} + \dots + \alpha_p Y_{t-p} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \dots - \theta_q \varepsilon_{t-q},$$

- δ η μέση τιμή της ARMA(p,q) διαδικασίας
- $\alpha_1, \alpha_2, \dots, \alpha_p$ οι συντελεστές του αυτοπαλίνδρομου μέρους (autoregressive, AR)
- $\theta_1, \theta_2, \dots, \theta_q$ οι συντελεστές των κινητών μέσων (moving average, MA)
- ε_t ο λευκός θόρυβος.

Θεωρείται ένα υπόδειγμα παλινδρόμησης με εξαρτημένη μεταβλητή Y_t και τις προηγούμενες τιμές της χρονοσειράς ως ερμηνευτικές μεταβλητές (υστερήσεις p-τάξεως της Y_t) καθώς και έναν αριθμό προηγούμενων λαθών (υστερήσεις q-τάξεως της τυχαίας μεταβλητής ε_t , αυτοπαλίνδρομο υπόδειγμα κινητών μέσων τάξης (p,q) και συμβολίζεται ως ARMA (p,q).

Ο τελεστής χρονικής υστέρησης συναντάται συχνά στις χρονοσειρές και ορίζεται ως εξής:

$$L^j Y_t = Y_{t-j}, \quad j = 1, 2, \dots, n.$$

Δηλαδή, ο L μετατοπίζει την χρονοσειρά j-χρονικές περιόδους προς τα πίσω

$$A(L)Y_t = \delta + \Theta(L)\varepsilon_t,$$

όπου τα A(L) και Θ(L) χαρακτηρίζουν τα πολυώνυμα του L τάξης p και q

$$A(L) = 1 - \alpha_1 L - \alpha_2 L^2 - \dots - \alpha_p L^p$$

και

$$\Theta(L) = 1 - \theta_1 L - \theta_2 L^2 - \dots - \theta_p L^p.$$

Το AR μέρος χαρακτηρίζει την στασιμότητα της χρονοσειράς ενώ το MA χαρακτηρίζει την αντιστρεψιμότητα της χρονοσειράς ARMA. Η στασιμότητα ορίζεται όταν το πολυώνυμο $A(L)$ έχει ρίζες εκτός του μοναδιαίου κύκλου, ενώ είναι αντιστρέψιμη όταν το χαρακτηριστικό πολυώνυμο έχει ρίζες εκτός του μοναδιαίου κύκλου.

1.5.6 Η γενική περίπτωση του ARIMA (p,d,q) υποδείγματος

Σε πολλές περιπτώσεις, οι μη στάσιμες χρονολογικές σειρές υφίστανται ανά διαστήματα μεταβολές στο μέσο διατηρώντας παράλληλα τον ομοιόμορφο χαρακτήρα τους. Αυτό σημαίνει ότι το επίπεδο του μέσου μεταβάλλεται στην πάροδο του χρόνου, αλλά η συμπεριφορά της σειράς παραμένει ενιαία στα μεμονωμένα χρονικά διαστήματα. Οι Box and Jenkins παρατήρησαν πρώτοι ότι το φαινόμενο αυτό παρουσιάζεται στις χρονολογικές σειρές, οι οποίες αφού λάβουν διαφορίση κατάλληλης τάξης, μετατρέπονται σε μία στάσιμη στοχαστική διαδικασία. Προκύπτει λοιπόν, μια ιδιαίτερη κατηγορία υποδειγμάτων στα οποία κατατάσσονται όσα υποδείγματα μετά από διαφορές τάξεως d ακολουθούν μια στάσιμη διαδικασία ARMA(p,q). Ο λόγος για τα **αυτοπαλίνδρομα ολοκληρωμένα κινητού μέσου υποδείγματα** (Autoregressive Integrated Moving Average) (p,d,q) που συμβολίζονται ως ARIMA(p,d,q). Αξίζει να σημειωθεί ότι η συγκεκριμένη περίπτωση αποτελεί διεύρυνση των μη αιτιωδών υποδειγμάτων πρόβλεψης.

Αν μελετήσουμε πιο προσεκτικά τη γενική μορφή του ARIMA υποδείγματος, θεωρώντας ότι η ρίζα της Y_t ισούται με τη μονάδα, υπάρχει η δυνατότητα της αλγεβρικής παραγοντοποίησης της ως εξής:

$$\Phi(L)Y_t = A(L)(1 - L)Y_t = \delta + \Theta(L)\varepsilon_t, \quad (4)$$

όπου το $\Phi(L) = A(L)(1 - L)$ αναπαριστά το “μη στάσιμο” πολυώνυμο, του οποίου οι ρίζες βρίσκονται εντός ή επί του μοναδιαίου κύκλου, ενώ το $A(L)$ συμβολίζει το “στάσιμο” πολυώνυμο, του οποίου οι ρίζες βρίσκονται εκτός αυτού. Το υπόδειγμα συχνά παρουσιάζεται και ως:

$$A(L)\nabla Y_t = \delta + \theta(L)\varepsilon_t.$$

Μια χρονολογική σειρά ονομάζεται ολοκληρωμένη (integrated) τάξεως d ($I(d)$) αν παίρνει στάσιμη μορφή μετά από το πλήθος διαφορίσεις. Το υπόδειγμα (4) χαρακτηρίζεται, λοιπόν, ως ένα ολοκληρωμένο $ARMA(p,q)$ υπόδειγμα πρώτης τάξης και συμβολίζεται ως $ARIMA(p,1,q)$.

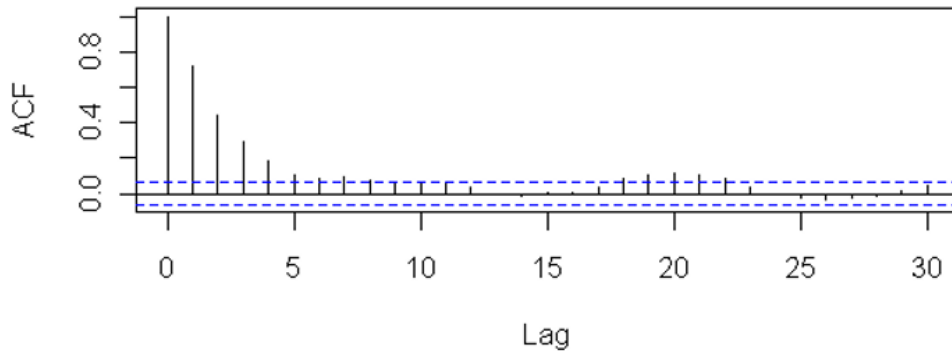
Καταλήγουμε ότι αναφερόμαστε σε αυτοπαλίνδρομο ολοκληρωμένο υπόδειγμα κινητών μέσων όταν το πολυώνυμο $\Phi(L)$ έχει d ρίζες ίσες με μονάδα, ενώ παράλληλα οι υπόλοιπες υπάρχουν εκτός του μοναδιαίου κύκλου,

$$\Phi(L)Y_t = A(L)(1 - L)^d Y_t = \delta + \theta(L)\varepsilon_t.$$

Διαπιστώνουμε, συνεπώς, ότι το μη στάσιμο $ARIMA$ υπόδειγμα της Y_t διακρίνεται από την εξής ιδιότητα: έχει την δυνατότητα να μετασχηματιστεί σε ένα στάσιμο $ARMA$ υπόδειγμα της στοχαστικής διαδικασίας $(1 - L)^d Y_t$, δηλαδή των διαφορών d τάξεως της Y_t . (Δημέλη, 2013)

1.5.7 Συνάρτηση αυτοσυσχέτισης (ACF)

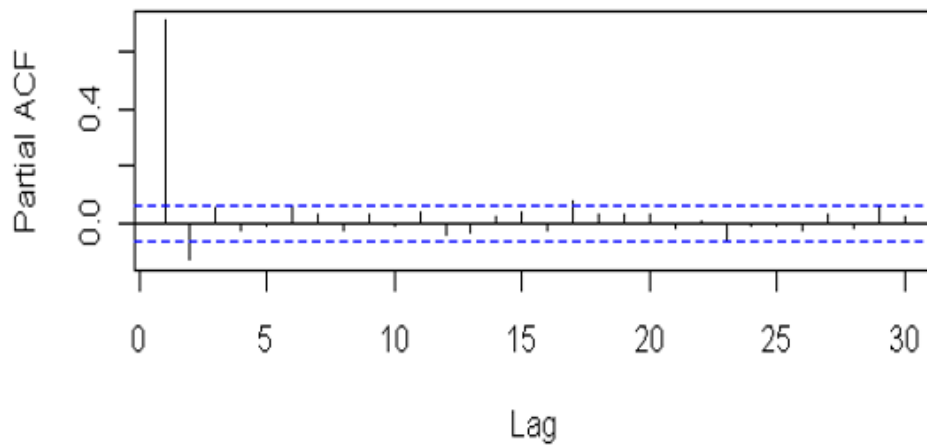
Η ACF περιγράφει πόσο σχετίζεται η σημερινή τιμή της σειράς με τις προηγούμενες τιμές της. Μια χρονοσειρά, όπως έγινε αναφορά και σε προηγούμενο υποκεφάλαιο, εμπεριέχει στοιχεία όπως αυτά της τάσης και της εποχικότητας. Μέσω του κορελογράμματος ο ερευνητής καλείται να διακρίνει και στη συνέχεια να εξαλείψει τα όποια προβλήματα. Για παράδειγμα, αν τα spikes της ACF φθίνουν αργά, τότε είναι πιθανό να υπάρχει παρουσία τάσης στα δεδομένα.



Γράφημα 1.5.7.1: Κορελόγραμμα ACF AR(2) προσομοίωσης

1.5.8 Συνάρτηση μερικής αυτοσυσχέτισης (PACF)

Ένας τρόπος για τον προσδιορισμό της τάξης της AR διαδικασίας είναι μέσω του κορελογράμματος μερικής αυτοσυσχέτισης. (Brockwell και Davis, 1996) Η συνάρτηση μερικής αυτοσυσχέτισης (PACF) παρέχει τόσες στατιστικά σημαντικές αυτοσυσχετίσεις όσος είναι και ο βαθμός της τάξης της AR(p) διαδικασίας. Η PACF έχει τα ίδια χαρακτηριστικά με την ACF με την μόνη διαφορά ότι αναφέρεται στην AR διαδικασία. (Μηλιώνης, 2017)



Γράφημα 1.5.8.1: Κορελόγραμμα PACF AR(2) προσομοίωσης

1.5.9 Μέθοδος ανάλυσης χρονοσειράς Box–Jenkins

Η μέθοδος Box–Jenkins δημιουργήθηκε το 1976 με σκοπό την εύρεση του καταλληλότερου μοντέλου για την πρόβλεψη χρονοσειρών και αποτελείται από μια σειρά τεσσάρων βημάτων:

1. **Ταυτοποίηση (Identification):** Με τον όρο ταυτοποίηση ενός υποδείγματος αναφερόμαστε στον προσδιορισμό της τάξης της μη στασιμότητας και της τάξης των AR και MA πολυωνύμων. Αυτό είναι εφικτό έπειτα από μελέτη και σύγκριση των δειγματικών συναρτήσεων αυτοσυσχέτισης και μερικής αυτοσυσχέτισης με άπειρο πλήθος όρων. Με αυτό τον τρόπο ταυτοποιείται ένα δοκιμαστικό υπόδειγμα. (Μηλιώνης, 2017) Αν η μη στασιμότητα οφείλεται στη δεύτερη ροπή, τότε η σειρά θα πρέπει αρχικά να μετατραπεί σε στάσιμη ως προς την δεύτερη ροπή και έπειτα να διεξαχθεί μελέτη ως προς τον βαθμό ολοκλήρωσης d .
2. **Εκτίμηση (Estimation):** Διεξάγεται εκτίμηση και κατ' επέκταση επιλογή των συντελεστών του ARIMA μοντέλου.
3. **Διαγνωστικός Έλεγχος (Diagnostic Checking):** Έλεγχος με την χρήση κριτηρίων πληροφορίας για τον βαθμό καταλληλότητας το μοντέλου ARIMA που επιλέχθηκε στο προηγούμενο στάδιο.
4. **Πρόβλεψη (Forecasting):** Τέλος λαμβάνει χώρα η πρόβλεψη των μελλοντικών τιμών της χρονοσειράς με βάση του επιλεχθέντος μοντέλου των προηγούμενων βημάτων.

1.6 Exponentially Weighted Moving Average–EWMA

Ο εκθετικά σταθμισμένος κινητός μέσος (EWMA), είναι μια σειρά πρώτης τάξης η οποία εφαρμόζει συντελεστές στάθμισης οι οποίοι μειώνονται εκθετικά. Η στάθμιση για κάθε παλαιότερο στοιχείο μειώνεται εκθετικά χωρίς όμως να φτάσει ποτέ στο μηδέν. (Chou, 1975) Τα παραπάνω επιτυγχάνονται με τη χρήση μιας ρυθμιστικής παραμέτρου λ η οποία φροντίζει να δώσει μεγαλύτερο βάρος στις πιο πρόσφατες παρατηρήσεις. Με αυτό τον τρόπο, ο εκθετικά σταθμισμένος κινητός μέσος είναι ο μέσος όλων των ιστορικών τιμών αλλά η επίδραση των ιστορικών τιμών μειώνεται εκθετικά σε συνάρτηση με τον

χρόνο. Έστω y_t η τιμή της χρονοσειράς την στιγμή t και $\lambda \in (0,1)$ ο παράγοντας εξομάλυνσης. Τότε:

$$\tilde{y}_t = \text{EWMA}(t) = (1 - \lambda)(y_t + \lambda y_{t-1} + \lambda^2 y_{t-2} + \dots).$$

Τα βάρη συνοψίζονται σε μια γεωμετρική σειρά

$$1 + \lambda + \lambda^2 + \lambda^3 + \dots = \frac{1}{1-\lambda},$$

όπου το λ παίρνει τιμές από το $(0,1)$. Ο παράγοντας εξομάλυνσης υπολογίζεται ως εξής:

$$\tilde{y}_t = (1 - \lambda)y_t + \lambda\tilde{y}_{t-1}, t=1, 2, \dots, n.$$

Όταν το λ παίρνει τιμές πλησίον του 0, τότε δίνεται μεγαλύτερη βαρύτητα στην πιο πρόσφατη τιμή ενώ όταν το λ παίρνει τιμές πλησίον του 1, δίνεται μεγαλύτερη βαρύτητα στις παρελθούσες τιμές.

1.7 Διαγνωστικοί Έλεγχοι

1.7.1 Έλεγχος στασιμότητας Dickey–Fuller

Οι έλεγχοι μοναδιαίας ρίζας χρησιμοποιούνται για τη διερεύνηση περί ύπαρξης ή μη στασιμότητας σε δεδομένα χρονολογικών σειρών. Αφορούν μη στάσιμες χρονοσειρές που επιδέχονται μετατροπή σε στάσιμες μετά από σχετική d -διαφορίση. Ως εκ τούτου, τέτοιου είδους έλεγχοι αναφέρονται σε μοντέλα τύπου ARIMA(p,d,q) όπως αυτά αναλύθηκαν στα προηγούμενα υποκεφάλαια.

Το 1979, οι Dickey και Fuller (Dickey και Fuller, 1979) δημιούργησαν ένα test το οποίο εξετάζει την ύπαρξη της μοναδιαίας ρίζας σε υποδείγματα ARIMA. Το test δημιουργήθηκε το 1979 από τους David Dickey και Wayne Fuller. Αρχικά το test δημιουργήθηκε επάνω στο αυτοπαλίνδρομο του πρώτου βαθμού και γενικεύτηκε στην οικογένεια των υποδειγμάτων ARIMA.

Ας θυμηθούμε το AR(1) υπόδειγμα ορίζεται ως εξής:

$$Y_t = \delta + aY_{t-1} + \varepsilon_t, t = 1, 2, \dots, n,$$

όπου ε_t ο λευκός θόρυβος. Αν το $\alpha < 1$, τότε η στοχαστική διαδικασία Y_t χαρακτηρίζεται στάσιμη. Αν όμως το α δέχεται τιμές κοντά στο 1, τότε η σειρά φαίνεται να εμφανίζει μια μη-στάσιμη συμπεριφορά. Ο στόχος του test είναι η εφαρμογή του ελέγχου ύπαρξης μοναδιαίας ρίζας (όχι στασιμότητα) έναντι μη ύπαρξης μοναδιαίας ρίζας (στασιμότητα):

$$H_0: \alpha = 1 \quad \text{vs} \quad H_a: \alpha < 1.$$

Όταν η H_0 δεν απορρίπτεται, δηλαδή $\alpha=1$, τότε η σειρά εμφανίζει μη-στάσιμη συμπεριφορά και έχει διακύμανση $t\sigma_\varepsilon^2$ η οποία τείνει στο άπειρο καθώς οι τιμές του t αυξάνονται.

1.7.2 Έλεγχος κανονικότητας Jarque–Bera

Το Jarque–Bera test (Jarque και Bera, 1987) αποτελεί μια τεχνική ελέγχου καλής προσαρμογής των δεδομένων σε κανονική κατανομή και δίνεται από τον ακόλουθο τύπο

$$JB = \frac{n}{6} \left(S^2 + \frac{1}{4}(K - 3)^2 \right),$$

όπου:

- n συμβολίζεται ο αριθμός των παρατηρήσεων
- S η ασυμμετρία

$$S = \frac{\hat{\mu}_3}{\hat{\sigma}^3} = \frac{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^3}{\left(\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2 \right)^{3/2}}$$

- K η κύρτωση

$$K = \frac{\hat{\mu}_4}{\hat{\sigma}^4} = \frac{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^4}{\left(\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2 \right)^2}.$$

1.7.3 Έλεγχος ανεξαρτησίας Ljung–Box

Το όνομα του δόθηκε προς τιμήν των Greta M. Ljung και George E. P. Box οι οποίοι πρότειναν τον έλεγχο και εφαρμόζεται ευρέως στην επιστημονική περιοχή των χρονοσειρών. Το Ljung–Box (Ljung και Box, 1978) test αποτελεί έλεγχο ανεξαρτησίας των δεδομένων, δηλαδή πότε μια ομάδα αυτοσυσχετίσεων μιας χρονοσειράς είναι διαφορετική ή όχι του μηδενός.

Θέλουμε να εφαρμόσουμε τον έλεγχο:

H_0 : τα δεδομένα είναι μεταξύ τους ανεξάρτητα

έναντι

H_a : τα δεδομένα είναι μεταξύ τους εξαρτημένα

Το Ljung–Box test ορίζεται ως εξής:

$$Q = n(n + 2) \sum_{k=1}^h \frac{\hat{\rho}_k^2}{(n-k)},$$

όπου:

- n : είναι το μέγεθος του δείγματος
- $\hat{\rho}_k^2$: η αυτοσυσχέτιση του δείγματος υστέρησης k
- h : ο αριθμός των υστερήσεων που έχουν ελεγχθεί,

Κάτω από την μηδενική υπόθεση το Q στατιστικό ακολουθεί ασυμπτωτικά ότι ακολουθεί την χ_h^2 και για συγκεκριμένο επίπεδο σημαντικότητας το αντίστοιχο χωρίο απόρριψης της H_0 είναι το ακόλουθο:

$$Q > \chi_{1-a,h}^2.$$

1.7.4 Έλεγχος ετεροσκεδαστικότητας Goldfeld–Quandt

Το test Goldfeld–Quandt (GQ) (Goldfeld και Quandt, 1965) είναι ένα οικονομετρικό test, το οποίο υποθέτει ότι υπάρχει ένα καθοριστικό σημείο το οποίο μπορεί να χρησιμοποιηθεί για τη διαφοροποίηση της διακύμανσης του όρου σφάλματος. Οι παρατηρήσεις των δειγμάτων διαχωρίζονται σε δύο ομάδες, και τα στοιχεία της ετεροσκεδαστικότητας βασίζονται στη σύγκριση των τετραγώνων των καταλοίπων χρησιμοποιώντας το F–στατιστικό το οποίο ορίζεται ως:

$$F = \frac{\frac{SSE_B}{df}}{\frac{SSE_A}{df}},$$

όπου:

- SSE_A, SSE_B τα αθροίσματα των τετραγώνων των σφαλμάτων των ομάδων A και B που χωρίστηκε το αρχικό δείγμα
- $df = \frac{N-c}{2} - p$, με p να αντιπροσωπεύει το πλήθος των παραμέτρων του υποδείγματος, c το πλήθος των παρατηρήσεων που αφαιρέθηκαν, N το μέγεθος του δείγματος.

Η μηδενική υπόθεση του test αναφέρεται στην ομοσκεδαστικότητα. Όσο πιο μεγάλη είναι η τιμή του F–στατιστικού ενισχύεται η υπόθεση της ύπαρξης ετεροσκεδαστικότητας, δηλαδή διαφορετική διακύμανση στις δύο ομάδες. Η μηδενική υπόθεση απορρίπτεται όταν $p\text{-value} < \alpha$. (Pedace, 2013)

1.8 Κριτήρια Επιλογής Μοντέλων

Τα τελευταία χρόνια η στατιστική επιστήμη έδωσε όλο και μεγαλύτερη έμφαση στα κριτήρια επιλογής των μοντέλων. Η ανάγκη δημιουργίας μέτρων/κριτηρίων αξιολόγησης και εν συνέχεια επιλογής του καταλληλότερου μοντέλου ήταν επιτακτική καθώς πολλές φορές ο ερευνητής έρχεται αντιμέτωπος με πολλά διαφορετικά υποδείγματα για την μοντελοποίηση του φαινομένου υπό εξέταση. Τα κριτήρια πληροφορίας, αποτελούν μια

ευρέως χρησιμοποιούμενη επιλογής του καταλληλότερου μοντέλου σε προβλήματα χρονοσειρών.

1.8.1 Akaike's Information Criterion

Το κριτήριο πληροφορίας του Akaike (AIC Akaike, 1978) είναι ένας τρόπος επιλογής ενός μοντέλου ανάμεσα στα υπό εξέταση μοντέλα για ένα συγκεκριμένο σύνολο δεδομένων.

Η γενική μορφή του Akaike's Information Criterion είναι η ακόλουθη:

$$AIC = -2\ln L + 2p,$$

όπου:

- L = η πιθανοφάνεια του μοντέλου
- p = ο αριθμός των συντελεστών παλινδρόμησης
- n = το μέγεθος του δείγματος.

1.8.2 Bayesian Information Criterion

Το Bayesian Information Criterion (BIC) χαρακτηρίζεται και αυτό ως ένα κριτήριο αξιολόγησης των κατ' επιλογή μοντέλων. (Schwarz, 1978) Η μορφή του BIC είναι η ακόλουθη:

$$BIC = -2\ln L + p\ln(n),$$

όπου:

- L = η πιθανοφάνεια του μοντέλου
- p = ο αριθμός των συντελεστών παλινδρόμησης
- n = το μέγεθος του δείγματος.

Το μοντέλο που εμφανίζει την πιο χαμηλή τιμή AIC ή BIC, είναι και αυτό που προτιμάται για χρήση. Σημειώνετε δε, ότι και τα δύο κριτήρια μπορούν να λάβουν αρνητικές τιμές. (Vrieze, 2012)

1.8.3 Hannan–Quinn Information Criterion

Ένα άλλο κριτήριο για την επιλογή του καλύτερου μοντέλου είναι αυτό των Hannan–Quinn (1979) το οποίο λαμβάνει υπόψιν τον αριθμό των ανεξάρτητων μεταβλητών που εμπεριέχουν τα υπό εξέταση μοντέλα καθώς και το μέγεθος του δείγματος. (Hannan και Quinn, 1979)

Το κριτήριο ορίζεται ως εξής:

$$HQC = -2L_{max} + 2p \ln(\ln(n)),$$

όπου:

- L_{max} είναι η log πιθανοφάνεια
- p ο αριθμός των παραμέτρων
- n το μέγεθος του δείγματος.

1.8.4 Μέτρα αξιολόγησης προβλεπτικής ικανότητας μοντέλων

Mean Squared Error (MSE): Το μέσο τετραγωνικό σφάλμα υπολογίζει την μέση τιμή των τετραγώνων των σφαλμάτων του μοντέλου. Πιο συγκεκριμένα, υπολογίζει τη μέση τιμή των τετραγώνων των διαφορών των παρατηρήσεων (observed values) Y_t από τις προβλεπόμενες τιμές (predict values) F_t .

Έτσι το MSE ορίζεται ως:

$$MSE = \frac{\sum_{t=1}^n \varepsilon_t^2}{n},$$

όπου:

$$\varepsilon_t = F_t - Y_t.$$

Root Mean Squared Error (RMSE): Το RMSE είναι μια μέτρηση που χρησιμοποιείται συχνά για την μέτρηση των διαφορών μεταξύ των προβλεπόμενων τιμών του μοντέλου και τις τιμές των παρατηρήσεων. Αποτελεί την τετραγωνική ρίζα του MSE. (Cort, Kenji και Willmott, 2005)

Δηλαδή,

$$RMSE = \sqrt{\frac{\sum_{t=1}^n \varepsilon_t^2}{n}}$$

Mean Absolute Percentage Error (MAPE): Το MAPE είναι ένα μέτρο της προβλεπτικής ακρίβειας όλων των μεθόδων πρόβλεψης.

$$MAPE = \frac{100 * \sum_{t=1}^n \frac{|\varepsilon_t|}{|Y_t|}}{n}, t = 1, 2, \dots, n.$$

Mean Absolute Error (MAE): Το MAE αναφέρεται στη μέτρηση ανάμεσα δύο διαδοχικές παρατηρήσεις.

$$MAE = \frac{\sum_{t=1}^n |\varepsilon_t|}{n}$$

Mean Absolute Scaler Error (MASE): Το MASE είναι ένα μέτρο της μέτρησης προβλεπτικής ικανότητας του υπό εξέτασης μοντέλου.

$$MASE = \frac{\frac{1}{n} \sum_{i=1}^n |\varepsilon_j|}{\frac{1}{n-c} \sum_{t=c+1}^n |y_t - y_{t-c}|}$$

Theil's U-statistic: Το U-στατιστικό του Theil είναι ένα μέτρο ακρίβειας το οποίο δίνει έμφαση στην σημαντικότητα των μεγάλων σφαλμάτων (όπως στο MSE) και στη σύγκριση με άλλες μεθόδους πρόβλεψης (Theil, 1996). Ο τύπος απλοποιήθηκε από τους (Makridakis και Wheelwright, 1998) και η μορφή του είναι η εξής:

$$U = \sqrt{\frac{\sum_{t=1}^{n-1} \left(\frac{(F_{t+1} - Y_{t+1})^2}{Y_t} \right)}{\sum_{t=1}^{n-1} \left(\frac{(F_{t+1} - Y_t)^2}{Y_t} \right)}}$$

όπου:

- F = τιμές πρόβλεψης
- Y = τιμές παρατήρησης.

1.9 Μετασχηματισμός Δεδομένων

Οι μετασχηματισμοί δεδομένων στοχεύουν στη βελτίωση της στατιστικής ανάλυσης μιας χρονοσειράς. Είναι πιθανό τα δεδομένα να μην πληρούν τις απαιτούμενες προϋποθέσεις, όπως αυτή της κανονικότητας και της ομοσκεδαστικότητας σε ένα πρόβλημα γραμμικής παλινδρόμησης, με αποτέλεσμα να απαιτείται κατάλληλος μετασχηματισμός των δεδομένων για την επίτευξη των παραπάνω.

Μερικοί από τους γνωστότερους μετασχηματισμούς αναλύονται στα υποκεφάλαια που ακολουθούν:

1.9.1 Λογαριθμικός μετασχηματισμός

Κάθε μεταβλητή x αντικαθίσταται με το $\log(x)$, όπου η επιλογή της βάσης του λογαρίθμου (Νεπέριος ή δεκαδικός) επαφίεται στον ερευνητή. Σκοπός του μετασχηματισμού αυτού είναι τα δεδομένα να βρεθούν πιο κοντά στην κανονική κατανομή.

Σε ορισμένες περιπτώσεις γίνεται χρήση μιας τροποποίησης του μετασχηματισμού:

$$g(Y_t) = \ln\left(\frac{Y_t}{1-Y_t}\right),$$

η οποία παρουσιάζει ασυμμετρία λόγω της ιδιότητας $\ln(Y_t) = -\ln\left(\frac{1}{Y_t}\right)$. Έτσι προκύπτει ότι

$$\ln\left(\frac{Y_t}{1-Y_t}\right) = -\ln\left(\frac{1-Y_t}{1-(1-Y_t)}\right)$$

και ως αποτέλεσμα οι μετασχηματισμοί της Y_t και της $1 - Y_t$ εμφανίζουν τις ίδιες συναρτήσεις αυτοσυσχέτισης. Σημειώνεται ότι ιδιότητα αυτή έχει ευρεία εφαρμογή σε οικονομικές εφαρμογές. (Δημέλη, 2013)

1.9.2 Μετασχηματισμός Box-Cox

Μια σημαντική κατηγορία μετασχηματισμών, η οποία είναι κατάλληλη για την χρήση στις χρονοσειρές που μετρούνται σε κλίμακα με αυστηρά θετική σχέση, είναι η μετασχηματισμένη ισχύς που προτάθηκε αρχικά από τον Tukey το 1957. Οι Box και Cox (1964) προσπάθησαν να δημιουργήσουν ένα μοντέλο με απλή δομή, κανονικά σφάλματα και σταθερή διακύμανση. Πρότειναν έναν μετασχηματισμό για την σταθεροποίηση της διασποράς μιας χρονοσειράς ο οποίος εμπεριέχει τον λογαριθμικό μετασχηματισμό για την αφαίρεση της ετεροσκεδαστικότητας (μη-συνεχής διακύμανση) μιας μεταβλητής κάνοντας έτσι την κατανομή της χρονοσειράς να μοιάζει περισσότερο σε αυτή της στην κανονικής κατανομή. (Box και Cox, 1964)

Ο μετασχηματισμός ορίζεται ως εξής:

$$w_t = \begin{cases} \log(y_t), & \lambda = 0 \\ \frac{y_t^\lambda - 1}{\lambda}, & \lambda \neq 0 \end{cases}.$$

Μια ικανοποιητική τιμή για το λ θεωρείται εκείνη που μετατρέπει την εποχική διακύμανση σε σταθερή σε ολόκληρη την χρονοσειρά, κάνοντας έτσι το μοντέλο πρόβλεψης απλούστερο.

Εφόσον η πρόβλεψη λάβει χώρα, έπειτα διεξάγεται ο αντίστροφος μετασχηματισμός των Box-Cox πάνω στα ήδη μετασχηματισμένα δεδομένα ώστε να επιστρέψουν στην αρχική τους μορφή. (Athanasopoulos και Hyndman, 2018)

Ο αντίστροφος μετασχηματισμός δίνεται από τον ακόλουθο τύπο:

$$y_t = \begin{cases} \exp(w_t), & \lambda = 0 \\ (\lambda w_t + 1)^{\frac{1}{\lambda}}, & \text{αλλιώς} \end{cases}.$$

2 Backtesting και Stresstesting

2.1 Μέθοδοι Υπολογισμού Αξίας σε Κίνδυνο (VaR)

Από τις αρχές του 1970 η διαχείριση κινδύνου έχει μετατραπεί σε ένα εργαλείο μείζονος σημασίας στην ευρύτερη περιοχή των χρηματοοικονομικών. Οι μαθηματικοί τύποι του VaR ξεκίνησαν να αναπτύσσονται θεωρητικά από τους Harry Markowitz το 1950, με τα χρηματοοικονομικά ιδρύματα να ακολουθούν τις επόμενες δεκαετίες και να αναπτύσσουν τα δικά τους μοντέλα για την διαχείριση κινδύνου.

Οι πρώτες προσπάθειες ποσοτικοποίησης του ρίσκου έγιναν την δεκαετία του 1980 σύμφωνα με το πλαίσιο της Βασιλείας. Το 1984 η J.P. Morgan με την γνωστοποίηση της εφαρμογής Riskmetrics, καθιέρωσε μια νέα μέθοδο η οποία εκτιμούσε τους χρηματοοικονομικούς κινδύνους και χρησιμοποιούσε ως μέτρο κινδύνου την **αξία σε κίνδυνο** (Value at Risk). (Morgan, 1996)

Η VaR ήταν το αποτέλεσμα των ζημιών υψηλών κεφαλαίων που πολλές τράπεζες και πολυεθνικές χρειάστηκε να αντιμετωπίσουν. Ο στόχος της VaR είναι η αμερόληπτη αξιολόγηση των κινδύνων ενός χαρτοφυλακίου διαχρονικά, με σκοπό την ανάλυση τους για την τροποποίηση και την εκτέλεση διαφορετικών στρατηγικών μεθόδων από τα ανώτερα στελέχη των επιχειρήσεων.

Η μέθοδος VaR υπολογίζεται σε κανονικές συνθήκες της αγοράς ως η μέγιστη πιθανή ζημιά της αξίας ενός χαρτοφυλακίου για δεδομένο χρονικό ορίζοντα ή για κάποια συγκεκριμένη χρονική περίοδο μέσα σε ένα προκαθορισμένο και συγκεκριμένο διάστημα στατιστικής εμπιστοσύνης. (Αγγελόπουλος, 2010)

Στη μεθοδολογία του VaR εντάσσονται 2 επιπλέον διαδικασίες, το Stresstesting και το Back-testing. Η πρώτη διαδικασία ελέγχει τη συμπεριφορά του υπό-εξέταση χαρτοφυλακίου σε ακραίες και αντίξοες μακροοικονομικές συνθήκες, ενώ η δεύτερη είναι υπεύθυνη για την επαλήθευσή της ορθότητας της υπολογίσιμης VaR. Εμείς στην παρούσα διπλωματική θα ασχοληθούμε μόνο με τη διαδικασία του Back-testing.

Έτσι, αν η τυχαία μεταβλητή X περιγράφει ένα πιθανό χαρτοφυλάκιο κερδών–ζημιών και το α συμβολίζει το ποσοστό το οποίο λαμβάνει υπόψιν τα χειρότερα σενάρια $\alpha \in (0,1)$, τότε η VaR εκφράζεται ως:

$$VaR_\alpha = -\sup\{x \mid P[X \leq x] < \alpha\},$$

$$VaR_\alpha = -x_\alpha(X).$$

Ο τρόπος υπολογισμού της VaR για είναι αποδεκτός, θα πρέπει οι αποδόσεις να είναι κανονικά κατανομημένες, κάτι τέτοιο όμως, πολλές φορές είναι ανέφικτο με αποτέλεσμα να έχουν δημιουργηθεί εναλλακτικοί τρόποι εύρεσης της VaR οι οποίοι δεν απαιτούν την κανονικότητα των δεδομένων. Ως εκ τούτου, οι μέθοδοι υπολογισμού της VaR διαχωρίζονται σε παραμετρικοί και μη–παραμετρικοί ανάλογα με την ύπαρξη ή με κανονικότητας στα υπό εξέταση δεδομένα. (Ammann και Reich, 2001)

Οι πιο διαδομένες μέθοδοι είναι οι ακόλουθες:

- Ιστορική προσομοίωση (μη–παραμετρική)
- Monte–Carlo προσομοίωση (μη–παραμετρική)
- Διακύμανσης–Διακύμανσης (παραμετρική)

2.1.1 Ιστορική προσομοίωση

Η μέθοδος της **ιστορικής προσομοίωσης** ανήκει στις μη παραμετρικές μεθόδους και θεωρείται η πιο εύκολη στη χρήση. Η βασική ιδέα πίσω από τον μηχανισμό της μεθόδου είναι η χρήση μόνο των ιστορικών τιμών για τον υπολογισμό της VaR στο παρών χαρτοφυλάκιο.

Η ιστορική προσομοίωση εμφανίζει κάποια πολύ σημαντικά πλεονεκτήματα λόγω της απλότητας της:

- Δεν είναι αναγκαία η υπόθεση της κανονικής κατανομής

- Δεν απαιτεί την εκτίμηση της μεταβλητότητας και κατ' επέκταση των συσχετίσεων των αποδόσεων.
- Σε σχέση με τις υπόλοιπες μεθόδους δίνει στον ερευνητή τη δυνατότητα εκτίμησης των βαριών ουρών της κατανομής.

Ωστόσο η ιστορική προσομοίωση χαρακτηρίζεται και από μερικά σημαντικά μειονεκτήματα:

- Απαιτείται πληθώρα δεδομένων τα οποία μερικές φορές δεν είναι διαθέσιμα, ώστε να λειτουργήσει με σχετική ακρίβεια. Το πρόβλημα εμφανίζεται ιδιαίτερα σε περιπτώσεις όπου νέα συμβόλαια και μετοχές, εισάγονται στο χαρτοφυλάκιο χωρίς όμως να βρίσκονται για αρκετό διάστημα στην αγορά. Υποθέτει ότι "οτιδήποτε συνέβη στο παρελθόν, θα συμβεί και στο μέλλον" παρόλο που φαίνεται λογικό, ενέχει τον κίνδυνο να οδηγήσει σε λανθασμένη εκτίμηση της VaR σε ορισμένες περιπτώσεις. (Damodaran, 2007)

2.1.2 Monte–Carlo προσομοίωση

Η μέθοδος **προσομοίωσης Monte–Carlo** ανήκει στις μη–παραμετρικές μεθόδους και χαρακτηρίζεται ως μια από τις ακριβέστερες προσεγγίσεις όταν δημιουργείται η ανάγκη για έναν εξελιγμένο και ισχυρό υπολογισμό της VaR, Ο βαθμός δυσκολίας όμως εφαρμογής της, πολλές φορές την καθιστά αποφευκτέα. (Dowd, 1998)

Η εφαρμογή της μεθόδου απαρτίζεται από τα ακόλουθα δύο βήματα:

Βήμα 1^ο

Καθορισμός μιας στοχαστικής διαδικασίας για τις χρηματοοικονομικές μεταβλητές και εκτίμηση της συσχέτισης της μεταβλητότητας, βασιζόμενοι στα δεδομένα της αγοράς ή στα ιστορικά στοιχεία.

Βήμα 2^ο

Προσομοίωση οικονομικών μεταβλητών και ενσωμάτωση των προσκοπτουσών τιμών σε μια κοινή κατανομή των αποδόσεων, από την οποία εκτιμάται η VaR. (Jorion, 2007)

Μερικά από τα πλεονέκτημα αυτής της μεθόδου είναι:

- Δεν υπάρχουν απαιτήσεις όσον αφορά την ύπαρξη της κανονικής κατανομής
- Οι παράμετροι μολονότι υπολογίζονται μέσω των ιστορικών δεδομένων, δίνεται η δυνατότητα να εισαχθούν επιπλέον πληροφορίες με σκοπό την βελτίωση της προβλεπόμενης νέας κατανομής προσομοίωσης.

Με αυτόν τον τρόπο, η δημιουργία μιας νέας κατανομής μπορεί να χρησιμοποιηθεί ως εργαλείο για τον υπολογισμό των απωλειών της VaR. (Dowd, 1998)

Ακολουθούν μερικά βασικά μειονεκτήματα της μεθόδου:

- Δυσκολία εφαρμογής
- Απαίτηση πολλών πόρων, ιδιαιτέρως σε μεγάλα χαρτοφυλάκια, με αποτέλεσμα η εφαρμογή της να είναι δαπανηρή.
- Πιθανότητα λανθασμένων υποθέσεων για τα μοντέλα που έχουν θεωρηθεί. (Dowd, 1998)

2.1.3 Η μέθοδος διακύμανσης–συνδιακύμανσης

Η μέθοδος διακύμανσης–συνδιακύμανσης αποτελεί μια παραμετρική μέθοδος η οποία βασίζεται στην υπόθεση ότι οι αλλαγές στις παραμέτρους της αγοράς και της αξίας ενός χαρτοφυλακίου ακολουθούν την κανονική κατανομή. Η εν λόγω υπόθεση θεωρείται ιδανική στην περίπτωση απλών χαρτοφυλακίων με λιγιστά στοιχεία. (Wiener, 1999)

Πλεονέκτημα της μεθόδου:

- Απλότητα του μοντέλου. Ο υπολογισμός της VaR, είναι σχετικά εύκολος με απαραίτητη προϋπόθεση ότι τα δεδομένα ακολουθούν την κανονική κατανομή η

οποία καθιστά εφικτή την εύκολη αλλαγή διαφορετικών επιπέδων εμπιστοσύνης και περιόδων κράτησης. (Dowd, 1998)

Μειονέκτημα της μεθόδου:

- Το μεγαλύτερο “αγκάθι” είναι αυτό της κανονικότητας. Τα περισσότερα χρηματοοικονομικά assets είναι γνωστό ότι παρουσιάζουν βαριές ουρές στη κατανομή τους, το οποίο σημαίνει ότι τα ακραία αποτελέσματα συμβαίνουν συχνότερα σε σχέση με το αν ίσχυε πράγματι η κανονική κατανομή. Ως αποτέλεσμα, οι εκτιμήσεις της VaR είναι πιθανό να είναι υποεκτιμημένες. (Jorion, 2007)

Πίνακας 2.1.3.1: Συνοπτικός πίνακας πλεονεκτημάτων–μειονεκτημάτων μεθόδων υπολογισμού VaR

Μέθοδος	Πλεονεκτήματα	Μειονεκτήματα
Ιστορική προσομοίωση	<ul style="list-style-type: none"> • Όχι αναγκαία κανονικότητα • Όχι ανάγκη εκτίμησης μεταβλητότητας • Δυνατότητα εκτίμησης βαριών ουρών της κατανομής 	<ul style="list-style-type: none"> • Απαίτηση πληθώρας δεδομένων για να επιτευχθεί ακρίβεια
Monte Carlo προσομοίωση	<ul style="list-style-type: none"> • Όχι απαίτηση κανονικότητας • Δυνατότητα εισαγωγής επιπλέον πληροφοριών κατά τη διαδικασία της προσομοίωσης 	<ul style="list-style-type: none"> • Δυσκολία εφαρμογής • Δαπανηρή • Πιθανότητα λανθασμένων υποθέσεων
Διακύμανσης–συνδιακύμανσης	<ul style="list-style-type: none"> • Απλότητα μοντέλου 	<ul style="list-style-type: none"> • Απαίτηση κανονικότητας • Υποεκτιμημένες εκτιμήσεις

2.2 Expected Shortfall

Η κύρια πληροφορία του VaR αναφέρεται στην αξία σε κίνδυνο στις περιπτώσεις $\alpha\%$, αλλά δεν προκύπτει κάποια πληροφορία για την αναμενόμενη απώλεια όταν η αξία του κινδύνου υπερβεί το VaR. Ένας τρόπος για να αποφευχθεί το προαναφερθέν πρόβλημα είναι με την βοήθεια του Expected Shortfall (ES). Ο κύριος στόχος του ES, είναι η κάλυψη της αδυναμίας του VaR μέσω των πληροφοριών που αντλούνται από την ουρά της κατανομής και ουσιαστικά δίνει πληροφορίες για το μέγεθος της απώλειας όταν η αξία του κινδύνου υπερβεί το VaR.

Έστω X το κέρδος–ζημιά ενός χαρτοφυλακίου σε έναν προκαθορισμένο χρονικό ορίζοντα T και $\alpha \in (0,1)$, ένα συγκεκριμένο προκαθορισμένο επίπεδο σημαντικότητας. Τότε η αναμενόμενη $\alpha\%$ απώλεια του χαρτοφυλακίου ορίζεται ως εξής: (Acerbi και Tasche, 2002)

- Διακριτές κατανομές

$$ES^{(\alpha)} = -\frac{1}{\alpha} \left(E \left[X 1_{\{X \leq x^{(\alpha)}\}} \right] - x^{(\alpha)} (P[X \leq x^{(\alpha)}] - \alpha) \right)$$

- Συνεχείς κατανομές

$$ES^{(\alpha)} = E[-X | -X \geq VaR_{\alpha}(X)] = \frac{E[-X | -X \geq VaR_{\alpha}(X)]}{P[-X \geq VaR_{\alpha}(X)]}$$

2.3 Backtesting και Τεχνικές

Σύμφωνα με τους Jeroen και Bertrand (Jeroen και Bertrand, 2004), το Backtesting αποτελεί διαγνωστικό έλεγχο για ένα μοντέλο κινδύνου το οποίο υλοποιείται μέσω ενός συνόλου στατιστικών διαδικασιών με στόχο τον έλεγχο του εάν παρατηρήθηκαν πραγματικές απώλειες. Συνήθως, η επιλογή της μεθόδου Backtesting εξαρτάται κυρίως από τις μεθόδους πρόβλεψης όπου σύμφωνα με τους (Emmer, Kratz και Tasche, 2013) είναι οι ακόλουθες:

- **Πρόβλεψη σε σημείο:** Η τιμή μιας τυχαίας μεταβλητής Y , ορίζεται συνήθως με όρους προσδοκίας, σε σχέση με τις διαθέσιμες πληροφορίες κατά τη στιγμή της πρόβλεψης:

$$E[Y_{t+k} | F(Y_s, s \leq t)], t = 1, 2, \dots, n,$$

όπου το $F(Y_s, s \leq t)$ αναπαριστά την διαθέσιμη πληροφορία σε μια χρονοσειρά Y , χρονικής περιόδου t .

- **Πρόβλεψη εύρους πιθανότητας ή προβλέψεις διαστήματος:** Δημιουργία ενός διαστήματος εμπιστοσύνης στο οποίο εμπεριέχεται η τιμή πρόβλεψης βάσει ενός προκαθορισμένου επιπέδου εμπιστοσύνης. Σε αυτή την κατηγορία υπάγονται το VaR και το ES.
- **Πρόβλεψη ολοκληρωμένης κατανομής πιθανότητας** (Forecasts of the complete probability distribution).

$$P[Y_{t+k} \leq \cdot | F(Y_s, s \leq t)].$$

Ως Backtesting χαρακτηρίζεται ο έλεγχος καλής πρόβλεψης ενός ήδη προσαρμοσμένου στατιστικού–οικονομετρικού υποδείγματος βάσει ενός αρχικού συνόλου δεδομένων. Στη διαχείριση κινδύνου συνήθως είναι ο έλεγχος καλής πρόβλεψης ενός ήδη προσαρμοσμένου στατιστικού–οικονομετρικού υποδείγματος βάσει ενός αρχικού συνόλου δεδομένων που αφορά συνήθως το μέγεθος του μεγέθους των ζημιών ενός χρηματοπιστωτικού οργανισμού πάνω από ένα συγκεκριμένο κατώφλι.

Το Backtesting είναι απαραίτητο για την ορθή διαχείριση του χρηματοοικονομικού κινδύνου, καθώς παρέχει τη δυνατότητα ελέγχου του πραγματικού επιπέδου εμπιστοσύνης της εκτιμώμενης VaR. Στις περιπτώσεις που διαπιστώνεται ότι το εκτιμώμενο VaR δεν χαρακτηρίζεται από το επίπεδο εμπιστοσύνης για το οποίο είχε υπολογιστεί, το μοντέλο υπολογισμού της VaR θα πρέπει να επανεξετάζεται για το ενδεχόμενο πιθανών εσφαλμένων υποθέσεων, λανθασμένων παραμέτρων ή ανακριβούς μοντελοποίησης.

Πίνακας 2.3.1: Επιλογές και αποφάσεις στατιστικού

	$H_0: p = 0.05$	$H_1: p \neq 0.05$
<u>Απόφαση 1</u> Αποδέχομαι το μοντέλο	Αποδέχομαι ορθά ένα σωστό μοντέλο Correct Decision	Type II Error Δεν απορρίπτω ένα μη ορθό μοντέλο
<u>Απόφαση 2</u> Δεν αποδέχομαι το μοντέλο	Type I Error Απορρίπτω λανθασμένα ένα σωστό μοντέλο	Δεν αποδέχομαι ορθά ένα λανθασμένο μοντέλο Correct Decision

2.3.1 Kupiec's Test

Το test του Kupiec ή POF–Test (Proportion of Failures) εξετάζει εάν ο αριθμός των εξαιρέσεων είναι συνεπής με το επίπεδο εμπιστοσύνης. Για να εξεταστεί αυτό χρειαζόμαστε το πλήθος των παρατηρήσεων $\{T\}$, το πλήθος των εξαιρέσεων $\{N\}$, καθώς και το επίπεδο εμπιστοσύνης (p).

Η μηδενική υπόθεση ορίζεται ως εξής:

$$H_0 := p = \hat{p} = \frac{N}{T}.$$

Ως αποτέλεσμα η ελεγχοσυνάρτηση δίνεται από τον ακόλουθο τύπο:

$$L_{Kup} = -2 \ln \left(\frac{(1-p)^{T-N} p^N}{\left[1 - \left(\frac{N}{T}\right)^{T-N} \left(\frac{N}{T}\right)^N\right]} \right).$$

Όταν ισχύει η μηδενική υπόθεση το μοντέλο θεωρείται σωστό και η ποσότητα L_{Kup} ασυπτωτικά ακολουθεί την X_1^2 (Christoffersen, 1998; Halilbegovic, 2016) που για $\alpha=5\%$ δίνει:

$$\text{Κρίσιμη περιοχή: } L_{Kup} > 3.84.$$

Από στατιστικής σκοπιάς, το test του Kupiec αποτελεί μια τροποποίηση του ελέγχου λόγου πιθανοφανειών.

Το test έχει δύο σημαντικά μειονεκτήματα:

- Είναι αδύναμο σε δείγματα μικρού μεγέθους
- Εμφανίζει μόνο τη συχνότητα των ζημιών και όχι το χρόνο εμφάνισής τους.

Αποτέλεσμα των παραπάνω είναι ότι ο κίνδυνος σφάλματος τύπου II είναι παραπάνω από εμφανής με αποτέλεσμα να μην απορριφθεί ένα μη ορθό μοντέλο. Η ιδέα του test είναι η εξέταση του εάν το $\frac{N}{T}$ (observed failure rate) είναι σημαντικά διαφορετικό από το p , σε ποσοστό αποτυχίας το οποίο καθορίζεται από το επίπεδο εμπιστοσύνης. Το 1998 ο Christoffersen όρισε τις δυο αναγκαίες συνθήκες ώστε να έχουμε επαρκή έλεγχο. Αυτές είναι:

- Unconditional Coverage Property (άνευ όρων κάλυψη)
- Independence Property (ανεξαρτησία)

Η σημασία της δεύτερης συνθήκης έγκειται στην άποψη ότι οι μαζεμένες ζημιές (ακόμα και αν είναι μικρές) είναι πιο επικίνδυνες από τις μεγάλες (σπάνιες). Η δεύτερη συνθήκη δεν καλύπτεται από το test του Kupiec. (Dowd, 2006)

2.3.2 Christoffersen's Test

Το test του Christoffersen συνδυάζει την ανεξαρτησία του στατιστικού του Kupiec's POF test με σκοπό την επίτευξη της κοινής δοκιμής. (Christoffersen, 1998). Για να προσπελάσει ο Christoffersen το πρόβλημα μη υπολογισμού του χρόνου εμφάνισης τροποποίησε το test του Kupiec ώστε να περιλαμβάνει και την ανεξαρτησία των εξαιρέσεων. Για να καταστεί κάτι τέτοιο εφικτό, θα πρέπει πρώτα να εξεταστεί αν η πιθανότητα εμφάνισης μια εξαίρεσης σχετίζεται με το αν υπάρχει εξαίρεση την προηγούμενη ημέρα. Ας θεωρήσουμε 4 μετρητές, έστω N_{ij} , όπου:

- i η δείτρια εμφάνισης (ή όχι) εξαίρεσης της i -οστής ημέρας
- j η δείτρια εμφάνισης (ή όχι) εξαίρεσης της προηγούμενης.

Ορίζουμε με π_i την πιθανότητα της i -οστής παρατήρησης ως:

$$\pi_0 = \frac{N_{01}}{N_{00} + N_{01}}, \quad \pi_1 = \frac{N_{11}}{N_{10} + N_{11}}, \quad \pi = \frac{N_{01} + N_{11}}{N_{00} + N_{01} + N_{10} + N_{11}},$$

όπου:

N_{00} : Δεν εμφανίστηκε εξαίρεση ούτε σήμερα ούτε εχθές

N_{01} : Δεν εμφανίστηκε εξαίρεση σήμερα ενώ εμφανίστηκε εχθές

N_{10} : Εμφανίστηκε εξαίρεση σήμερα ενώ δεν εμφανίστηκε εχθές

N_{11} : Εμφανίστηκε εξαίρεση και σήμερα και εχθές.

Αν το μοντέλο είναι ακριβές, τότε τα π_0, π_1 θα πρέπει να είναι ίσα. Το test για την ανεξαρτησία των εξαιρέσεων δίνεται από τη σχέση:

$$L_{ind} = -2 \ln \left(\frac{(1-\pi)^{N_{00}+N_{10}} \pi^{N_{01}+N_{11}}}{(1-\pi_0)^{N_{00}} \pi_0^{N_{01}} (1-\pi_1)^{N_{10}} \pi_1^{N_{11}}} \right),$$

όπου κάτω από τη μηδενική υπόθεση το L_{ind} ακολουθεί την X_1^2 . (Christoffersen και Pelletier, 2004)

Συνδυάζοντας το παραπάνω test με αυτό του Kupiec, λαμβάνουμε μια κοινή δοκιμή που εξετάζει τις ιδιότητες ενός καλού μοντέλου VaR, το σωστό failure rate και την ανεξαρτησία των εξαιρέσεων. Αυτή η κοινή δοκιμή L_{common} , ακολουθεί την X_2^2 και ορίζεται ως:

$$L_{common} = L_{Kup} + L_{ind},$$

που για $\alpha=5\%$ δίνει:

Κρίσιμη περιοχή: $L_{common} > 5.99$.

2.4 Stresstesting και Τεχνικές

Λόγω των παγκόσμιων χρηματοπιστωτικών κρίσεων, θεωρείται αναγκαία η διερεύνηση των αιτιών που είναι ικανά να απαντήσουν στο ερώτημα ‘‘Ποιο θα ήταν το κόστος καθώς και οι πιθανές έμμεσες ή άμεσες επιπτώσεις των αρνητικών οικονομικών επιδόσεων;’’. Έτσι οι υπεύθυνοι των χρηματοπιστωτικών ιδρυμάτων για να είναι σε θέση να αντιμετωπίσουν τα οικονομικά προβλήματα που προκύπτουν, θα πρέπει να γνωρίζουν τις αδυναμίες ολόκληρου του χρηματοπιστωτικού ιδρύματος. Το Stresstesting μπορεί να συμβάλει σε αυτό καθώς αποτελεί ένα ποσοτικό εργαλείο το οποίο δύναται να αξιοποιηθεί από τους υπευθύνους των τραπεζών, με σκοπό την εκτίμηση της καταλληλότητας των χρηματοπιστωτικών ιδρυμάτων σε ακραίες καταστάσεις.

Η έννοια του Stresstesting δεν είναι καινούργια, δεδομένου ότι εφαρμόζεται από τις αρχές του 1990. Κυρίως από μεγάλες ιδιωτικές τράπεζες, εξαιτίας της φύσης των κινδύνων που συνοδεύουν την τεχνική με αποτέλεσμα μόνο λίγοι ερευνητές να είναι σε θέση να κατανοήσουν την χρησιμότητα της. Η χρήση του test ήταν κυρίως συμπληρωματική ως προς την ήδη υπάρχουσα ανάλυση των μοντέλων διαχείρισης κινδύνου, όπως αυτής της αξίας σε κίνδυνο. Σε αντίθεση με την VaR, η οποία δίνει πληροφορίες για το επίπεδο της αρνητικής επίδοσης, το stress-test παρείχε πιο χρήσιμες πληροφορίες δεδομένης της δυνατότητας του να επεξηγεί ακραία γεγονότα.

Το 1996, οι τράπεζες και οι επενδυτικές επιχειρήσεις είχαν την υποχρέωση να διεξάγουν stress-tests ως μέρος του μοντέλου που σχετίζεται με τον υπολογισμό των απαιτήσεων σε κεφάλαια. Ωστόσο εδώ και αρκετά χρόνια τα stress-tests διεξάγονται όλο και πιο συχνά και συνήθως καθίστανται υποχρεωτικά. (Quagliariello, 2009)

Από πλευράς στατιστικής, το stress-test ορίζονται ως: το ποσό των προβλέψεων που διακρατείται από τις τράπεζες είναι υπεύθυνο για την κάλυψη της **αναμενόμενης ζημιάς** (expected loss), ενώ τα εποπτικά ίδια κεφάλαια είναι υπεύθυνα για την κάλυψη των **μη αναμενόμενων ζημιών** (unexpected loss) όπου τοποθετούνται στο 99% της καμπάνας της κανονικής κατανομής. Τα stress-test είναι αυτά, που ερμηνεύουν την πιθανότητα ζημιάς από ακραία και ξαφνικά γεγονότα.

Οι τεχνικές Stresstesting οι οποίες χρησιμοποιούνται συχνά από τα τραπεζικά ιδρύματα και τις επιχειρήσεις είναι οι ακόλουθες:

2.4.1 Ανάλυση ευαισθησίας

Μια μεγάλη μεταβολή σε έναν παράγοντα κινδύνου έχει καθοριστική επιρροή στην αξία ενός χαρτοφυλακίου και μπορεί να εκτιμηθεί μέσω της ανάλυσης ευαισθησίας, χωρίς όμως να προσδιορίζεται η αιτία της. Πρόκειται για ένα γρήγορο τρόπο υπολογισμού της

επιρροής μια τέτοιας αλλαγής με τον οποίο εντοπίζονται οι συσχετιζόμενοι παράγοντες κινδύνου και υπολογίζεται ο προσδοκώμενος κίνδυνος με σαφή διάκριση από τους άλλους παράγοντες.

Ένα σημείο στο οποίο υστερεί, είναι ότι δεν αναγνωρίζει την αμοιβαία επίδραση αυτών των παραγόντων που συντελείται στη πραγματικότητα, δηλαδή το γεγονός ότι κάποιος παράγοντας κινδύνου ενδέχεται να μην μοιάζει σημαντικός από μόνος του αλλά αν συσχετιστεί με κάποιες άλλες αλλαγές να αποσταθεροποιήσει το σύνολο του συστήματος. Η ανάλυση ευαισθησίας εφαρμόζεται συνήθως σε πρώιμο στάδιο της ανάλυσης ώστε να ερευνηθούν οι συνέπειες στην αξία ενός χαρτοφυλακίου που επιφέρει μια μεγάλη αλλαγή στην αγορά.

2.4.2 Ανάλυση σεναρίου

Το εύρος των προοπτικών διευρύνεται στην περίπτωση της ανάλυσης σεναρίου. Σε αντίθεση με την ανάλυση ευαισθησίας η εν λόγω μέθοδος ανταποκρίνεται περισσότερο στην πραγματικότητα και είναι σε κάποιο βαθμό πολυπλοκότερη καθώς ελέγχει παραπάνω από έναν κινδύνους και προσμετρά την επιρροή που ασκούν μεγάλες αλλαγές στα μεγέθη συγχρόνως και στην αξία του χαρτοφυλακίου αλλά και της επιχείρησης. Σε ένα μακροοικονομικό περιβάλλον οι μεταβολές που παρουσιάζονται αλληλοεπιδρούν. Συνεπώς, η ανάλυση σεναρίου αποδεικνύεται συχνά καταλληλότερη και χρησιμοποιείται στις ασκήσεις προσομοίωσης ακραίων καταστάσεων.

Οι παράγοντες κινδύνου που υφίστανται μεταβολές, προέρχονται είτε από στοιχεία του χαρτοφυλακίου είτε από άλλα γεγονότα που ενδέχεται να έχουν επίδραση σε αυτό. Αν η επιλογή γίνεται μέσα από τα στοιχεία του χαρτοφυλακίου, ακολουθείται ένα σενάριο σύμφωνα με το οποίο δίνονται τιμές στους παράγοντες κινδύνου πέρα των σταθερών τιμών του VaR και στη συνέχεια μπορούν να εκτιμηθούν τα αποτελέσματα. Στην περίπτωση όμως που οι παράγοντες επιλέγονται από τα γεγονότα που επηρεάζουν το σενάριο, αφού το τελευταίο στηριχτεί σε ρεαλιστικά και ασυνήθη γεγονότα, θα ελέγξει με ποιο τρόπο αυτά μπορούν να επιφέρουν αλλαγές στους παράγοντες κινδύνου που έχουν σχέση με το χαρτοφυλάκιο ή την επιχείρηση. Εξωτερικοί παράγοντες λαμβάνονται υπόψη σε περιπτώσεις μακροοικονομικών ασκήσεων προσομοίωσης ακραίων καταστάσεων, διότι συνεπάγονται καθοριστικές αλλαγές σε μεταβλητές του χρηματοοικονομικού τομέα.

2.4.3 Αντίστροφη προσομοίωση ακραίων καταστάσεων

Η έλλειψη βιωσιμότητας ενός επιχειρηματικού σχεδίου και η αφερεγγυότητα ενός χρηματοπιστωτικού ιδρύματος μπορούν να ανιχνευτούν μέσω της αντίστροφης προσομοίωσης ακραίων καταστάσεων, η οποία στοχεύει στον καθορισμό ενός ή περισσότερων σεναρίων, αναδεικνύοντας ταυτόχρονα το κατά πόσο πραγματοποιήσιμα

είναι τα σενάρια αυτά. Η συγκεκριμένη τεχνική αποδεικνύεται εξαιρετικά χρήσιμη κατά τη διαδικασία διαχείρισης κινδύνων και δύναται να εκτελεστεί με ποικίλους τρόπους, είτε ως μια απλή ποιοτική συζήτηση για τους κύριους παράγοντες κινδύνου είτε ως μια πιο σύνθετη ποσοτική προσέγγιση. Στη τελευταία περίπτωση, αφού οριστεί το σημείο που εμφανίζεται η αφερεγγυότητα του χρηματοπιστωτικού ιδρύματος, ακολουθείται μια αντίστροφη ποσοτική πορεία με σκοπό τον εντοπισμό των μακροοικονομικών παραγόντων που την επιφέρουν.

Όλα τα χρηματοπιστωτικά ιδρύματα είναι αναγκαίο να την εφαρμόζουν ανά τακτά χρονικά διαστήματα και ειδικότερα τα πιο μεγάλα και περίπλοκα καθώς η εστίαση με πιο ποιοτικό τρόπο σε γεγονότα και κινδύνους συμβάλει στη βιωσιμότητα της επιχείρησης. Σημαντικό πλεονέκτημα της αντίστροφης προσομοίωσης ακραίων καταστάσεων, αποτελεί το γεγονός ότι αξιολογεί τις παραδοχές του επιχειρηματικού μοντέλου.

Σημείωση 1:

Το Stresstesting και το Backtesting είναι έννοιες ταυτόσημες. Αυτό διότι σε κάθε τέτοια μέθοδο προσαρμόζεται αρχικά κάποιο υπόδειγμα χρονοσειράς και στη συνέχεια γίνεται πρόβλεψη βάσει αυτού για τον υπολογισμό του μεγέθους των υπερβάσεων.

Οι οποίες μπορεί να είναι είτε υπερβάσεις απόδοσης ενός χαρτοφυλακίου είτε υπερβάσεις ζημιάς πάνω από κάποιο προσδιορισμένο κατώφλι (threshold). Στην εργασία αυτή ασχολούμαστε με υπερβάσεις αποδόσεων ενός χαρτοφυλακίου χρηματοοικονομικών τίτλων. Η απόδοση λαμβάνεται ως ημερήσια βάσει των τιμών κλεισίματος των τίτλων που απαρτίζουν το χαρτοφυλάκιό μας. Για λόγους απλότητας τα βάρη του χαρτοφυλακίου θεωρούνται σταθερά και οι θέσεις μας είναι long, δηλαδή θέσεις αγοράς των τίτλων.

3 Ένα Τροποποιημένο ARIMA(1,1,0) για την Μοντελοποίηση και Πρόβλεψη Αποδόσεων Χαρτοφυλακίου

Η μοντελοποίηση και κατ' επέκταση πρόβλεψη χρηματιστηριακών αποδόσεων, αποτελεί τις τελευταίες δεκαετίες ένα ενδιαφέρον ερευνητικό αντικείμενο, κυρίως λόγω της ανάγκης των επενδυτών να παίρνουν τις σωστές επενδυτικές αποφάσεις. Μια από τις τεχνικές προσέγγισης των προαναφερθέντων, είναι αυτή των τυπικών ARIMA(p, d, q) διαδικασιών.

3.1 Ο Αλγόριθμος – Το Τροποποιημένο ARIMA μοντέλο

Στην παρούσα εφαρμογή, προτείνουμε μια αλγοριθμική διαδικασία 5 βημάτων η οποία βασίζεται σε ένα ARIMA(1,1,0) μοντέλο με EWMA βαρύτητα:

Βήμα 1 Επιλογή μετοχών, εύρεση των αντίστοιχων αποδόσεων και εν συνεχεία δημιουργία χαρτοφυλακίου.

Βήμα 2 Διαχωρισμός του δείγματος σε k -υποδείγματα ίσου μεγέθους (η επιλογή του k επαφίεται στον ερευνητή).

Βήμα 3 Εκτίμηση μέσω της μεγίστης πιθανοφάνειας των EWMA παραμέτρων λ_i για καθένα από τα k υποδείγματα, βάση του μοντέλου:

$$Y_t - Y_{t-1} = \lambda_i(Y_{t-1} - Y_{t-2}) + \varepsilon_t, \quad t = 3, \dots, n, \quad i = 1, 2, \dots, k,$$

με

$$\varepsilon_t \sim i.i.d WN(0, \sigma^2).$$

Βήμα 4 Το **εξεταζόμενο μοντέλο** για την εκτίμηση των χρηματιστηριακών αποδόσεων είναι το ακόλουθο:

$$\hat{Y}_t - \hat{Y}_{t-1} = \hat{w}_\lambda(\hat{Y}_{t-1} - \hat{Y}_{t-2}), \quad t = 3, \dots, n, \quad i = 1, 2, \dots, k,$$

όπου

$$\hat{w}_\lambda = \frac{\sum_{i=1}^k \hat{\lambda}_i}{k},$$

η σταθμισμένη, από τα k -υποδείγματα, EWMA παράμετρος.

Βήμα 5 Αξιολόγηση εκτιμητικής/προβλεπτικής ικανότητας του προτεινόμενου μοντέλου.

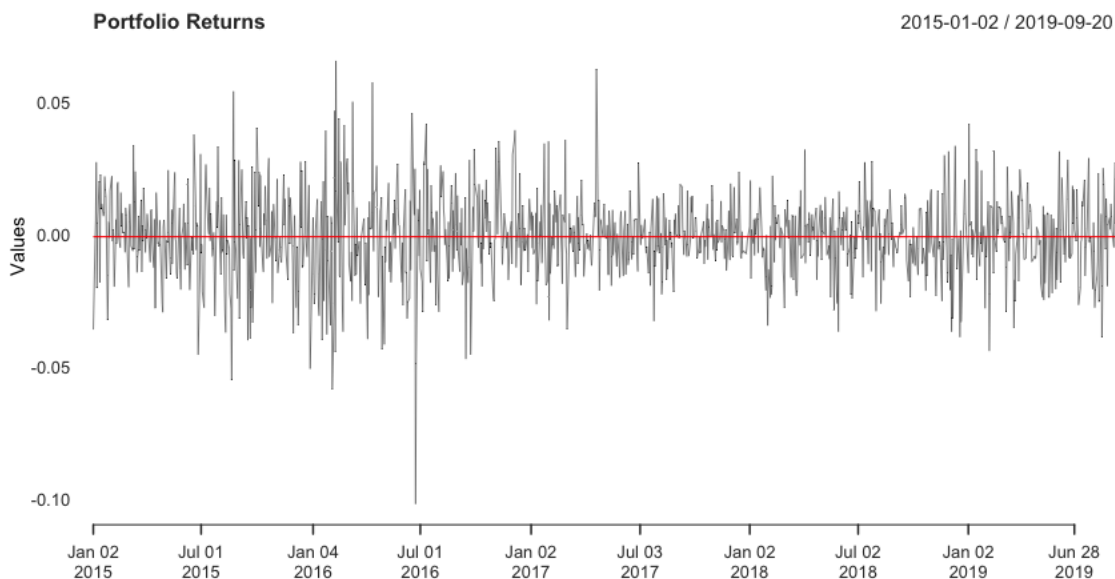
3.2 Πειραματική Μελέτη

Οι μετοχές οι οποίες χρησιμοποιήθηκαν είναι οι *Siemens*, *Deutsche Bank AG* και *BMW*, ενώ ο χρονικός ορίζοντας των δεδομένων είναι από τις 02/01/2015 έως τις 20/09/2019 (1204 παρατηρήσεις). Βασιζόμενοι στις τιμές κλεισίματος της κάθε μετοχής, δημιουργήθηκαν οι ημερήσιες αποδόσεις τους και εν συνεχεία έγινε σύσταση ενός χαρτοφυλακίου αποτελούμενο από το 33.33%, αντίστοιχα της κάθε μιας από τις προαναφερθείσες μετοχές.

Στόχος μας είναι η διερεύνηση της εκτιμητικής και προβλεπτικής ικανότητας του εξεταζόμενου τροποποιημένου ARIMA (1,1,0) όπως αυτό παρουσιάστηκε στο Βήμα 4 της προτεινόμενης αλγοριθμικής διαδικασίας.

Σημειώνεται ότι, από τις 1204 παρατηρήσεις, για την ανάλυση γίνεται χρήση των πρώτων 1200, με σκοπό οι 4 τελευταίες να χρησιμοποιηθούν για σκοπούς πρόβλεψης.

Στο Γράφημα 3.2.1 αναπαρίστανται οι τιμές των αποδόσεων του χαρτοφυλακίου με μέση τιμή περίπου -0.0003 (κόκκινη ευθεία) και διασπορά 0.0002.



Γράφημα 3.2.1: Χρονοσειρά αποδόσεων χαρτοφυλακίου 2015–2019. Με την κόκκινη γραμμή, η μέση τιμή των αποδόσεων.

Χωρίζουμε τις 1200 παρατηρήσεις σε $k = 12$ δείγματα 100 παρατηρήσεων έκαστο και εφαρμόζουμε τα Βήματα 3 και 4.

Πίνακας 3.2.1 Εκτιμώμενες τιμές παραμέτρου λ_i για καθένα από τα k -υποδείγματα

Υπόδειγμα	$\hat{\lambda}_i$
1	0.25153409694
2	0.09425002919
3	0.03468238676
4	0.01567128507
5	0.05647333168
6	0.06233176314
7	0.10924680350
8	0.00005625841
9	0.02233985555
10	0.00006610696
11	0.03777516866
12	0.12712864290

Υπολογίζοντας το

$$\hat{w}_\lambda = \frac{\sum_{i=1}^{12} \hat{\lambda}_i}{12} = 0.06762964,$$

το εκτιμώμενο μοντέλο είναι το ακόλουθο:

$$\hat{Y}_t - \hat{Y}_{t-1} = 0.068(\hat{Y}_{t-1} - \hat{Y}_{t-2}), \quad t = 3, \dots, 1200.$$

Σημειώνεται δε, ότι μέσω των διαγνωστικών ελέγχων *Ljung-Box* ($Pvalue = 0.5431$) και *Goldfeld-Quandt* ($Pvalue = 0.999$), διαπιστώθηκε ότι τα $\hat{\varepsilon}_t, t = 3, \dots, 1200$ είναι λευκός θόρυβος.

Επιπλέον, για σκοπούς σύγκρισης, προσαρμόζουμε στις 1200 παρατηρήσεις και το τυπικό ARIMA(1,1,0) μοντέλο:

$$\hat{Y}_t - \hat{Y}_{t-1} = -0.487(\hat{Y}_{t-1} - \hat{Y}_{t-2}), \quad t = 3, \dots, 1200,$$

το οποίο δεν έχει περάσει μέσα από την προτεινόμενη αλγοριθμική διαδικασία.

Τα αποτελέσματα των MSE, AIC και BIC των δύο μοντέλων παρουσιάζονται στον Πίνακα 3.2.2.

Πίνακας 3.2.2 Τιμές MSE, AIC και BIC εξεταζόμενου και τυπικού ARIMA(1,1,0) μοντέλου

MSE		AIC		BIC	
Εξεταζόμενο Μοντέλο	Τυπικό Μοντέλο	Εξεταζόμενο Μοντέλο	Τυπικό Μοντέλο	Εξεταζόμενο Μοντέλο	Τυπικό Μοντέλο
0.0002467958	0.0003574071	-6560.887	-6110.141	-6555.797	-6099.963

Παρατηρούμε ότι οι τιμές των MSE, AIC και BIC παρουσιάζουν σημαντική βελτίωση μέσω της προτεινόμενης διαδικασίας. Το αυτό ισχύει και για την προβλεπτική ικανότητα του προτεινόμενου μοντέλου όπως αυτή αποτυπώνεται στον Πίνακα 3.2.3.

Πίνακας 3.2.3 Τιμή MSE πρόβλεψης εξεταζόμενου και τυπικού ARIMA(1,1,0) μοντέλου

Εξεταζόμενο Μοντέλο	Τυπικό Μοντέλο
0.0001404239	0.0004713606

3.3 Συμπεράσματα

Στην παρούσα εφαρμογή προτείναμε μια αλγοριθμική διαδικασία 5 βημάτων, η οποία ενσωμάτωνε ένα τροποποιημένο ARIMA(1,1,0) με EWMA βαρύτητα, με σκοπό την μοντελοποίηση και πρόβλεψη χρηματιστηριακών αποδόσεων. Την εν λόγω διαδικασία την εφαρμόσαμε σε ένα χαρτοφυλάκιο αποτελούμενο από συνολικά 1200 αποδόσεις 3 μετοχών. Στις ίδιες 1200 παρατηρήσεις, εφαρμόσαμε και το τυπικό ARIMA(1,1,0) μοντέλο με τα προκύπτοντα αποτελέσματα να παρουσιάζουν ιδιαίτερο ενδιαφέρον καθώς μέσω της προτεινόμενης διαδικασίας οδηγηθήκαμε σε ένα μοντέλο με MSE, AIC και BIC καλύτερο από το τυπικό μοντέλο.

Το σημείο που θα πρέπει κανείς να δώσει μεγάλη προσοχή, είτε στην περίπτωση του τροποποιημένου μοντέλου είτε σε αυτή του τυπικού, είναι οι πολύ χαμηλές τιμές του MSE. Η τόσο καλή προσαρμογή των μοντέλων στα δεδομένα, είναι πιθανό να οδηγήσει στο φαινόμενο της υπερπροσαρμογής με αποτέλεσμα να οδηγηθούμε σε μεροληπτικές εκτιμήσεις/προβλέψεις. Το συγκεκριμένο χρήζει περεταίρω διερεύνησης μέσω μιας διαδικασίας παρόμοιας με αυτή του Backtesting προσαρμοσμένης όμως σε μοντέλα ARIMA (Cross-Validation).

Βιβλιογραφία

1. Acerbi, C., & Tasche, D. (2002). On the Coherence of Expected Shortfall. *Journal of Banking and Finance*, 7, 1487–1503.
2. Akaike, H. (1978). A Bayesian Analysis of the Minimum AIC Procedure. *Annals of the Institute of Statistical Mathematics*, 9–14.
3. Ammann, M., & Reich, C. (2001). Value-at-Risk for Nonlinear Financial Instruments—Linear Approximation or Full Monte-Carlo;. *Financial Markets and Portfolio Management*, 3, 363–378.
4. Athanasopoulos, G., & Hyndman, R. (2018). *Forecasting: Principles and Practice*. Texts.
5. Box, G., & Cox, D. (1964). An Analysis of Transformations. *Journal of the Royal Statistical Society*, 26(2), 211–252.
6. Brockwell, P., & Davis, R. (1996). *Introduction to Time Series and Forecasting*—New York Springer. Springer.
7. Chatfield, C. (1984). *The Analysis of time series: An Introduction*, Third Edition. Springer.
8. Christoffersen, P., & Pelletier, D. (2004). Backtesting Value at Risk; A Duration Bases Approach. *Journal of Financial Econometrics*, 84–108.
9. Christoffersen, P. (1998). Evaluating Interval Forecasts. *International Economic Review*, 841–862.
10. Cort, J., Kenji, M., & Willmott, M. (2005). Advantages of the Mean Absolute Error (MAE) Over the Mean Absolute Error (MAE) Over The Root Mean Square Error (RMSE) in Assessing Average Model Performance. *Climate Research*, 1, 79–82.
11. Damodaran, A. (2007). *Strategic Risk Taking: A Framework for Risk Management*. Yoram Jerry Wind.
12. Dickey, D. A. & Fuller, W. A. (1979). Distribution of the Estimators for Autoregressive Time Series with a Unit Root, *American Statistical Association*. 74(310), 427–431.
13. Dowd, K. (1998). *Beyond Value at Risk, The New Science of Risk Management*. Wiley Frontiers of Finance.
14. Dowd, K. (2006). *Retrospective Assessment of Value-at-Risk*. Risk Management. Michael K. Ong.
15. Emmer, S., Kratz, M., & Tasche, D. (2013). What is the Best Risk Measure in Practise. a Comparison of Standard Measures. *Journal of Risk*, 2, 31–60.
16. Garrick, R., & Wong, R. (2002). *The Validity of Forecasting 2002* PRRESS Garrick Small.
17. Goldfeld, S., & Quandt, R. (1965). Some Tests for Homoskedasticity. *Journal of the American Statistical Association*, 60(310), 539–547.
18. Halilbegovic, S. (2016). Backtesting Value at Risk Forecast: the Case of Kupiec Pof-Test. *European Journal of Economic Studies*, 17(3), 393–404.
19. Hannan, E., & Quinn, B. (1979). The Determination of the Order of an Autoregression. *Journal of the Royal Statistical Society*, 41(2), 190–195.

20. Jarque, C. M., & Bera, A. K. (1987). A Test for Normality of Observations and Regression Residuals. *International Statistical Review*, 55(2), 163–172.
21. Jereon, K., & Bertrand, M. (2004). Backtesting for Risk–Based Regulatory Capital. *Journal of Banking and Finance*, 8, 1845–1865.
22. Jorion, P. (2007). *Value at Risk, The New Benchmark for Managing Financial Risk*, 2nd Edition. The McGraw–Hill Companies.
23. Ljung, G. M., & Box, G. E. (1978). On a Measure of a Lack of Fit in Time Series Models. *Biometrika*, 2, 297–303.
24. Makridakis, S., & Wheelwright, S. (1984). Forecasting Methods and Applications. *Journal of the Operational Research Society*, 35(1), 79.
25. Montgomery, D. C. (2008). *Introduction to Time Series Analysis and Forecasting*. Wiley.
26. Morgan, J. (1996). Technical Document (4). Risk Metrics.
27. Pedace, R. (2013). <https://www.dummies.com/education/economics/econometrics/test-for-heteroskedasticity-with-the-goldfeld-quandt-test/>.
28. Quagliariello, M. (2009). *Stresstesting the Banking System*. Cambridge University Press, 18–36.
29. Theil, H. & Rudd, R. W. (1967). Applied Economic Forecasting. *Journal of Farm Economics*, 49(1), 241–243.
30. Schwarz, G. (1978). Estimating the Dimension of a Model. *Annals of Statistics*, 2, 461–464.
31. Sugiura, N. (2000). Akaike's Information Criterion and Recent Developments in Information Complexity. *Journal of Mathematical Psychology*, 62–98.
32. Vrieze, S. I. (2012). <http://en.wikipedia.org/w/index.php?oldid=496433604>.
33. Wiener, Z. (1999). *Introduction to VaR (Value–at–Risk), Risk Management and Regulation in Banking*. Kluwer Academic Publishers.
34. Αγγελόπουλος, Π. (2010). *Τράπεζες και Χρηματοπιστωτικό Σύστημα, Αγορές – Προϊόντα – Κίνδυνοι*, (Γ' Έκδοση). Αθήνα: Εκδόσεις Σταμούλη.
35. Δημέλη, Σ. (2013). *Σύγχρονες Μέθοδοι Ανάλυσης Χρονολογικών Σειρών* (ΟΠΑ εκδ.).
36. Μηλιώνης, Α. (2017). *Σημειώσεις χρονοσειρές*. Πανεπιστήμιο Αιγαίου.

Παράρτημα

```
rm(list = ls()) #Clear Environment
dev.off() # Clear Plots
cat("\014") #Clear console

#INSTALL/LOAD PACKAGES----

# install.packages("readxl")
# install.packages("xts")
# install.packages("xtsExtra")
# install.packages("tseries")
# install.packages("forecast")
# install.packages("lmtest")
# install.packages("plotrix")
# install.packages("qcc")
# install.packages("writexl")

library(readxl)
library(xts)
library(tseries)
library(forecast)
library(lmtest)
library(plotrix)
library(qcc)
library(writexl)

#LOAD DATA----
data <- read_excel("m1.xlsx", col_names = T)

options(max.print = 1000000)

dataset <- ts(data$pr)
pr.for.prediction <- dataset[1201:1204]
pr <- dataset[1:1200]
date.pr <- as.Date(as.character(data$date[1:1200]))

mean.1200.data <- mean(pr)

var.1200.data <- var(pr)

segments <- 100
```

```

difs <-
  0

digits <-
  7

#TIME-SERIES ANALYSIS-----

pr.ts <- xts(pr, date.pr)
colnames(pr.ts)[1] <- "Returns"

pr.ts$horizontal_line <- mean(pr.ts)

plot(
  pr.ts,
  grid.col = NA,
  lwd = 0.5,
  ylab = "Values",
  yaxis.right = FALSE,
  main = "Portfolio Returns"
)
lines(pr.ts[, "horizontal_line"], col = "red")

autocor.f <- acf(pr , NULL, main = "Autocorrelation
Function")

parautocor.f <-
  pacf(pr , NULL, main = "Partial Autocorrelation
Function")

#___SEGMENT DATA-----
new.analysis.pr <- pr[1:segments]

ewma.smoothing <-
  HoltWinters(new.analysis.pr, beta = FALSE, gamma = FALSE)

smoothed.values <- c()
smoothed.values[1] <- mean(new.analysis.pr)

lambda <- c()

for (t in 2:(length(as.numeric(ewma.smoothing$fitted[, 1]))
+ 1)) {

```

```

    smoothed.values[t] <- as.numeric(ewma.smoothing$fitted[,
1])[t - 1]
}

lambda[1] <- ewma.smoothing$alpha

options("scipen" = 10)
options("digits" = 7)

stored.final.est <- c()

f <- 1
l <- 1

estimations <- c()

smoothed.values <- round(smoothed.values, digits = digits)

t <- 2

count.for.lambda <- 2

for (j in 1:((1200 / segments))) {
  for (k in 3:segments) {
    estimations[k] <-
      difs * new.analysis.pr[k - 1] +
lambda[length(lambda)] * (new.analysis.pr[k - 1] - (difs *
new.analysis.pr[k - 2]))
  }

  k <- 0

  estimations[(t - 1):t] <- mean(tail(estimations, -2))

  stored.final.est[f:(l * segments)] <-
    estimations

  if (j < 12) {
    f <- l * segments + 1
    l <- l + 1

    estimations <- c()
  }
}

```

```

new.analysis.pr <- pr[f:(l * segments)]
ewma.smoothing <-
  HoltWinters(new.analysis.pr, beta = FALSE, gamma =
FALSE)

lambda[count.for.lambda] <- ewma.smoothing$alpha
count.for.lambda <- count.for.lambda + 1

smoothed.values <- c()
smoothed.values[1] <- mean(new.analysis.pr)

for (t in 2:(length(as.numeric(ewma.smoothing$fitted[,
1])) + 1)) {
  smoothed.values[t] <-
as.numeric(ewma.smoothing$fitted[, 1])[t - 1]
}

smoothed.values <- round(smoothed.values, digits =
digits)

  t <- 2
}

}

estimated.values <- stored.final.est

overall.lambda <- sum(lambda) / 12

final.MSE <- c()
final.AIC <- c()
final.BIC <- c()

final.estimations <- c()

for (k in 3:1200) {
  final.estimations[k] <-
  difs * pr[k - 1] + overall.lambda * (pr[k - 1] - (difs
* pr[k - 2]))
}

final.estimations[1:2] <- mean(tail(final.estimations, -2))

final.error <- c()

```

```

for (p in 1:1200) {
  final.error[p] <-
    pr[p] - final.estimate[p]
}

sqrd.error <- final.error ^ 2

final.MSE <- sum(sqrd.error) / (length(sqrd.error) - 1)

final.AIC <- 2 + (1200) * (log(2 * pi) + log(final.MSE) +
1)

final.BIC <- log(1200) + (1200) * (log(2 * pi) +
log(final.MSE) + 1)

#TYPICAL ARIMA APPROACH----

fitted.arima <- Arima(pr[1:1200], order = c(1, 1, 0))
arima.coefficient <- fitted.arima$coef[1]
arima.MSE <- fitted.arima$sigma2
arima.AIC <- fitted.arima$aic
arima.BIC <- fitted.arima$bic

ewma.smoothing.real <-
  HoltWinters(pr, beta = FALSE, gamma = FALSE)

smoothed.values.real <- c()
smoothed.values.real[1] <- mean(new.analysis.pr)

for (t in 2:(length(as.numeric(ewma.smoothing.real$fitted[,
1])) + 1)) {
  smoothed.values.real[t] <-
    as.numeric(ewma.smoothing.real$fitted[, 1])[t - 1]
}

# _____CHECK FOR INDEPENDENCY OF THE RESIDUALS (LJUNG-BOX
TEST)-----
ind <- Box.test(final.error, type = "Ljung-Box")

```

```

#_____CHECK FOR HOMOSCEDASTICITY OF THE RESIDUALS (GQ TEST)-
---
time.gq <- 1:1200
hom <- gqtest(final.error ~ time.gq)

#FORECASTING WITH MODIFIED ARIMA-----
forecast.mod.values <- c()

counter <- 1200
for (i in 1:4) {
  forecast.mod.values[i] <-
    difs * pr[counter - 1] + overall.lambda * (pr[counter -
1] - (difs *
pr[counter - 2]))
  pr[counter + 1] <- forecast.mod.values[i]
  counter <- counter + 1
}

forecast.mod.error <- c()

for (i in 1:4) {
  forecast.mod.error[i] <-
  forecast.mod.values[i] - pr.for.prediction[i]
}

forecast.mod.error.MSE <- sum(forecast.mod.error ^ 2) / 4

#FORECASTING WITH TYPICAL ARIMA-----
forc.typ.arima <- forecast(fitted.arima, h = 4)

forecast.error <- c()

for (i in 1:4) {
  forecast.error[i] <- forc.typ.arima$fitted[i] -
pr.for.prediction[i]
}

forecast.error.MSE <- sum(forecast.error ^ 2) / 4

cat("\014") #Clear console

```

####MEAN AND VARIANCE OF ORIGINAL 1200 DATA####

mean.1200.data

var.1200.data

DIAGNOSTICS MODIFIED ARIMA####

ind

hom

MODIFIED ARIMA RESULTS

lambda

overall.lambda

final.MSE

final.AIC

final.BIC

forecast.mod.error.MSE

TYPICAL ARIMA RESULTS

arma.coefficient

arma.MSE

arma.AIC

arma.BIC

forecast.error.MSE