

# ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΕΙΡΑΙΩΣ

Σχολή Χρηματοοικονομικής και Στατιστικής



Τμήμα Στατιστικής και Ασφαλιστικής Επιστήμης

ΜΕΤΑΠΤΥΧΙΑΚΟ ΠΡΟΓΡΑΜΜΑ ΣΠΟΥΔΩΝ ΣΤΗΝ  
ΑΝΑΛΟΓΙΣΤΙΚΗ ΕΠΙΣΤΗΜΗ ΚΑΙ ΔΙΑΧΕΙΡΙΣΗ ΚΙΝΔΥΝΩΝ

## ΜΕΛΕΤΗ ΑΚΡΑΙΩΝ ΠΑΡΑΤΗΡΗΣΕΩΝ ΣΕ ΧΡΗΜΑΤΟΟΙΚΟΝΟΜΙΚΕΣ ΧΡΟΝΟΣΕΙΡΕΣ ΜΕ ΣΤΟΧΑΣΤΙΚΗ ΜΕΤΑΒΛΗΤΟΤΗΤΑ

Γρηγόριος Ν. Τσοπανάκης

Διπλωματική Εργασία

που υποβλήθηκε στο Τμήμα Στατιστικής και  
Ασφαλιστικής Επιστήμης του Πανεπιστημίου  
Πειραιώς ως μέρος των απαιτήσεων για την απόκτηση  
του Μεταπτυχιακού Διπλώματος Ειδίκευσης στην  
*Αναλογιστική Επιστήμη και Διαχείριση Κινδύνων*

Πειραιάς

Σεπτέμβρης 2022



# ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΕΙΡΑΙΩΣ

Σχολή Χρηματοοικονομικής και Στατιστικής



Τμήμα Στατιστικής και Ασφαλιστικής Επιστήμης

ΜΕΤΑΠΤΥΧΙΑΚΟ ΠΡΟΓΡΑΜΜΑ ΣΠΟΥΔΩΝ ΣΤΗΝ  
ΑΝΑΛΟΓΙΣΤΙΚΗ ΕΠΙΣΤΗΜΗ ΚΑΙ ΔΙΑΧΕΙΡΙΣΗ ΚΙΝΔΥΝΩΝ

## ΜΕΛΕΤΗ ΑΚΡΑΙΩΝ ΠΑΡΑΤΗΡΗΣΕΩΝ ΣΕ ΧΡΗΜΑΤΟΟΙΚΟΝΟΜΙΚΕΣ ΧΡΟΝΟΣΕΙΡΕΣ ΜΕ ΣΤΟΧΑΣΤΙΚΗ ΜΕΤΑΒΛΗΤΟΤΗΤΑ

Γρηγόριος Ν. Τσοπανάκης

Διπλωματική Εργασία

που υποβλήθηκε στο Τμήμα Στατιστικής και  
Ασφαλιστικής Επιστήμης του Πανεπιστημίου  
Πειραιώς ως μέρος των απαιτήσεων για την απόκτηση  
του Μεταπτυχιακού Διπλώματος Ειδίκευσης στην  
*Αναλογιστική Επιστήμη και Διαχείριση Κινδύνων*

Πειραιάς

Σεπτέμβρης 2022

Η παρούσα Διπλωματική Εργασία εγκρίθηκε ομόφωνα από την Τριμελή Εξεταστική Επιτροπή που ορίστηκε από τη ΓΣ του Τμήματος Στατιστικής και Ασφαλιστικής Επιστήμης του Πανεπιστημίου Πειραιώς στην υπ' αριθμ. .... συνεδρίασή του σύμφωνα με τον Εσωτερικό Κανονισμό Λειτουργίας του Προγράμματος Μεταπτυχιακών Σπουδών στην Αναλογιστική Επιστήμη και Διαχείριση Κινδύνων.

Τα μέλη της Επιτροπής ήταν:

- Μιχαήλ Μπούτσικας, Αναπληρωτής Καθηγητής (Επιβλέπων)
- Γεώργιος Τζαβελάς, Αναπληρωτής Καθηγητής
- Γεώργιος Ψαρράκος, Αναπληρωτής Καθηγητής

Η έγκριση της Διπλωματικής Εργασίας από το Τμήμα Στατιστικής και Ασφαλιστικής Επιστήμης του Πανεπιστημίου Πειραιώς δεν υποδηλώνει αποδοχή των γνωμών του συγγραφέα.

**UNIVERSITY OF PIRAEUS**  
**School of Finance and Statistics**



**Department of Statistics and Insurance Science**

**POSTGRADUATE PROGRAM IN ACTUARIAL  
SCIENCE AND RISK MANAGEMENT**

**EXTREMES IN FINANCIAL TIME  
SERIES WITH STOCHASTIC  
VOLATILITY**

Grigorios Tsopanakis

MSc Dissertation

Submitted to the department of Statistics and Insurance  
Science of the University of Piraeus in partial fulfillment  
of the requirements for the degree of Master *in Actuarial  
Science and Risk Management*

Piraeus, Greece

September 2022



## ΕΥΧΑΡΙΣΤΙΕΣ

Ευχαριστώ τον καθηγητή κύριο Μ. Μπούτσικα για την πολύτιμη βοήθεια και καθοδήγηση που μου προσέφερε στη συγγραφή της παρούσας διπλωματικής εργασίας, καθώς και για την προθυμία που επέδειξε να απαντήσει στις ερωτήσεις που ανέκυψαν κατά τη μελέτη του συγκεκριμένου θέματος.





## ΠΕΡΙΛΗΨΗ

Στην παρούσα διπλωματική εργασία μελετάται η εκτίμηση της αξίας σε κίνδυνο ( $VaR$  και  $TVaR$ ) ενός χρηματοοικονομικού στοιχείου (π.χ. μετοχής, χρηματιστηριακού δείκτη, κρυπτονομίσματος κ.α.), χρησιμοποιώντας ένα υπόδειγμα διαφορετικό σε σχέση με το απλούστερο μοντέλο της αγοράς, σύμφωνα με το οποίο οι λογαριθμικές αποδόσεις κατανέμονται κανονικά. Συγκεκριμένα, η υπόθεση της κανονικότητας των λογαριθμικών αποδόσεων με σταθερή μεταβλητότητα παραβιάζεται τις περισσότερες φορές, όπως υποδηλώνουν τα δεδομένα που λαμβάνονται από την αγορά. Επομένως, η χρήση γενικευμένων αυτοπαλίδρομων μοντέλων δεσμευμένης ετεροσκεδαστικότητας ( $GARCH$ ), που περιγράφουν ικανοποιητικά τη μεταβλητότητα των λογαριθμικών αποδόσεων, σε συνδυασμό με τη Θεωρία Ακραίων Τιμών ( $EVT$ ), για την περίπτωση δεδομένων από μια κατανομή με βαριά ουρά, οδηγούν σε καλύτερα αποτελέσματα.

Στο πρώτο κεφάλαιο, γίνεται μια σύντομη αναφορά στα βασικά στοιχεία που συνθέτουν μια χρονοσειρά, ενώ στη συνέχεια εισάγονται τα μοντέλα  $ARCH$  και  $GARCH$  καθώς και οι μέθοδοι που απαιτούνται για το χειρισμό αυτών.

Στο δεύτερο κεφάλαιο, εισάγεται η Θεωρία Ακραίων Τιμών. Παρουσιάζονται τα βασικά πιθανοθεωρητικά αποτελέσματα σε συνδυασμό με τις μεθόδους  $Block Maxima$  και  $POT$  που χρησιμοποιούνται για τη στατιστική ανάλυση των δεδομένων και την εκτίμηση των παραμέτρων των κατανομών  $GEV$  και  $GPD$ .

Στο τρίτο κεφάλαιο, συνδυάζονται το υπόδειγμα  $GARCH$  με τη Θεωρία Ακραίων Τιμών, με σκοπό την παρουσίαση της μεθοδολογίας που ακολουθείται για τον υπολογισμό των μέτρων κινδύνου  $VaR$  και  $TVaR$ .

Στο τέταρτο κεφάλαιο, παρουσιάζεται η μεθοδολογία μέσω πρακτικής εφαρμογής σε δεδομένα που αφορούν το κρυπτονόμισμα *Bitcoin*. Γίνεται έλεγχος αποτελεσματικότητας της μεθοδολογίας έναντι άλλων απλούστερων υποδειγμάτων, λαμβάνοντας ιστορικά δεδομένα για τις περιπτώσεις των μετοχών των εταιριών *Amazon* και *Tesla*, του κρυπτονομίσματος *Bitcoin* καθώς και των χρηματιστηριακών δεικτών *Dow Jones*, *FTSE* (χρηματιστηριακός δείκτης Λονδίνου) και *Nasdaq*.

Τέλος, αναφέρεται πως για την πρακτική εφαρμογή της μεθοδολογίας κρίθηκε απαραίτητη η χρήση του στατιστικού πακέτου  $R$ , τόσο για τη δημιουργία των

κατάλληλων αλγορίθμων όσο και για τη στατιστική επεξεργασία και παρουσίαση των δεδομένων.



## ABSTRACT

The main aim of this Master Thesis is to present and numerically evaluate an improved methodology for the estimation of the Value at Risk (VaR and TVaR) of a financial asset (e.g. stock, stock market index, cryptocurrency, etc.) using an alternative model compared to the simpler market one. According to the simpler model logarithmic returns are assumed to be normally distributed with constant volatility. However, this assumption is violated most of the time, as suggested by the data obtained from the market. The alternative approach presented in this thesis, combines Generalized Autoregressive Conditional Heteroskedasticity models (*GARCH*) with Extreme Value Theory (*EVT*) to obtain better *VaR* estimates assuming volatility clustering and heavy tails for the distribution of the log-returns.

In the first chapter, the main characteristics of time series are presented and the *ARCH* and *GARCH* models are introduced along with the appropriate methods required for their exposition.

In the second chapter, Extreme Value Theory is introduced along with its fundamental probabilistic results. Also, the main two statistical methods, Block Maxima and POT are presented for the statistical analysis of extreme value data and the parameters estimation of *GEV* and *GPD* distributions respectively.

In the third chapter, *GARCH* models are combined with the Extreme Value Theory, in order to obtain improved estimates for *VaR* and *TVaR*.

In the fourth chapter, the aforementioned methodology is illustrated by applying the *VaR* and *TVaR* estimation *GARCH – EVT* methods to cryptocurrency data (*Bitcoin*). Furthermore, we back-test the proposed model using stock-market and cryptocurrency data (*Amazon, Tesla, Bitcoin, Dow Jones Index, FTSE Index, Nasdaq Index,*) and numerically assess its effectiveness against simpler models.

The practical application of the 4<sup>th</sup> Chapter was carried out using the *R* package, both for the statistical processing and the presentation of the results.



# ΠΕΡΙΕΧΟΜΕΝΑ

<b>ΕΥΧΑΡΙΣΤΙΕΣ.....</b>	<b>VII</b>
<b>ΠΕΡΙΛΗΨΗ .....</b>	<b>IX</b>
<b>ABSTRACT .....</b>	<b>XII</b>
<b>ΠΙΝΑΚΑΣ ΣΧΗΜΑΤΩΝ.....</b>	<b>XVI</b>
<b>ΚΕΦΑΛΑΙΟ 1 ΜΟΝΤΕΛΑ ARCH - GARCH .....</b>	<b>1</b>
1.1 ΕΙΣΑΓΩΓΗ.....	1
1.2 ΒΑΣΙΚΑ ΧΑΡΑΚΤΗΡΙΣΤΙΚΑ ΧΡΟΝΟΣΕΙΡΩΝ.....	1
1.3 ΧΑΡΑΚΤΗΡΙΣΤΙΚΑ ΧΡΗΜΑΤΟΟΙΚΟΝΟΜΙΚΩΝ ΧΡΟΝΟΣΕΙΡΩΝ.....	9
1.4 ΑΥΤΟΠΑΛΙΔΡΟΜΑ ΜΟΝΤΕΛΑ ΔΕΣΜΕΥΜΕΝΗΣ ΕΤΕΡΟΣΚΕΔΑΣΤΙΚΟΤΗΤΑΣ - ARCH MODELS.....	10
1.5 ΙΔΙΟΤΗΤΕΣ ΜΟΝΤΕΛΟΥ ARCH.....	12
1.6 ΕΛΕΓΧΟΣ ΥΠΑΡΕΞΗΣ ΕΤΕΡΟΣΚΕΔΑΣΤΙΚΟΤΗΤΑΣ .....	14
1.7 ΕΠΙΛΟΓΗ ΤΑΞΗΣ ΥΠΟΔΕΙΓΜΑΤΟΣ.....	16
1.8 QUASI MAXIMUM LIKELIHOOD .....	17
1.9 ΜΕΙΟΝΕΚΤΗΜΑΤΑ ΥΠΟΔΕΙΓΜΑΤΟΣ ARCH.....	19
1.10 ΓΕΝΙΚΕΥΜΕΝΑ ΑΥΤΟΠΑΛΙΔΡΟΜΑ ΜΟΝΤΕΛΑ ΔΕΣΜΕΥΜΕΝΗΣ ΕΤΕΡΟΣΚΕΔΑΣΤΙΚΟΤΗΤΑΣ - GARCH MODELS.....	20
1.11 ΙΔΙΟΤΗΤΕΣ ΥΠΟΔΕΙΓΜΑΤΟΣ GARCH.....	22
1.12 ΕΠΙΛΟΓΗ ΤΑΞΗΣ ΥΠΟΔΕΙΓΜΑΤΟΣ.....	25
1.13 ΕΚΤΙΜΗΣΗ ΠΑΡΑΜΕΤΡΩΝ ΜΟΝΤΕΛΟΥ .....	27
<b>ΚΕΦΑΛΑΙΟ 2 ΘΕΩΡΙΑ ΑΚΡΑΙΩΝ ΤΙΜΩΝ.....</b>	<b>30</b>
2.1 ΕΙΣΑΓΩΓΗ.....	30
2.2 ΘΕΩΡΙΑ ΑΚΡΑΙΩΝ ΤΙΜΩΝ .....	30
2.3 ΜΕΘΟΔΟΣ ΤΟΥ ΜΕΡΙΚΟΥ ΜΕΓΙΣΤΟΥ (BLOCK MAXIMA METHOD).....	35
2.4 ΣΤΑΘΜΗ ΑΠΟΔΟΣΗΣ .....	36
2.5 ΕΚΤΙΜΗΣΗ ΠΑΡΑΜΕΤΡΩΝ ΜΟΝΤΕΛΟΥ .....	38
2.6 ΕΛΕΓΧΟΣ ΚΑΛΗΣ ΠΡΟΣΑΡΜΟΓΗΣ.....	39
2.7 ΜΕΘΟΔΟΣ ΡΟΤ (PEAK OVER THRESHOLD) .....	40
2.8 ΣΤΑΘΜΗ ΑΠΟΔΟΣΗΣ .....	42
2.9 ΕΠΙΛΟΓΗ ΚΑΤΑΛΛΗΛΟΥ ΚΑΤΩΦΛΙΟΥ $U$ .....	43
2.10 ΜΕΘΟΔΟΣ ΕΚΤΙΜΗΣΗΣ ΠΑΡΑΜΕΤΡΩΝ .....	46
2.11 ΕΛΕΓΧΟΣ ΚΑΛΗΣ ΠΡΟΣΑΡΜΟΓΗΣ.....	47
<b>ΚΕΦΑΛΑΙΟ 3 ΑΞΙΑ ΣΕ ΚΙΝΔΥΝΟ ΚΑΙ ΥΠΟ-ΣΥΝΘΗΚΗ ΑΞΙΑ ΣΕ ΚΙΝΔΥΝΟ ΑΞΙΟΓΡΑΦΟΥ.....</b>	<b>48</b>
3.1 ΕΙΣΑΓΩΓΗ.....	48
3.2 ΜΕΛΕΤΗ ΑΠΟΔΟΣΗΣ ΑΞΙΟΓΡΑΦΟΥ .....	49

3.3 ΑΞΙΑ ΣΕ ΚΙΝΔΥΝΟ (VALUE AT RISK - VAR) ΚΑΙ ΥΠΟ ΣΥΝΘΗΚΗ ΑΞΙΑ ΣΕ ΚΙΝΔΥΝΟ (TAIL VALUE AT RISK - TVAR).....	51
<b>ΚΕΦΑΛΑΙΟ 4 ΕΦΑΡΜΟΓΗ .....</b>	<b>54</b>
4.1 ΕΙΣΑΓΩΓΗ.....	54
4.2 ΠΡΟΒΛΕΨΗ VAR ΚΑΙ TVAR.....	54
4.3 ΕΛΕΓΧΟΣ ΜΕΘΟΔΟΛΟΓΙΑΣ .....	64
<b>ΠΑΡΑΡΤΗΜΑ Ι.....</b>	<b>74</b>
<b>ΒΙΒΛΙΟΓΡΑΦΙΑ .....</b>	<b>87</b>

## ΠΙΝΑΚΑΣ ΣΧΗΜΑΤΩΝ

<b>Σχήμα 1.1:</b> Γραφικές παραστάσεις ενός Gaussian λευκού θορύβου, ενός υποδείγματος $ARMA(1,1)$ και ενός τυχαίου περιπάτου.....	3
<b>Σχήμα 1.2:</b> Στιγμιότυπα από την πραγματοποίηση του ίδιου στοχαστικού φαινομένου. Στην περίπτωση όπου υπήρχε ένα επαρκές δείγμα πραγματοποιήσεων (στιγμιότυπων) του στοχαστικού φαινομένου, θα ήταν δυνατή η μελέτη της τ.μ. $X_t$ , χρησιμοποιώντας μεθόδους της κλασικής στατιστικής.....	4
<b>Σχήμα 1.3:</b> Γραφική παράσταση των τιμών κλεισίματος του δείκτη $S\&P500$ . Η μόνη διαθέσιμη πληροφορία που υπάρχει για την τ.μ. $X_t$ είναι η τιμή που έχει λάβει τη χρονική στιγμή 3/9/2019. ....	5
<b>Σχήμα 1.4:</b> Παράδειγμα χρονοσειράς όπου οι τιμές κινούνται (κατά μέσο όρο) γύρω από τη μονάδα. Στη συγκεκριμένη περίπτωση μπορούμε να ισχυριστούμε ότι ο μέσος της χρονοσειράς $\mu = 1$ .....	5
<b>Σχήμα 1.5:</b> Στο πρώτο γράφημα ο μέσος της χρονοσειράς αυξάνεται με το πέρασμα του χρόνου (ύπαρξη τάσης), ενώ παρόμοια συμπεριφορά παρουσιάζει και η διασπορά. Μια τέτοια χρονοσειρά χαρακτηρίζεται ως μη-στάσιμη χρονοσειρά. Αντίθετα, στο δεύτερο γράφημα τόσο ο μέσος όσο και η διασπορά δεν μεταβάλλονται σημαντικά. Στην περίπτωση αυτή, η χρονοσειρά είναι στάσιμη.....	7
<b>Σχήμα 1.6:</b> Γραφική παράσταση αυτοσυσχετίσεων ενός υποδείγματος $AR(1)$ και ενός $MA(2)$ . Αρκετές φορές η μορφή ενός διαγράμματος αυτοσυσχέτισης μπορεί να βοηθήσει στην ορθή επιλογή ενός υποδείγματος καθώς και στην ανίχνευση τάσης και εποχικότητας στα δεδομένα.....	8
<b>Σχήμα 1.7:</b> Σύγκριση γραφικών παραστάσεων μιας χρονοσειράς λευκού θορύβου σταθερής διασποράς και της χρονοσειράς των αποδόσεων του δείκτη $S\&P500$ . ....	10
<b>Σχήμα 1.8:</b> Γραφική παράσταση της χρονοσειράς $x_t = 0.5x_{t-1} + 0.2x_{t-2} + e_t$ , με $e_t \sim N(0,1)$ . Στο δεύτερο γράφημα απεικονίζεται η γραφική παράσταση της συνάρτησης μερικής αυτοσυσχέτισης της παραπάνω χρονοσειράς. Σύμφωνα με τη θεωρία, η γραφική παράσταση της συνάρτησης μερικής αυτοσυσχέτισης ενός $AR(2)$ υποδείγματος, αναμένεται να λάβει σχεδόν μηδενικές τιμές μετά την εκτίμηση της δεύτερης παραμέτρου του μοντέλου.....	17
<b>Σχήμα 2.1:</b> Γραφική παράσταση κατανομής πιθανότητας των συναρτήσεων <i>Frechet</i> , ( <i>reversed</i> ) <i>Weibull</i> και <i>Gumbel</i> . ....	33



<b>Σχήμα 2.2:</b> Γραφική παράσταση της στάθμης απόδοσης για τιμές της παραμέτρου θέσης $\xi = -0.5, 0, 0.5$ . Όπως φαίνεται στο γράφημα, όταν η παράμετρος θέσης λαμβάνει τιμές κοντά στο μηδέν η στάθμη απόδοσης είναι γραμμική συνάρτηση, ενώ όταν λαμβάνει θετικές (αντ. αρνητικές) τιμές η στάθμη απόδοσης είναι κοίλη (αντ. κυρτή).	37
<b>Σχήμα 2.3:</b> $QQ - plot$ για την περίπτωση προσαρμογής της κατανομής $GEV(\mu = 0.13, \sigma = 1.1, \xi = 0.2)$ σε τυχαίο δείγμα. Στο γράφημα επίσης έχουν σχεδιαστεί και τα αντίστοιχα όρια των διαστημάτων εμπιστοσύνης για το κάθε ποσοστιαίο σημείο.	40
<b>Σχήμα 2.4:</b> Γραφική παράσταση κατανομής $GPD$ για τις διάφορες τιμές του $\xi$ .	42
<b>Σχήμα 2.5:</b> Γραφική παράσταση μέσης υπερβάλλουσας ζημίας. Παρατηρούμε ότι στην τιμή $u = 0.2$ η γραφική παράσταση της συνάρτησης φαίνεται να γίνεται αύξουσα. Επομένως, σαν κατώφλι μπορούμε να επιλέξουμε την τιμή $u_0 = 0.2$ .	44
<b>Σχήμα 2.6:</b> Στο πρώτο γράφημα φαίνεται η γραφική παράσταση της παραμέτρου κλίμακας $\sigma^*$ ως προς το κατώφλι $u$ , ενώ στο δεύτερο φαίνεται η γραφική παράσταση της παραμέτρου σχήματος $\xi$ ως προς το κατώφλι $u$ . Η ελάχιστη τιμή $u_0$ για την οποία παρατηρείται ότι οι ποσότητες $\xi$ και $\sigma^*(u)$ παραμένουν σταθερές όταν το $u > u_0$ , αποτελεί τη ζητούμενη ποσότητα. Τη δεδομένη περίπτωση, μια επιλογή θα μπορούσε να ήταν η τιμή $u_0 = 0.7$ .	45
<b>Σχήμα 2.7:</b> Γραφικές παραστάσεις, $PP - plot, QQ - plot$ , συνάρτησης κατανομής πιθανότητας και στάθμης απόδοσης για τον έλεγχο καλής προσαρμογής του μοντέλου.	47
<b>Σχήμα 4.1:</b> Γραφική παράσταση ημερήσιων τιμών Bitcoin για το χρονικό διάστημα 01-01-2017 έως 03-01-2021.	55
<b>Σχήμα 4.2:</b> Στο πρώτο γράφημα φαίνεται η εξέλιξη του αρνητικού λογάριθμου των αποδόσεων, ενώ στο δεύτερο και τρίτο σχεδιάζονται οι αυτοσυσχετίσεις της χρονοσειράς του αρνητικού λογάριθμου και του τετραγώνου αυτού.	56
<b>Σχήμα 4.3:</b> Στο πρώτο γράφημα απεικονίζεται η γραφική παράσταση των καταλοίπων $r_t$ σε συνδυασμό με το διπλάσιο ( $\pm 2$ ) της εκτιμώμενης υπό-συνθήκη διασποράς, όπως προκύπτει μετά την εφαρμογή ενός $GARCH(1,1)$ μοντέλου. Αντίστοιχα, στο δεύτερο γράφημα απεικονίζονται οι πραγματικές τιμές της εκτιμώμενης υπό-συνθήκη διασποράς.	58
<b>Σχήμα 4.4:</b> Γραφικές παραστάσεις των αυτοσυσχετίσεων των χρονοσειρών $\hat{\epsilon}_t$ και $\hat{\epsilon}_t^2$ , αντίστοιχα.	59

<b>Σχήμα 4.5:</b> Γραφικός έλεγχος κανονικότητας χρονοσειράς $\hat{e}_t$ . Η απόκλιση των σημείων από την ευθεία (πράσινη) γραμμή, υποδηλώνει την μη κανονικότητα της κατανομής των $\hat{e}_t$ .....	60
<b>Σχήμα 4.6:</b> Σύγκριση της εκτιμώμενης κατανομής της χρονοσειράς $e_t$ (πράσινη γραμμή) σε σχέση με την κανονική κατανομή (μπλε γραμμή). Όπως απεικονίζεται στο πρώτο γράφημα, η κατανομή των $e_t$ φαίνεται να είναι λεπτόκυρτη και επεκτείνεται προς τα δεξιά (θετικά ασύμμετρη). Μάλιστα, το γεγονός ότι η ουρά της κατανομής των $e_t$ φαίνεται να είναι βαρύτερη από αυτή της κανονικής, διακρίνεται καλύτερα στο δεύτερο γράφημα.....	61
<b>Σχήμα 4.7:</b> Γραφική παράσταση των τιμών της χρονοσειράς $e_t$ , για την οποία θα εφαρμοστεί η Θεωρία Ακραίων Τιμών.....	62
<b>Σχήμα 4.8:</b> Γραφικές παραστάσεις των παραμέτρων κλίμακας $\sigma$ και $\xi$ ως προς το κατώφλι $u$ . Σαν ελάχιστο κατώφλι για το οποίο οι γραφικές παραστάσεις φαίνεται να σταθεροποιούνται επιλέγεται η τιμή $u = 0.5$ . .....	62
<b>Σχήμα 4.9:</b> <i>QQ – plot</i> για τον έλεγχο καλής προσαρμογής της εκτιμώμενης <i>GPD</i> κατανομής στα δεδομένα. ....	64
<b>Σχήμα 4.10:</b> Γραφικές παραστάσεις των αρνητικών λογαριθμικών αποδόσεων των χρηματοοικονομικών στοιχείων Amazon, Bitcoin, Tesla, Dow Jones, FTSE και Nasdaq για τις περιόδους 1/4/2010 έως 30/12/2020, 17/9/2014 έως 31/12/2021, 29/6/2010 έως 30/12/2020, 3/1/2000 έως 30/12/2013, 3/1/2012 έως 30/12/2021 και αντίστοιχα 2/1/2014 έως 30/12/2021.....	66

# ΚΕΦΑΛΑΙΟ 1 ΜΟΝΤΕΛΑ ARCH - GARCH

## 1.1 ΕΙΣΑΓΩΓΗ

Στο παρόν κεφάλαιο γίνεται μια σύντομη παρουσίαση των βασικών ιδιοτήτων των χρονολογικών σειρών, ενώ δίνεται ιδιαίτερη έμφαση στο Αυτοπαλίνδρομο Μοντέλο Δεσμευμένης Ετεροσκεδαστικότητας (*AutoRegressive Conditional Heteroscedasticity - ARCH*) και στο Γενικευμένο Αυτοπαλίνδρομο Μοντέλο Δεσμευμένης Ετεροσκεδαστικότητας (*Generalized AutoRegressive Conditionally Heteroscedastic - GARCH*).

## 1.2 ΒΑΣΙΚΑ ΧΑΡΑΚΤΗΡΙΣΤΙΚΑ ΧΡΟΝΟΣΕΙΡΩΝ

Μια χρονοσειρά είναι μια ακολουθία παρατηρήσεων που περιγράφει την χρονική εξέλιξη κάποιου ποσοτικού χαρακτηριστικού ενός στοχαστικού φαινομένου.

Οι χρονοσειρές συναντώνται σε αρκετούς κλάδους της επιστήμης όπως για παράδειγμα στην οικονομία, τη μετεωρολογία και τη σεισμολογία. Επομένως, η εύρεση κατάλληλων τεχνικών καθίσταται αναγκαία τόσο για την μελέτη αυτών όσο και για την κατανόηση των υπό εξέταση στοχαστικών φαινομένων.

Ένα ιδιαίτερο χαρακτηριστικό που απαντάται στις περισσότερες χρονοσειρές είναι ότι οι διαδοχικές παρατηρήσεις που περιγράφουν ένα στοχαστικό φαινόμενο παρουσιάζουν (συνήθως) μια εξάρτηση. Συνεπώς, η πολυπλοκότητα του υπό εξέταση προβλήματος γίνεται μεγαλύτερη, καθώς τεχνικές της κλασσικής στατιστικής που αφορούν την μελέτη ανεξάρτητων τυχαίων δειγμάτων δεν είναι δυνατό να εφαρμοστούν.

Τα παραπάνω οδηγούν στο συμπέρασμα ότι η ανάλυση μιας χρονοσειράς έγκειται στην εύρεση ενός κατάλληλου μαθηματικού υποδείγματος, το οποίο θα περιγράφει όσο το δυνατόν καλύτερα τα διαθέσιμα δεδομένα, χρησιμοποιώντας ένα σύνολο τεχνικών που θα αναλυθούν στην συνέχεια.

Εδώ είναι σημαντικό να διευκρινιστεί ότι τα μαθηματικά πρότυπα που χρησιμοποιούνται στην ανάλυση χρονοσειρών ανήκουν στην κατηγορία των στοχαστικών μοντέλων. Δηλαδή, μαθηματικών υποδειγμάτων μέσω των οποίων η

μελλοντική τιμή του υπό εξέταση χαρακτηριστικού δεν μπορεί να υπολογιστεί με ακρίβεια, αλλά με μια πιθανότητα αυτό να λάβει τιμές εντός ενός διαστήματος.

Το σύνολο των παρατηρήσεων μιας χρονοσειράς συμβολίζεται με  $X$ , ενώ αναφερόμαστε σε αυτήν ως μια χρονοσειρά  $X$ .

Οι παρατηρήσεις του συνόλου  $X$  συμβολίζονται με  $x_t$ ,  $t \in T$  όπου το  $T$  εκφράζει το μετρούμενο χρόνο, ο οποίος μπορεί να λάβει είτε συνεχείς είτε διακριτές τιμές.

Στην παρούσα εργασία, η μελέτη και παρουσίαση των χρονοσειρών γίνεται λαμβάνοντας υπόψη ότι το  $T \subset \mathbb{N}$  ή  $T \subset \mathbb{Z}$ .

Μερικά κλασικά παραδείγματα χρονοσειρών που εξηγούν αρκετά φαινόμενα, είτε αποτελούν μέρος ενός γενικότερου μαθηματικού υποδείγματος είναι τα εξής:

**Θόρυβος (Noise):** Μια χρονοσειρά  $X = \{X_t, t \in T\}$  που αποτελείται από ανεξάρτητες και ισόνομες τυχαίες μεταβλητές (**Independent and Identically Distributed - IID**) με μέση τιμή  $E[X_t] = 0$  και διασπορά  $V[X_t] = \sigma^2 < \infty$ , συχνά αναφέρεται στην βιβλιογραφία ως θόρυβος ή ισχυρός λευκός θόρυβος (**Strong White Noise**) και συμβολίζεται με  $IIDN(0, \sigma^2)$ .

**Λευκός Θόρυβος (White Noise):** Στην περίπτωση που στην προηγούμενη χρονοσειρά  $X$  αντικαταστήσουμε το «ανεξάρτητες τ.μ.» με «ασυσχέτιστες τ.μ.», τότε λέμε ότι η  $X$  αποτελεί μια χρονοσειρά λευκού θορύβου και συμβολίζεται με  $WN(0, \sigma^2)$ .

Εύκολα γίνεται αντιληπτό ότι, ο ορισμός της χρονοσειράς λευκού θορύβου είναι γενικότερος του θορύβου, διότι δυο ανεξάρτητες τ.μ. είναι ασυσχέτιστες ενώ το αντίθετο δεν ισχύει πάντοτε.

**Τυχαίος Περίπατος (Random Walk):** Έστω η χρονοσειρά  $X = \{X_t, t \in T\}$  με:

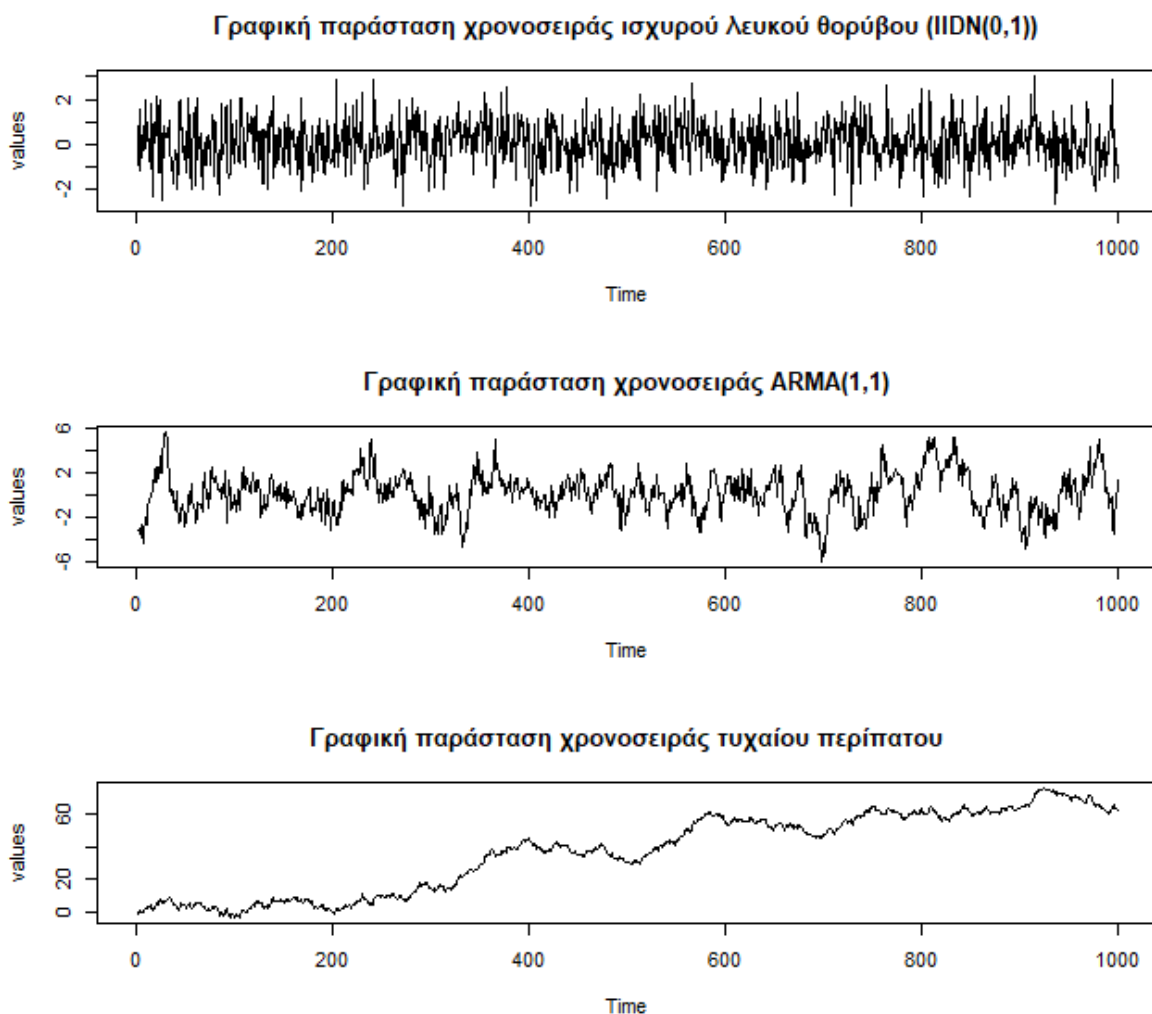
$$X_t = \sum_{i=1}^t W_i, t \in \{1, 2, \dots\} \text{ και } W_i \sim WN(0, \sigma^2). \quad (1.1)$$

Μια τέτοια χρονοσειρά καλείται τυχαίος περίπατος και έχει αρκετές εφαρμογές τόσο στις φυσικές επιστήμες όσο και στην οικονομία. Για παράδειγμα, η περιγραφή της κίνησης της τιμής μιας μετοχής, αποτελεί ένα τυχαίο περίπατο.

**Χρονοσειρά ARMA:** Μια χρονοσειρά που συνδυάζει ένα αυτοπαλίνδρομο μοντέλο  $AR(p)$  με ένα μοντέλο κινούμενου μέσου  $MA(q)$  ονομάζεται  $ARMA(p,q)$  χρονοσειρά και περιγράφεται από το εξής υπόδειγμα:

$$X_t - \sum_{i=1}^p a_i X_{t-i} = w + e_t + \sum_{j=1}^q b_j e_{t-j}, \quad t \in \mathbb{Z} \text{ και } e_t \sim WN(0, \sigma^2). \quad (1.2)$$

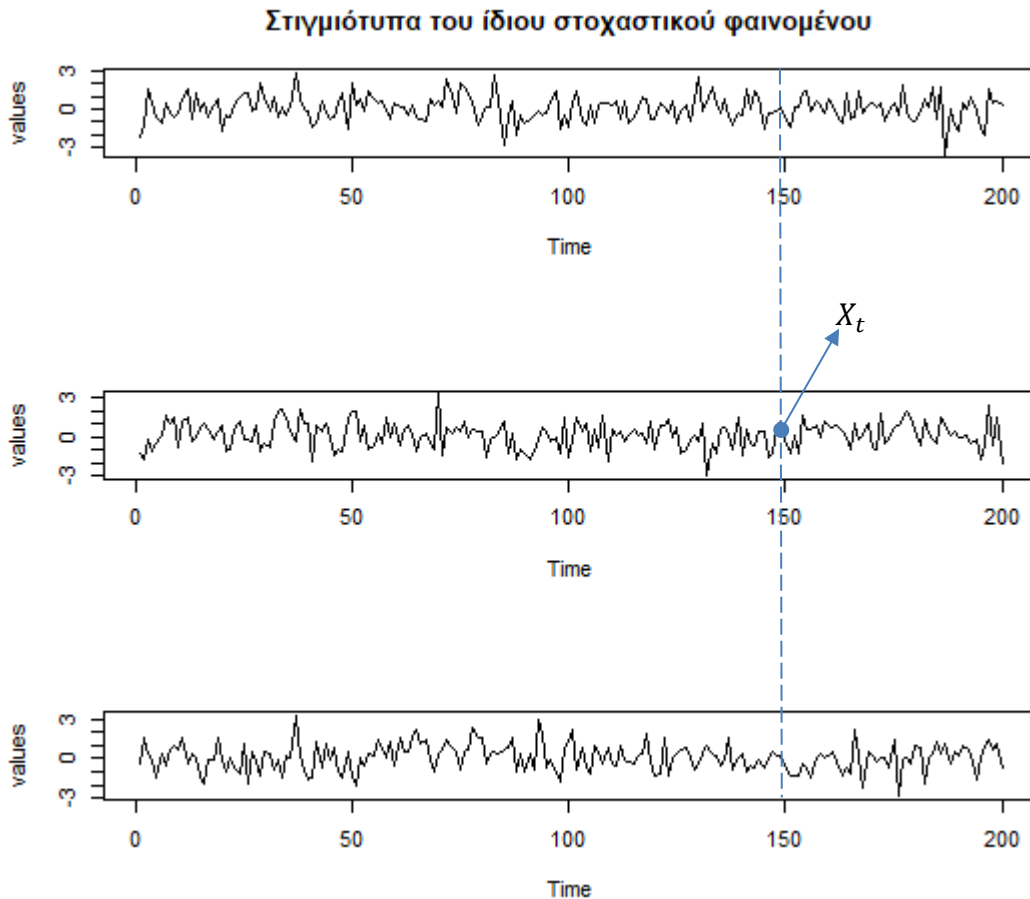
Όπως είναι φανερό, η τιμή της χρονοσειράς τη χρονική στιγμή  $t$  εξαρτάται τόσο από τις  $p$  προηγούμενες τιμές, όσο και από τα  $q$  το πλήθος προηγούμενα σφάλματα.



**Σχήμα 1.1:** Γραφικές παραστάσεις ενός Gaussian λευκού θορύβου, ενός υποδείγματος  $ARMA(1,1)$  και ενός τυχαίου περιπάτου.

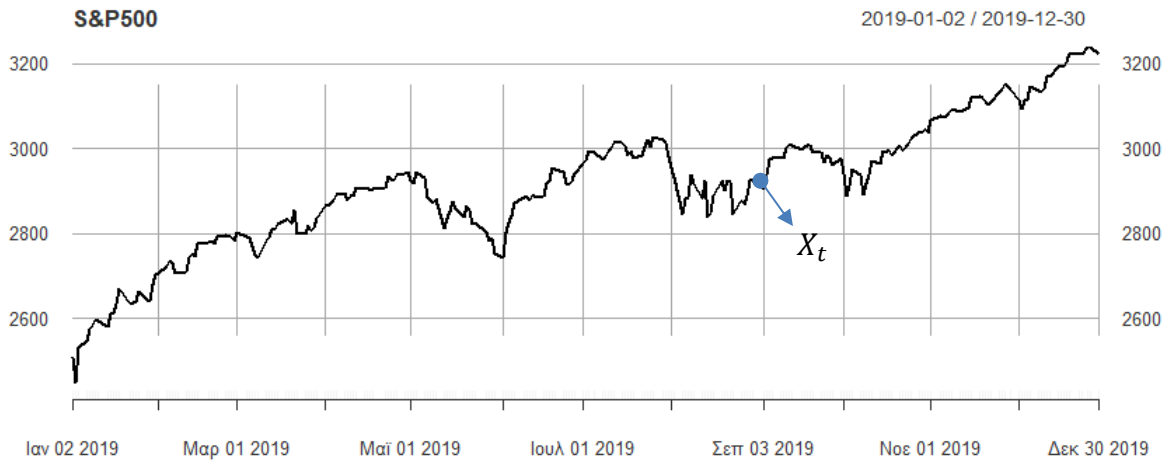
Καθώς μια χρονοσειρά εκφράζει την πραγματοποίηση ενός στοχαστικού φαινομένου, θα μπορούσαμε να φανταστούμε ότι αποτελεί ένα «στιγμιότυπο» μιας στοχαστικής ανέλιξης. Έτσι, θεωρώντας ότι μια στοχαστική ανέλιξη μπορεί να είναι μια ακολουθία

τ.μ.  $X_t$ , αν είχαμε αρκετές πραγματοποιήσεις (ή στιγμιότυπα) του στοχαστικού φαινομένου, θα μπορούσαμε να εξάγουμε κάποιο ποσοτικό συμπεράσματα για κάθε τ.μ.  $X_t$  και εν συνεχεία για ολόκληρη την ανάλυση, π.χ. βλ. Cowpertwait, P. S. P. and Metcalfe, A. V. (2009).



**Σχήμα 1.2:** Στιγμιότυπα από την πραγματοποίηση του ίδιου στοχαστικού φαινομένου. Στην περίπτωση όπου υπήρχε ένα επαρκές δείγμα πραγματοποιήσεων (στιγμιότυπων) του στοχαστικού φαινομένου, θα ήταν δυνατή η μελέτη της τ.μ.  $X_t$ , χρησιμοποιώντας μεθόδους της κλασικής στατιστικής.

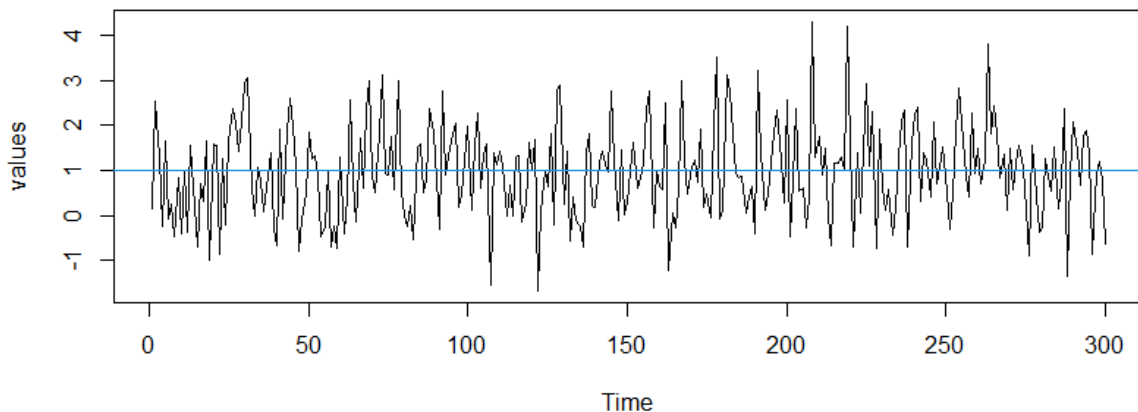
Στην πραγματικότητα όμως είναι αρκετές οι περιπτώσεις όπου η διεξαγωγή ενός (στοχαστικού) πειράματος δεν δύναται να επαναληφθεί περισσότερες από μια φορές. Ένα τέτοιο παράδειγμα αποτελεί η κίνηση της τιμής μια μετοχής. Στις περιπτώσεις αυτές, η ανάλυση που διεξάγεται περιορίζεται σε μια μόνο πραγματοποίηση της ανάλυσης που μελετάμε. Επομένως, η μόνη διαθέσιμη πληροφορία για κάθε τ.μ.  $X_t$ , είναι η τιμή που αυτή έχει λάβει!



**Σχήμα 1.3:** Γραφική παράσταση των τιμών κλεισίματος του δείκτη S&P500. Η μόνη διαθέσιμη πληροφορία που υπάρχει για την τ.μ.  $X_t$  είναι η τιμή που έχει λάβει τη χρονική στιγμή 3/9/2019.

Για να αντιπαρέλθουμε την δυσκολία αυτή, μια τεχνική που χρησιμοποιείται είναι η μελέτη της συμπεριφοράς των όρων της χρονοσειράς και των πιθανών σχέσεων που τους διέπουν. Για παράδειγμα, αν παρατηρηθεί ότι οι τιμές  $x_t, t \in T$  μιας χρονοσειράς  $X$  κινούνται γύρω από μια σταθερή τιμή, θα μπορούσαμε να υποθέσουμε ότι όλες οι τ.μ.  $X_t$  έχουν την ίδια μέση τιμή. Επομένως, λαμβάνοντας σαν μέτρο το δειγματικό μέσο μπορούμε να υπολογίσουμε το μέσο της χρονοσειράς  $X$ .

#### Παράδειγμα χρονοσειράς όπου οι τιμές κινούνται γύρω από τη μονάδα



**Σχήμα 1.4:** Παράδειγμα χρονοσειράς όπου οι τιμές κινούνται (κατά μέσο όρο) γύρω από τη μονάδα. Στη συγκεκριμένη περίπτωση μπορούμε να ισχυριστούμε ότι ο μέσος της χρονοσειράς  $\mu = 1$ .

Συμπερασματικά, μελετώντας χαρακτηριστικές ποσότητες μιας χρονοσειράς, (όπως για παράδειγμα τις ροπές πρώτης, δεύτερης ή και μεγαλύτερης τάξης) οι οποίες παραμένουν αμετάβλητες ή «στάσιμες» στο χρόνο, μπορεί να διευκολύνουν αρκετά την ανάλυση της. Αυτό μας οδηγεί στην εισαγωγή μιας εξαιρετικά χρήσιμης και συνάμα σημαντικής έννοιας των χρονοσειρών, αυτής της στασιμότητας (**stationarity**).

**Αυστηρή στασιμότητα (Strict Stationarity):** Η χρονοσειρά  $X = (X_t, t \in T)$  καλείται αυστηρώς στάσιμη αν οι πολυδιάστατες κατανομές των τυχαίων διανυσμάτων  $(X_{t_1}, X_{t_2}, \dots, X_{t_k})'$  και  $(X_{t_1+h}, X_{t_2+h}, \dots, X_{t_k+h})'$  είναι ίδιες για κάθε  $t_1, t_2, \dots, t_k \in T$  και  $h \in \mathbb{N}$ .

Ο ορισμός αυτός αναφέρει ότι σε μια αυστηρώς στάσιμη χρονοσειρά η πιθανότητα να παρατηρηθεί μια συγκεκριμένη τροχιά, παραμένει αμετάβλητη στο χρόνο. Δηλαδή είτε το πείραμα ξεκινήσει τη χρονική στιγμή  $t$  είτε την  $t + h$ , η πιθανότητα να πάρουμε σαν αποτέλεσμα του πειράματος ακριβώς την ίδια διαδρομή (τροχιά), δεν αλλάζει.

Στην πράξη, η απόδειξη της αυστηρής στασιμότητας μπορεί να αποβεί μια κοπιαστική ή και αδύνατη κάποιες φορές διαδικασία. Αυτό δικαιολογείται από το γεγονός ότι η από κοινού κατανομή των όρων που απαρτίζουν την χρονοσειρά, δεν είναι πάντοτε εφικτό να προσδιοριστεί. Ακόμα όμως και αν υπάρχει ένας κλειστός τύπος της από κοινού κατανομής, μπορεί να είναι αρκετά δύσκολος στον χειρισμό του.

Επομένως, «χαλαρώνοντας» κάπως τον παραπάνω ορισμό, μελετώντας μια χρονοσειρά όχι βάσει της από κοινού κατανομής των τ.μ.  $X_t$ , αλλά βάσει των ροπών τους, προκύπτει η έννοια της ασθενούς στασιμότητας (**weak stationarity**).

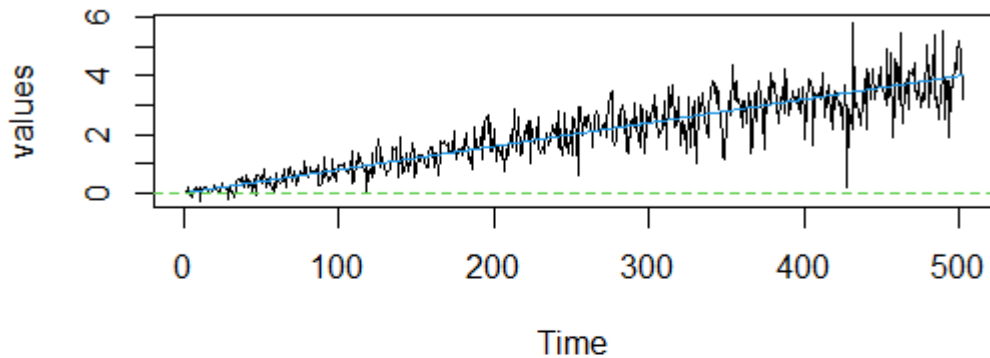
**Ασθενής στασιμότητα (Weak Stationarity):** Μια χρονοσειρά  $X = (X_t, t \in T)$  καλείται ασθενώς στάσιμη αν:

$$\begin{cases} E[X_t^2] < \infty, & \forall t \in \mathbb{Z} \\ E[X_t] = \mu, & \forall t \in \mathbb{Z} \\ Cov[X_t, X_{t+h}] = \gamma_X(h), & \forall t, h \in \mathbb{Z}. \end{cases} \quad (1.3)$$

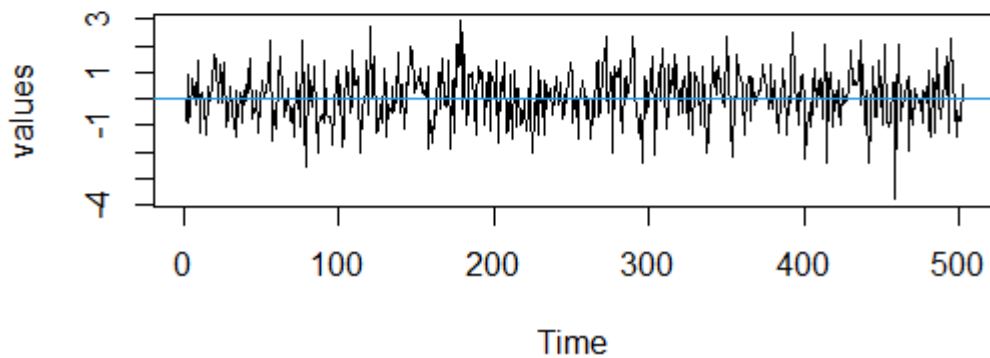
Διαισθητικά παρατηρώντας το γράφημα μιας ασθενώς στάσιμης χρονοσειράς (βλ. σχήμα 1.5) γίνεται αντιληπτό ότι οι τιμές της κυμαίνονται γύρω από ένα σταθερό μέσο, με σταθερή διασπορά.



## Μη-στάσιμη χρονοσειρά



## Στάσιμη χρονοσειρά



**Σχήμα 1.5:** Στο πρώτο γράφημα ο μέσος της χρονοσειράς αυξάνεται με το πέρασμα του χρόνου (ύπαρξη τάσης), ενώ παρόμοια συμπεριφορά παρουσιάζει και η διασπορά. Μια τέτοια χρονοσειρά χαρακτηρίζεται ως μη-στάσιμη χρονοσειρά. Αντίθετα, στο δεύτερο γράφημα τόσο ο μέσος όσο και η διασπορά δεν μεταβάλλονται σημαντικά. Στην περίπτωση αυτή, η χρονοσειρά λέγεται ότι είναι στάσιμη.

Αξίζει να σημειωθεί ότι η αυστηρή στασιμότητα μιας χρονοσειράς δεν συνεπάγεται απαραίτητα και την ασθενή στασιμότητα. Το γεγονός αυτό δικαιολογείται από τον ορισμό της αυστηρής στασιμότητας, σύμφωνα με τον οποίο η διασπορά δεν επιβάλλεται να είναι πεπερασμένη, σε αντίθεση με αυτόν της ασθενούς στασιμότητας. Βέβαια, στην περίπτωση όπου μια αυστηρώς στάσιμη χρονοσειρά έχει πεπερασμένη διασπορά, τότε είναι και ασθενώς στάσιμη.

Στη συνέχεια ορίζονται οι συναρτήσεις αυτοσυνδιακύμανσης (**autocovariance function**) και αυτοσυσχέτισης (**autocorrelation function - acf**) οι οποίες κατέχουν κεντρικό ρόλο στην ανάλυση χρονοσειρών και δεν είναι τίποτα περισσότερο από τις γνωστές συναρτήσεις συνδιακύμανσης και συσχέτισης.

Έστω μια (ασθενώς) στάσιμη χρονοσειρά  $X = (X_t, t \in T)$ , τότε:

1) η συνάρτηση αυτοσυνδιακύμανσης ορίζεται ως:

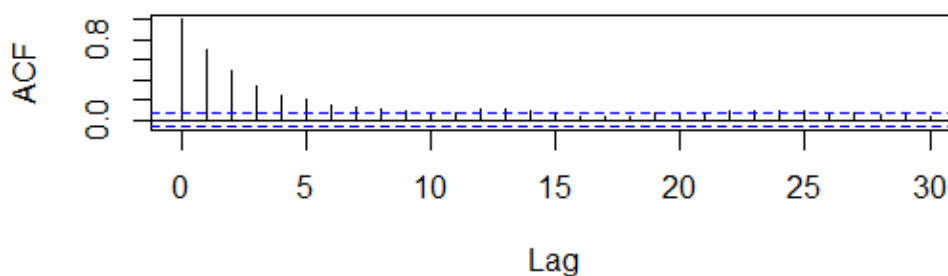
$$\gamma_X(h) = Cov(X_{t+h}, X_t), \quad t \in T \text{ και } h \in \mathbb{Z} \quad (1.4)$$

2) η συνάρτηση αυτοσυσχέτισης ορίζεται ως:

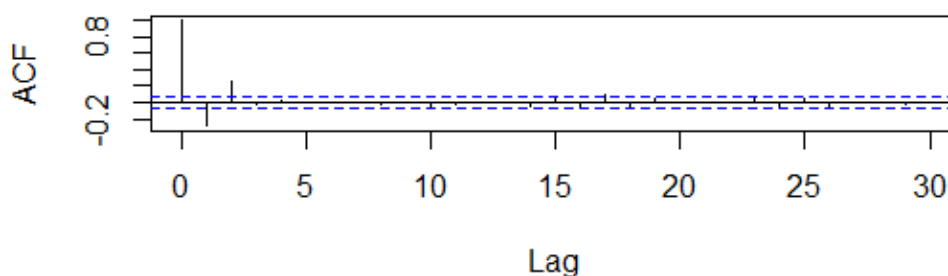
$$\rho_X(h) = \frac{Cov[X_{t+h}, X_t]}{\sqrt{Var[X_{t+h}]Var[X_t]}} = \frac{\gamma_X(h)}{\gamma_X(0)}, \quad t \in T \text{ και } h \in \mathbb{Z}. \quad (1.5)$$

Το γράφημα της συνάρτησης αυτοσυσχέτισης  $\rho_X$ , γνωστό και ως κορελόγραμμα (*correlogram*), μπορεί να δώσει χρήσιμες πληροφορίες σχετικά με την επιλογή του μοντέλου που ταιριάζει καλύτερα στα δεδομένα. Για παράδειγμα, μια εκθετική μείωση της συνάρτησης αυτοσυσχέτισης αποτελεί μια πρώτη ένδειξη ότι ένα αυτοπαλίνδρομο μοντέλο  $AR(p)$  θα μπορούσε να εφαρμοστεί στα δεδομένα.

**Γράφημα συνάρτησης αυτοσυσχέτισης  
χρονοσειράς AR(1)**



**Γράφημα συνάρτησης αυτοσυσχέτισης  
χρονοσειράς MA(2)**



**Σχήμα 1.6:** Γραφική παράσταση αυτοσυσχετίσεων ενός υποδείγματος  $AR(1)$  και ενός  $MA(2)$ . Αρκετές φορές η μορφή ενός διαγράμματος αυτοσυσχέτισης μπορεί να βοηθήσει στην ορθή επιλογή ενός υποδείγματος καθώς και στην ανίχνευση τάσης και εποχικότητας στα δεδομένα.

### 1.3 ΧΑΡΑΚΤΗΡΙΣΤΙΚΑ ΧΡΗΜΑΤΟΟΙΚΟΝΟΜΙΚΩΝ ΧΡΟΝΟΣΕΙΡΩΝ

Μια σημαντική εφαρμογή των χρονοσειρών συναντάται στην ανάλυση χρηματοοικονομικών στοιχείων, όπως οι τιμές και οι αποδόσεις μετοχών και χρηματιστηριακών δεικτών.

Στην παρούσα διπλωματική εργασία γίνεται εφαρμογή των μοντέλων *ARCH* και *GARCH* σε χρηματοοικονομικά στοιχεία (μετοχές, χρηματιστηριακούς δείκτες και κρυπτονομίσματα), με σκοπό τη μελέτη των αποδόσεων τους. Προτού όμως προχωρήσουμε στην περιγραφή των μοντέλων αυτών, κρίνεται σκόπιμη η παρουσίαση κάποιων βασικών χαρακτηριστικών που παρατηρούνται σε τέτοιου είδους χρηματοοικονομικές σειρές.

**Συσταδοποίηση μεταβλητότητας (*volatility clustering*):** Ιστορικά δεδομένα έχουν δείξει ότι η διακύμανση (*volatility*) της απόδοσης ενός χρηματοοικονομικού στοιχείου, φαίνεται να αυξομειώνεται με την πάροδο του χρόνου. Συνεπώς, μια μεγάλη μεταβολή της τιμής του τείνει να συνεχιστεί και τις επόμενες χρονικές περιόδους, ενώ μια μικρή μεταβολή (συνήθως) συμπληρώνεται από την μετέπειτα εμφάνιση μικρών τιμών.

**Βαριά ουρά (*fat tails*):** Η κατανομή των λογαριθμικών αποδόσεων των υποκείμενων μέσων (π.χ. μετοχές), συνήθως φαίνεται να έχει βαριά ουρά, αντιβαίνοντας της υπόθεσης της κανονικότητας που ενσωματώνουν αρκετά μοντέλα (π.χ. μοντέλο Black – Scholes).

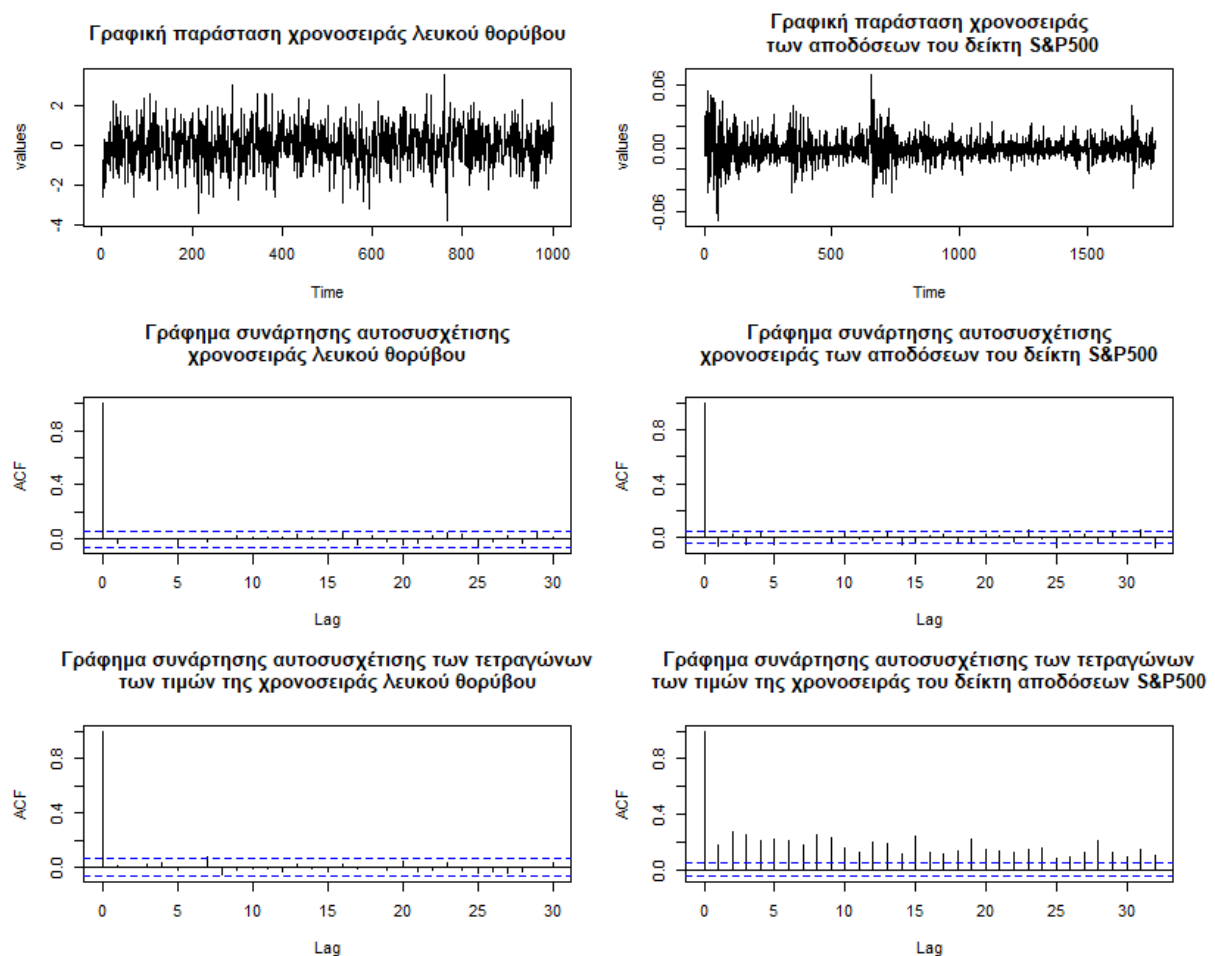
**Επενέργεια μόχλευσης (*leverage effect*):** Η μεταβολή της διακύμανσης των τιμών μιας μετοχής έχει παρατηρηθεί ότι συνδέεται με τη μεταβολή των τιμών που λαμβάνει αυτή τις προηγούμενες χρονικές περιόδους. Μάλιστα παρατηρείται μια αρνητική γραμμική συσχέτιση. Συγκεκριμένα, όταν μειώνεται η τιμή μιας μετοχής η διακύμανση της έχει την τάση να αυξάνεται και το αντίθετο.

Για περαιτέρω πληροφορίες ο ενδιαφερόμενος αναγνώστης μπορεί π.χ. να ανατρέξει στα συγγράμματα των Βρόντος, Ι. (2006), Tsay, R.S. (2002) (2006) και Xekalaki, E. and Degiannakis, S. (2010).

## 1.4 ΑΥΤΟΠΑΛΙΔΡΟΜΑ ΜΟΝΤΕΛΑ ΔΕΣΜΕΥΜΕΝΗΣ ΕΤΕΡΟΣΚΕΔΑΣΤΙΚΟΤΗΤΑΣ - ARCH MODELS

Στο παρόν εδάφιο, εισάγονται τα μοντέλα δεσμευμένης ετεροσκεδαστικότητας όπου αποτελούν μια πρώτη προσέγγιση μοντελοποίησης της χρονοσειράς λευκού θορύβου.

Για την καλύτερη κατανόηση της κατασκευής του μοντέλου, παρακάτω εικονίζονται οι γραφικές παραστάσεις και τα αντίστοιχα γραφήματα αυτοσυσχέτισης των χρονοσειρών  $X = (x_t, t \in T)$  και  $r = (r_t, t \in T)$ , με την πρώτη να εκφράζει ένα Γκαουσιανό λευκό θόρυβο (*Gaussian White Noise*), ενώ η δεύτερη τις αποδόσεις του δείκτη S&P500.



**Σχήμα 1.7:** Σύγκριση γραφικών παραστάσεων μιας χρονοσειράς λευκού θορύβου σταθερής διασποράς και της χρονοσειράς των αποδόσεων του δείκτη S&P500.

Παρατηρώντας το γράφημα της συνάρτησης αυτοσυσχέτισης της κάθε χρονοσειράς, γίνεται σαφές ότι και στις δύο περιπτώσεις οι όροι είναι (γραμμικά) ασυσχέτιστοι μεταξύ τους. Επομένως, η εφαρμογή ενός γραμμικού μοντέλου (π.χ.  $ARMA(p, q)$ ), τόσο στη χρονοσειρά  $X$  όσο και στην  $r$ , θα ήταν ακατάλληλη.

Εξετάζοντας προσεκτικότερα τις ανωτέρω χρονοσειρές, παρατηρείται μια εγγενής διαφορά μεταξύ των, ανεξαρτήτως του γεγονότος ότι το γράφημα της συνάρτησης αυτοσυσχέτισης τους συμπεριφέρεται με παρόμοιο τρόπο.

Ενώ στη χρονοσειρά  $X$  παρατηρείται μια «συνέπεια» στη μεταβολή των τιμών της, στην περίπτωση της  $r$  παρατηρούνται περιοχές όπου οι διασπορά των αποδόσεων εμφανίζεται αυξημένη (π.χ. γύρω από το  $t = 700$ ) και αντίστοιχα κάποιες φορές μειωμένη (π.χ. γύρω από το  $t = 1000$ )<sup>1</sup>. Μάλιστα, όταν η διασπορά (κάποια τυχαία χρονική στιγμή) λάβει μεγάλη τιμή, αυτό τείνει να συνεχιστεί τις επόμενες χρονικές περιόδους (και το αντίθετο).

Έτσι, οδηγούμαστε στο συμπέρασμα ότι η δεσμευμένη διασπορά των αποδόσεων  $Var[r_t|r_0, \dots, r_{t-1}]$  είναι αυτή που μεταβάλλεται στο χρόνο, με αποτέλεσμα τη δημιουργία συστάδων μεγάλης και μικρής μεταβλητότητας.

Συνεχίζοντας την περαιτέρω ανάλυση για την εύρεση βαθύτερων αιτιών που διαφοροποιούν τις δυο αυτές χρονοσειρές, υποθέτοντας (για λόγους απλούστευσης) ότι  $E[r_t] = 0, \forall t \in T$ , παρατηρούμε ότι το γράφημα αυτοσυσχέτισης του τετραγώνου των τιμών  $r_t$ , δηλαδή της χρονοσειράς  $r^2 = (r_t^2, t \in T)$ , παρουσιάζει μια γραμμική συσχέτιση των όρων μεταξύ τους (κάτι το οποίο δεν φαίνεται να ισχύει για τους όρους της  $X$ ). Το τελευταίο μας οδηγεί στα εξής δύο συμπεράσματα:

1. οι όροι της  $r$  είναι εξαρτημένες τ.μ.<sup>2</sup>,
2. εφόσον ισχύει ότι  $Var[r_t|r_0, \dots, r_{t-1}] = E[r_t^2|r_0, \dots, r_{t-1}]$  (όπου  $E[r_t] = 0$ ), σε συνδυασμό με το γεγονός ότι οι όροι  $r_t^2$  είναι (γραμμικά) συσχετισμένοι, τότε υπάρχει ένα γραμμικό μοντέλο που συνδέει τα τετράγωνα των τιμών της  $r$ .

Συνοψίζοντας τα παραπάνω, ένα από τα πρώτα υποδείγματα που αναπτύχθηκαν για να ερμηνεύσουν την συμπεριφορά της χρονοσειράς  $r$ , δηλαδή μιας χρονοσειράς όπου:

- ❖ είναι στάσιμη,
- ❖ οι όροι πρώτης τάξης είναι ασυσχέτιστοι,
- ❖ οι όροι δεύτερης τάξης συσχετισμένοι,

---

<sup>1</sup> Όταν η διασπορά των δεδομένων παραμένει σταθερή στο χρόνο, λέγεται ότι υπάρχει ομοσκεδαστικότητα (**homoscedasticity**) στα δεδομένα, ενώ στην αντίθετη περίπτωση ετεροσκεδαστικότητα (**heteroscedasticity**).

<sup>2</sup> Όπως είναι γνωστό από τη θεωρία πιθανοτήτων, αν οι τ.μ.  $r_t, t \in T$  είναι ανεξάρτητες μεταξύ τους, τότε και οι όροι υψηλότερης τάξης (π.χ.  $r_t^2$ ) θα πρέπει να είναι ανεξάρτητες τ.μ.

❖ η (δεσμευμένη) διασπορά μεταβάλλεται στο χρόνο δημιουργώντας ετεροσκεδαστικότητα στα δεδομένα,

δόθηκε από τον Engle (1982) και είναι γνωστό ως υπόδειγμα *ARCH*.

**Ορισμός:** Μια χρονοσειρά  $r = (r_t, t \in T)$  καλείται *ARCH*( $p$ ) αν είναι αυστηρά στάσιμη και ικανοποιεί τις εξισώσεις:

$$\begin{cases} r_t = \sigma_t e_t & , e_t \sim IID(0,1) \\ \sigma_t^2 = a_0 + \sum_{i=1}^p a_i r_{t-i}^2 & , t \in T \end{cases} \quad (1.6)$$

όπου  $a_0 > 0, a_i \geq 0$  για  $i = 1, \dots, p$  και κάθε  $e_t$  θεωρείται ανεξάρτητο των  $r_s, s < t$  και  $\sigma_s, s \leq t$ .

Σημειώνεται πως στις εφαρμογές η κατανομή των  $e_t$  επιλέγεται συνήθως να ακολουθεί είτε την κανονική κατανομή  $N(0,1)$  είτε την τυποποιημένη *Student*( $v$ ) στην περίπτωση που επιβάλλεται μια κατανομή με βαρύτερη ουρά από την κανονική.

## 1.5 ΙΔΙΟΤΗΤΕΣ ΜΟΝΤΕΛΟΥ ARCH

Μια *ARCH* χρονοσειρά είναι ασθενώς και ισχυρά στάσιμη όταν ικανοποιείτε η συνθήκη:

$$\sum_{i=1}^p a_i < 1. \quad (1.7)$$

Αναφέρεται ότι η παραπάνω ανισότητα αποτελεί ικανή και αναγκαία συνθήκη ώστε μια *ARCH* χρονοσειρά να είναι ασθενώς και ισχυρά στάσιμη (π.χ. βλ. Francq, C. and Zakoian, J. M. (2019)).

Έστω  $\mathcal{F}_{t-1}$  η πληροφορία που είναι γνωστή τη χρονική στιγμή  $t - 1$ , δηλαδή το σύνολο όλων των συναρτήσεων των  $r_0, \dots, r_{t-1}$ .

Τόσο η δεσμευμένη μέση τιμή όσο και η μέση τιμή της χρονοσειράς  $r$  δίνονται από τις σχέσεις:

$$E[r_t | \mathcal{F}_{t-1}] = E[\sigma_t e_t | \mathcal{F}_{t-1}] = \sigma_t E[e_t | \mathcal{F}_{t-1}] = \sigma_t E[e_t] = 0, t \in T \quad (1.8)$$

$$E[r_t] = E[E[r_t|\mathcal{F}_{t-1}]] = 0. \quad (1.9)$$

Το γεγονός ότι η ποσότητα  $\sigma_t^2$ , στον ορισμό (βλ. (1.6)), εκφράζει τη δεσμευμένη διακύμανση της  $r$ , δεδομένου της πληροφορίας  $\mathcal{F}_{t-1}$ , αποδεικνύεται εύκολα ως εξής:

$$Var[r_t|\mathcal{F}_{t-1}] = E[(\sigma_t e_t)^2|\mathcal{F}_{t-1}] = \sigma_t^2 E[e_t^2|\mathcal{F}_{t-1}] = \sigma_t^2, t \in T. \quad (1.10)$$

Όσον αφορά την διασπορά  $\sigma^2$  της χρονοσειράς  $r$ , λαμβάνοντας υπόψη τη στασιμότητα της (βλ. σχέση (1.7)), προκύπτει ότι:

$$\begin{aligned} \sigma^2 &= Var[r_t] = E[\sigma_t^2 e_t^2] \\ &\stackrel{\sigma_i, \varepsilon_i}{=} E[\sigma_t^2] E[e_t^2] = a_0 + \sum_{i=1}^q a_i Var[r_{t-1}] \\ &\stackrel{\text{ανεξ.}}{=} a_0 + \sum_{i=1}^p a_i \sigma^2 \\ &\stackrel{r_i}{=} \text{στάσιμη} \end{aligned} \quad (1.11)$$

από την οποία καταλήγουμε ότι θα πρέπει να ισχύει ότι:

$$\sigma^2 = \frac{a_0}{1 - \sum_{i=1}^p a_i}.$$

Τέλος, η συνάρτηση αυτοσυνδιακύμανσης υπολογίζεται ως εξής:

$$\begin{aligned} \gamma_r(h) &= Cov[r_{t+h}, r_t] = E[r_{t+h} r_t] \\ &= E[E[r_{t+h} r_t | \mathcal{F}_{t+h-1}]] = E[r_t E[r_{t+h} | \mathcal{F}_{t+h-1}]] \\ &= 0. \end{aligned} \quad (1.12)$$

Τα παραπάνω οδηγούν στο συμπέρασμα ότι μια χρονοσειρά  $ARCH(p)$  είναι στην πραγματικότητα μια χρονοσειρά λευκού θορύβου:

$$r = (r_t, t \in T) \sim WN \left( E[r_t] = 0, Var[r_t] = \frac{a_0}{1 - \sum_{i=1}^p a_i} \right). \quad (1.13)$$

Ο υπολογισμός της κατανομής της  $r$  μπορεί να αποτελέσει μια αρκετά πολύπλοκη διαδικασία. Έτσι, μέσω της χρήσης στατιστικών μέτρων, είναι εφικτό να εξαχθούν χρήσιμα συμπεράσματα για αυτή, ακόμα και αν δεν είναι γνωστός ο ακριβής της τύπος.

Εστιάζοντας κυρίως στη συμπεριφορά της ουράς της, ένα χρήσιμο μέτρο που δίνει μια αίσθηση του κατά πόσο η ουρά μιας συμμετρικής κατανομής είναι βαριά, είναι ο συντελεστής κύρτωσης  $\kappa$ , υπό την προϋπόθεση βέβαια ότι η τέταρτη ροπή της  $r$  είναι πεπερασμένη.

Έτσι, υποθέτοντας (για απλούστευση και ολοκλήρωση των υπολογισμών) ότι  $e_t \sim N(0,1)$ , προκύπτει ότι:

$$\begin{aligned} \kappa &= \frac{E[r_t^4]}{(E[r_t^2])^2} = \frac{E[\sigma^4 e_t^4]}{(E[\sigma_t^2 e_t^2])^2} \\ E[e_t^4] &= 3 \quad \text{Anισ.} \\ &= 3 \frac{E[\sigma_t^4]}{(E[\sigma_t^2])^2} \geq 3. \quad \text{Jensen} \end{aligned} \quad (1.14)$$

Το αποτέλεσμα της σχέσης (1.14) οδηγεί στο συμπέρασμα ότι η κατανομή μιας *ARCH* χρονοσειράς έχει βαρύτερη ουρά από την κανονική κατανομή, κάτι το οποίο δικαιολογεί την εμφάνιση περισσότερων ακραίων τιμών όπως φαίνεται και στο σχήμα 1.7.

Στη συνέχεια, για την καλύτερη κατανόηση της δομής του υποδείγματος (βλ. (1.6)) θεωρούμε την τ.μ.  $\eta_t = r_t^2 - \sigma_t^2$ . Με αντικατάσταση του  $\sigma_t^2$  από στη σχέση (1.6) λαμβάνουμε ότι:

$$\eta_t = r_t^2 - a_0 - \sum_{i=1}^p a_i r_{t-i}^2 \quad \text{συνεπάγεται ότι} \quad r_t^2 = a_0 + \sum_{i=1}^p a_i r_{t-i}^2 + \eta_t. \quad (1.15)$$

Η σχέση (1.15) αποδεικνύει ότι σε ένα υπόδειγμα *ARCH*( $p$ ), η μοντελοποίηση της δεσμευμένης ετεροσκεδαστικότητας έχει τη μορφή ενός *AR*( $p$ ) υποδείγματος. Για το λόγο αυτό τα *ARCH* μοντέλα ονομάστηκαν και Αυτοπαλίδρομα Μοντέλα Δεσμευμένης Ετεροσκεδαστικότητας.

## 1.6 ΕΛΕΓΧΟΣ ΥΠΑΡΞΗΣ ΕΤΕΡΟΣΚΕΔΑΣΤΙΚΟΤΗΤΑΣ

Για την εφαρμογή ενός υποδείγματος *ARCH*, είναι απαραίτητο να γίνει έλεγχος στα διαθέσιμα δεδομένα για την ύπαρξη φαινομένων ετεροσκεδαστικότητας. Επομένως, αυτό που χρήζει εξέτασης είναι το κατά πόσο οι όροι  $r_t^2$ , σχετίζονται γραμμικά (υποθέτοντας ότι  $E[r_t] = 0$ ).



Αν και υπάρχει μια ευρεία γκάμα τέτοιων ελέγχων όπως τα Ljung-Box statistics, Engle's Lagrange test, Tse's residual based diagnostic test, στην παρούσα εργασία επιλέχθηκαν να παρουσιαστούν τα δύο πρώτα, καθώς είναι και τα συχνότερα στις εφαρμογές.

**Ljung-Box statistics:** Για τον ταυτόχρονο έλεγχο των  $T$  πρώτων αυτοσυσχετίσεων (όπου  $T \approx \ln(n)$ ), δηλαδή για τον έλεγχο της υπόθεσης  $H_0 : \rho(i) = 0, i = 1, \dots, T$  έναντι της  $H_1 : \rho(i) \neq 0$  για κάποιο  $i \in \{1, \dots, T\}$  το στατιστικό που χρησιμοποιείται είναι το εξής:

$$Q(T) = n(n+2) \sum_{i=1}^T \frac{\hat{\rho}(i)^2}{n-i} \sim_{H_0} \chi_T^2$$

$$\text{όπου: } \hat{\rho}(i) = \frac{n}{n-i} \frac{\sum_{j=i+1}^n (\hat{r}_j^2 - \bar{r}^2)(\hat{r}_{j-i}^2 - \bar{r}^2)}{\sum_{j=1}^n (\hat{r}_j^2 - \bar{r}^2)^2}$$

$$\bar{r}^2 = \frac{\sum_{j=1}^n r_j^2}{n}$$

$$n : \text{ μέγεθος δείγματος.}$$
(1.16)

**Engle's Lagrange test:** Πρόκειται για το ανάλογο  $F$  – test όπως είναι γνωστό από τα μοντέλα γραμμικής παλινδρόμησης. Συγκεκριμένα ο έλεγχος μπορεί να πραγματοποιηθεί ακολουθώντας τα εξής βήματα:

1. Εφαρμόζεται ένα υπόδειγμα της μορφής:

$$r_t^2 = a_0 + a_1 r_{t-1}^2 + \dots + a_T r_{t-T}^2 + \varepsilon_t, \quad t = T, \dots, n \text{ και } \varepsilon_t \sim N(0, \sigma_\varepsilon^2) \quad (1.17)$$

όπου γίνεται a priori επιλογή του πλήθους των όρων που εισάγονται στο μοντέλο.

2. Ελέγχεται η στατιστική υπόθεση  $H_0 : a_i = 0, i = 1, \dots, T$  έναντι της  $H_1 : a_i \neq 0$  για κάποιο  $i = 1, \dots, T$  χρησιμοποιώντας το στατιστικό:

$$F = \frac{\sum_{j=T+1}^n ((r_j^2 - \bar{r}^2)^2 - \hat{\varepsilon}_j^2) / T}{\sum_{j=T+1}^n \hat{\varepsilon}_j^2 / (n - 2T - 1)} \sim \chi_T^2 \quad (1.18)$$

όπου  $n$  το μέγεθος του δείγματος,  $\bar{r}^2$  ο δειγματικός μέσος των τετραγώνων των  $r_t$  και  $\hat{\varepsilon}_t$  η εκτίμηση του σφάλματος  $\varepsilon_t$ .

Για επιπλέον πληροφορίες και ελέγχους που μπορούν να πραγματοποιηθούν, ο ενδιαφερόμενος αναγνώστης μπορεί να ανατρέξει στα συγγράμματα π.χ. των Tsay, R.S. (2002) και Xekalaki, E. and Degiannakis, S. (2010).

## 1.7 ΕΠΙΛΟΓΗ ΤΑΞΗΣ ΥΠΟΔΕΙΓΜΑΤΟΣ

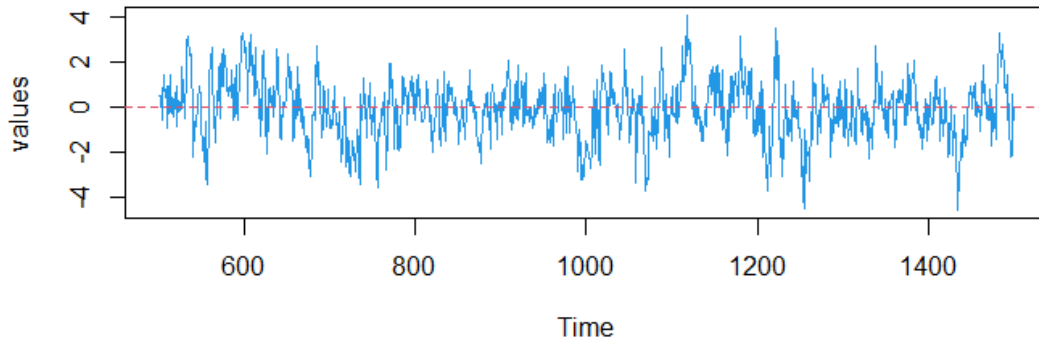
Στη συνέχεια, εφόσον τα δεδομένα επιβάλουν τη χρήση ενός *ARCH* υποδείγματος, πρέπει να προσδιοριστεί η τάξη του  $p$ , δηλαδή το πλήθος των όρων που πρέπει να συμπεριληφθούν στο μοντέλο.

Μια τεχνική που χρησιμοποιείται για το σκοπό αυτό, εφόσον το δείγμα είναι επαρκώς μεγάλο, είναι η θεώρηση του υποδείγματος (1.6) σαν ένα  $AR(p)$  υπόδειγμα (βλ. (1.15)), για το οποίο πρέπει να υπολογιστεί η συνάρτηση μερικής αυτοσυσχέτισης (***partial autocorrelation function - PACF***).

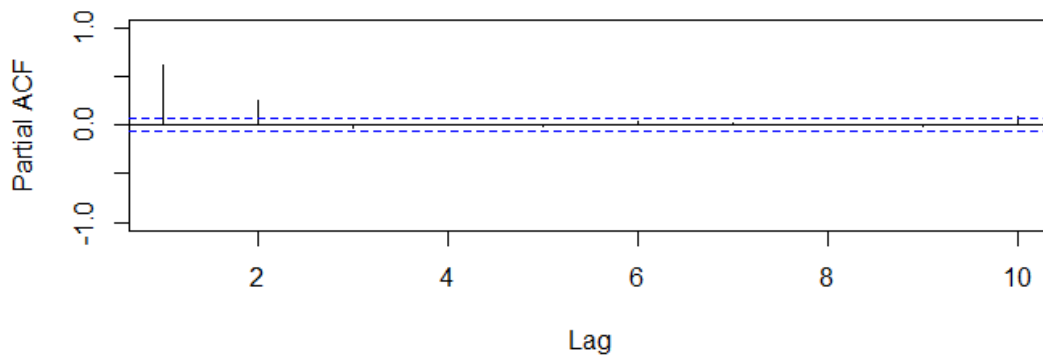
Η κατασκευή της συνάρτησης μερικής αυτοσυσχέτισης στηρίζεται στην εξής παρατήρηση:

Αν υποθέσουμε ότι μια χρονοσειρά  $X = (x_t: t \in T)$  ακολουθεί πράγματι ένα υπόδειγμα  $AR(P)$ , τότε οι συντελεστές κάθε υποδείγματος μικρότερη τάξης ( $p < P$ ), θα διαφέρουν σημαντικά από το μηδέν. Αντίθετα, αν εφαρμοστεί ένα μοντέλο μεγαλύτερης τάξης σε σχέση με το πραγματικό, για παράδειγμα το  $AR(P + 1)$ , ο επιπλέον όρος που εισάγεται στο μοντέλο αναμένεται να λάβει μια τιμή κοντά στο μηδέν. Για περαιτέρω εμβάθυνση ο ενδιαφερόμενος αναγνώστης μπορεί να ανατρέξει π.χ. στο σύγγραμμα του Μπούτσικα, Μ. (2019-2020).

### Γραφική παράσταση χρονοσειράς AR(2)



### Γραφική παράσταση της συνάρτησης μερικής αυτοσυσχέτισης του υποδείγματος AR(2)



**Σχήμα 1.8:** Γραφική παράσταση της χρονοσειράς  $x_t = 0.5x_{t-1} + 0.2x_{t-2} + e_t$ , με  $e_t \sim N(0,1)$ . Στο δεύτερο γράφημα απεικονίζεται η γραφική παράσταση της συνάρτησης μερικής αυτοσυσχέτισης της παραπάνω χρονοσειράς. Σύμφωνα με τη θεωρία, η γραφική παράσταση της συνάρτησης μερικής αυτοσυσχέτισης ενός AR(2) υποδείγματος, αναμένεται να λάβει σχεδόν μηδενικές τιμές μετά την εκτίμηση της δεύτερης παραμέτρου του μοντέλου.

## 1.8 QUASI MAXIMUM LIKELIHOOD

Οι πλέον συνήθεις μέθοδοι εκτίμησης των παραμέτρων του μοντέλου (1.6) είναι αυτή της μέγιστης πιθανοφάνειας (Maximum Likelihood Estimation – MLE) και η Quasi Maximum Likelihood - QMLE ή όπως κάποιες φορές αναφέρεται στη βιβλιογραφία Pseudo Maximum Likelihood.

Η διαφορά των δύο παραπάνω μεθόδων έγκειται στο γεγονός ότι για την εφαρμογή της MLE είναι απαραίτητη η γνώση της (ακριβούς) συνάρτησης κατανομής των σφαλμάτων  $e_t$ , ενώ στην QMLE είναι αρκετή η υπόθεση της κανονικότητας αυτών ( $e_t \sim N(0,1)$ ), ακόμα και αν κάτι τέτοιο μπορεί να μην ισχύει στην πραγματικότητα.

Τονίζεται πως η βασική αιτία που «νομιμοποιούμαστε» να χρησιμοποιήσουμε τη μέθοδο QMLE, είναι ότι μέσω αυτής παράγονται συνεπείς<sup>3</sup> αλλά μη αποδοτικές εκτιμήσεις<sup>4</sup> των παραμέτρων του μοντέλου.

Η από κοινού συνάρτηση πυκνότητας πιθανότητας, υπό την υπόθεση της κανονικότητάς των  $e_t$  και για δεδομένο δείγμα  $n$ , δίνεται από σχέση:

$$\begin{aligned} f(r_n, \dots, r_p, \dots, r_1) &= f(r_n | \mathcal{F}_{n-1}) f(r_{n-1} | \mathcal{F}_{n-2}) \cdots f(r_{p+1} | \mathcal{F}_p) f(r_p, \dots, r_1) \\ &= \prod_{t=p+1}^n \frac{1}{\sqrt{2\pi\sigma_t^2}} \exp\left(-\frac{r_t^2}{2\sigma_t^2}\right) \times f(r_p, \dots, r_1) \end{aligned} \quad (1.19)$$

όπου χρησιμοποιείται η ιδιότητα ότι  $E[r_t | \mathcal{F}_{t-1}] = 0$  και  $Var[r_t | \mathcal{F}_{t-1}] = \sigma_t^2$ .

Καθώς ο υπολογισμός της συνάρτησης  $f(r_p, \dots, r_1)$  είναι αρκετά πολύπλοκος, ένα τέχνασμα που χρησιμοποιείται για την απλοποίηση της διαδικασίας, είναι ο υπολογισμός της δεσμευμένης ως προς τα  $(r_p, \dots, r_1)$  συνάρτησης πιθανοφάνειας, υπό την προϋπόθεση ενός επαρκούς μεγάλου δείγματος. Επομένως, καταλήγουμε ότι η συνάρτηση πιθανοφάνειας παίρνει τη μορφή:

$$L_0(a_0, a_1, \dots, a_p) = f(r_n, \dots, r_{p+1} | r_p, \dots, r_1) = \prod_{t=p+1}^n \frac{1}{\sqrt{2\pi\sigma_t^2}} \exp\left(-\frac{r_t^2}{2\sigma_t^2}\right) \quad (1.20)$$

όπου το  $\sigma_t^2 = a_0 + a_1 r_{t-1}^2 + \cdots + a_p r_{t-p}^2$  υπολογίζεται αναδρομικά.

Η βελτιστοποίηση της σχέσης (1.20) ισοδυναμεί με τη βελτιστοποίηση του λογαρίθμου αυτής, δηλαδή της λογαριθμικής συνάρτησης πιθανοφάνειας,

$$l_0(a_0, a_1, \dots, a_p) = \sum_{t=p+1}^n \left( -\frac{1}{2} \ln(2\pi) - \frac{1}{2} \ln(\sigma_t^2) - \frac{1}{2} \frac{r_t^2}{\sigma_t^2} \right). \quad (1.21)$$

<sup>3</sup> Ένας αμερόληπτος εκτιμητής  $T_n$  μιας παραμέτρου  $\theta$  ονομάζεται συνεπής εάν συγκλίνει κατά πιθανότητα στην πραγματική τιμή της παραμέτρου. Δηλαδή, για κάθε  $\varepsilon > 0$  ισχύει ότι  $\lim_{n \rightarrow \infty} \Pr(|T_n - \theta| > \varepsilon) = 0$ .

<sup>4</sup> Ένας αμερόληπτος εκτιμητής ονομάζεται αποδοτικός αν είναι ελάχιστης διασποράς.

Σημειώνεται ότι η μεγιστοποίηση της παραπάνω συνάρτησης πιθανοφάνειας ως προς  $a_0, a_1, \dots, a_p$  γίνεται με την χρήση αριθμητικών μεθόδων, καθώς δεν υπάρχουν κλειστοί τύποι.

Αποδεικνύεται ότι οι εκτιμητές των παραμέτρων  $a_0, a_1, \dots, a_p$ , σύμφωνα με τη μέθοδο QMLE, είναι συνεπείς αλλά μη αποδοτικοί. Έτσι, μια αναπροσαρμογή της εκτίμησης της διασποράς αυτών είναι απαραίτητη.

Τέλος, αναφέρεται ότι στην περίπτωση που η κατανομή των σφαλμάτων  $e_t$  είναι συμμετρική, οι προσεγγίσεις που παράγονται μέσω της μεθόδου QMLE είναι αρκετά κοντά με αυτές που δίνονται μέσω της μεθόδου MLE. Αντίθετα, στην περίπτωση μη συμμετρικών κατανομών, άλλες μέθοδοι εκτίμησης των παραμέτρων του μοντέλου θα πρέπει να χρησιμοποιηθούν, καθώς παρατηρείται μεγάλη αύξηση της μη αποδοτικότητας των εκτιμητών π.χ. βλ. Xekalaki, E. and Degiannakis, S. (2010).

## 1.9 ΜΕΙΟΝΕΚΤΗΜΑΤΑ ΥΠΟΔΕΙΓΜΑΤΟΣ ARCH

Αν και τα ARCH υποδείγματα αποτελούν μια φαινομενικά καλή προσέγγιση μοντελοποίησης της ετεροσκεδαστικότητας, εντούτοις εμφανίζουν τα εξής μειονεκτήματα:

- Ανατρέχοντας στο μοντέλο (1.6) παρατηρείται ότι η (δεσμευμένη) διασπορά εξαρτάται από το τετράγωνο των τιμών των  $r_t$ . Αυτό έχει σαν συνέπεια ότι τόσο οι θετικές όσο και οι αρνητικές τιμές που λαμβάνει η χρονοσειρά  $r_t$  επιδρούν ακριβώς με τον ίδιο τρόπο στον υπολογισμό της διασποράς  $\sigma_t$ , κάτι το οποίο δε συνάδει με τα πραγματικά δεδομένα που παρατηρούνται στις χρηματοοικονομικές χρονοσειρές (leverage effect).
- Επίσης, η εφαρμογή του μοντέλου μπορεί να οδηγήσει, κάποιες φορές, στην υπερεκτίμηση της (δεσμευμένης) διασποράς, κάτι το οποίο οφείλεται στην εσωτερική του δομή. Συγκεκριμένα, ο τρόπος με τον οποίο ανταποκρίνεται το μοντέλο σε μεμονωμένες ακραίες τιμές  $r_t$  είναι σχετικά αργός, καθώς αυτές συνεχίζουν να εμφανίζονται στις επόμενες εκτιμήσεις της  $\sigma_t$  λόγω της ύπαρξης του αθροίσματος  $\sum_{i=1}^p a_i r_{t-i}^2$ . Μάλιστα το φαινόμενο αναμένεται εντονότερο όσο μεγαλύτερη είναι η τάξη  $p$  του υποδείγματος.

- Τέλος, τα μοντέλα *ARCH* δεν δίνουν κάποια περαιτέρω πληροφορία σχετικά με την πηγή της μεταβλητότητας της (δεσμευμένης) διασποράς παρά μόνο ένα μαθηματικό υπόδειγμα που την περιγράφει.

## 1.10 ΓΕΝΙΚΕΥΜΕΝΑ ΑΥΤΟΠΑΛΙΔΡΟΜΑ ΜΟΝΤΕΛΑ ΔΕΣΜΕΥΜΕΝΗΣ ΕΤΕΡΟΣΚΕΔΑΣΤΙΚΟΤΗΤΑΣ – GARCH MODELS

Το μοντέλο *ARCH* (βλ. 1.6), αν και έχει απλή δομή, μπορεί να καταστεί δύσχρηστο στην πράξη, καθώς απαιτεί συνήθως την εισαγωγή αρκετών όρων για την περιγραφή της μεταβλητότητας μιας χρηματοοικονομικής χρονοσειράς.

Το πρόβλημα αυτό αίρεται με την γενίκευση που προτάθηκε από τον Bollerslev (1986), επεκτείνοντας τα *ARCH* μοντέλα στα επονομαζόμενα Γενικευμένα Αυτοπαλίδρομα Μοντέλα Δεσμευμένης Ετεροσκεδαστικότητας (**Generalized ARCH Models**). Η βασική πλέον διαφορά είναι ότι η δεσμευμένη διασπορά  $\sigma_t^2$  δεν εξαρτάται μόνο από τις  $q$  προηγούμενες τιμές  $r_t^2$ , αλλά και από τις  $p$  προηγούμενες τιμές που έχει αυτή λάβει.

**Ορισμός:** Μια χρονοσειρά  $r = (r_t, t \in T)$  καλείται *GARCH*( $p, q$ ) αν υπάρχουν οι δύο πρώτες δεσμευμένες ροπές (δηλαδή είναι πεπερασμένες) και σταθερές  $w, a_i, i = 1, \dots, q$  και  $b_j, j = 1, \dots, p$  έτσι ώστε να ικανοποιούνται οι σχέσεις:

$$1. \quad E[r_t | r_s, s < t] = 0, t \in \mathbb{Z} \quad (1.22)$$

$$2. \quad \sigma_t^2 = \text{Var}[r_t | r_s, s < t] = w + \sum_{i=1}^q a_i r_{t-i}^2 + \sum_{j=1}^p b_j \sigma_{t-j}^2, t \in \mathbb{Z}. \quad (1.23)$$

### Παρατηρήσεις

Αν  $b_j = 0, j = 1, \dots, p$  τότε καταλήγουμε σε ένα *ARCH*( $q$ ) μοντέλο, καθώς όπως έχει αναφερθεί τα *GARCH* μοντέλα αποτελούν γενίκευση των *ARCH*.

Θεωρώντας την τ.μ.  $\eta_t = r_t^2 - \sigma_t^2$  η οποία εκφράζει την ανέλιξη των σφαλμάτων, τότε το μοντέλο (βλ. 1.23) μπορεί να γραφτεί στη μορφή:

$$r_t^2 = w + \sum_{i=1}^q a_i r_{t-i}^2 + \eta_t + \sum_{j=1}^p b_j (r_{t-j}^2 - \eta_{t-j}), t \in \mathbb{Z} \quad (1.24)$$

επίσης, θέτοντας όπου  $\delta = \max\{p, q\}$ , προκύπτει ότι:

$$r_t^2 = w + \sum_{i=1}^{\delta} (a_i + b_i)r_{t-i}^2 - \sum_{j=1}^p b_j \eta_{t-j} + \eta_t, t \in \mathbb{Z}. \quad (1.25)$$

Η παράσταση (1.25) φανερώνει ότι σε ένα υπόδειγμα *GARCH* η σχέση που συνδέει τα  $r_t^2$  έχει τη δομή ενός *ARMA*( $\delta, p$ ) υποδείγματος.

Σχετικά με τον ορισμό του υποδείγματος *GARCH* που δόθηκε παραπάνω (βλ. 1.22 και 1.23), παρατηρείται ότι δεν γίνεται καμία αναφορά σε κάποια επιπρόσθετη σχέση που να εκφράζει τον τρόπο εξέλιξης της  $r$  χρονοσειράς. Αυτό θα μπορούσε να ξεπεραστεί με την εισαγωγή αυστηρότερων συνθηκών, όπως περιγράφεται στον παρακάτω ορισμό.

**Ορισμός:** Μια χρονοσειρά  $r = (r_t, t \in T)$  καλείται **αυστηρώς *GARCH*( $p, q$ ) χρονοσειρά (*strong GARCH*( $p, q$ ) process)** αν ικανοποιούνται οι εξής συνθήκες:

1. 
$$r_t = \sigma_t e_t \quad (1.26)$$

2. 
$$\sigma_t^2 = w + \sum_{i=1}^q a_i r_{t-i}^2 + \sum_{j=1}^p b_j \sigma_{t-j}^2 \quad (1.27)$$

όπου  $(e_t, t \in T) \sim IID(0,1)$ ,  $e_t$  ανεξάρτητη των  $(r_s, s < t)$ ,  $w > 0$ ,  $a_i, b_j \geq 0$  και  $\sigma_t > 0$ .

Αναφέρεται πως μια αυστηρά *GARCH* χρονοσειρά ικανοποιεί τη σχέση (1.22), του προηγούμενου γενικότερου ορισμού.

Αντικαθιστώντας στη σχέση (1.27) το  $r_{t-i}$  με  $\sigma_{t-i} e_{t-i}$ , προκύπτει ότι σε ένα αυστηρά *GARCH* υπόδειγμα η δεσμευμένη διασπορά είναι ένα *AR*( $\delta$ ) μοντέλο όπου οι συντελεστές των  $\sigma_{t-i}^2$  είναι τυχαίες μεταβλητές:

$$\sigma_t^2 = w + \sum_{i=1}^{\delta} (a_i e_{t-i}^2 + b_i) \sigma_{t-i}^2, \quad \delta = \max\{p, q\}. \quad (1.28)$$

Τέλος, τονίζεται πως όταν αναφερόμαστε σε μια *GARCH* χρονοσειρά, θα εννοείται ότι είναι μια αυστηρά *GARCH* χρονοσειρά.

## 1.11 ΙΔΙΟΤΗΤΕΣ ΥΠΟΔΕΙΓΜΑΤΟΣ GARCH

Για να χαρακτηριστεί μια αυστηρά  $GARCH(p, q)$  χρονοσειρά ως ασθενώς στάσιμη, μια επαρκής και αναγκαία συνθήκη είναι:

$$\sum_{i=1}^q a_i + \sum_{j=1}^p b_j < 1. \quad (1.29)$$

Η μέση τιμή της  $r$  είναι μηδέν και υπολογίζεται με τη βοήθεια του θεωρήματος της διπλής μέσης τιμής:

$$E[r_t] = E[E[r_t | r_{t-1}, \dots, r_0]] = 0. \quad (1.30)$$

Η διασπορά της  $r$ , όταν ισχύει η (1.29), αποδεικνύεται ότι δίνεται από τη σχέση:

$$\sigma^2 = \frac{w}{1 - \sum_{i=1}^q a_i - \sum_{j=1}^p b_j}. \quad (1.31)$$

Επίσης, η συνάρτηση αυτοσυνδιακύμανσης  $Cov[r_{t+h}, r_t]$  και αυτοσυσχέτισης  $\gamma_r(h)$ ,  $h = 1, 2, \dots$  αποδεικνύεται ότι είναι μηδενικές.

Συνοψίζοντας τα παραπάνω χαρακτηριστικά μια  $GARCH$  χρονοσειρά  $r = (r_t, t \in T)$  είναι ουσιαστικά μια χρονοσειρά λευκού θορύβου με  $E[r_t] = 0$  και  $Var[r_t] = \sigma^2$ .

Στη συνέχεια, για τη μελέτη της ουράς της κατανομής και της δεσμευμένης κατανομής ενός  $GARCH$  υποδείγματος, αναλύεται ο συντελεστή κύρτωσης και για τις δυο περιπτώσεις. Τονίζεται ότι για να ισχύει η ανάλυση, απαραίτητη προϋπόθεση είναι η ύπαρξη των ροπών τέταρτης τάξης. Συγκεκριμένα:

- ο συντελεστής κύρτωσης  $\kappa_{r_t | \mathcal{F}_{t-1}}$  για την υπό συνθήκη κατανομή της  $r$  (βλ. σχ. 1.26, 1.27), δίνεται από τη σχέση:

$$\kappa_{r_t | \mathcal{F}_{t-1}} = \frac{E[r_t^4 | \mathcal{F}_{t-1}]}{(E[r_t^2 | \mathcal{F}_{t-1}])^2} = \frac{\sigma_t^4 E[e_t^4]}{(\sigma_t^2 E[e_t^2])^2} = E[e_t^4] \quad (1.32)$$



- ο συντελεστής κύρτωσης  $\kappa_{r_t}$  της κατανομής της  $r$  είναι:

$$\kappa_{r_t} = \frac{E[E[r_t^4|\mathcal{F}_{t-1}]]}{(E[E[r_t^2|\mathcal{F}_{t-1}]])^2} = \frac{E[\sigma_t^4]}{(E[\sigma_t^2])^2} E[e_t^4]. \quad (1.33)$$

Συμβολίζοντας με  $\kappa_{e_t}$  τον συντελεστή κύρτωσης της κατανομής της χρονοσειράς  $e_t$  (δηλαδή  $\kappa_{e_t} = E[e_t^4]$ ), αν η διασπορά της  $\sigma_t^2$  είναι μεγαλύτερη από το τετράγωνο της αναμενόμενης τιμής της, τότε από την (1.33) προκύπτει ότι:

$$\kappa_{r_t} \geq \kappa_{e_t} \quad (1.34)$$

όπου η ισότητα επιτυγχάνεται στην περίπτωση που έχουμε έλλειψη του φαινομένου της ετεροσκεδαστικότητας.

Η τελευταία σχέση αναφέρει ότι η κατανομή μιας αυστηρά *GARCH* χρονοσειράς (αν ισχύουν οι παραπάνω υποθέσεις) έχει βαρύτερη ουρά από την κατανομή των σφαλμάτων της  $e_t$ . Μάλιστα το πόσο μακριά βρισκόμαστε από την κανονικότητα, διατυπώνεται από το παρακάτω θεώρημα.

**Θεώρημα:** Έστω μια αυστηρά *GARCH*( $p, q$ ) χρονοσειρά  $r = (r_t, t \in T)$  με  $E[r_t^4] < \infty$ . Η διαφορά του συντελεστή κύρτωσης  $\kappa_{r_t}$  ως προς αυτόν της κανονικής κατανομής, συμβολίζεται με  $\kappa_{r_t}^*$  και δίνεται από τη σχέση:

$$\kappa_{r_t}^* = \kappa_{r_t} - 3 = \frac{6\lambda + (\kappa_{e_t} - 3)(1 + 3\lambda)}{1 - \lambda(\kappa_{e_t} - 1)} \quad (1.35)$$

όπου  $\lambda = \sum_{i=1}^{\infty} \psi_i^2$  και οι συντελεστές  $\psi_i$  προκύπτουν από την γεννήτρια συνάρτηση:

$$1 + \sum_{i=1}^{\infty} \psi_i x^i = \left(1 - \sum_{i=1}^p b_i x^i\right) / \left(1 - \sum_{i=1}^{\delta} (a_i + b_i) x^i\right)$$

και  $\delta = \max\{p, q\}$ .

Το επόμενο θεώρημα μας πληροφορεί για τις τιμές που λαμβάνει η συνάρτηση αυτοσυσχέτισης του τετραγώνου των τιμών μιας *GARCH*( $p, q$ ) χρονοσειράς.

**Θεώρημα:** Έστω  $r = (r_t, t \in T)$  μια  $GARCH(p, q)$  χρονοσειρά. Εάν  $E[r_t^4] < \infty$  τότε:

$$\gamma_{r^2}(h) = Cov[r_t^2, r_{t-h}^2] \geq 0, h \geq 0. \quad (1.36)$$

Επίσης, εάν επιπλέον ισχύει ότι  $a_1 > 0$  τότε:

$$\gamma_{r^2}(h) > 0, h \geq 0. \quad (1.37)$$

Όπως έχει αποδειχθεί παραπάνω, μια  $GARCH$  χρονοσειρά είναι μια χρονοσειρά λευκού θορύβου. Επομένως, για να γίνει δυνατή η εφαρμογή ενός υποδείγματος  $GARCH(p, q)$  σε μια τυχαία χρονοσειρά, θα πρέπει αρχικά να ελέγξουμε αν αυτή είναι πράγματι λευκός θόρυβος.

Το επόμενο θεώρημα, το οποίο αποτελεί μια διόρθωση του ελέγχου *Ljung – Box*, παρέχει ένα στατιστικό μέσω του οποίου μπορούμε να ελέγξουμε αν τα δεδομένα προέρχονται από μια  $GARCH$  χρονοσειρά.

**Θεώρημα (corrected portmanteau test):** Έστω  $r = (r_t, t \in T)$  μια  $GARCH(p, q)$  χρονοσειρά με  $E[r_t^4] < \infty$ , τότε για επαρκώς μεγάλο δείγμα ( $n \rightarrow \infty$ ), ισχύει ότι:

$$Q_m = n\hat{\rho}'_m \hat{\Sigma}_{\hat{\rho}_m}^{-1} \hat{\rho}_m \sim \chi_m^2 \quad (1.38)$$

όπου με  $\chi_m^2$  συμβολίζεται η χι-τετράγωνο κατανομή με  $m$  βαθμούς ελευθερίας,  $\hat{\rho}_m = (\hat{\rho}_1, \hat{\rho}_2, \dots, \hat{\rho}_m)'$  το διάνυσμα των  $m$  πρώτων δειγματικών αυτοσυσχετίσεων,  $\hat{\Sigma}_{\hat{\rho}_m} = \frac{\hat{\Sigma}_{\gamma_m}}{\gamma_0^2}$

και

$$\Sigma_{\gamma_m} = \begin{pmatrix} E[r_t^2 r_{t-1}^2] & \cdots & E[r_t^2 r_{t-1} r_{t-2}] \\ E[r_t^2 r_{t-1} r_{t-2}] & \cdots & E[r_t^2 r_{t-2} r_{t-m}] \\ \vdots & \ddots & \vdots \\ E[r_t^2 r_{t-1} r_{t-2}] & \cdots & E[r_t^2 r_{t-m}^2] \end{pmatrix} \quad (1.39)$$

όπου  $\hat{\Sigma}_{\gamma_m}$  είναι η εκτίμηση του πίνακα  $\Sigma_{\gamma_m}$  από τα δεδομένα. Σημειώνεται ότι στην περίπτωση που απορρίπτεται η μηδενική υπόθεση, η οποία εκφράζει ότι τα δεδομένα προέρχονται από ένα υπόδειγμα  $GARCH$ , θα πρέπει να δοκιμαστεί η εφαρμογή ενός  $ARMA - GARCH$  υποδείγματός όπου το  $ARMA$  μέρος θα περιγράφει τον τρόπο εξέλιξης του μέσου της χρονοσειράς, ενώ το  $GARCH$  τον τρόπο εξέλιξης των σφαλμάτων.

## 1.12 ΕΠΙΛΟΓΗ ΤΑΞΗΣ ΥΠΟΔΕΙΓΜΑΤΟΣ

Η επιλογή τάξης ενός υποδείγματος *GARCH* αποτελεί μια πολυσύνθετη διαδικασία, καθώς μπορεί να συνδυάζει πλήθος γραφημάτων και στατιστικών ελέγχων. Συγκεκριμένα, για την εφαρμογή του υποδείγματος (βλ. 1.26 και 1.27) σε μια χρονοσειρά  $r = (r_t \in T)$  μια απλή μεθοδολογία που θα μπορούσε κάποιος να ακολουθήσει είναι η εξής:

1. Από μια πληθώρα *GARCH*( $p, q$ ) μοντέλων, επιλέγεται εκείνο με το μικρότερο δείκτη *AIC* ή *BIC*.
2. Ελέγχεται η σημαντικότητα των όρων που εισάγονται στο μοντέλο μέσω των  $t$  – *tests*. Στην περίπτωση που κάποιος ή κάποιοι όροι εμφανίζουν μικρότερη τιμή  $p$  – *value* από το επίπεδο σημαντικότητάς  $\alpha$  (στατιστικά σημαντικός έλεγχος), απαλείφουμε τον όρο εκείνο με το μικρότερο  $p$  – *value* και εφαρμόζουμε ξανά το βήμα 2 έως ότου να προκύψει ένα μοντέλο όπου κανένας όρος δεν απορρίπτεται από τον παραπάνω έλεγχο.

Μια επιπλέον μέθοδος που μπορεί να χρησιμοποιηθεί για την εύρεση της τάξης ενός *GARCH* υποδείγματος, είναι η μέθοδος της γωνίας (***corner method***) που αναπτύχθηκε από τους Beguin, J. M., Gouriéroux, C., Monfort, A. (1980) και αφορά την εύρεση της τάξης ενός *ARMA* υποδείγματος.

Συγκεκριμένα, χρησιμοποιώντας την *ARMA* αναπαράσταση των τετραγώνων των τιμών ενός *GARCH* μοντέλου (βλ. 1.26), μπορούμε να προσδιορίσουμε το πλήθος των όρων που πρέπει να εισαχθούν τελικά σε αυτό.

Για την περιγραφή της μεθόδου, θεωρούμε μια  $X = (x_t, t \in T)$  *ARMA*( $p, q$ ) χρονοσειρά η οποία περιγράφεται από τη σχέση:

$$x_t + \sum_{i=1}^p \varphi_i x_{t-i} = e_t + \sum_{j=1}^q \theta_j e_{t-j}, e_t \sim WN(0,1) \quad (1.40)$$

ενώ για την συνάρτηση αυτοσυσχέτισης ισχύει ότι:

1. 
$$\rho_X(h) \neq \sum_{i=1}^p \varphi_i \rho_{h-i}, h = 1, 2, \dots, q \quad (1.41)$$

$$2. \quad \rho_X(h) = \sum_{i=1}^p \varphi_i \rho_{h-i}, h \geq q + 1 \quad (1.42)$$

Στη συνέχεια ορίζεται ο  $j \times j$  πίνακας *Toeplitz* ο οποίος κατασκευάζεται ως ακολούθως:

$$D(i, j) = \begin{pmatrix} \rho_i & \rho_{i-1} & \cdots & \rho_{i-j+1} \\ \rho_{i+1} & \rho_i & \cdots & \rho_{i-j+2} \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{i+j-1} & \rho_{i+j-2} & \cdots & \rho_i \end{pmatrix}, \text{ όπου } i, j \geq 1 \quad (1.43)$$

Αν με  $\Delta(i, j)$  συμβολίσουμε την ορίζουσα του πίνακα  $D(i, j)$ , οι Beguin, Gourieroux και Monfort απέδειξαν ότι η ελάχιστη τάξη  $p, q$  ενός *ARMA* υποδείγματος προσδιορίζεται αν ισχύουν οι συνθήκες:

$$\begin{cases} i) & \Delta(i, j) = 0, \forall i \geq q + 1 \text{ και } j \geq p + 1 \\ ii) & \Delta(i, p) \neq 0, \forall i \geq q \\ iii) & \Delta(q, j) \neq 0, \forall j \geq p \end{cases} \quad (1.44)$$

Επομένως, ο αλγόριθμος εύρεσης της τάξης μιας *ARMA* χρονοσειράς διαμορφώνεται ως εξής:

### **Βήμα 1<sup>ο</sup>**

Επιλέγεται τυχαία ένας μεγάλος ακέραιος αριθμός έστω  $M$ , ώστε  $M > \max\{p, q\}$ , και υπολογίζονται οι ορίζουσες  $\Delta(i, j)$  για  $i, j = 1, 2, \dots, M$ .

### **Βήμα 2<sup>ο</sup>**

Σε ένα  $K$  πίνακα διάστασης  $M \times M$  απεικονίζουμε τις υπολογιζόμενες ορίζουσες τους βήματος 1:

$$K = \begin{pmatrix} \Delta(1,1) & \Delta(1,2) & \cdots & \Delta(1,p) & \Delta(1,p+1) & \cdots & \Delta(1,M) \\ \Delta(2,1) & \Delta(2,2) & \cdots & \Delta(2,p) & \Delta(2,p+1) & \cdots & \Delta(2,M) \\ \vdots & \vdots & \ddots & \vdots & \vdots & \cdots & \vdots \\ \Delta(q,1) & \Delta(q,2) & \cdots & \Delta(q,p) & \Delta(q,p+1) & \cdots & \Delta(q,M) \\ \Delta(q+1,1) & \Delta(q+1,2) & \cdots & \Delta(q+1,p) & 0 & 0 & 0 \\ \vdots & \vdots & \cdots & \vdots & 0 & 0 & 0 \\ \Delta(M,1) & \Delta(M,2) & \cdots & \Delta(M,p) & 0 & 0 & 0 \end{pmatrix} \quad (1.45)$$

### Βήμα 3<sup>ο</sup>

Οι γραμμές του πίνακα (1.45) αναφέρονται στην τάξη του  $MA$  μέρους, ενώ οι στήλες στην τάξη του  $AR$  μέρους του υποδείγματος  $ARMA$ .

Έτσι, η τάξη του μοντέλου προσδιορίζεται εύκολα από τη θέση του στοιχείου  $\Delta(q,p)$  που αντιστοιχεί στο τελευταίο μη-μηδενικό στοιχείο της κύριας διαγωνίου του πίνακα  $K$ , προτού ξεκινήσει ο μηδενικός υποπίνακας στο δεξιό κάτω άκρο.

Είναι σημαντικό να αναφέρουμε ότι επειδή στην πράξη μόνο ένας πεπερασμένος αριθμός  $m$  το πλήθος αυτοσυσχετίσεων είναι διαθέσιμος, ο υπολογισμός των οριζουσών  $\Delta(i,j)$  γίνεται μόνο για  $i, j \geq 1$  και  $i + j \leq m + 1$ . Επομένως, ο πίνακας  $K$  τις περισσότερες φορές είναι τριγωνικός.

Στην περίπτωση αυτή, η επιλογή των  $p, q$  γίνεται με βάση την περιγραφή που δόθηκε στο τρίτο βήμα, με τη διαφορά ότι αντί να υπάρχει ένας υποπίνακας με μηδενικά στοιχεία, αυτός αποτελείται από αρκετά μικρές τιμές.

Αναφέρεται πως για τον ακριβέστερο προσδιορισμό της τάξης του μοντέλου, μπορεί να εφαρμοστεί και ένας στατιστικός έλεγχος.

Για περαιτέρω πληροφορίες σχετικά με τη μέθοδο *corner*, ο ενδιαφερόμενος αναγνώστης μπορεί να ανατρέξει π.χ. στην εργασία των Gooijer, J. G. and Heuts, R. M. J. (1981).

## 1.13 ΕΚΤΙΜΗΣΗ ΠΑΡΑΜΕΤΡΩΝ ΜΟΝΤΕΛΟΥ

Όπως στα μοντέλα  $ARCH$  (βλ. Κεφ. 1.8) έτσι και στα  $GARCH$  υποδείγματα μπορεί να εφαρμοστεί η μέθοδος  $QMLE$ . Συνοπτικά η μέθοδος περιγράφεται ως εξής:

Έστω  $\theta \in (0, +\infty) \times [0, +\infty)^{p+q}$  το διάνυσμα των αγνώστων παραμέτρων και  $\theta_0$  το διάνυσμα που περιέχει τις πραγματικές τιμές αυτών.

Η υπό – συνθήκη *QMLE* συνάρτηση δοθείσης των αρχικών τιμών  $r_0, \dots, r_{1-q}, \tilde{\sigma}_0^2, \dots, \tilde{\sigma}_{1-p}^2$  δίνεται από τη σχέση:

$$L(\theta; r_1, \dots, r_n) = \prod_{k=1}^n \frac{1}{\sqrt{2\pi\tilde{\sigma}_k^2}} e^{-\frac{r_k^2}{2\tilde{\sigma}_k^2}} \quad (1.46)$$

όπου η διασπορά υπολογίζεται για  $k \geq 1$  μέσω του επαναληπτικού τύπου:

$$\tilde{\sigma}_k^2 = w + \sum_{i=1}^q a_i r_{k-i}^2 + \sum_{j=1}^p b_j \tilde{\sigma}_{k-j}^2 \quad (1.47)$$

ενώ οι αρχικές συνθήκες επιλέγονται (εφόσον ισχύει η ασθενής στασιμότητα και για δεδομένη τιμή του  $\theta$ , άρα και των παραμέτρων) έτσι ώστε να ικανοποιούν τη σχέση:

$$r_0^2 = r_1^2 = \dots = r_{1-q}^2 = \tilde{\sigma}_0^2 = \dots = \tilde{\sigma}_{1-p}^2 = \frac{w}{1 - \sum_{i=1}^q a_i - \sum_{j=1}^p b_j} \quad (1.48)$$

Επομένως, οποιοδήποτε  $\hat{\theta}$  μεγιστοποιεί την (1.46), θα περιέχει ως στοιχεία τις εκτιμήσεις των παραμέτρων του μοντέλου.

## ΚΕΦΑΛΑΙΟ 2 ΘΕΩΡΙΑ ΑΚΡΑΙΩΝ ΤΙΜΩΝ

### 2.1 ΕΙΣΑΓΩΓΗ

Στο παρόν κεφάλαιο γίνεται εισαγωγή της Θεωρίας Ακραίων Τιμών (***Extreme Value Theory***) και των βασικών αποτελεσμάτων της, ενώ παρουσιάζονται αναλυτικά οι μέθοδοι ***Block Maxima*** και ***Peaks Over Threshold***.

### 2.2 ΘΕΩΡΙΑ ΑΚΡΑΙΩΝ ΤΙΜΩΝ

Ιδιαίτερο ενδιαφέρον κατά τη μελέτη στοχαστικών φαινομένων, παρουσιάζουν οι (σπάνιες) περιπτώσεις εμφάνισης αρκετά μικρών ή μεγάλων σε τιμή παρατηρήσεων, σε σχέση με το σύνολο των δεδομένων. Τέτοιου είδους παρατηρήσεις ονομάζονται ακραίες τιμές (***extreme values***) και η σημαντικότητα τους έγκειται στο γεγονός ότι υποδηλώνουν (συνήθως) την πραγματοποίηση ακραίων συμβάντων. Μερικά χαρακτηριστικά παραδείγματα τέτοιων ακραίων περιπτώσεων είναι μια μεγάλη σεισμική δόνηση, τα ακραία καιρικά φαινόμενα, καθώς και οι μεγάλες χρηματικές απώλειες που μπορεί να προκύψουν σε ένα χαρτοφυλάκιο τίτλων, οι τιμές των οποίων εξαρτώνται από άλλους ευμετάβλητους παράγοντες.

Θέτοντας το πρόβλημα υπό την σκοπιά των πιθανοτήτων, γίνεται εύκολα αντιληπτό ότι οι ακραίες παρατηρήσεις σχετίζονται άμεσα με τα άκρα της συνάρτησης κατανομής πιθανότητας ενός στοχαστικού φαινομένου. Επομένως, η μελέτη των ακραίων παρατηρήσεων αντιστοιχεί στη μελέτη και γνώση της (αριστερής ή δεξιάς) ουράς της συνάρτησης κατανομής του φαινομένου.

Ένα χρήσιμο παράδειγμα που αντικατοπτρίζει το παραπάνω είναι το πρόβλημα εύρεσης εκείνης της τιμής, για την οποία καμία παρατήρηση δεν θα την υπερβεί, για δεδομένη (μικρή) πιθανότητα και για συγκεκριμένο χρονικό διάστημα. Μια πρώτη προσέγγιση σε προβλήματα αυτού του είδους ή γενικότερα προβλήματα που σχετίζονται με ακραίες παρατηρήσεις, θα μπορούσε να γίνει φυσικά μέσω της θεωρίας που παρέχει η κλασική στατιστική. Συγκεκριμένα, αφού γίνει η κατάλληλη επιλογή της κατανομής του στοχαστικού φαινομένου και εκτιμηθούν οι παράμετροι της, μπορούμε στη συνέχεια να προβούμε στην οποιαδήποτε συμπερασματολογία αφορά τη συμπεριφορά των ακραίων



παρατηρήσεων. Αν και φαινομενικά αυτή η πρακτική φαίνεται σωστή, στην πραγματικότητα ο κίνδυνος να οδηγηθούμε σε «κακές» προβλέψεις είναι μεγάλος. Η ιδιομορφία αυτή, απορρέει από το γεγονός ότι τις περισσότερες φορές η πραγματική κατανομή των δεδομένων δεν είναι γνωστή. Επομένως, ακόμα και η επιλογή μιας κατάλληλης (από στατιστικής άποψης) κατανομής, θα μπορούσε να οδηγήσει σε τελείως διαφορετικά αποτελέσματα σε σχέση με την πραγματικότητά. Επίσης, η εκτίμηση των παραμέτρων ποτέ δεν είναι απόλυτα ακριβής, με συνέπεια ακόμα και μικρές διαφορές να έχουν μεγάλο αντίκτυπο στην μελέτη και πρόβλεψη των ακραίων παρατηρήσεων.

Προβλήματα αυτού του είδους θα μπορούσε κανείς να ισχυριστεί ότι επιλύονται, σε ένα ικανοποιητικό βαθμό, χρησιμοποιώντας τεχνικές της Θεωρίας Ακραίων Τιμών που ως κύριο στόχο έχει την ανάπτυξη στοχαστικών μοντέλων που ερμηνεύουν τη συμπεριφορά ακραίων παρατηρήσεων.

Εδώ είναι σημαντικό να αποσαφηνιστεί ότι αν και οι ακραίες παρατηρήσεις αναφέρονται σε πολύ μεγάλες ή πολύ μικρές τιμές, ο τρόπος με τον οποίο αυτές ορίζονται στην πράξη οδηγεί σε μια διαφορετική προσέγγιση από πλευράς τεχνικών και στατιστικής συμπερασματολογίας. Συγκεκριμένα, η θεωρία μας εφοδιάζει με δύο διαφορετικές (στατιστικές) προσεγγίσεις μελέτης ακραίων παρατηρήσεων, αυτή της μέγιστης παρατήρησης ανά ομάδα (**block maxima method**), η οποία οφείλεται στον Gumbel, και αυτή των υπερβάσεων πάνω από ένα συγκεκριμένο κατώφλι (**peaks over threshold method**). Το ποιά από τις δύο θεωρείται η καταλληλότερη, εξαρτάται σε μεγάλο βαθμό από τον τρόπο που καταγράφονται και γίνονται γνωστά τα δεδομένα. Προτού βέβαια γίνει η παρουσίαση των δύο αυτών στατιστικών μεθόδων, κρίνεται σκόπιμο να παρουσιαστούν τα βασικά αποτελέσματα της Θεωρίας Ακραίων Τιμών, από πλευράς πιθανοτήτων, και συγκεκριμένα θα αναλυθεί το βασικότερο μοντέλο της θεωρίας αυτής που αφορά τη μέγιστη παρατήρηση ενός δείγματος.

Έστω  $X_1, X_2, \dots, X_n$  ανεξάρτητες και ισόνομες τυχαίες μεταβλητές, με συνάρτηση κατανομής πιθανότητας  $F$ . Η μέγιστη παρατήρηση, που στο εξής θα συμβολίζεται με  $M_n$ , ισούται με:

$$M_n = \max \{X_1, \dots, X_n\}, \quad (2.01)$$

όπου εύκολα μπορούμε να βρούμε, λόγω της ανεξαρτησίας, ότι η συνάρτηση κατανομής πιθανότητας της  $M_n$  είναι:

$$\Pr[M_n \leq x] = (F(x))^n. \quad (2.02)$$

Όπως αναφέρθηκε προηγουμένως, επειδή η  $F$  δεν δύναται τις περισσότερες φορές να είναι γνωστή, η βασική ιδέα για τον προσδιορισμό της κατανομής της  $M_n$  είναι η μελέτη της οριακής περίπτωσης αυτής. Δηλαδή, της περίπτωσης όπου έχουμε ένα μεγάλο δείγμα παρατηρήσεων και στοχεύουμε να καταλήξουμε σε κάποια κατανομή η οποία, αν είναι δυνατόν, δεν θα εξαρτάται από καθαυτή την αρχική κατανομή  $F$ . Συγκεκριμένα, οι Roland Fisher και Leonard Tippett, απέδειξαν ότι η οριακή κατανομή του κανικοποιημένου μέγιστου, αν υπάρχει, θα συγκλίνει σε μια από τις τρεις οικογένειες κατανομών *Frechet*, (*reversed*) *Weibull* ή *Gumbel*.

Αναλυτικότερα το θεώρημα των Fisher-Tippett έχει ως εξής:

**Θεώρημα (Fisher-Tippett):** Έστω  $X_1, X_2, \dots, X_n$  μια ακολουθία ανεξάρτητων και ισόνομων τυχαίων μεταβλητών. Αν υπάρχουν ακολουθίες  $\{a_n > 0\}$  και  $\{b_n \in \mathbb{R}\}$  και μια μη-εκφυλισμένη συνάρτηση κατανομής  $G$ , ώστε να ισχύει ότι:

$$\Pr \left[ \frac{M_n - b_n}{a_n} \leq x \right] \xrightarrow{n \rightarrow \infty} G(x) \quad (2.03)$$

τότε η  $G$  θα είναι του ίδιου τύπου<sup>5</sup> με μια από τις παρακάτω κατανομές:

$$\Phi_a(x) = \begin{cases} 0, & x \leq 0 \\ e^{-x^{-a}}, & x > 0 \end{cases}, \quad a > 0 \quad (2.04)$$

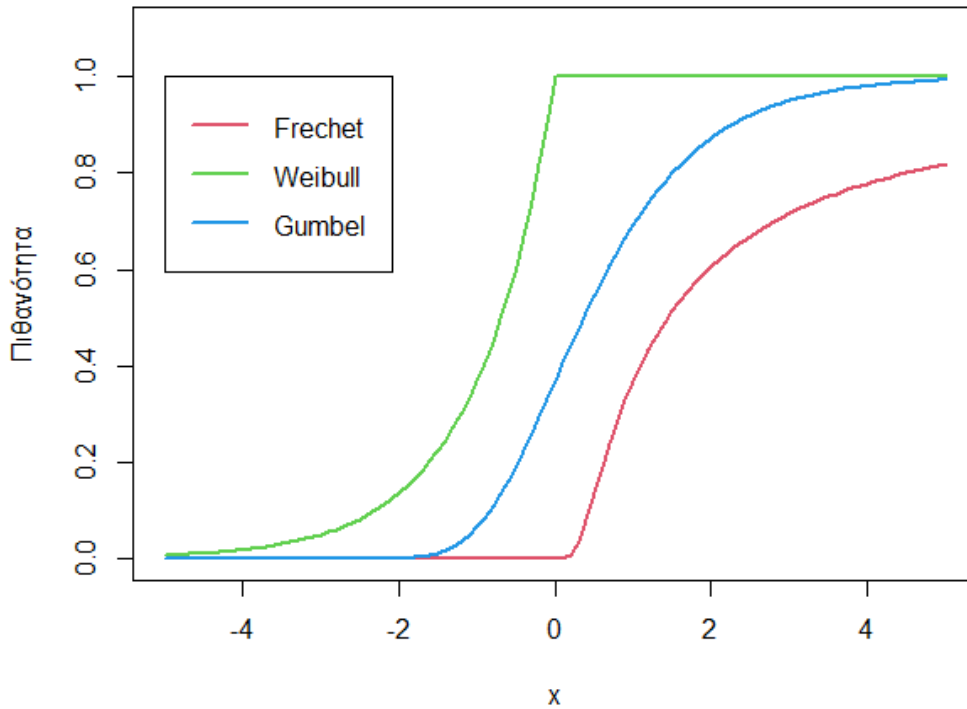
$$\Psi_a(x) = \begin{cases} e^{-(-x)^a}, & x \leq 0 \\ 1, & x > 0 \end{cases}, \quad a > 0 \quad (2.04)$$

$$\Lambda(x) = e^{-e^{-x}}, \quad x \in \mathbb{R}. \quad (2.05)$$

---

<sup>5</sup> Δύο κατανομές  $F, G$  θα λέμε ότι είναι του ίδιου τύπου, αν υπάρχουν σταθερές  $a > 0, b \in \mathbb{R}$  ώστε  $G(x) = F(ax + b)$ . Δηλαδή, οι δύο κατανομές είναι ίδιες και διαφέρουν μόνο ως προς την παράμετρο θέσης και κλίμακας.

### Γραφική παράσταση κατανομής πιθανότητας των συναρτήσεων Frechet, Weibull και Gumbel



**Σχήμα 2.1:** Γραφική παράσταση κατανομής πιθανότητας των συναρτήσεων Frechet, (reversed) Weibull και Gumbel.

Για την καλύτερη κατανόηση του θεωρήματος αναφέρεται ότι ο λόγος που υπολογίζεται η κατανομή του κανονικοποιημένου μέγιστου και όχι απλά του μέγιστου, είναι διότι η κατανομή  $(F(x))^n$  εκφυλίζεται για  $n \rightarrow \infty$ . Αναλυτικότερα, αν συμβολίσουμε με  $x_+$  το στήριγμα της  $F$ , δηλαδή το  $x_+ = \sup\{x \in \mathbb{R} : F(x) < 1\}$ , τότε ισχύει ότι:

$$\Pr[M_n \leq x] = (F(x))^n \xrightarrow{n \rightarrow \infty} \begin{cases} 0, & x < x_+ \\ 1, & x > x_+ \end{cases} \quad (2.06)$$

Είναι σημαντικό να αναφερθεί ότι η κανονικοποίηση του μέγιστου, επομένως και η ύπαρξη των ακολουθιών  $\{a_n > 0\}$  και  $\{b_n \in \mathbb{R}\}$ , δεν είναι πάντοτε εφικτή, διότι εξαρτάται από την ταχύτητα σύγκλισης του λόγου  $\frac{\bar{F}(x)}{\bar{F}(x-)}$ , καθώς το  $x \rightarrow x_+$ . Αναλυτικότερα, ισχύει το εξής θεώρημα:

**Θεώρημα:** Έστω μια συνάρτηση κατανομής  $F$  με  $x_+ \leq \infty$ , τότε ισχύει ότι:

$$\Pr \left[ \frac{M_n - b_n}{a_n} \leq x \right] \xrightarrow{n \rightarrow \infty} G(x) \text{ αν και μόνο αν } \lim_{x \rightarrow x_+} \frac{\bar{F}(x)}{\bar{F}(x-)} = 1.$$

Τέλος, η επιλογή της κατανομής (Frechet, (reversed) Weibull, Gumbel) που θα συγκλίνει η οριακή κατανομή του κανικοποιημένου μέγιστου, αν αυτή υπάρχει, εξαρτάται από τη μορφή της δεξιάς (ή αριστερής) ουράς της  $F$  και συγκεκριμένα από το πόσο γρήγορα αυτή συγκλίνει στο μηδέν, όταν το  $x \rightarrow x_+$ . Ο ενδιαφερόμενος αναγνώστης μπορεί να ανατρέξει π.χ. στα συγγράμματα των Embrechts, P., Kluppelberg, C. and Mikosch, T. (1997) και Haan, L. and Ferreira, A (2006) όπου γίνεται μια εκτεταμένη ανάλυση των συναρτήσεων ομαλής κύμανσης, προσφέροντας έναν αποτελεσματικό τρόπο διαχωρισμού των κατανομών στις περιοχές έλξης των τριών κατανομών του θεωρήματος των Fisher-Tippett.

Το θεώρημα των Fisher-Tippett, όπως διατυπώθηκε παραπάνω, εξασφαλίζει ότι αν υπάρχει η κατανομή του οριακού μέγιστου, αυτή δεν μπορεί να είναι κάποια άλλη εκτός των *Frechet*, *Weibull* ή *Gumbel*. Στην πράξη όμως, για να προβούμε στην οποιαδήποτε συμπερασματολογία, θα πρέπει αρχικά να γίνει επιλογή εκείνης της κατανομής που ταιριάζει καλύτερα στα δεδομένα, με κίνδυνο πάντα η επιλογή αυτή να είναι εσφαλμένη. Επομένως, αν ήταν δυνατόν να εκφραστούν οι παραπάνω τρεις κατανομές μέσω μιας και μοναδικής, όπου μια παράμετρος σχήματος θα καθόριζε ποια από τις τρεις οικογένειες κατανομών έχουμε, αυτό θα απλοποιούσε αρκετά την όλη διαδικασία και θα περιοριζε τον κίνδυνο της λανθασμένης επιλογής. Κάτι τέτοιο είναι εφικτό στην πράξη. Συγκεκριμένα, οι παραπάνω κατανομές συνοψίζονται μέσω μιας οικογένειας κατανομών γνωστή και ως *GEV (Generalized Extreme Value Distribution)* με συναρτησιακή σχέση:

$$G(x) = \exp\left(-\left[1 + \xi\left(\frac{x-\mu}{\sigma}\right)\right]^{-\frac{1}{\xi}}\right) \quad (2.07)$$

όπου  $\{x: 1 + \xi\frac{x-\mu}{\sigma} > 0\}$ ,  $\mu, \xi \in \mathbb{R}$  και  $\sigma > 0$ .

Το παραπάνω επιτρέπει την αναδιατύπωση του θεωρήματος των Fisher-Tippett ως ακολούθως:

**Θεώρημα:** Έστω  $X_1, X_2, \dots, X_n$  μια ακολουθία ανεξάρτητων και ισόνομων τυχαίων μεταβλητών. Αν υπάρχουν ακολουθίες  $\{a_n > 0\}$  και  $\{b_n \in \mathbb{R}\}$  και μια μη-εκφυλισμένη συνάρτηση κατανομής  $G$ , ώστε να ισχύει ότι:

$$\Pr \left[ \frac{M_n - b_n}{a_n} \leq x \right] \xrightarrow{n \rightarrow \infty} G(x) \quad (2.08)$$

τότε η  $G$  θα είναι της μορφής:

$$G(x) = \exp \left( - \left[ 1 + \xi \left( \frac{x-\mu}{\sigma} \right) \right]^{-\frac{1}{\xi}} \right), \left\{ x: 1 + \xi \frac{(x-\mu)}{\sigma} > 0 \right\}, \text{ όπου } \mu, \xi \in \mathbb{R} \text{ και } \sigma > 0. \quad (2.09)$$

Το γεγονός ότι η κατανομή  $GEV$  περιέχει τις κατανομές που αναφέραμε στο θεώρημα των *Fisher – Tippett*, προκύπτει μέσα από κατάλληλες τιμές των παραμέτρων  $\mu$ ,  $\xi$  και  $\sigma$ . Επίσης τονίζεται ιδιαίτερα ότι, ανάλογα με τις τιμές που λαμβάνει η παράμετρος σχήματος  $\xi$ , μπορούμε να καταλάβουμε σε ποια από τις τρεις κατανομές απευθυνόμαστε. Συγκεκριμένα για:

- $\xi > 0$  προκύπτει η κατανομή *Frechet*
- $\xi < 0$  προκύπτει η κατανομή *Weibull*
- $\xi \rightarrow 0$  προκύπτει η κατανομή *Gumbel*.

**Παρατήρηση:** Οι κατανομές που έχουν παρουσιαστεί έως τώρα, αναφέρονται στο κανονικοποιημένο μέγιστο και όχι στο μέγιστο  $M_n$  για το οποίο θέλουμε να εξάγουμε συμπεράσματα. Θέτοντας όμως όπου  $x = a_n u + b_n$ , προκύπτει άμεσα ότι και η κατανομή του μέγιστου ακολουθεί την  $GEV$  κατανομή, αλλά με διαφορετικές παραμέτρους θέσης και κλίμακας, σε σχέση με αυτές του κανονικοποιημένου μέγιστου:

$$\Pr[M_n \leq x] = \Pr \left[ \frac{M_n - b_n}{a_n} \leq \frac{x - b_n}{a_n} \right] = G \left( \frac{x - b_n}{a_n}; \mu', \xi, \sigma' \right) = e^{- \left( 1 + \xi \frac{x - b_n - a_n \mu}{a_n \sigma} \right)^{\frac{1}{\xi}}} \quad (2.10)$$

όπου  $\mu' = a_n \mu + b_n$  και  $\sigma' = a_n \sigma$ .

### 2.3 ΜΕΘΟΔΟΣ ΤΟΥ ΜΕΡΙΚΟΥ ΜΕΓΙΣΤΟΥ (Block Maxima Method)

Για τη στατιστική μελέτη της κατανομής  $GEV$ , όπως συμβαίνει σε ανάλογες περιπτώσεις, κρίνεται απαραίτητη η κατάλληλη επιλογή ενός δείγματος, με σκοπό να εκτιμηθούν οι παράμετροι του μοντέλου. Η μέθοδος που προτείνεται από τη βιβλιογραφία ονομάζεται μέθοδος των μερικών μέγιστων (**block maxima method**). Η συγκεκριμένη μέθοδος

κρίνεται κατάλληλη όταν τα δεδομένα που έχουμε στη διάθεση μας είναι οι μέγιστες παρατηρήσεις από διαφορετικά σύνολα δεδομένων (**blocks**) του εξεταζόμενου φαινομένου, που αναφέρονται σε ξεχωριστές χρονικές περιόδους, συγκεκριμένης χρονικής διάρκειας (ημέρα, εβδομάδα, μήνα, έτος). Αναλυτικότερα η μέθοδος μπορεί να περιγραφεί με τον ακόλουθο τρόπο. Έστω ένα σύνολο παρατηρήσεων  $X_1, X_2, \dots, X_n$  το οποίο χωρίζεται σε  $k$  το πλήθος υποσύνολα (**blocks**) των  $r$  παρατηρήσεων το κάθε ένα, έτσι ώστε  $n = kr$ . Έστω επίσης  $M_1, M_2, \dots, M_k$  οι μέγιστες παρατηρήσεις από τα  $k$  υποσύνολα, οι οποίες διαφορετικά καλούνται μερικά μέγιστα (**block maxima**). Τότε, εάν το πλήθος των  $r$  παρατηρήσεων σε κάθε υποσύνολο είναι επαρκώς μεγάλο, μπορούμε να υποθέσουμε ότι τα μέγιστα  $M_1, M_2, \dots, M_k$  προέρχονται από την κατανομή *GEV*. Επομένως, χρησιμοποιώντας κατάλληλες στατιστικές μεθόδους, γίνεται εφικτή η εκτίμηση των παραμέτρων της κατανομής αυτής.

Σημειώνεται ότι για την επιλογή του αριθμού  $r$  (πλήθος παρατηρήσεων σε κάθε υποσύνολο), θα πρέπει να είμαστε ιδιαίτερα προσεκτικοί. Ένα μεγάλο σε τιμή  $r$ , οδηγεί σε ένα μικρό πλήθος υποσυνόλων  $k$  άρα και μέγιστων παρατηρήσεων, με αποτέλεσμα να αυξάνεται σημαντικά η διασπορά των εκτιμήσεων. Από την άλλη, εάν το  $r$  είναι αρκετά μικρό, υπάρχει ο κίνδυνος τα μέγιστα να μην προσεγγίζονται από την κατανομή *GEV* πχ. βλ. Coles, S. (2001).

## 2.4 ΣΤΑΘΜΗ ΑΠΟΔΟΣΗΣ

Ένα πρόβλημα ιδιαίτερης σημασίας που συναντάται σε αρκετές εφαρμογές, αναφέρεται ως στάθμη απόδοσης (**return level**). Το συγκεκριμένο πρόβλημα συνίσταται στην εύρεση εκείνης της τιμής  $u_p$  της κατανομής, την οποία η μέγιστη παρατήρηση  $M_n$ , ενός συνόλου δεδομένων (block), μπορεί να υπερβεί με πιθανότητα  $p$ .

$$\Pr[M_n > u_p] = p. \quad (2.11)$$

Η σημαντικότητα του προβλήματος αυτού δεν έγκειται μόνο στην εύρεση της στάθμης απόδοσης  $u_p$ , αλλά και στην εύρεση της μέσης χρονικής διάρκειας όπου η μέγιστη παρατήρηση θα υπερβεί το κατώφλι αυτό, π.χ. βλ. Μ. Μπούτσικας (2008). Συγκεκριμένα,

το μέσο αυτό χρονικό διάστημα ισούται με  $1/p$  χρονικές περιόδους και ονομάζεται περίοδος απόδοσης (**return period**).

Συνοψίζοντας, αναφέρεται ότι η στάθμη απόδοσης για  $1/p$  χρονικές περιόδους, αν θέσουμε όπου  $y_p = -\ln(1 - p)$ , ισούται με:

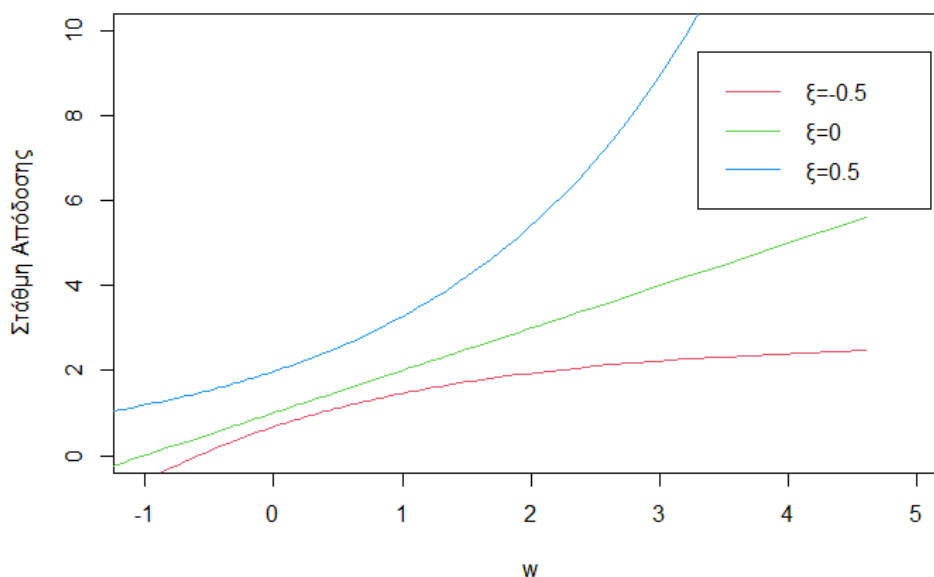
$$u_p = \begin{cases} \mu - \frac{\sigma}{\xi} (1 - y_p^{-\xi}), & \xi \neq 0 \\ \mu - \sigma \ln y_p, & \xi = 0 \end{cases} \quad (2.12)$$

Από τη σχέση (βλ. 2.12) εύκολα προκύπτει ότι θέτοντας όπου  $w = -\ln y_p$ , τότε:

- Για  $\xi = 0$  το  $u_p$  είναι γραμμική συνάρτηση του  $w$ .
- Για  $\xi > 0$  το  $u_p$  είναι κοίλη συνάρτηση του  $w$ .
- Για  $\xi < 0$  το  $u_p$  είναι κυρτή συνάρτηση του  $w$ .

Η παραπάνω παρατήρηση είναι ιδιαίτερα χρήσιμη, καθώς συμπεραίνουμε τί τιμές λαμβάνει η παράμετρος σχήματος  $\xi$ , ενώ συνάμα μπορούμε να ελέγξουμε και την καλή προσαρμογή του μοντέλου.

**Γράφημα στάθμης απόδοσης**



**Σχήμα 2.2:** Γραφική παράσταση της στάθμης απόδοσης για τιμές της παραμέτρου θέσης  $\xi = (-0.5, 0, 0.5)$ . Όπως φαίνεται στο γράφημα, όταν η παράμετρος θέσης λαμβάνει τιμές κοντά στο μηδέν η στάθμη απόδοσης είναι γραμμική συνάρτηση, ενώ όταν λαμβάνει θετικές (αντ. αρνητικές) τιμές η στάθμη απόδοσης είναι κοίλη (αντ. κυρτή).

## 2.5 ΕΚΤΙΜΗΣΗ ΠΑΡΑΜΕΤΡΩΝ ΜΟΝΤΕΛΟΥ

Η εκτίμηση των παραμέτρων της  $GEV$ , μπορεί να πραγματοποιηθεί μέσω μιας πληθώρας αριθμητικών (μέθοδος ροπών, μέγιστης πιθανοφάνειας) ή και γραφικών μεθόδων. Στην παρούσα διπλωματική εργασία επιλέχθηκε να παρουσιαστεί η μέθοδος της μέγιστης πιθανοφάνειας, καθώς είναι αυτή που θα χρησιμοποιηθεί στη συνέχεια για την εκτίμηση των παραμέτρων του μοντέλου στα πρακτικά προβλήματα που θα εξεταστούν.

Είναι χρήσιμο να αναφερθεί ότι ο εκτιμητής μέγιστης πιθανοφάνειας είναι συνεπής και ασυμπτωτικά κανονικός, όταν ισχύουν οι συνθήκες ομαλότητας. Μάλιστα μια εκ των συνθηκών ομαλότητας, επιβάλλει το στήριγμα της συνάρτησης κατανομής να μην εξαρτάται από τις παραμέτρους της. Όπως είναι φανερό από τη σχέση (βλ. 2.09) κάτι τέτοιο δεν ισχύει για την περίπτωση της κατανομής  $GEV$ . Επομένως, χρειάζεται ιδιαίτερη προσοχή η χρήση της μεθόδου. Συγκεκριμένα, ο Richard L. Smith (1985) απέδειξε ότι:

1. Για  $\xi > -1/2$  οι εκτιμητές μέγιστης πιθανοφάνειας έχουν τη συνήθη ασυμπτωτική συμπεριφορά.
2. Για  $-1 \leq \xi < -1/2$  οι εκτιμήσεις μέγιστης πιθανοφάνειας μπορούν να βρεθούν, αλλά οι εκτιμητές δεν έχουν τις ιδιότητες της ασυμπτωτικής συμπεριφοράς.
3. Για  $\xi < -1$  δεν δύναται να υπολογιστούν οι εκτιμήσεις μέγιστης πιθανοφάνειας.

Τέλος, αναφέρεται ότι η φύση των προβλημάτων τα οποία πραγματεύεται η παρούσα εργασία, καλύπτονται από την περίπτωση ένα, διότι οι περιπτώσεις για  $\xi < -1/2$  αντιστοιχούν σε κατανομές με πεπερασμένο στήριγμα. Επομένως, η μέθοδος μέγιστης πιθανοφάνειας μπορεί να εφαρμοστεί χωρίς κάποιο ιδιαίτερο πρόβλημα.

Έστω  $M_1, M_2, \dots, M_k$  ανεξάρτητες και ισόνομες τυχαίες μεταβλητές που προέρχονται από την κατανομή  $GEV$ . Η λογαριθμική συνάρτηση πιθανοφάνειας δίνεται από τη σχέση:

$$l(x; \mu, \sigma, \xi) = \begin{cases} -k \ln \sigma - \left(1 + \frac{1}{\xi}\right) \sum_{i=1}^k \left[ \ln \left(1 + \xi \left(\frac{x_i - \mu}{\sigma}\right)\right) - \left(1 + \xi \left(\frac{x_i - \mu}{\sigma}\right)\right)^{\frac{1}{\xi}} \right] & , \quad \xi \neq 0 \\ -k \ln \sigma - \sum_{i=1}^k \left[ \left(\frac{x_i - \mu}{\sigma}\right) - \exp\left(-\left(\frac{x_i - \mu}{\sigma}\right)\right) \right] & , \quad \xi = 0 \end{cases} \quad (2.13)$$

όπου,  $\{x: 1 + \xi \frac{x-\mu}{\sigma} > 0\}$ .

Για την μεγιστοποίηση της εξίσωσης (βλ. 2.13) δεν υπάρχουν αναλυτικές λύσεις. Επομένως, η εύρεση των εκτιμητών γίνεται χρησιμοποιώντας αριθμητικές μεθόδους,



προσέχοντας πάντα οι λύσεις που προκύπτουν να ικανοποιούν το πεδίο ορισμού της εξίσωσης.

Επίσης, χρησιμοποιώντας τη μέθοδο Δέλτα (**delta method**), εύκολα μπορούμε να εκτιμήσουμε τη στάθμη απόδοσης  $u_p$  που αντιστοιχεί κατά μέσο όρο σε  $1/p$  χρονικές περιόδους με  $0 < p < 1$ . Συγκεκριμένα, αντικαθιστώντας τις εκτιμήσεις  $\hat{\mu}, \hat{\sigma}, \hat{\xi}$  των παραμέτρων  $\mu, \sigma$  και  $\xi$  στην εξίσωση (βλ. 2.12), προκύπτει ότι:

$$\hat{u}_p = \begin{cases} \hat{\mu} - \frac{\hat{\sigma}}{\hat{\xi}} (1 - y_p^{-\hat{\xi}}) & , \quad \hat{\xi} \neq 0 \\ \hat{\mu} - \hat{\sigma} \ln y_p & , \quad \hat{\xi} = 0 \end{cases} \quad (2.14)$$

## 2.6 ΕΛΕΓΧΟΣ ΚΑΛΗΣ ΠΡΟΣΑΡΜΟΓΗΣ

Ένας γρήγορος και σχετικά εύκολος τρόπος ελέγχου του μοντέλου που δείχνει το πόσο καλά προσαρμόζεται αυτό στα (ιστορικά) δεδομένα, είναι μέσω του  $QQ - plot$ .

Το  $QQ - plot$  είναι ένα γράφημα το οποίο συνδυάζει τα ποσοστιαία σημεία ενός διατεταγμένου συνόλου δεδομένων, με τα αντίστοιχα ποσοστιαία σημεία της κατανομής που θέλουμε να ελέγξουμε.

Αναλυτικότερα, έστω ένα δείγμα ανεξάρτητων, διατεταγμένων παρατηρήσεων από πληθυσμό με εκτιμώμενη συνάρτηση κατανομής  $\hat{F}$ .

$$x_{(1)} \leq x_{(2)} \leq \dots \leq x_{(k)}$$

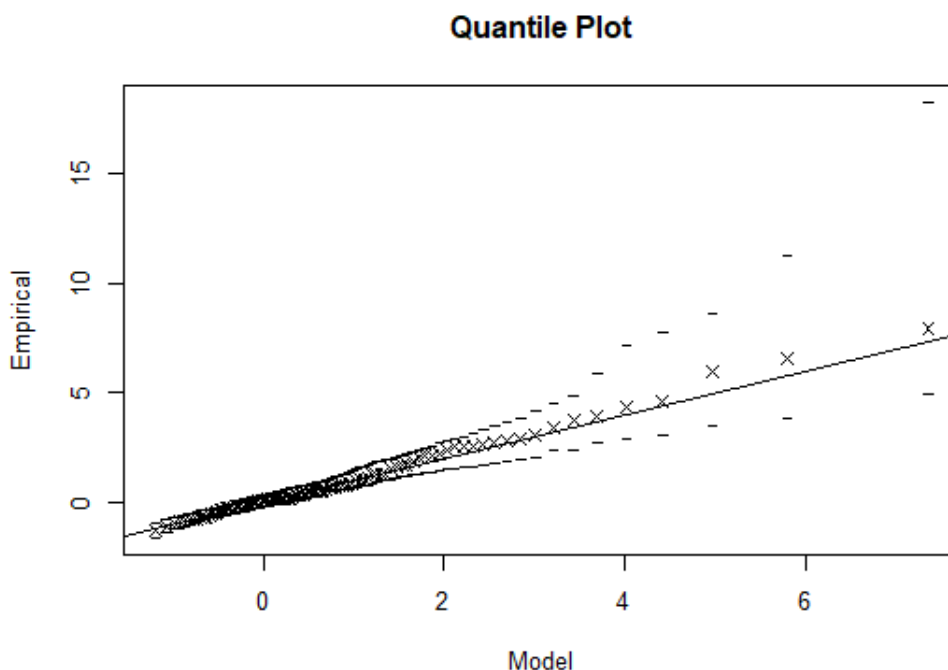
Το  $QQ - plot$  είναι το γράφημα που αποτελείται από τα σημεία:

$$\left\{ \left( \hat{F}^{-1} \left( \frac{i}{n+1} \right), x_{(i)} \right) : i = 1, \dots, k \right\} \quad (2.15)$$

Στην περίπτωση όπου η εκτιμώμενη κατανομή προσαρμόζεται «καλά» στα δεδομένα, τα σημεία του παραπάνω γραφήματος αναμένεται να κατανέμονται γύρω από την ευθεία  $y = x$ , ενώ τυχόν μεγάλες αποκλίσεις από αυτή υποδηλώνουν την μη-καταλληλότητα της κατανομής.

Γενικότερα, η κατασκευή του  $QQ - plot$  βασίζεται στο γεγονός ότι αν  $X_1, \dots, X_n$  ανεξάρτητες και ισόνομες τυχαίες μεταβλητές με κατανομή  $F$ , τότε οι τυχαίες μεταβλητές

$Y_i = F(X_i) \sim U(0,1)$ , επομένως και οι διατεταγμένες παρατηρήσεις  $Y_{(1)} < \dots < Y_{(n)}$  θα προέρχονται από την  $U(0,1)$ . Όμως, για  $n \rightarrow \infty$  ισχύει ότι  $Y_{(i)} \approx \frac{i}{n+1}$  ή ισοδύναμα  $X_{(i)} \approx F^{-1}\left(\frac{i}{n+1}\right)$ .



**Σχήμα 2.3:** QQ-plot για την περίπτωση προσαρμογής της κατανομής  $GEV(\mu=0.13, \sigma=1.1, \xi=0.2)$  σε τυχαίο δείγμα. Στο γράφημα επίσης έχουν σχεδιαστεί και τα αντίστοιχα όρια των διαστημάτων εμπιστοσύνης για το κάθε ποσοστιαίο σημείο.

## 2.7 ΜΕΘΟΔΟΣ POT (Peak Over Threshold)

Μια εναλλακτική μέθοδος μελέτης ακραίων παρατηρήσεων, είναι η μέθοδος **POT (Peak Over Threshold)**. Η μέθοδος αυτή βασίζεται στη μελέτη ενός συνόλου παρατηρήσεων που βρίσκονται πάνω από ένα προκαθορισμένο κατώφλι  $u$ . Η POT προτιμάται έναντι της *Block – Maxima*, όταν το δείγμα που έχουμε στις διάθεση μας περιέχει περισσότερη πληροφορία (παρατηρήσεις) εκτός από τις μέγιστες παρατηρήσεις ανά *block*. Στην περίπτωση αυτή, η κατανομή η οποία παρουσιάζει ενδιαφέρον είναι η δεσμευμένη κατανομή των υπερβάσεων. Συγκεκριμένα, έστω  $X_1, X_2, \dots, X_n$  ανεξάρτητες και ισόνομες τυχαίες μεταβλητές, με συνάρτηση κατανομής  $F$  και  $u$  ένα (αρκετά) μεγάλο κατώφλι. Η δεσμευμένη κατανομή των υπερβάσεων, για δεδομένο κατώφλι  $u$ , δίνεται από τη σχέση:

$$\Pr[X_i > u + y | X_i > u] = \frac{1 - F(u + y)}{1 - F(u)}, y > 0. \quad (2.16)$$

Καθώς και σε αυτή την περίπτωση, η αρχική κατανομή  $F$  δεν είναι γνωστή, στόχος είναι η εύρεση μιας οριακής κατανομής  $H$ , όπου για μεγάλες τιμές του  $u$ , αυτή να αποτελεί μια προσέγγιση της δεσμευμένης κατανομής (βλ. 2.16). Το επόμενο θεώρημα, δίνει μια λύση στο συγκεκριμένο πρόβλημα μέσω της γενικευμένης κατανομής Pareto.

**Θεώρημα:** Έστω  $X_1, X_2, \dots, X_n$  ανεξάρτητες και ισόνομες τυχαίες μεταβλητές, με συνάρτηση κατανομής  $F$  τέτοια ώστε η κατανομή του κανικοποιημένου μέγιστου  $M_n$  να είναι η  $GEV$ :

$$\Pr[M_n \leq x] \rightarrow G(x; \mu, \sigma, \xi). \quad (2.17)$$

Τότε, για αρκετά μεγάλες τιμές του  $u$  ισχύει ότι:

$$\Pr[X - u \leq y | X > u] = H(y; \tilde{\sigma}, \xi) = 1 - \left(1 + \frac{\xi y}{\tilde{\sigma}}\right)^{-\frac{1}{\xi}} \quad (2.18)$$

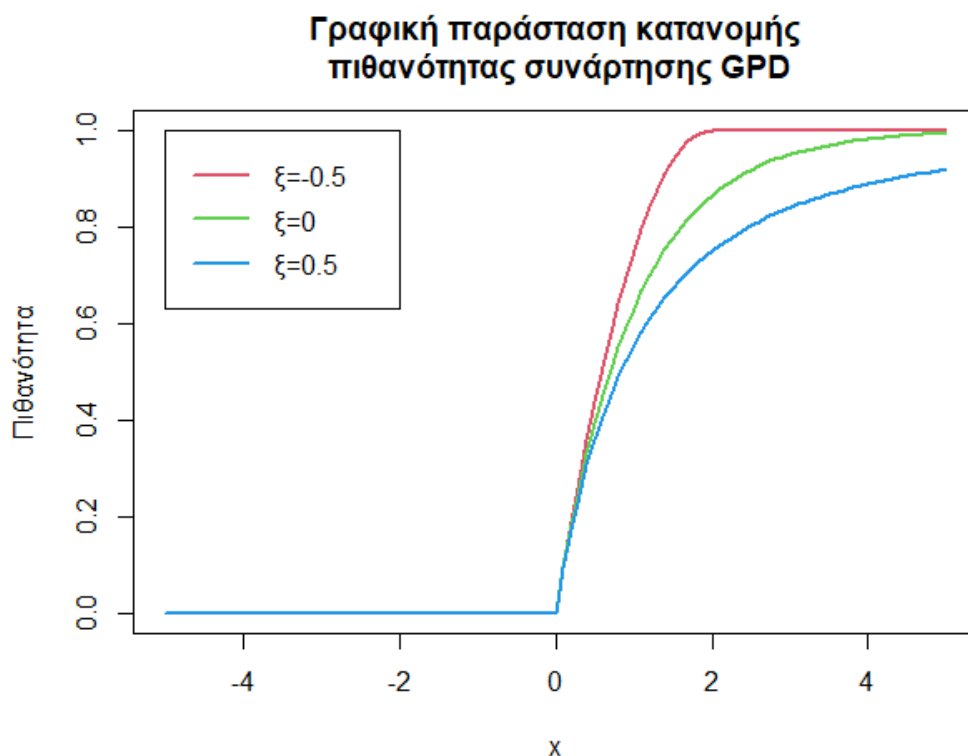
όπου  $y \in (0, \infty)$  και  $1 + \frac{\xi y}{\tilde{\sigma}} > 0$ .

Η κατανομή  $H$  αποτελεί στην πραγματικότητα μια οικογένεια κατανομών, που είναι γνωστή με την ονομασία *Γενικευμένη Κατανομή Pareto (Generalized Pareto Distribution - GPD)*.

Επισημαίνεται ότι οι παράμετροι της κατανομής  $GPD$  συνδέονται με τις αντίστοιχες της  $GEV$ . Συγκεκριμένα, το  $\tilde{\sigma}(u) = \sigma + \xi(u - \mu)$ , ενώ η παράμετρος σχήματος  $\xi$  λαμβάνει την ίδια ακριβώς τιμή και για τις δύο κατανομές. Επίσης, ανάλογα με την τιμή της παραμέτρου  $\xi$  διακρίνονται οι εξής τρεις περιπτώσεις:

- Για  $\xi < 0$  η κατανομή  $GPD$  έχει ως άνω φράγμα (στήριγμα) την ποσότητα  $u - \tilde{\sigma}/\xi$ .
- Για  $\xi > 0$  η κατανομή εκτείνεται στο άπειρο.
- Όταν το  $\xi$  τείνει στο μηδέν, η  $GPD$  εκφράζει την εκθετική κατανομή με παράμετρο  $1/\tilde{\sigma}$

$$H(x) = 1 - \exp\left(-\frac{x}{\tilde{\sigma}}\right), x > 0. \quad (2.19)$$



**Σχήμα 2.4:** Γραφική παράσταση κατανομής GPD για τις διάφορες τιμές του  $\xi$ .

## 2.8 ΣΤΑΘΜΗ ΑΠΟΔΟΣΗΣ

Έστω ότι η κατανομή των υπερβάσεων  $Y_i$  για δεδομένο κατώφλι  $u$  είναι η GPD, τότε η στάθμη απόδοσης  $x_p$  (με  $x_p > u$ ) που αντιστοιχεί σε  $1/p$  χρονικές περιόδους δίνεται από τη σχέση:

$$x_p = \begin{cases} u + \frac{\tilde{\sigma}}{\xi} ((pw)^\xi - 1) & , \quad \xi \neq 0 \\ u + \tilde{\sigma} \ln(pw) & , \quad \xi = 0 \end{cases} \quad (2.20)$$

όπου  $w = \Pr[X_i > u]$  με αντίστοιχη εκτιμήτρια το λόγο του αριθμού των υπερβάσεων προς το συνολικό αριθμό των παρατηρήσεων:

$$\hat{w} = \frac{k}{n} \quad , \quad k = \text{card}(i : x_i > u). \quad (2.21)$$

## 2.9 ΕΠΙΛΟΓΗ ΚΑΤΑΛΛΗΛΟΥ ΚΑΤΩΦΛΙΟΥ $u$

Σύμφωνα με το προηγούμενο θεώρημα, είδαμε ότι η κατανομή των υπερβάσεων  $Y_i =_d X_i - y | X_i > u$ , για ένα δεδομένο (αρκετά μεγάλο) κατώφλι  $u$ , ακολουθεί την κατανομή  $GPD$ . Επομένως, η επιλογή του κατωφλιού  $u$  είναι κρίσιμης σημασίας. Μια αρκετά μεγάλη τιμή  $u$ , θα μπορούσε να έχει ως αποτέλεσμα ένα μικρό αριθμό παρατηρήσεων και κατά συνέπεια «κακές» εκτιμήσεις (μεγάλης διασποράς) των παραμέτρων της κατανομής, π.χ. βλ. Μ. Μπούτσικας (2008). Από την άλλη, η επιλογή ενός μικρού κατωφλιού, θα οδηγούσε στην μη καταλληλότητα της συγκεκριμένης μεθόδου. Συμπερασματικά, η βέλτιστη επιλογή του  $u$  γίνεται κατά τρόπο ώστε από τη μια να ισχύει το μοντέλο και από την άλλη οι εκτιμήσεις που λαμβάνονται να έχουν όσο το δυνατό μικρότερη διασπορά. Για το σκοπό αυτό μια γραφική μέθοδος, που στηρίζεται στη γραμμικότητα της μέση υπερβάλλουσας συνάρτησης (**mean excess function**) των υπερβάσεων  $Y_i$ , μπορεί να χρησιμοποιηθεί. Συγκεκριμένα, αν υποθέσουμε ότι  $Y_i \sim GPD(\tilde{\sigma}, \xi)$ , τότε για  $\xi < 1$  η μέση υπερβάλλουσα συνάρτηση  $e(u)$  είναι:

$$e(u) = E[X - y | X > u] = \frac{\sigma + \xi(u - \mu)}{1 - \xi}. \quad (2.22)$$

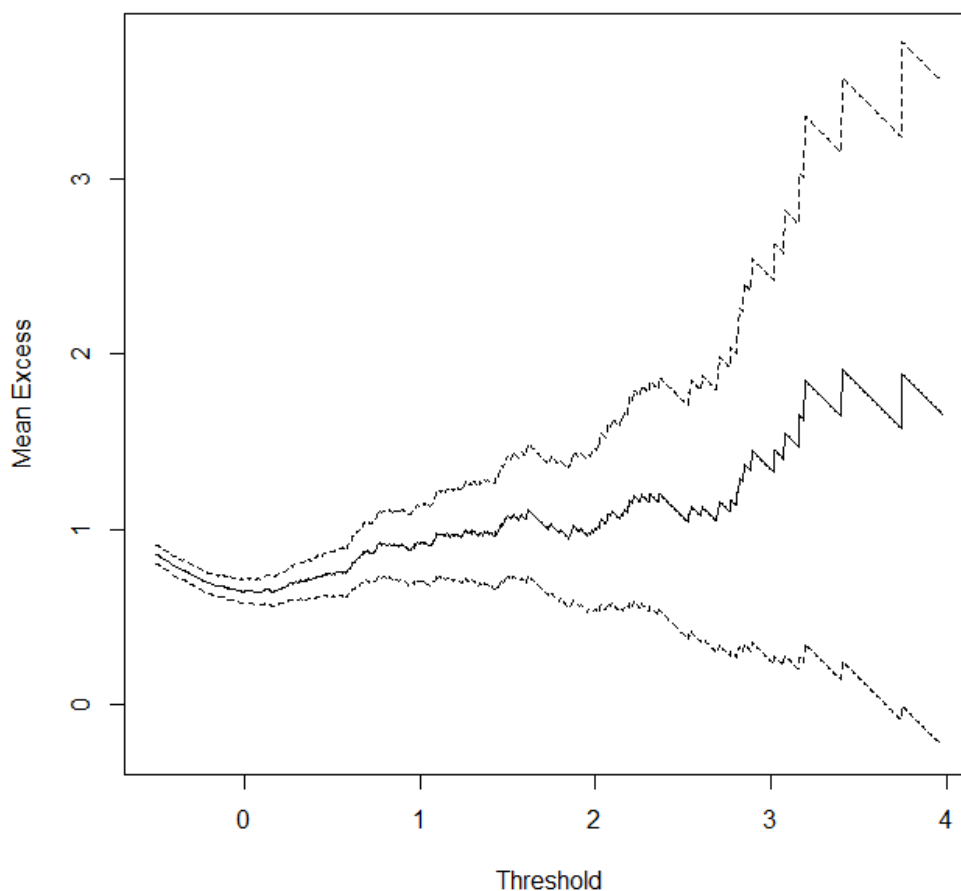
Η γραμμικότητα της παραπάνω σχέσης ως προς της μεταβλητή  $u$ , επιτρέπει την επιλογή του (μικρότερου) κατωφλιού  $u_0$ , ώστε η  $e(u)$  να απεικονίζει μια ευθεία γραμμή, για τιμές του  $u \geq u_0$ .

Ο γραφικός έλεγχος επιτυγχάνεται αφού πρώτα εκτιμηθεί η  $e(u)$  από την εμπειρική μέση υπερβάλλουσα συνάρτηση που δίνεται από τη σχέση:

$$\hat{e}(u) = \frac{1}{n(u)} \sum_{i: x_i > u} (x_i - u) \quad (2.23)$$

όπου το  $n(u) = \text{card}(i : i = 1, \dots, n, x_i > u)$  εκφράζει το πλήθος των  $x_i$  που υπερβαίνουν το κατώφλι  $u$ .

### Γράφημα μέσης υπερβάλλουσας συνάρτησης



**Σχήμα 2.5:** Γραφική παράσταση μέσης υπερβάλλουσας ζημίας. Παρατηρούμε ότι στην τιμή  $u = 0.2$  η γραφική παράσταση της συνάρτησης φαίνεται να γίνεται αύξουσα. Επομένως, σαν κατώφλι μπορούμε να επιλέξουμε την τιμή  $u_0 = 0.2$ .

Μια επιπλέον γραφική μέθοδος που μπορεί να χρησιμοποιηθεί για τον προσδιορισμό του κατώφλιού  $u$  στηρίζεται στην παρατήρηση ότι εάν η κατανομή των υπερβάσεων, για ένα δεδομένο κατώφλι  $u_0$ , είναι η *GPD* αυτή θα συνεχίσει να είναι *GPD* και για τιμές του  $u$  μεγαλύτερες του  $u_0$ . Μάλιστα, η τιμή της παραμέτρου  $\xi$  παραμένει σταθερή, ενώ η παράμετρος κλίμακας  $\tilde{\sigma}(u)$  μεταβάλλεται γραμμικά όπως περιγράφει η ακόλουθη σχέση:

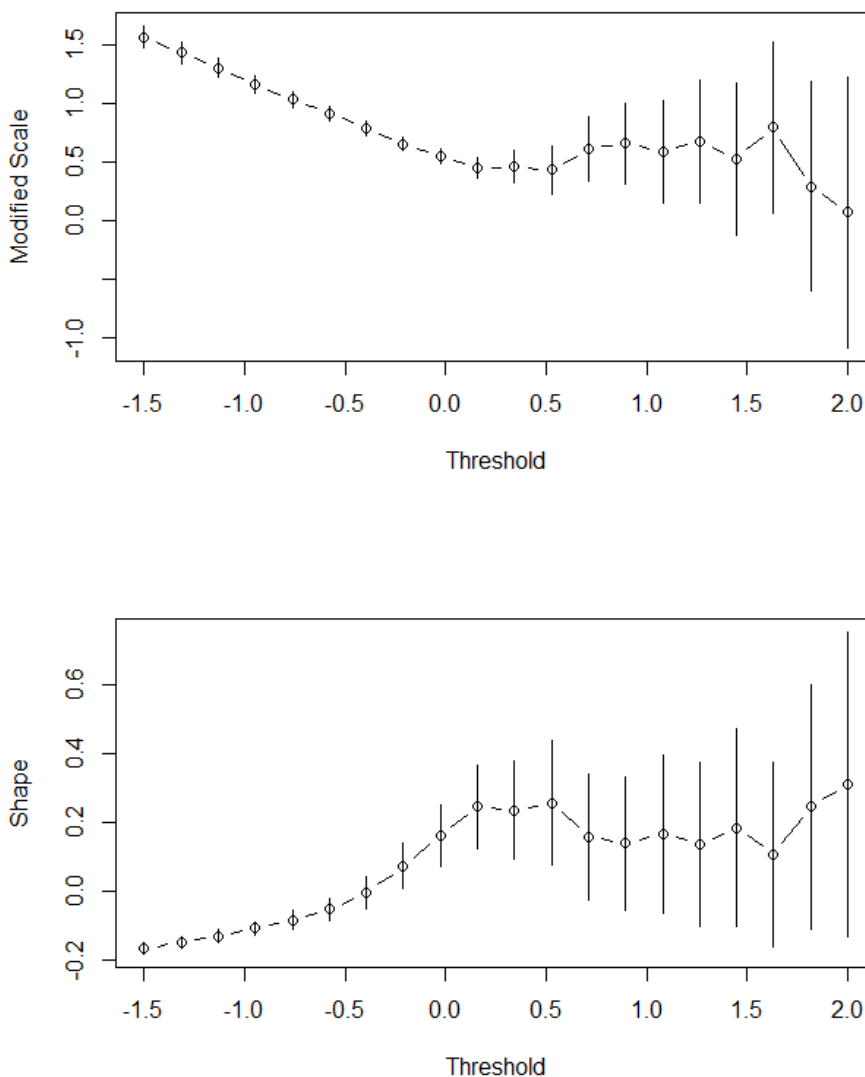
$$\tilde{\sigma}(u) = \tilde{\sigma}(u_0) + \xi(u - u_0), u > u_0. \quad (2.24)$$

Η συγκεκριμένη ισότητα προκύπτει εύκολα παρατηρώντας ότι για  $u = u_0$  ισχύει η σχέση  $\tilde{\sigma}(u_0) = \sigma + \xi(u_0 - \mu)$ , ενώ για  $u > u_0$  ισχύει ότι  $\tilde{\sigma}(u) = \sigma + \xi(u - \mu)$ . Αφαιρώντας κατά μέλη τις δυο αυτές ισότητες, προκύπτει η σχέση (2.24).

Τονίζεται ότι στην περίπτωση που η παράμετρος  $\xi$  ισούται με μηδέν ( $\xi = 0$ ), η ποσότητα  $\tilde{\sigma}(u)$  λαμβάνει τη σταθερή τιμή  $\tilde{\sigma}(u_0)$ . Έτσι, για να μη δημιουργείται σύγχυση ως προς τη συμπεριφορά της εξίσωσης (2.24) ανάλογα της τιμής που λαμβάνει το  $\xi$ , μπορούμε να αναπροσαρμόσουμε τη σχέση αυτή (βλ. 2.24) ως εξής:

$$\sigma^*(u) = \tilde{\sigma}(u) - \xi u = \tilde{\sigma}(u_0) - \xi u_0 = \text{σταθ.}, u > u_0. \quad (2.25)$$

Συνοψίζοντας, η εύρεση του ελάχιστου κατώφλιού  $u_0$  επιτυγχάνεται σχεδιάζοντας τις γραφικές παραστάσεις των σημείων  $(\xi, u)$  και  $(\sigma^*(u), u)$ . Η ελάχιστη τιμή  $u_0$  για την οποία παρατηρείται ότι τα  $\xi$  και  $\sigma^*(u)$  παραμένουν σταθερά όταν το  $u > u_0$ , αποτελεί τη ζητούμενη ποσότητα.



**Σχήμα 2.6:** Στο πρώτο γράφημα φαίνεται η γραφική παράσταση της παραμέτρου κλίμακας  $\sigma^*$  ως προς το κατώφλι  $u$ , ενώ στο δεύτερο φαίνεται η γραφική παράσταση της παραμέτρου σχήματος  $\xi$  ως προς το κατώφλι  $u$ . Η ελάχιστη τιμή

$u_0$  για την οποία παρατηρείται ότι οι ποσότητες  $\xi$  και  $\sigma^*(u)$  παραμένουν σταθερές όταν το  $u > u_0$ , αποτελεί τη ζητούμενη ποσότητα. Τη δεδομένη περίπτωση, μια επιλογή θα μπορούσε να ήταν η τιμή  $u_0 = 0.7$ .

## 2.10 ΜΕΘΟΔΟΣ ΕΚΤΙΜΗΣΗΣ ΠΑΡΑΜΕΤΡΩΝ

Όπως και στην περίπτωση της μεθόδου *Block Maxima*, έτσι και εδώ επιλέχθηκε να χρησιμοποιηθεί η μέθοδος της Μέγιστης Πιθανοφάνειας για την εκτίμηση των παραμέτρων της κατανομής *GPD*.

Έστω  $Y = (Y_1, \dots, Y_n)$  ανεξάρτητες και ισόνομες τυχαίες μεταβλητές με κοινή συνάρτηση κατανομής  $H$  (*GPD*). Για δεδομένο δείγμα  $y = (y_1, \dots, y_n)$  η συνάρτηση πιθανοφάνειας θα δίνεται από τη σχέση:

$$l(\xi, \tilde{\sigma}; y) = \begin{cases} -n \ln \tilde{\sigma} - \left(1 + \frac{1}{\xi}\right) \sum_{i=1}^n \ln \left(1 + \frac{\xi}{\tilde{\sigma}} y_i\right) & , \quad \xi \neq 0 \\ -n \ln \tilde{\sigma} - \frac{1}{\tilde{\sigma}} \sum_{i=1}^n y_i & , \quad \xi = 0 \end{cases} \quad (2.26)$$

υπό τον περιορισμό  $1 + \frac{\xi}{\tilde{\sigma}} y_i > 0$  για κάθε  $i = 1, \dots, n$  όταν το  $\xi \neq 0$ .

Αντικαθιστώντας τις εκτιμήσεις μέγιστης πιθανοφάνειας  $\hat{\xi}$ ,  $\hat{\tilde{\sigma}}$  στην εξίσωση (2.20), προκύπτει εύκολα η εκτίμηση της στάθμης απόδοσης  $x_p$ :

$$\hat{x}_p = \begin{cases} u + \frac{\hat{\tilde{\sigma}}}{\hat{\xi}} \left( (p\hat{w})^{\hat{\xi}} - 1 \right) & , \quad \hat{\xi} \neq 0 \\ u + \hat{\tilde{\sigma}} \ln(p\hat{w}) & , \quad \hat{\xi} = 0 \end{cases} \quad (2.27)$$

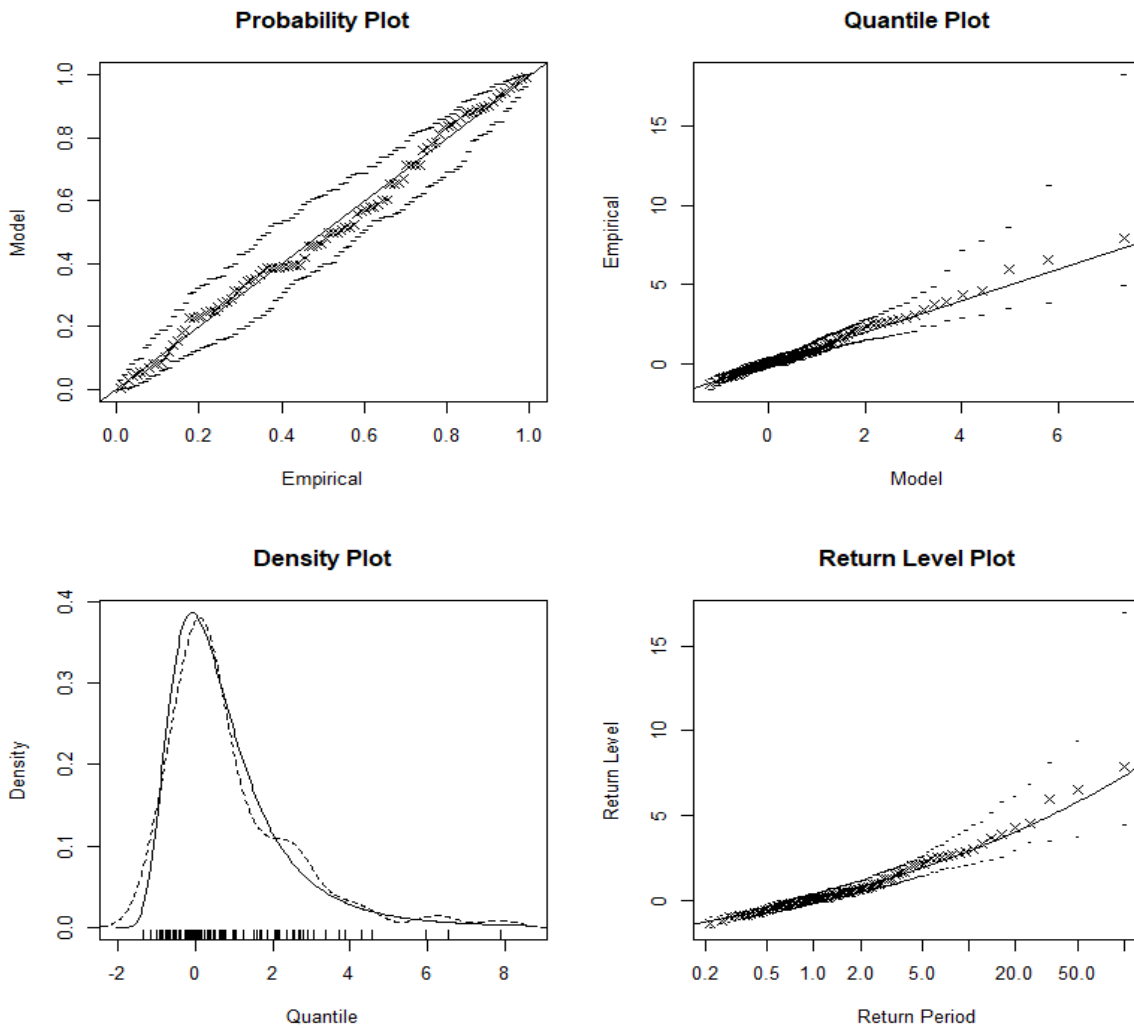
Τονίζεται πως και σε αυτή την περίπτωση, η εύρεση των εκτιμήσεων  $\hat{\xi}$  και  $\hat{\tilde{\sigma}}$  δεν προκύπτει μέσα από κλειστούς τύπους, αλλά μέσω της χρήσης αριθμητικών μεθόδων. Τέλος, διαστήματα εμπιστοσύνης για τις παραμέτρους  $\xi$  και  $\tilde{\sigma}$  μπορούν να κατασκευαστούν για τιμές του  $\xi > -\frac{1}{2}$  καθώς τότε ισχύουν οι ιδιότητες της συνέπειας και επάρκειας των εκτιμητών μέγιστης πιθανοφάνειας.



## 2.11 ΕΛΕΓΧΟΣ ΚΑΛΗΣ ΠΡΟΣΑΡΜΟΓΗΣ

Ο έλεγχος για το κατά πόσο η κατανομή *GPD* προσαρμόζεται ικανοποιητικά στο σύνολο δεδομένων που έχουμε στη διάθεση μας, μπορεί να επιτευχθεί μέσω μιας σειράς γραφικών μεθόδων όπως τα *PP – plot*, *QQ – plot*, το ιστόγραμμα και το γράφημα της στάθμης απόδοσης.

Για τα γραφήματα *PP – plot* και *QQ – plot*, όσο πιο κοντά βρίσκονται τα σημεία στη διαγώνιο (βλ. Κεφ. 2.6), τόσο καλύτερη προσαρμογή, της εκτιμώμενης κατανομής, έχουμε στα δεδομένα. Ανάλογα, στο γράφημα της στάθμης απόδοσης, όσο πιο κοντά βρίσκονται οι εκτιμήσεις  $\hat{x}_p$  (για τις διάφορες τιμές του  $p$ ) στις αντίστοιχες θεωρητικές τιμές  $x_p$ , τόσο καλύτερη είναι η προσαρμογή.



**Σχήμα 2.7:** Γραφικές παραστάσεις, *PP-plot*, *QQ-plot*, συνάρτησης κατανομής πιθανότητας και στάθμης απόδοσης για τον έλεγχο καλής προσαρμογής του μοντέλου.

## ΚΕΦΑΛΑΙΟ 3 ΑΞΙΑ ΣΕ ΚΙΝΔΥΝΟ ΚΑΙ ΥΠΟ-ΣΥΝΘΗΚΗ ΑΞΙΑ ΣΕ ΚΙΝΔΥΝΟ ΑΞΙΟΓΡΑΦΟΥ

### 3.1 ΕΙΣΑΓΩΓΗ

Στο παρόν κεφάλαιο αναλύεται ο τρόπος με τον οποίο οι χρονολογικές σειρές, σε συνδυασμό με τη Θεωρία Ακραίων Τιμών, μπορούν να χρησιμοποιηθούν για τον υπολογισμό της αξίας σε κίνδυνο (**Value at Risk - VaR**) και της υπό συνθήκης αξία σε κίνδυνο (**Tail Value at Risk - TVaR**) ενός αξιογράφου.

Για τον υπολογισμό των μέτρων κινδύνου ενός αξιογράφου, είναι σημαντικό να προσδιοριστεί η κατανομή των δεδομένων που έχουμε στη διάθεση μας. Για το σκοπό αυτό, τρεις μέθοδοι μπορούν να χρησιμοποιηθούν π.χ. βλ. McNeil and Frey (2000):

1. η μη παραμετρική μέθοδος της ιστορικής προσομοίωσης,
2. οι παραμετρικές μέθοδοι που λαμβάνουν υπόψη τους τη μεταβλητότητα της διασποράς των αποδόσεων (όπως μοντέλα *ARCH/GARCH*), υπό την προϋπόθεση ότι οι αποδόσεις των αξιογράφων ακολουθούν κανονική κατανομή· και
3. η μέθοδος ακραίων τιμών.

Η μέθοδος της ιστορικής προσομοίωσης είναι εύκολη στην εφαρμογή της, όμως μειονεκτεί σημαντικά στον υπολογισμό ποσοστημορίων που αφορούν την περίπτωση ακραίων γεγονότων. Αυτό εξηγείται από το γεγονός ότι η εκτιμώμενη κατανομή δεν είναι δυνατόν να επεκταθεί πέραν των ιστορικών παρατηρήσεων.

Από την άλλη, τα μοντέλα που λαμβάνουν υπόψη τους την μεταβλητότητα της διασποράς των αποδόσεων, υιοθετώντας την υπόθεση της κανονικότητας, μειονεκτούν στο γεγονός ότι οι αποδόσεις φαίνεται να προέρχονται από κατανομές με βαρύτερη ουρά σε σχέση με την κανονική (βλ. Κεφ. 1.3).

Τέλος, η μέθοδο ακραίων τιμών, δεν λαμβάνει καθόλου υπόψη η μεταβλητότητα της διασποράς κατά τον υπολογισμό της κατανομής της χρονοσειράς των αποδόσεων.

Τα παραπάνω οδηγούν στο συμπέρασμα ότι ένας συνδυασμός των μεθόδων αυτών, θα μπορούσε να δώσει πιθανών καλύτερα αποτελέσματα, καθώς τα πλεονεκτήματα της

μιας θα μπορούσαν να βελτιώσουν τα μειονεκτήματα της άλλης, λειτουργώντας με αυτό τον τρόπο συμπληρωματικά.

Στην παρούσα διπλωματική εργασία επιλέχθηκε να γίνει ένας συνδυασμός των μοντέλων δεσμευμένης ετεροσκεδαστικότητας (*ARCH/GARCH*) και Θεωρίας Ακραίων Τιμών. Συγκεκριμένα, αίρεται η υπόθεση της κανονικότητας των λογαριθμικών αποδόσεων, υποθέτοντας ότι αυτές προέρχονται από μια άγνωστη κατανομή, της οποίας η ουρά θα προσδιοριστεί χρησιμοποιώντας τη Θεωρία Ακραίων Τιμών.

### 3.2 ΜΕΛΕΤΗ ΑΠΟΔΟΣΗΣ ΑΞΙΟΓΡΑΦΟΥ

Η μελέτη ενός αξιογράφου, εκτός από την παρακολούθηση της τιμής του, μπορεί να περιλαμβάνει και άλλες χρήσιμες ποσοτικές μεταβλητές, όπως αυτή της απόδοσης. Μάλιστα, σε αρκετές οικονομικές μελέτες, δίνεται ιδιαίτερη βαρύτητα στη μελέτη της λογαριθμικής απόδοσης και όχι της τιμής του αξιογράφου π.χ. βλ. Tsay, R.S. (2002). Αυτό αποδίδεται στο γεγονός ότι μια χρονοσειρά λογαριθμικών αποδόσεων, έχει καλύτερες στατιστικές ιδιότητες σε σχέση με αυτή της χρονοσειράς τιμών. Επίσης, η χρονοσειρά λογαριθμικών αποδόσεων έχει το πλεονέκτημα να είναι απαλλαγμένη από μια μονάδα μέτρησης, κάνοντας ευκολότερη τη σύγκριση διαφορετικών αξιογράφων.

Έστω  $S_{t-1}$  η τιμή ενός αξιογράφου τη χρονική στιγμή  $t - 1$ . Αν η απόδοση μετά το πέρας μια χρονικής περιόδου είναι  $R_t$ , τότε η τιμή, την χρονική στιγμή  $t$ , θα ισούται με:

$$S_t = S_{t-1}(1 + R_t) \quad (3.01)$$

ή ισοδύναμα,

$$1 + R_t = \frac{S_t}{S_{t-1}}. \quad (3.02)$$

Αν στην εξίσωση (βλ. 3.02) εφαρμόσουμε το φυσικό λογάριθμο και στα δύο μέλη, τότε η λογαριθμική απόδοση, που συμβολίζεται με  $r_t$ , ισούται με:

$$r_t = \ln(1 + R_t) = \ln \frac{S_t}{S_{t-1}}. \quad (3.03)$$

Σύμφωνα με το απλούστερο μοντέλο της αγοράς, η χρονοσειρά των λογαριθμικών αποδόσεων αποτελείται από ανεξάρτητες και ισόνομες τυχαίες μεταβλητές, που ακολουθούν κανονική κατανομή  $N(\mu, \sigma^2)$  (ή διαφορετικά, οι λογαριθμικές τιμές σχηματίζουν μια χρονοσειρά τυχαίου περιπάτου):

$$r_t = \ln S_t - \ln S_{t-1} = \mu + e_t, \text{ όπου } (e_t, t \in T) \sim N(0, \sigma^2). \quad (3.04)$$

Βέβαια τα ιστορικά δεδομένα (βλ. Κεφ. 1.3) αποδεικνύουν ότι το παραπάνω μοντέλο εμφανίζει αρκετές ασυνέπειες, σε σχέση με την πραγματικότητα. Συγκεκριμένα, οι τυχαίες μεταβλητές  $r_t$  φαίνεται να προέρχονται από κατανομές που έχουν βαρύτερη ουρά από την κανονική, είναι εξαρτημένες και η διασπορά τους μεταβάλλεται με την πάροδο του χρόνου. Επομένως, μια κοινή πρακτική που ακολουθείται στη μοντελοποίηση των λογαριθμικών αποδόσεων, είναι η χρήση μοντέλων *ARCH* και *GARCH* τα οποία μπορούν να ενσωματώσουν την μεταβλητότητα της διασποράς. Αναλυτικότερα η μέθοδος που μπορεί να υιοθετηθεί περιγράφεται ως ακολούθως.

Έστω ότι οι καθημερινές αποδόσεις ενός αξιογράφου εκφράζονται μέσω μιας χρονοσειράς  $r = (r_t, t \in \mathbb{Z})$ , που αποτελείται από έναν όρο  $\mu_t$  ο οποίος εκφράζει την εξέλιξη του μέσου, και έναν όρο  $e_t$  ο οποίος περιγράφει τα σφάλματα και θεωρείται μια χρονοσειρά λευκού θορύβου:

$$r_t = \mu_t + e_t. \quad (3.05)$$

Συμβολίζοντας με  $\mathcal{F}_{t-1}$  την διαθέσιμη πληροφορία τη χρονική στιγμή  $t - 1$ , η  $\mu_t = E[r_t | \mathcal{F}_{t-1}]$  μπορεί να θεωρηθεί ως ο μέσος μιας χρονοσειράς τύπου *ARMA*( $p, q$ ), ενώ ο όρος  $e_t$  ως μια χρονοσειρά τύπου *GARCH*( $p', q'$ ), η οποία μπορεί να πάρει τη μορφή:

$$e_t = \sigma_t \varepsilon_t \text{ με } \varepsilon_t \sim IID(0,1). \quad (3.06)$$

Σημειώνεται ότι καθώς η κατανομή των λογαριθμικών αποδόσεων έχει βαριά ουρά, αποφεύγουμε στο σημείο αυτό να θεωρήσουμε μια εκ των προτέρων κατανομή για τις τυχαίες μεταβλητές  $\varepsilon_t$  (για παράδειγμα μια *Student*( $n$ ) κατανομή). Αυτό αν και αποτελεί έναν περιοριστικό παράγοντα για την εκτίμηση των παραμέτρων του μοντέλου, η

εφαρμογή της μεθόδου Quasi-Maximum Likelihood στη χρονοσειρά  $(e_t, t \in \mathbb{Z})$ , παρέχει συνεπείς εκτιμητές για όλες τις παραμέτρους του μοντέλου (βλ. Κεφ. 1.8 και 1.13).

Στη συνέχεια, αφού γίνει η προσαρμογή του μοντέλου και υπολογιστούν τα κατάλοιπα  $\hat{\varepsilon}_t$ , μπορεί να χρησιμοποιηθεί η Θεωρία Ακραίων Τιμών για την μοντελοποίηση της ουράς της άγνωστης κατανομής των  $\varepsilon_t$ .

Αναφορικά με την κατανομή των λογαριθμικών αποδόσεων, δύο είναι οι περιπτώσεις που χρίζουν ιδιαίτερης σημασίας ως προς την πληροφορία που παρέχουν. Μια είναι η δεσμευμένη, ως προς την υπάρχουσα πληροφορία  $(\mathcal{F}_{t-1})$ , κατανομή των αποδόσεων και η δεύτερη αφορά την κατανομή της στάσιμης χρονοσειράς  $r_t$ . Κάθε μια από τις παραπάνω κατανομές χρησιμοποιείται για να δοθούν απαντήσεις σε διαφορετικού τύπου ερωτήματα. Έτσι, για να απαντηθεί το ερώτημα που αφορά το ποια θα είναι η πιθανή ζημιά ενός δεδομένου χαρτοφυλακίου αξιογράφων την επόμενη ημέρα (ή τις επόμενες ημέρες), λαμβάνοντας υπόψη την υπάρχουσα μεταβλητότητα των αποδόσεων, θα πρέπει να είναι γνωστή η δεσμευμένη κατανομή των αποδόσεων. Από την άλλη, για να απαντηθεί το ερώτημα του ποια θα είναι η πιθανή ημερήσια απώλεια για ένα ακραίο γεγονός το οποίο συμβαίνει κατά μέσο όρο μετά από  $k$ -έτη, απαιτείται η γνώση της (μη δεσμευμένης) κατανομής της χρονοσειράς των αποδόσεων.

Σημειώνεται ότι στην παρούσα διπλωματική εργασία θα ασχοληθούμε μόνο με την δεσμευμένη κατανομή των αποδόσεων.

### 3.3 ΑΞΙΑ ΣΕ ΚΙΝΔΥΝΟ (Value at Risk - VaR) ΚΑΙ ΥΠΟ ΣΥΝΘΗΚΗ ΑΞΙΑ ΣΕ ΚΙΝΔΥΝΟ (Tail Value at Risk - TVaR)

Η εύρεση της αξίας σε κίνδυνο (**Value at Risk-VAR**), από μαθηματικής άποψης, δεν είναι τίποτα περισσότερο από τον προσδιορισμό ενός συγκεκριμένου ποσοστιαίου σημείου μιας συνάρτησης κατανομής, για δεδομένη πιθανότητα.

Από την πλευρά ενός χρηματοοικονομικού οργανισμού η αξία σε κίνδυνο, μπορεί να ερμηνευθεί ως η μέγιστη απώλεια της χρηματοοικονομικής θέσης της επιχείρησης, για ένα δεδομένο χρονικό ορίζοντα και για δεδομένη πιθανότητα.

Συχνά, ο προσδιορισμός του  $VaR$  μπορεί να μην είναι τόσο διαφωτιστικός για ένα χρηματοοικονομικό οργανισμό, υπό την έννοια ότι πολλές φορές το ενδιαφέρον δεν

εστιάζεται στο ποια θα είναι η μέγιστη ζημιά την οποία μπορεί να υποστεί, αλλά το πόσο (κατά μέσο όρο) μεγάλη θα είναι η απώλεια αυτή, στην περίπτωση που πραγματοποιηθεί το ακραίο γεγονός. Έτσι, ένα «καλύτερο» μέτρο που προσδιορίζει την ποσότητα αυτή, είναι η υπό-συνθήκη αξία σε κίνδυνο (**Tail Value at Risk-TVaR**). Συγκεκριμένα, αν με  $L$  συμβολίσουμε την τυχαία μεταβλητή που εκφράζει τη ζημιά ενός χρηματοοικονομικού οργανισμού, τότε το  $TVaR$  εκφράζεται μέσω της σχέσης  $E[L|L > VaR]$ .

Τονίζεται ότι, τόσο το  $VaR$  όσο και το  $TVaR$  σχετίζεται άμεσα με τις ακραίες τιμές της συνάρτησης κατανομής. Επομένως, ο προσδιορισμός και γνώση της ουράς αυτής κρίνεται απαραίτητο.

Έστω  $F_r$  η κατανομή της χρονοσειράς των αποδόσεων ( $r_t$ ) και  $F_{r_{t+1}|\mathcal{F}_t}$  η κατανομή δεδομένης της πληροφορίας  $\mathcal{F}_t$ . Τότε, τα μέτρα κινδύνου  $VaR$  και  $TVaR$ , περιγράφονται από τις εξής σχέσεις:

1. Αν είναι γνωστή η  $F_r$

$$\bullet \quad VaR(q) = \inf\{x \in \mathbb{R} : F_r(x) \geq q\} \quad (3.07)$$

$$\bullet \quad TVaR(q) = E[r|r > VaR(q)] \quad (3.08)$$

2. Αν είναι γνωστή η  $F_{r_{t+1}|\mathcal{F}_t}$

$$\bullet \quad VaR(q) = \inf\{x \in \mathbb{R} : F_{r_{t+1}|\mathcal{F}_t}(x) \geq q\} \quad (3.09)$$

$$\bullet \quad TVaR(q) = E[r_{t+1}|r_{t+1} > VaR(q), \mathcal{F}_t] \quad (3.10)$$

Καθώς στην παρούσα διπλωματική εργασία θα ασχοληθούμε μόνο με την περίπτωση που είναι γνωστή η υπό συνθήκη κατανομή, χρησιμοποιώντας το γεγονός ότι:

$$\begin{aligned} F_{r_{t+1}|\mathcal{F}_t}(x) &= P[\sigma_{t+1}\varepsilon_{t+1} + \mu_{t+1} \leq x|\mathcal{F}_t] \\ &= F_\varepsilon\left(\frac{x - \mu_{t+1}}{\sigma_{t+1}}\right). \end{aligned} \quad (3.11)$$

οι σχέσεις (βλ. 3.09 και 3.10) μπορούν να πάρουν τη μορφή:

$$\bullet \quad VaR(q) = \mu_{t+1} + \sigma_{t+1}\varepsilon_q \quad (3.12)$$

- $$TVaR(q) = \mu_{t+1} + \sigma_{t+1}E[\varepsilon | \varepsilon > \varepsilon_q] \quad (3.13)$$

όπου το  $\varepsilon_q$  είναι το  $q$ -ποσοστιαίο σημείο της χρονοσειράς των σφαλμάτων  $\varepsilon = (\varepsilon_t, t \in \mathbb{N})$  που εκτιμάται μέσω των σχέσεων 2.14 ή 2.27, ανάλογα με το ποια μέθοδο χρησιμοποιείται (βλ. Κεφ. 2).

## ΚΕΦΑΛΑΙΟ 4 ΕΦΑΡΜΟΓΗ

### 4.1 ΕΙΣΑΓΩΓΗ

Στο παρόν κεφάλαιο παρουσιάζεται η εφαρμογή των μοντέλων *GARCH* σε συνδυασμό με τη Θεωρία Ακραίων Τιμών, χρησιμοποιώντας δεδομένα χρηματοοικονομικών σειρών.

Αρχικά, οι μέθοδοι εφαρμόζονται σε ένα σύνολο από ημερήσια δεδομένα του κρυπτονομίσματος *Bitcoin*, με σκοπό την εκτίμηση του *VaR* και *TVaR* της απόδοσης του, τη χρονική στιγμή  $t + 1$ . Για τον έλεγχο της αποτελεσματικότητας της μεθοδολογίας, πραγματοποιείται ένας εκ των υστέρων έλεγχος (***back-testing***), συγκρίνοντας τα αποτελέσματα της μεθόδου που εφαρμόζουμε σε σχέση με άλλες τεχνικές εύρεσης της αξίας σε κίνδυνο (βλ. Κεφ. 4.3).

Για το σκοπό της εφαρμογής, κρίθηκε απαραίτητη η χρήση του στατιστικού πακέτου *R*. Συγκεκριμένα, δημιουργήθηκαν δύο κώδικες *R* (ΠΑΡΑΡΤΗΜΑ Ι) για την διαχείριση των δεδομένων, την εφαρμογή των μεθόδων και τον έλεγχο της μεθοδολογίας.

### 4.2 ΠΡΟΒΛΕΨΗ *VaR* ΚΑΙ *TVaR*

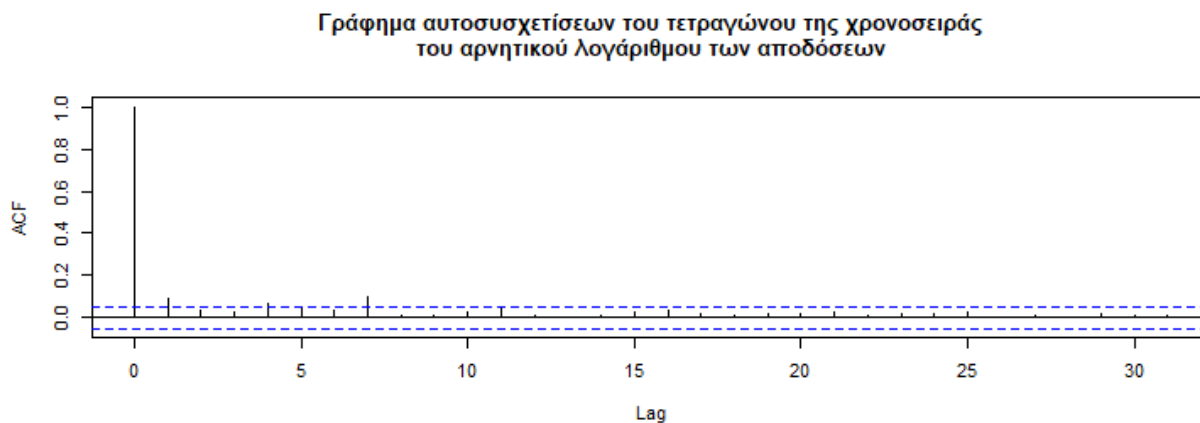
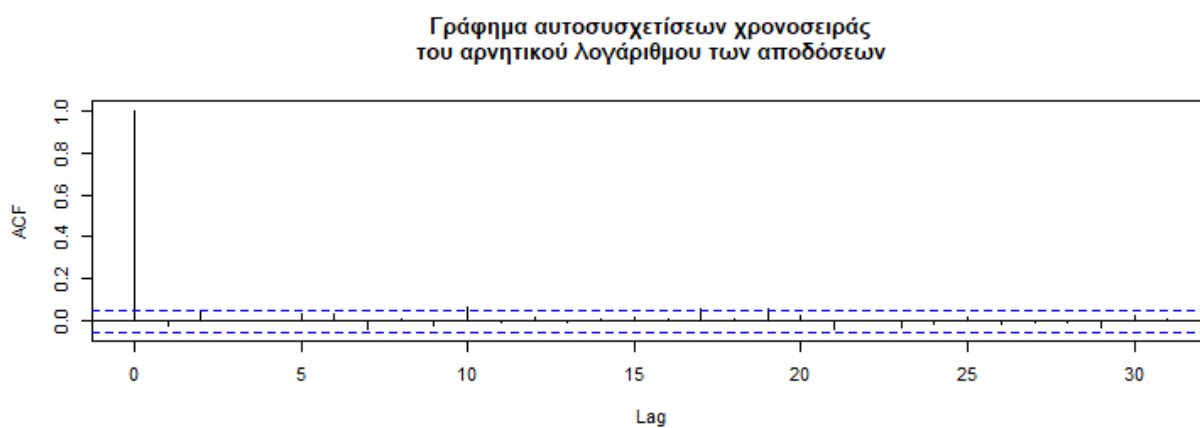
Η εφαρμογή της μεθοδολογίας (βλ. Κεφ. 3) ενδείκνυται για χρηματοοικονομικές σειρές που εμφανίζουν έντονες διακυμάνσεις. Μια τέτοια χαρακτηριστική περίπτωση, μεταξύ άλλων, αποτελούν τα κρυπτονομίσματα. Συγκεκριμένα, η μεθοδολογία εφαρμόζεται για την περίπτωση του *Bitcoin*, με τα δεδομένα των ημερήσιων τιμών να έχουν ληφθεί από την ιστοσελίδα «<https://finance.yahoo.com/>» και αφορούν την περίοδο 01-01-2017 έως 03-01-2021.





**Σχήμα 4.1:** Γραφική παράσταση ημερήσιων τιμών Bitcoin για το χρονικό διάστημα 01-01-2017 έως 03-01-2021.

Αρχικά, ελέγχεται η καταλληλότητα των δεδομένων της χρονοσειράς των αποδόσεων  $r_t$ . Συγκεκριμένα, μέσω της γραφικής παράστασης των τιμών και των αποδόσεων της  $r_t$  (βλ. σχήμα 4.1 και 4.2), εξετάζονται φαινόμενα ύπαρξης τάσης και περιοδικότητας, όπως και η μεταβλητότητα της διασποράς των αποδόσεων. Επίσης, μέσω των γραφικών παραστάσεων των αυτοσυσχετίσεων των χρονοσειρών  $r_t$  και  $r_t^2$ , δίνεται μια πρώτη αίσθηση της σχέσης εξάρτησης που εμφανίζουν οι όροι μεταξύ τους, καθώς και το είδος των υποψήφιων μοντέλων που πρέπει να ληφθούν υπόψη για την περιγραφή των δεδομένων. Επιπλέον, εκτελούνται διαγνωστικοί έλεγχοι (*Ljung – Box test*) με σκοπό τη στατιστική επιβεβαίωση των οπτικών παρατηρήσεων.



**Σχήμα 4.2:** Στο πρώτο γράφημα φαίνεται η εξέλιξη του αρνητικού λογάριθμου των αποδόσεων, ενώ στο δεύτερο και τρίτο σχεδιάζονται οι αυτοσυσχετίσεις της χρονοσειράς του αρνητικού λογάριθμου και του τετραγώνου αυτού.

**Αποτελέσματα Ljung-Box ελέγχου σε ε.σ. 5%**

Χρονοσειρά	X-squared	df	p-value
Χρονοσειρά αρνητικού λογάριθμου	17	15	0.31891
Χρονοσειρά τετραγώνου του αρνητικού λογάριθμου	46.81	15	$3.6 \cdot 10^{-5}$

Οι γραφικές παραστάσεις (βλ. σχήμα 4.2) αποτελούν χαρακτηριστική περίπτωση ύπαρξης φαινομένου ετεροσκεδαστικότητας στα δεδομένα και κατ' επέκταση τη χρήση *GARCH* υποδειγμάτων. Συγκεκριμένα, από το γράφημα αυτοσυσχέτισης των αρνητικών λογαριθμικών αποδόσεων, παρατηρούμε έλλειψη (γραμμικής) συσχέτισης μεταξύ αυτών. Χωρίς βέβαια να προβούμε σε περαιτέρω ελέγχους, θα μπορούσε κάποιος, λανθασμένα, να εξαγάγει και μια σχέση ανεξαρτησίας για τους όρους της  $r_t$ . Αντιθέτως, παρατηρώντας τη γραφική παράσταση των τετραγώνων των αρνητικών λογαριθμικών αποδόσεων, οδηγούμαστε στο συμπέρασμα ότι οι όροι της χρονοσειράς  $r_t$  είναι εξαρτημένοι. Μάλιστα, αυτό ενισχύεται και από το αποτέλεσμα του *Ljung – Box* ελέγχου, όπου για την περίπτωση της χρονοσειράς  $r_t$ , η υπόθεση της γραμμικής ανεξαρτησίας δεν απορρίπτεται ( $p - value > 0.05$ ), ενώ για την περίπτωση του τετραγώνου του αρνητικού λογαρίθμου, ο έλεγχος αποδεικνύεται στατιστικά σημαντικός ( $p - value < 0.05$ ).

Τα παραπάνω, οδηγούν στο συμπέρασμα ότι ένα γενικό μοντέλο της μορφής

$$r_t = d_t + \mu_t \quad (4.01)$$

όπου:

$$\begin{aligned} d_t &= \gamma t, \quad \gamma \neq 0 \\ \mu_t &= \sum_{i=1}^2 a_i \mu_{t-i} + \sum_{j=1}^2 b_j \varepsilon_{t-j} + \varepsilon_t \\ \varepsilon_t &= \sigma_t e_t \\ \sigma_t^2 &= c_0 + \sum_{i=1}^2 c_i \varepsilon_{t-i}^2 + \sum_{j=1}^2 g_j \sigma_{t-j}^2 \end{aligned} \quad (4.02)$$

θα μπορούσε να εφαρμοστεί στα δεδομένα.

Συγκεκριμένα, μέσω του αλγορίθμου που αναπτύχθηκε (βλ. ΠΑΡΑΡΤΗΜΑ Ι) χρησιμοποιώντας τη συνάρτηση *predict\_VaR*, ελέγχονται τα εξής:

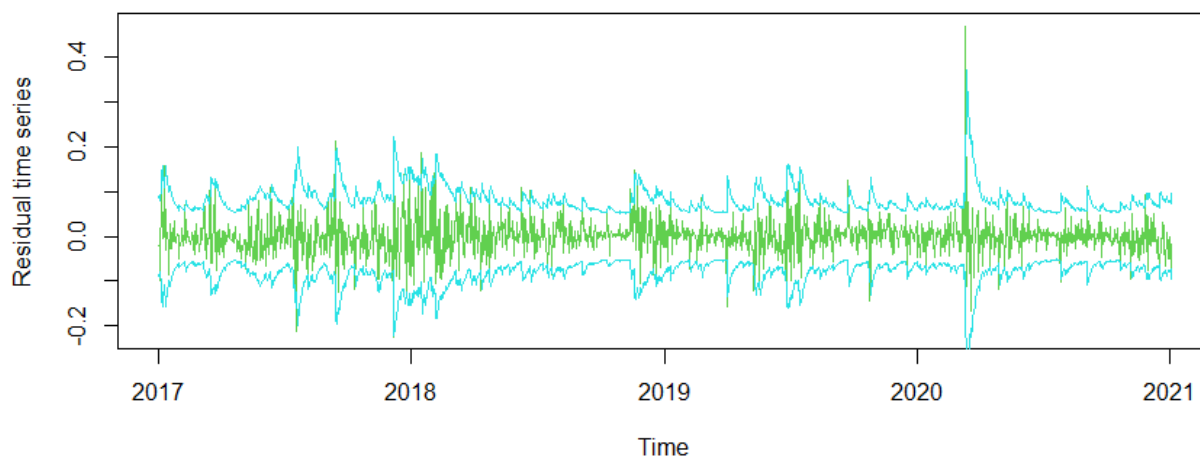
1. Ύπαρξη φαινομένου τάσης στα δεδομένα, δηλαδή αν ο όρος  $d_t$  είναι σημαντικός.
2. Έλεγχος αν κάποιο από τα μοντέλα  $ARMA(p, q)$  με  $p, q \in \{0, 1, 2\}$ , πρέπει να προσαρμοστεί στα δεδομένα, με σκοπό την περιγραφή της εξέλιξης της χρονοσειράς του μέσου  $\mu_t$ .

3. Έλεγχος μεταξύ των μοντέλων  $GARCH(p', q')$  με  $p', q' \in \{1, 2\}$ , ώστε να βρεθεί αυτό που προσαρμόζεται καλύτερα στην χρονοσειρά των καταλοίπων  $\hat{\epsilon}_t$ .

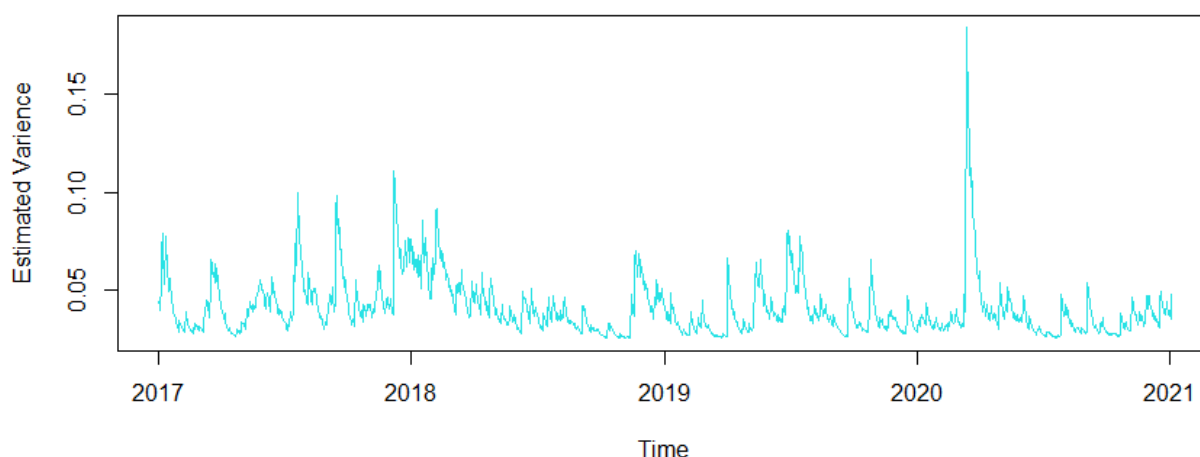
Από την παραπάνω ανάλυση, προκύπτει ότι η χρονοσειρά  $r_t$  περιγράφεται «καλύτερα» μέσω ενός υποδείγματος  $GARCH(1,1)$ . Επομένως:

$$\begin{aligned} r_t &= \varepsilon_t \\ \varepsilon_t &= \sigma_t e_t \\ \sigma_t^2 &= c_0 + c_1 \varepsilon_{t-1}^2 + g_1 \sigma_{t-1}^2 \end{aligned} \quad (4.03)$$

**Γραφική παράσταση χρονοσειράς για την οποία εφαρμόζεται ένα μοντέλο GARCH σε συνδυασμό με την εκτιμώμενη (υπό συνθήκη) διασπορά**



**Γραφική παράσταση εκτιμώμενης (υπό συνθήκη) διασποράς**



**Σχήμα 4.3:** Στο πρώτο γράφημα απεικονίζεται η γραφική παράσταση των καταλοίπων  $r_t$  σε συνδυασμό με το διπλάσιο ( $\pm 2$ ) της εκτιμώμενης υπό-συνθήκη διασποράς, όπως προκύπτει μετά την εφαρμογή ενός  $GARCH(1,1)$  μοντέλου. Αντίστοιχα, στο δεύτερο γράφημα απεικονίζονται οι πραγματικές τιμές της εκτιμώμενης υπό-συνθήκη διασποράς.

Στη συνέχεια, εξετάζοντας τόσο τη χρονοσειρά των καταλοίπων  $\hat{e}_t$ , όσο και την χρονοσειρά των τετραγώνων των καταλοίπων  $\hat{e}_t^2$  μέσω του ελέγχου *Ljung – Box*, οδηγούμαστε στο συμπέρασμα ότι η  $e_t$  αποτελεί θόρυβο (**noise**).

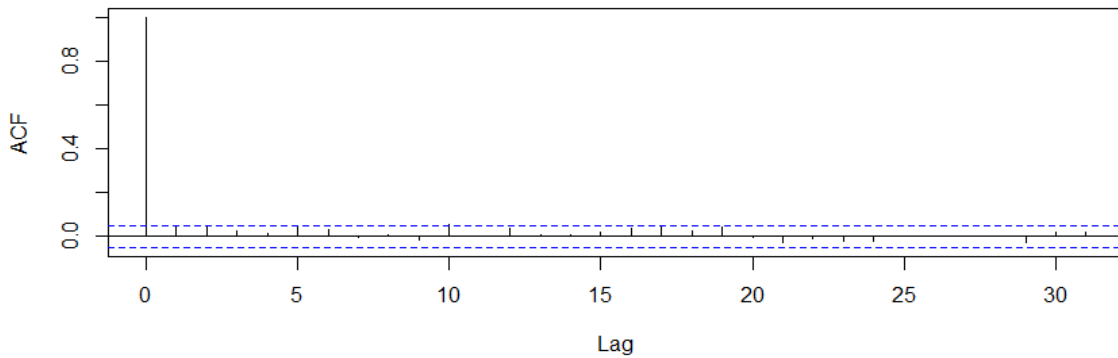
$$\varepsilon_t = \sigma_t e_t, \text{ όπου } e_t \sim IID(0,1) \quad (4.04)$$

**Αποτελέσματα Ljung-Box ελέγχου σε ε.σ. 5%**

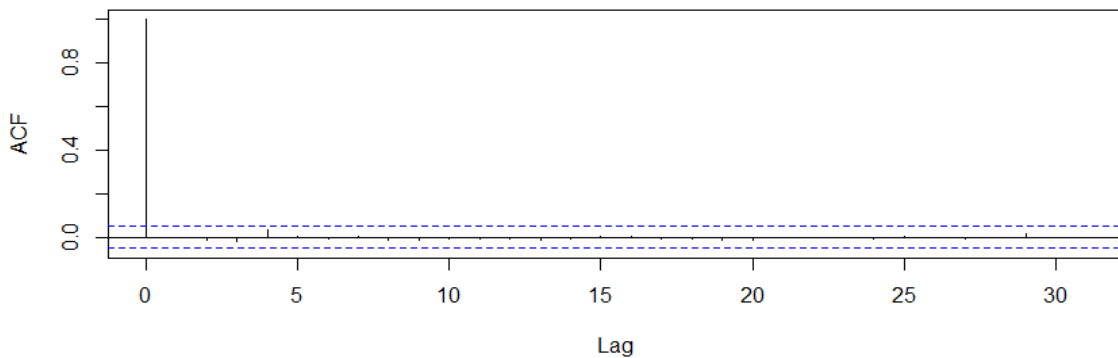
Χρονοσειρά	X-squared	df	p-value
Χρονοσειρά καταλοίπων $\hat{e}_t$	19.63	15	0.1866
Χρονοσειρά τετραγώνου των καταλοίπων $\hat{e}_t^2$	3.62	15	0.9987

Αντίστοιχα συμπεράσματα, εξάγονται ελέγχοντας τα διαγράμματα αυτοσυσχετίσεων των χρονοσειρών  $\hat{e}_t$  και  $\hat{e}_t^2$ .

**Γράφημα αυτοσυσχετίσεων χρονοσειράς καταλοίπων αφού έχει εφαρμοστεί ένα μοντέλο GARCH**



**Γράφημα αυτοσυσχετίσεων χρονοσειράς του τετραγώνου των καταλοίπων αφού έχει εφαρμοστεί ένα μοντέλο GARCH**



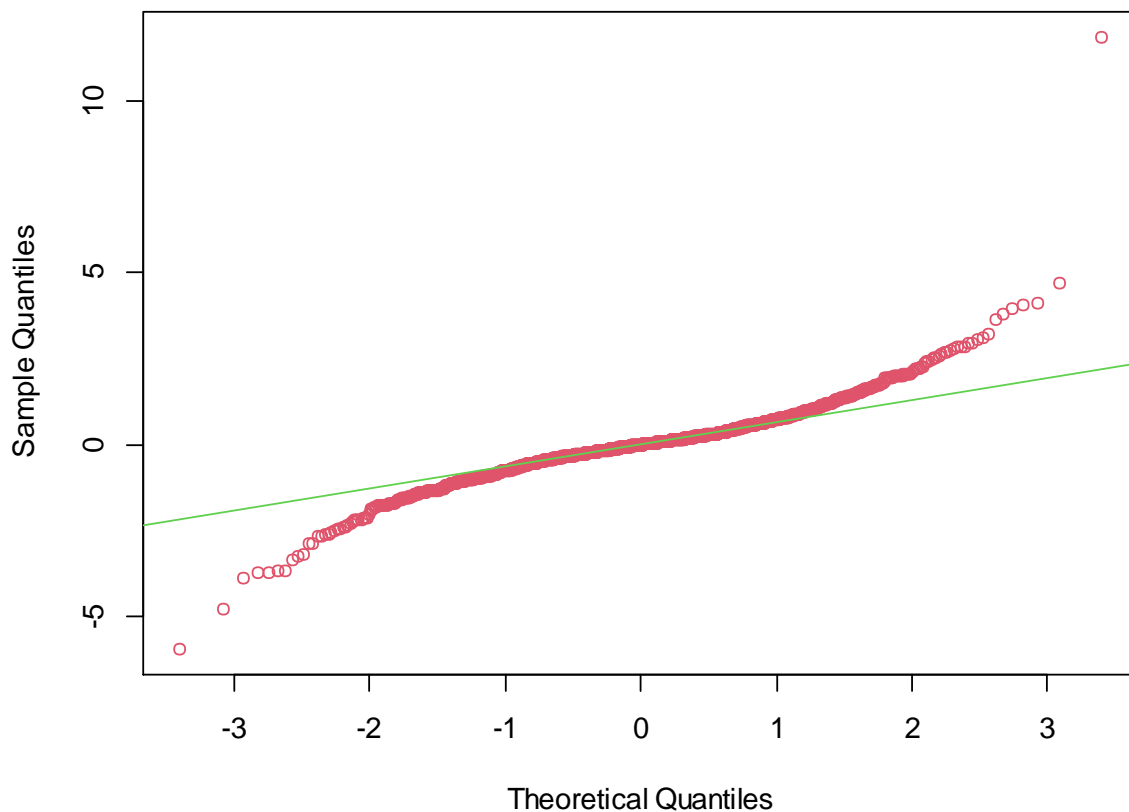
**Σχήμα 4.4:** Γραφικές παραστάσεις των αυτοσυσχετίσεων των χρονοσειρών  $\hat{e}_t$  και  $\hat{e}_t^2$ , αντίστοιχα.

Όπως έχει ήδη αναφερθεί (βλ. Κεφ. 1.3), η φύση των χρηματοοικονομικών σειρών επιβάλλει τα δεδομένα να προέρχονται από μια κατανομή με βαριά ουρά, και όχι συνήθως από την κανονική, όπως για παράδειγμα υποθέτει το απλούστερο μοντέλο τιμολόγησης. Συγκεκριμένα, τόσο ο γραφικός, όσο και ο στατιστικός έλεγχος της χρονοσειράς των καταλοίπων  $\hat{\epsilon}_t$ , απορρίπτουν την υπόθεση της κανονικότητας.

**Έλεγχος κανονικότητας ε.σ. 5%**

Χρονοσειρά $\hat{\epsilon}_t$	Ελεγκοσυνάρτηση	p-value
Shapiro-Wilk	0.90	$7.48 \cdot 10^{-30}$
Kolmogorov-Smirnov	0.09	$2.79 \cdot 10^{-11}$

**Normal Q-Q Plot**



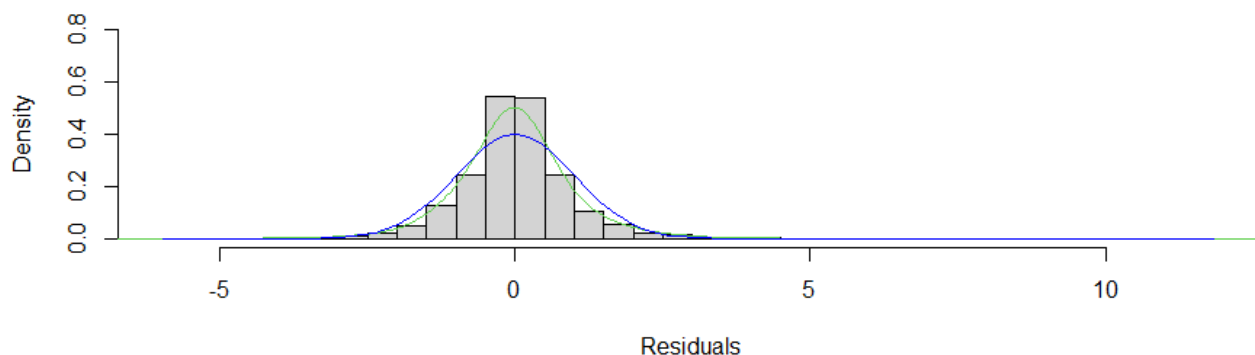
**Σχήμα 4.5:** Γραφικός έλεγχος κανονικότητας χρονοσειράς  $\hat{\epsilon}_t$ . Η απόκλιση των σημείων από την ευθεία (πράσινη) γραμμή, υποδηλώνει την μη κανονικότητα της κατανομής των  $\hat{\epsilon}_t$ .

Μάλιστα, όπως διακρίνεται στα γραφήματα (βλ. 4.6) και σε συνδυασμό με τις τιμές που λαμβάνουν τα αριθμητικά μέτρα λοξότητας και κυρτότητας, η κατανομή της  $\hat{\epsilon}_t$  είναι θετικά ασύμμετρη ( $s = 1.1 > 0$ ) και λεπτόκυρτη ( $k = 16.25 > 3$ ).

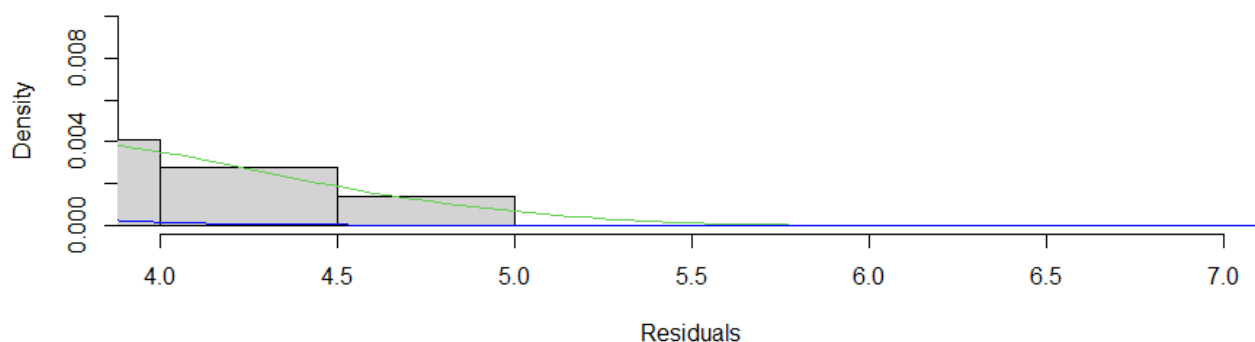
### Μορφή κατανομής χρονοσειράς $\hat{e}_t$

Αριθμητικά μέτρα	Τιμή	Χαρακτηρισμός
Λοξότητα	1.1	Θετική ασυμμετρία
Κυρτότητα	16.25	Λεπτόκυρτη

### Ιστόγραμμα χρονοσειράς καταλοίπων και σύγκριση με την κανονική κατανομή



### Ιστόγραμμα χρονοσειράς καταλοίπων και σύγκριση με την κανονική κατανομή δίνοντας έμφαση στην ουρά της κατανομής



**Σχήμα 4.6:** Σύγκριση της εκτιμώμενης κατανομής της χρονοσειράς  $e_t$  (πράσινη γραμμή) σε σχέση με την κανονική κατανομή (μπλε γραμμή). Όπως απεικονίζεται στο πρώτο γράφημα, η κατανομή των  $e_t$  φαίνεται να είναι λεπτόκυρτη και επεκτείνεται προς τα δεξιά (θετικά ασύμμετρη). Μάλιστα, το γεγονός ότι η ουρά της κατανομής των  $e_t$  φαίνεται να είναι βαρύτερη από αυτή της κανονικής, διακρίνεται καλύτερα στο δεύτερο γράφημα.

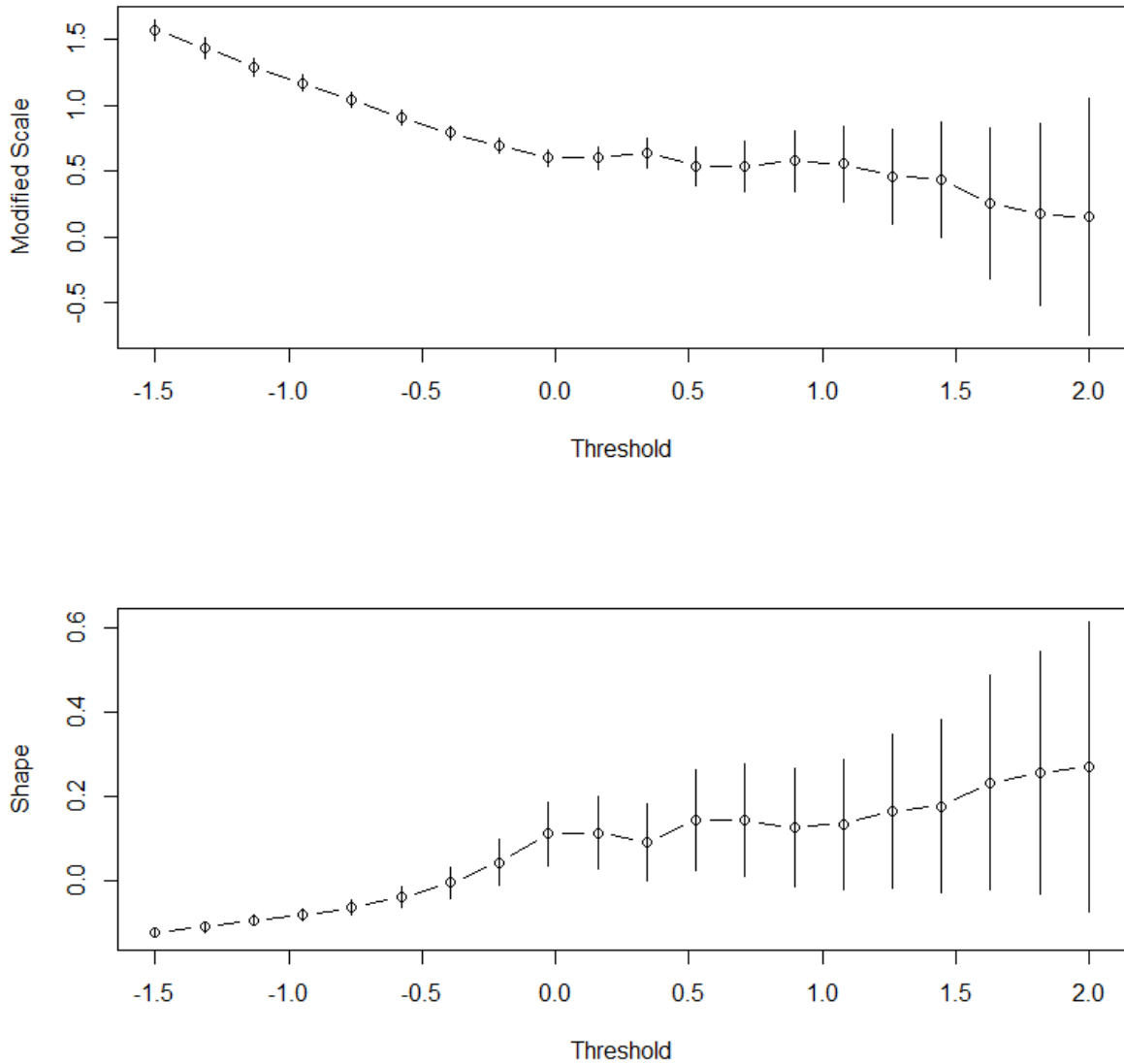
Καθώς το ενδιαφέρον εστιάζεται στην πρόβλεψη της μέγιστης απώλειας που μπορεί να υποστεί ένας επενδυτής τη χρονική στιγμή  $t + 1$ , για δεδομένη πιθανότητα, μελετάται η μορφή της δεξιάς ουράς της συνάρτησης κατανομής των σφαλμάτων  $e_t$ , χρησιμοποιώντας τη Θεωρία Ακραίων Τιμών. Συγκεκριμένα, για τη στατιστική ανάλυση των δεδομένων χρησιμοποιείται η μέθοδος *POT*, η οποία επιβάλλει ότι οι υπερβάσεις των καταλοίπων  $\hat{e}_t$ , πάνω από ένα κατώφλι  $u$ , θα πρέπει να ακολουθούν την κατανομή *GPD*. Αρχικά από τα διαγράμματα των παραμέτρου κλίμακας  $\sigma$  και σχήματος  $\xi$  (βλ. σχήμα 4.8), παρατηρούμε ότι για  $u \approx 0.5$ , οι γραφικές παραστάσεις σταθεροποιούνται. Επομένως, το

κατώφλι που χρησιμοποιείται για την εκτίμηση των παραμέτρων της συνάρτησης κατανομής *GPD*, επιλέχθηκε να είναι το  $u = 0.5$ .

**Γραφική παράσταση χρονοσειράς καταλοίπων για τα οποία εφαρμόζεται η EVT**



**Σχήμα 4.7:** Γραφική παράσταση των τιμών της χρονοσειράς  $\hat{\epsilon}_t$ , για την οποία θα εφαρμοστεί η Θεωρία Ακραίων Τιμών.



**Σχήμα 4.8:** Γραφικές παραστάσεις των παραμέτρων κλίμακας  $\sigma$  και  $\xi$  ως προς το κατώφλι  $u$ . Σαν ελάχιστο κατώφλι για το οποίο οι γραφικές παραστάσεις φαίνεται να σταθεροποιούνται επιλέγεται η τιμή  $u = 0.5$ .



Οι εκτιμήσεις των παραμέτρων κλίμακας και σχήματος της κατανομής *GPD*, όπως προέκυψαν μετά την εφαρμογή της μεθόδου *MLE*, δίνονται στον παρακάτω πίνακα:

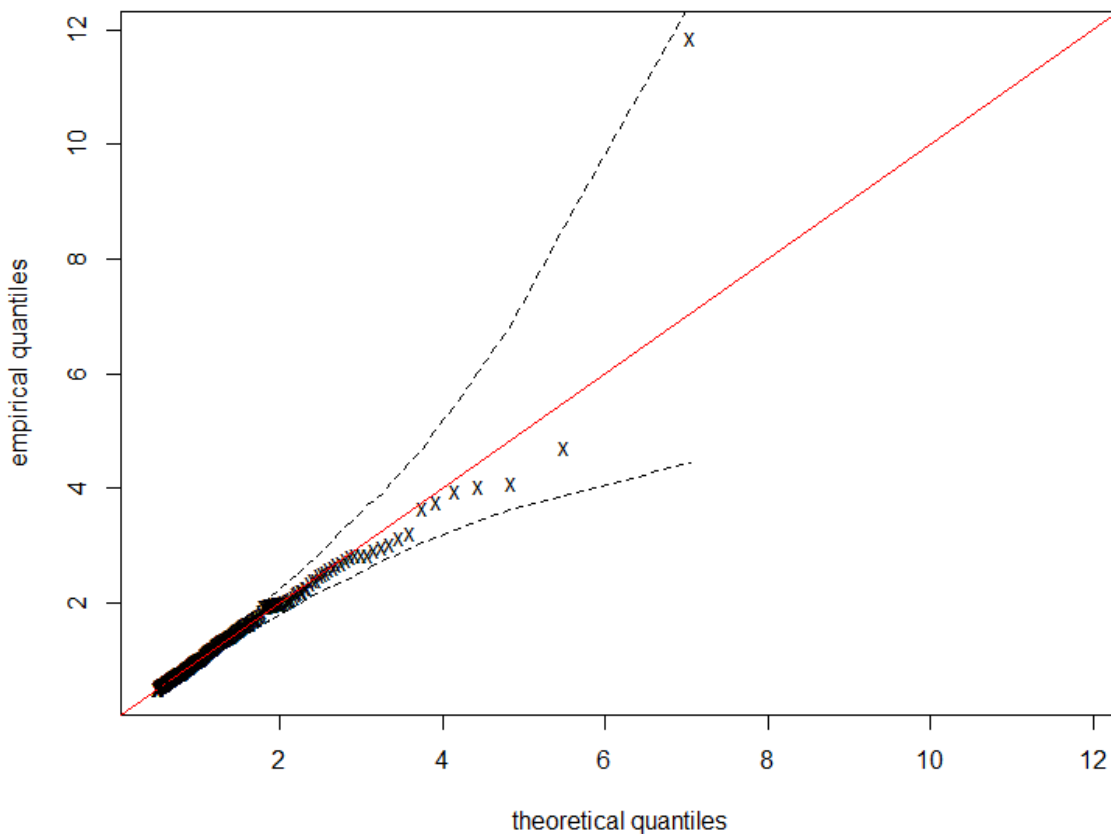
**Εκτίμηση παραμέτρων *GPD* με χρήση της μεθόδου *MLE***

	Εκτίμηση	s.d.	z-value	p-value
Παράμετρος κλίμακας $\sigma$	0.62224	0.051605	12.0579	$1.76 \cdot 10^{-33}$
Παράμετρος σχήματος $\xi$	0.13830	0.062567	2.2104	$2.71 \cdot 10^{-2}$

Από το γράφημα *QQ – plot*, όπου γίνεται σύγκριση των εμπειρικών ποσοστιαίων σημείων ως προς τα αντίστοιχα θεωρικά, παρατηρούμε ότι έχουμε μια αρκετά καλή προσαρμογή της κατανομής στα δεδομένα. Μάλιστα αυτό επισφραγίζεται και μέσα από τον έλεγχο *Cramer – von Mises*, όπου ελέγχεται αν τα δεδομένα πάνω από το επιλεγμένο κατώφλι *u*, ακολουθούν κατανομή *GPD* με παραμέτρους όμοιες με τις παραπάνω.

**Έλεγχος καλής προσαρμογής *Cramer – von Mises***

Υπόθεση	Ελεγχοσυνάρτηση	p-value
Τα δεδομένα προέρχονται από μια κατανομή <i>GPD</i>	0.03	0.92



Σχήμα 4.9: QQ-plot για τον έλεγχο καλής προσαρμογής της εκτιμώμενης GPD κατανομής στα δεδομένα.

Τέλος, εφόσον εκτιμήθηκαν οι παράμετροι της κατανομής GPD, είμαστε σε θέση να εκτιμήσουμε ποια θα μπορούσε να είναι η μέγιστη απώλεια που θα μπορούσε να βιώσει ένας επενδυτής, για δεδομένη πιθανότητα. Συγκεκριμένα, ενδιαφερόμαστε για το 99% ποσοστιαίο σημείο ή διαφορετικά για το σημείο εκείνο όπου οι αρνητικές λογαριθμικές αποδόσεις λαμβάνουν την τιμή αυτή ή και μεγαλύτερη (κατά μέσο όρο) κάθε εκατό ημέρες ( $p = 1\%$ ). Χρησιμοποιώντας τις σχέσεις (βλ. σχ. 2.27, 3.11 και 3.12) καταλήγουμε ότι:

Εκτίμηση VaR και TVaR	
Μέτρα κινδύνου	Εκτίμηση αρνητικής λογαριθμικής απόδοσης (04-01-2021)
VaR(1%)	0.1282629
TVaR(1%)	0.1778372

Ο ενδιαφερόμενος αναγνώστης μπορεί να αναπαράγει τα παραπάνω αποτελέσματα στην R, αφού πρώτα εγκαταστήσει τον κώδικα του ΠΑΡΑΡΤΗΜΑΤΟΣ I και στη συνέχεια πληκτρολογήσει την εντολή:

```
predict_VaR(start_date = "2017-01-01",end_date = "2021-01-03",asset_name = "BTC-USD",sign_level = 0.99,garch_zero = FALSE,graph = TRUE,period = 365,threshold_val = 0.5)
```

### 4.3 ΕΛΕΓΧΟΣ ΜΕΘΟΔΟΛΟΓΙΑΣ

Στην παρούσα παράγραφο πραγματοποιείται ένας εκ των υστέρων έλεγχος (**back-testing**) με σκοπό να εξεταστεί κατά πόσο η χρήση GARCH μοντέλων σε συνδυασμό με τη Θεωρία Ακραίων Τιμών (γενικευμένη Pareto κατανομή), υπερτερεί έναντι άλλων υποδειγμάτων. Αναλυτικότερα, αν θεωρήσουμε ότι ένα σύνολο δεδομένων έστω  $(x_1, \dots, x_t)$ , περιγράφεται μέσω ενός συγκεκριμένου υποδείγματος, τότε ο εκ των υστέρων έλεγχος συνίσταται στη σύγκριση της εκτιμώμενης τιμής  $\hat{x}_{t+1}$  του υποδείγματος, σε σχέση με την πραγματική  $x_{t+1}$ .

Για να γίνει σαφής η διάκριση ανάμεσα στα πλεονεκτήματα και μειονεκτήματα της μεθοδολογίας που εφαρμόζουμε, σε σχέση με άλλες παραλλαγές αυτής, εξετάζονται οι παρακάτω περιπτώσεις:

1. Υπολογισμός  $VaR$  μέσω ενός γενικού υποδείγματος της μορφής:

$$x_t = \text{Τάση} + ARMA(p, q) + GARCH(p', q')$$

όπου η ουρά της χρονοσειράς των σφαλμάτων  $e_t$  του μοντέλου  $GARCH$  ακολουθεί μια κατανομή  $GPD(\sigma, \xi)$ .

2. Υπολογισμός  $VaR$  μέσω ενός γενικού υποδείγματος της μορφής:

$$x_t = \text{Τάση} + ARMA(p, q) + GARCH(p', q')$$

όπου η ουρά της χρονοσειράς των σφαλμάτων  $e_t$  του μοντέλου  $GARCH$  ακολουθεί κανονική κατανομή  $N(0,1)$ .

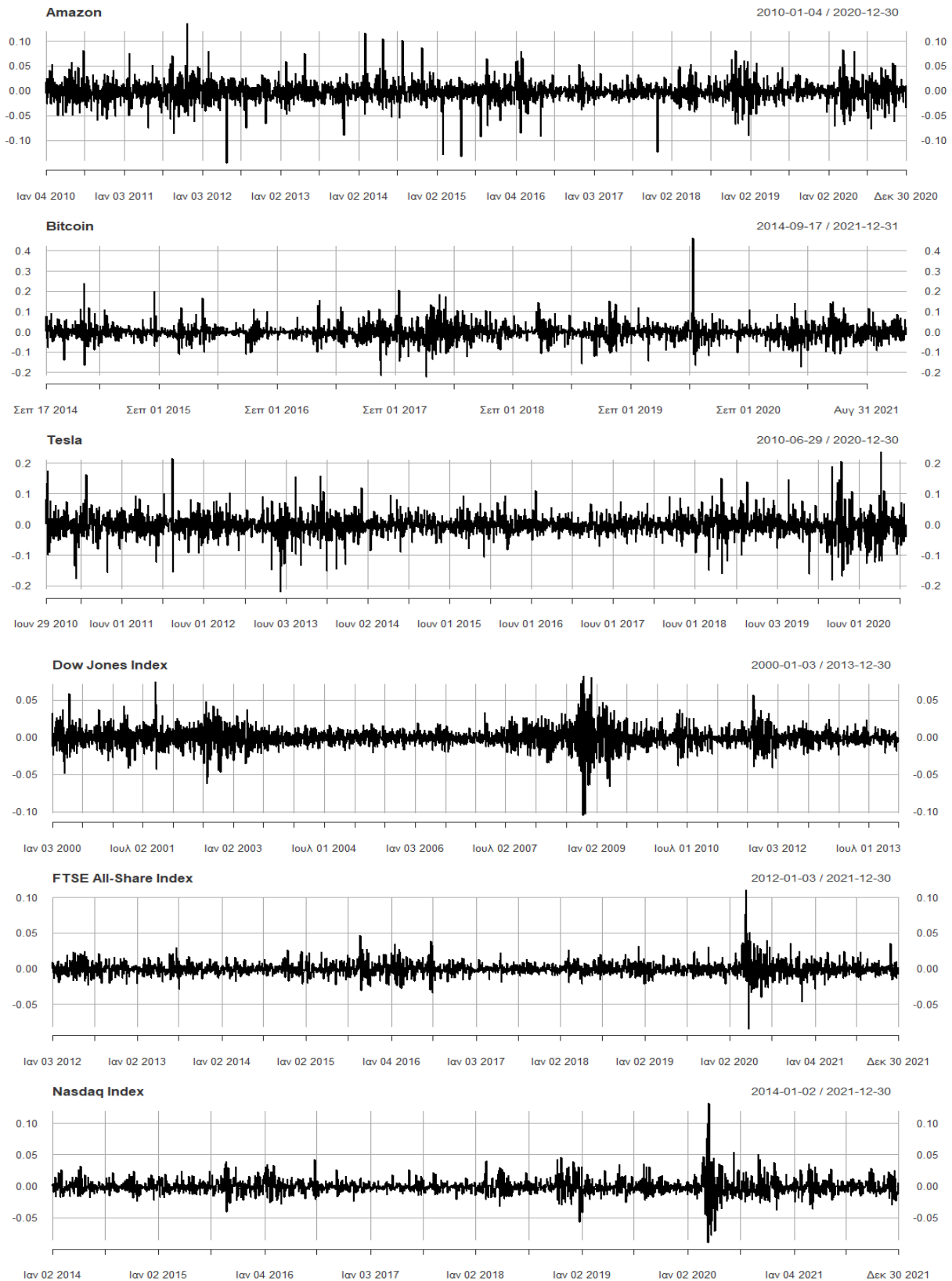
3. Υπολογισμός  $VaR$  θεωρώντας ότι τα δεδομένα ακολουθούν κανονική κατανομή  $N(\mu, \sigma^2)$ .

4. Υπολογισμός  $VaR$  μέσω του δειγματικού ποσοστιαίου σημείου.

Τα χρηματοοικονομικά στοιχεία που επιλέχθηκαν για τον έλεγχο των παραπάνω περιπτώσεων συνίσταται από εξής:

1. Τη μετοχή της *Amazon* για την περίοδο από 1/4/2010 έως 30/12/2020.
2. Το κρυπτονόμισμα *Bitcoin* για την περίοδο από 17/9/2014 έως 31/12/2021.
3. Τη μετοχή της *Tesla* για την περίοδο από 29/6/2010 έως 30/12/2020.
4. Το χρηματιστηριακό δείκτη *Dow Jones* για την περίοδο 3/1/2000 έως 30/12/2013.
5. Το χρηματιστηριακό δείκτη *FTSE* του χρηματιστήριου του Λονδίνου για την περίοδο 3/1/2012 έως 30/12/2021.
6. Το χρηματιστηριακό δείκτη *Nasdaq* για την περίοδο 2/1/2014 έως 30/12/2021.

Τονίζεται πως οι χρονικές περίοδοι επιλέχθηκαν με βάση τα διαθέσιμα δεδομένα, την οικονομική κατάσταση της εκάστοτε περιόδου και την (έντονη) μεταβλητότητα των λογαριθμικών αποδόσεων.



**Σχήμα 4.10:** Γραφικές παραστάσεις των αρνητικών λογαριθμικών αποδόσεων των χρηματοοικονομικών στοιχείων Amazon, Bitcoin, Tesla, Dow Jones, FTSE και Nasdaq για τις περιόδους 1/4/2010 έως 30/12/2020, 17/9/2014 έως 31/12/2021, 29/6/2010 έως 30/12/2020, 3/1/2000 έως 30/12/2013, 3/1/2012 έως 30/12/2021 και αντίστοιχα 2/1/2014 έως 30/12/2021.

Η μεθοδολογία που ακολουθείται για τον εκ των υστέρων έλεγχο έχει ως ακολούθως. Έστω  $(x_m, \dots, x_t, \dots, x_n)$  η χρονοσειρά αποδόσεων ενός χρηματοοικονομικού στοιχείου, με  $t \ll n$ . Επιλέγοντας το  $t$  ώστε η χρονοσειρά  $(x_{m+l}, \dots, x_{t+l})$  με  $l = 0, 1, \dots, n - t - 1$  να αντιστοιχεί κάθε φορά σε δεδομένα τεσσάρων ετών, εκτιμάται μέσω του αλγορίθμου του ΠΑΡΑΡΤΗΜΑΤΟΣ I, το  $VaR$  (για  $p = 5\%, 1\%, 0.5\%$ ) της αρνητικής λογαριθμικής απόδοσης του χρηματοοικονομικού στοιχείου, για την αμέσως επόμενη χρονική περίοδο  $t + l + 1$ , για κάθε ένα από τα τέσσερα υποδείγματα που περιεγράφηκαν παραπάνω. Στη συνέχεια, η εκτιμώμενη (με βάση τα δεδομένα  $x_{m+l}, \dots, x_{t+l}$ ) τιμή του  $VaR$  για το χρόνο  $t + l + 1$ , συγκρίνεται με την πραγματική απόδοση  $x_{t+l+1}$  που έλαβε το χρηματοοικονομικό στοιχείο, και αυτό επαναλαμβάνεται για  $l = 0, 1, \dots, n - t - 1$ . Σημειώνεται ότι σε κάθε επανάληψη ( $l = 0, 1, \dots, n - t - 1$ ), για τις περιπτώσεις των μοντέλων ένα και δύο, ελέγχεται μέσω του αλγορίθμου (ΠΑΡΑΡΤΗΜΑ I) αν υπάρχουν φαινόμενα τάσης, αν ο μέσος ακολουθεί κάποιο από τα μοντέλα  $ARMA(p, q)$  με  $p, q \in \{0, 1, 2\}$  και ποιο ανάμεσα από τα υποδείγματα  $GARCH(p', q')$  με  $p', q' \in \{1, 2\}$  προσαρμόζεται καλύτερα στη χρονοσειρά των καταλοίπων.

Αφού ολοκληρωθούν όλες οι επαναλήψεις για τις οποίες πραγματοποιείται ο εκ των υστέρων έλεγχος, στην ιδανική περίπτωση, αναμένεται ότι το ποσοστό των πραγματικών αποδόσεων που υπερβαίνουν το εκάστοτε εκτιμημένο  $VaR(p)$  να είναι ακριβώς ίσο με  $p$ . Εάν αυτό το ποσοστό των υπερβάσεων έχει μέση τιμή μικρότερη του  $p$  (π.χ. είναι 1% αντί  $p = 5\%$ ) τότε έχει γίνει υπερεκτίμηση του  $VaR$  με συνέπεια π.χ. να κρατείται μεγαλύτερο κεφάλαιο ασφαλείας από ότι χρειάζεται. Στην αντίθετη περίπτωση, αν το ποσοστό αυτό έχει μέση τιμή μεγαλύτερη του  $p$  (π.χ. είναι 7% αντί  $p = 5\%$ ) τότε γίνεται υποεκτίμηση του  $VaR$  με συνέπεια να εκτίθεται η επένδυση σε μεγαλύτερο κίνδυνο από τον προβλεπόμενο.

Για να ελεγχθεί η υπόθεση:

$$H_0: \text{το ποσοστό των υπερβάσεων} = p,$$

καταγράφεται το πλήθος των υπερβάσεων του εκτιμημένου  $VaR$ , έστω  $w$ , το οποίο θα πρέπει υπό την  $H_0$  να ακολουθεί Διωνυμική Κατανομή με πιθανότητα επιτυχίας  $p$ , ( $W \sim B(n - t, p)$ ). Συνεπώς το  $p - value$  του ελέγχου θα είναι ίσο με την πιθανότητα:

$$P(|W - p \cdot (n - t)| \geq |w - p \cdot (n - t)| | H_0)$$

Αναφέρεται ότι για την εφαρμογή της μεθόδου *POT*, το κατώφλι  $u$  δεν επιλέγεται μέσω της χρήσης των διαγραμμάτων  $\sigma(u)$  και  $\xi(u)$  (βλ. Κεφ. 2), αλλά λαμβάνοντας υπόψη τις παρατηρήσεις που βρίσκονται πάνω από το 90% του συνόλου του διατεταγμένου δείγματος. Αυτό αποτελεί ένα επαρκές δείγμα για την εκτίμηση των παραμέτρων της κατανομής *GPD*, καθώς περιέχει «αρκετά» σε πλήθος δεδομένα (περίπου 100 παρατηρήσεις) που αντιπροσωπεύουν ακραίες περιπτώσεις. Για μια πληρέστερη περιγραφή της μεθοδολογίας βέλτιστης επιλογής κατωφλιού, ο ενδιαφερόμενος αναγνώστης μπορεί να ανατρέξει στις εργασίες των McNeil, A. J. and Frey, R. (2000) και Xiangxia, Z. and Wenlei, G. (2009).

Στους παρακάτω πίνακες συνοψίζονται τα αποτελέσματα του back-testing, όπου για κάθε χρηματοοικονομικό στοιχείο και για κάθε υπόδειγμα αναφέρεται το πλήθος των ημερών που ελέγχθηκαν, το σύνολο των επιτυχιών, δηλαδή το σύνολο των ημερών όπου οι εκτιμηθείσες τιμές *VaR* ήταν μεγαλύτερες από την πραγματική απόδοση, τα αντίστοιχα ποσοστά επιτυχίας καθώς και τα  $p - values$  των διωνυμικών ελέγχων.

#### Back-testing μετοχής Amazon

Μοντέλο		<i>VaR</i> <sub>5%</sub>	<i>VaR</i> <sub>1%</sub>	<i>VaR</i> <sub>0.5%</sub>	<i>VaR</i> <sub>0.1%</sub>
<i>Τάση+ARMA+GARCH+EVT</i>	Πλήθος ημερών	1763	1763	1763	1763
	Πλήθος επιτυχιών	1668	1742	1749	1757
	Ποσοστό επιτυχιών	<b>0.946</b>	<b>0.988</b>	<b>0.992</b>	<b>0.997</b>
	$p - value$	0.4441	0.4006	0.0884	0.009
<i>Τάση+ARMA+GARCH + N(0,1)</i>	Πλήθος ημερών	1763	1763	1763	1763
	Πλήθος επιτυχιών	1683	1735	1741	1749
	Ποσοστό επιτυχιών	<b>0.955</b>	<b>0.984</b>	<b>0.988</b>	<b>0.992</b>
	$p - value$	0.4122	0,0217	0.0001	$6 \cdot 10^{-9}$
$N(\mu, \sigma^2)$	Πλήθος ημερών	1763	1763	1763	1763
	Πλήθος επιτυχιών	1675	1725	1731	1742
	Ποσοστό επιτυχιών	<b>0.950</b>	<b>0.979</b>	<b>0.982</b>	<b>0.988</b>
	$p - value$	1	$1.9 \cdot 10^{-5}$	$1.2 \cdot 10^{-9}$	$4.9 \cdot 10^{-16}$
Δειγματικό ποσοστιαίο σημείο	Πλήθος ημερών	1763	1763	1763	1763
	Πλήθος επιτυχιών	1654	1737	1750	1758
	Ποσοστό επιτυχιών	<b>0.938</b>	<b>0.985</b>	<b>0.993</b>	<b>0.997</b>
	$p - value$	0.025	0.0541	0.1717	0.034

### Back-testing κρυπτονομίσματος Bitcoin

Μοντέλο		VaR <sub>5%</sub>	VaR <sub>1%</sub>	VaR <sub>0.5%</sub>	VaR <sub>0.1%</sub>
Τάση+ARMA+GARCH+EVT	Πλήθος ημερών	1461	1461	1461	1461
	Πλήθος επιτυχιών	1386	1443	1453	1458
	Ποσοστό επιτυχιών	<b>0.949</b>	<b>0.988</b>	<b>0.995</b>	<b>0.998</b>
	<i>p</i> – value	0.8101	0.3555	0.7095	0.181
Τάση+ARMA+GARCH + N(0,1)	Πλήθος ημερών	1461	1461	1461	1461
	Πλήθος επιτυχιών	1393	1433	1437	1445
	Ποσοστό επιτυχιών	<b>0.954</b>	<b>0.981</b>	<b>0.984</b>	<b>0.989</b>
	<i>p</i> – value	0.5890	0.0014	$7.4 \cdot 10^{-7}$	$4.9 \cdot 10^{-12}$
N( $\mu, \sigma^2$ )	Πλήθος ημερών	1461	1461	1461	1461
	Πλήθος επιτυχιών	1390	1429	1437	1452
	Ποσοστό επιτυχιών	<b>0.951</b>	<b>0.978</b>	<b>0.984</b>	<b>0.994</b>
	<i>p</i> – value	0.8571	$5.7 \cdot 10^{-5}$	$7.4 \cdot 10^{-7}$	$2.2 \cdot 10^{-5}$
Δειγματικό ποσοστιαίο σημείο	Πλήθος ημερών	1461	1461	1461	1461
	Πλήθος επιτυχιών	1397	1449	1452	1460
	Ποσοστό επιτυχιών	<b>0.956</b>	<b>0.992</b>	<b>0.994</b>	<b>0.999</b>
	<i>p</i> – value	0.3072	0.5983	0.4575	1

### Back-testing μετοχής Tesla

Μοντέλο		VaR <sub>5%</sub>	VaR <sub>1%</sub>	VaR <sub>0.5%</sub>	VaR <sub>0.1%</sub>
Τάση+ARMA+GARCH+EVT	Πλήθος ημερών	1763	1763	1763	1763
	Πλήθος επιτυχιών	1672	1741	1749	1758
	Ποσοστό επιτυχιών	<b>0.948</b>	<b>0.988</b>	<b>0.992</b>	<b>0.997</b>
	<i>p</i> – value	0.7429	0.2803	0.0884	0.034
Τάση+ARMA+GARCH + N(0,1)	Πλήθος ημερών	1763	1763	1763	1763
	Πλήθος επιτυχιών	1685	1730	1734	1744
	Ποσοστό επιτυχιών	<b>0.956</b>	<b>0.981</b>	<b>0.984</b>	<b>0.989</b>
	<i>p</i> – value	0.2988	0.0011	$5.5 \cdot 10^{-8}$	$6.8 \cdot 10^{-14}$
N( $\mu, \sigma^2$ )	Πλήθος ημερών	1763	1763	1763	1763
	Πλήθος επιτυχιών	1681	1719	1731	1745
	Ποσοστό επιτυχιών	<b>0.954</b>	<b>0.975</b>	<b>0.982</b>	<b>0.99</b>
	<i>p</i> – value	0.5476	$9.6 \cdot 10^{-8}$	$1.2 \cdot 10^{-9}$	$7.4 \cdot 10^{-13}$
Δειγματικό ποσοστιαίο σημείο	Πλήθος ημερών	1763	1763	1763	1763
	Πλήθος επιτυχιών	1667	1736	1746	1755
	Ποσοστό επιτυχιών	<b>0.946</b>	<b>0.985</b>	<b>0.990</b>	<b>0.995</b>
	<i>p</i> – value	0.3819	0.0306	0.0105	0.0005

### Χρηματιστηριακός δείκτης Dow Jones

Μοντέλο		VaR <sub>5%</sub>	VaR <sub>1%</sub>	VaR <sub>0.5%</sub>	VaR <sub>0.1%</sub>
Τάση+ARMA+GARCH+EVT	Πλήθος ημερών	2517	2517	2517	2517
	Πλήθος επιτυχιών	2384	2486	2499	2513
	Ποσοστό επιτυχιών	<b>0.947</b>	<b>0.988</b>	<b>0.993</b>	<b>0.998</b>
	<i>p</i> – value	0.493	0.229	0.12	0.327
Τάση+ARMA+GARCH + N(0,1)	Πλήθος ημερών	2517	2517	2517	2517
	Πλήθος επιτυχιών	2369	2458	2482	2500
	Ποσοστό επιτυχιών	<b>0.941</b>	<b>0.977</b>	<b>0.986</b>	<b>0.993</b>
	<i>p</i> – value	0.049	$5.4 \cdot 10^{-9}$	$1.4 \cdot 10^{-7}$	$1.7 \cdot 10^{-9}$
N( $\mu, \sigma^2$ )	Πλήθος ημερών	2517	2517	2517	2517
	Πλήθος επιτυχιών	2394	2448	2462	2481
	Ποσοστό επιτυχιών	<b>0.951</b>	<b>0.973</b>	<b>0.978</b>	<b>0.986</b>
	<i>p</i> – value	0.855	$3.2 \cdot 10^{-13}$	$7.5 \cdot 10^{-19}$	$5.1 \cdot 10^{-29}$
Δειγματικό ποσοστιαίο σημείο	Πλήθος ημερών	2517	2517	2517	2517
	Πλήθος επιτυχιών	2392	2475	2486	2504
	Ποσοστό επιτυχιών	<b>0.95</b>	<b>0.983</b>	<b>0.988</b>	<b>0.995</b>
	<i>p</i> – value	1	0.0018	$1.2 \cdot 10^{-5}$	$2.5 \cdot 10^{-6}$

### Χρηματιστηριακός δείκτης FTSE (Χρηματιστήριο Λονδίνου)

Μοντέλο		VaR <sub>5%</sub>	VaR <sub>1%</sub>	VaR <sub>0.5%</sub>	VaR <sub>0.1%</sub>
Τάση+ARMA+GARCH+EVT	Πλήθος ημερών	1518	1518	1518	1518
	Πλήθος επιτυχιών	1447	1496	1504	1513
	Ποσοστό επιτυχιών	<b>0.953</b>	<b>0.986</b>	<b>0.991</b>	<b>0.997</b>
	<i>p</i> – value	0.637	0.091	0.028	0.019
Τάση+ARMA+GARCH + N(0,1)	Πλήθος ημερών	1518	1518	1518	1518
	Πλήθος επιτυχιών	1435	1478	1487	1503
	Ποσοστό επιτυχιών	<b>0.945</b>	<b>0.974</b>	<b>0.98</b>	<b>0.99</b>
	<i>p</i> – value	0.409	$7.2 \cdot 10^{-8}$	$1.3 \cdot 10^{-10}$	$9.2 \cdot 10^{-11}$
N( $\mu, \sigma^2$ )	Πλήθος ημερών	1518	1518	1518	1518
	Πλήθος επιτυχιών	1449	1480	1486	1495
	Ποσοστό επιτυχιών	<b>0.955</b>	<b>0.975</b>	<b>0.979</b>	<b>0.985</b>
	<i>p</i> – value	0.444	$7.5 \cdot 10^{-7}$	$3 \cdot 10^{-11}$	$1.2 \cdot 10^{-19}$
Δειγματικό ποσοστιαίο σημείο	Πλήθος ημερών	1518	1518	1518	1518
	Πλήθος επιτυχιών	1431	1494	1501	1513
	Ποσοστό επιτυχιών	<b>0.943</b>	<b>0.984</b>	<b>0.989</b>	<b>0.997</b>
	<i>p</i> – value	0.195	0.028	0.003	0.019



### Χρηματιστηριακός δείκτης Nasdaq

Μοντέλο		VaR <sub>5%</sub>	VaR <sub>1%</sub>	VaR <sub>0.5%</sub>	VaR <sub>0.1%</sub>
Τάση+ARMA+GARCH+EVT	Πλήθος ημερών	1007	1007	1007	1007
	Πλήθος επιτυχιών	945	993	996	1003
	Ποσοστό επιτυχιών	<b>0.938</b>	<b>0.986</b>	<b>0.989</b>	<b>0.996</b>
	<i>p</i> – value	0.096	0.203	0.021	0.019
Τάση+ARMA+GARCH + N(0,1)	Πλήθος ημερών	1007	1007	1007	1007
	Πλήθος επιτυχιών	935	970	979	994
	Ποσοστό επιτυχιών	<b>0.929</b>	<b>0.963</b>	<b>0.972</b>	<b>0.987</b>
	<i>p</i> – value	0.003	$3.8 \cdot 10^{-11}$	$9 \cdot 10^{-13}$	$6.5 \cdot 10^{-11}$
N( $\mu, \sigma^2$ )	Πλήθος ημερών	1007	1007	1007	1007
	Πλήθος επιτυχιών	927	964	969	982
	Ποσοστό επιτυχιών	<b>0.92</b>	<b>0.957</b>	<b>0.962</b>	<b>0.975</b>
	<i>p</i> – value	$6.4 \cdot 10^{-5}$	$7.2 \cdot 10^{-15}$	$4 \cdot 10^{-21}$	$2.2 \cdot 10^{-26}$
Δειγματικό ποσοστιαίο σημείο	Πλήθος ημερών	1007	1007	1007	1007
	Πλήθος επιτυχιών	929	984	992	997
	Ποσοστό επιτυχιών	<b>0.92</b>	<b>0.977</b>	<b>0.985</b>	<b>0.99</b>
	<i>p</i> – value	0.0002	0.0003	0.0002	$1.2 \cdot 10^{-7}$

Στον επόμενο πίνακα συνοψίζονται τα αποτελέσματα του back-testing ελέγχου, όπου οι λόγοι εκφράζουν το ποσοστό των περιπτώσεων που εξετάστηκαν, για τις οποίες ο έλεγχος δεν έδωσε κάποιο στατιστικά σημαντικό αποτέλεσμα σε ε.σ. 5%.

### Συγκεντρωτικά αποτελέσματα back-testing

	Μοντέλο	VaR <sub>5%</sub>	VaR <sub>1%</sub>	VaR <sub>0.5%</sub>	VaR <sub>0.1%</sub>
1 <sup>ο</sup>	Τάση+ARMA+GARCH+EVT	$\frac{6}{6}$ (100%)	$\frac{6}{6}$ (100%)	$\frac{4}{6}$ (67%)	$\frac{2}{6}$ (33%)
2 <sup>ο</sup>	Τάση+ARMA+GARCH + N(0,1)	$\frac{4}{6}$ (67%)	$\frac{0}{6}$ (0%)	$\frac{0}{6}$ (0%)	$\frac{0}{6}$ (0%)
3 <sup>ο</sup>	N( $\mu, \sigma^2$ )	$\frac{5}{6}$ (83%)	$\frac{0}{6}$ (0%)	$\frac{0}{6}$ (0%)	$\frac{0}{6}$ (0%)
4 <sup>ο</sup>	Δειγματικό ποσοστιαίο σημείο	$\frac{4}{6}$ (67%)	$\frac{2}{6}$ (33%)	$\frac{2}{6}$ (33%)	$\frac{1}{6}$ (17%)

## Συμπεράσματα

- Όπως αποδεικνύεται από το back-testing, το πρώτο μοντέλο (δηλαδή το μοντέλο όπου λαμβάνεται υπόψη η μεταβλητότητα της διασποράς των λογαριθμικών αποδόσεων μέσω ενός υποδείγματος *GARCH*, σε συνδυασμό με την υπόθεση ότι η ουρά της κατανομής των καταλοίπων ακολουθεί *GPD* κατανομή), εκτιμά ορθότερα την αξία σε κίνδυνο ( $VaR(p)$ ) για  $p = (5\%, 1\%, 0.5\%, 0.1\%)$ , σε σχέση με τα υπόλοιπα υποδείγματα. Συγκεκριμένα, από ολόκληρο το σύνολο των περιπτώσεων που εξετάστηκαν, στην περίπτωση του  $VaR(0.5\%)$  σε δύο από τα έξι χρηματοοικονομικά στοιχεία, ο διωνυμικός έλεγχος απέρριψε την υπόθεση (σε ε.σ. 5%) ότι το ποσοστό των υπερβάσεων του εκτιμημένου  $VaR(p)$  είναι ίσο με  $p$ , ενώ για την περίπτωση του  $VaR(0.1\%)$  σε τέσσερα από τα έξι χρηματοοικονομικά στοιχεία ο διωνυμικός έλεγχος απέρριψε την μηδενική υπόθεση. Αναφέρεται πως σε κάθε περίπτωση απόρριψης της μηδενικής υπόθεσης, το μοντέλο φαίνεται να υποεκτιμά το  $VaR(p)$ .
- Το δεύτερο μοντέλο (δηλαδή το απλούστερο μοντέλο της αγοράς όπου λαμβάνεται υπόψη η μεταβλητότητα της διασποράς των λογαριθμικών αποδόσεων μέσω ενός υποδείγματος *GARCH*, σε συνδυασμό με την υπόθεση ότι η κατανομή των καταλοίπων ακολουθεί κανονική κατανομή) φαίνεται ότι δίνει αρκετά καλά αποτελέσματα για «μικρές» τιμές του επιπέδου σημαντικότητας κατά τον υπολογισμό του  $VaR$ , ενώ αντίθετα υποεκτιμά τον υπολογισμό των λογαριθμικών αποδόσεων για επίπεδο σημαντικότητας 1%, 0.5% και 0.1% αντίστοιχα (και στις τρεις περιπτώσεις ο διωνυμικός έλεγχος απορρίπτει την καταλληλότητα του μοντέλου).
- Στο τρίτο μοντέλο, στην περίπτωση δηλαδή που υποθέτουμε ότι οι λογαριθμικές αποδόσεις ακολουθούν κανονική κατανομή, τα αποτελέσματα εμφανίζουν μια ανάλογη συμπεριφορά όπως αυτή του δεύτερου μοντέλου. Επισημαίνεται πως αυτό οφείλεται στο γεγονός ότι τα δεδομένα, στην πραγματικότητα, προέρχονται από μια κατανομή με βαρύτερη ουρά σε σχέση με αυτή της κανονικής, με αποτέλεσμα η αξία σε κίνδυνο να υποεκτιμάται για μεγάλες τιμές του επιπέδου σημαντικότητας.
- Στην περίπτωση υπολογισμού της αξίας σε κίνδυνο ( $VaR$ ) μέσω του πληθυσμιακού ποσοστιαίου σημείου, παρατηρούμε ότι η μέθοδος αυτή δίνει καλύτερα

αποτελέσματα σε σχέση με το δεύτερο και τρίτο μοντέλο, καθώς σε αρκετές από τις υπό εξεταζόμενες περιπτώσεις, ο διωνυμικός έλεγχος δεν απορρίπτει το μοντέλο. Τέλος, αναφέρεται πως οι περιπτώσεις στις οποίες απορρίπτεται η καταλληλότητα του πληθυσμιακού ποσοστιαίου σημείου, ως μέθοδος υπολογισμού της αξίας σε κίνδυνο, οφείλεται σε υποεκτίμηση του  $VaR$ .

Ο ενδιαφερόμενος αναγνώστης μπορεί να αναπαράγει τα παραπάνω αποτελέσματα στο στατιστικό πακέτο  $R$ , χρησιμοποιώντας τον κώδικα του ΠΑΡΑΡΤΗΜΑΤΟΣ Ι και τις εντολές:

### 1. Amazon

- `back_testing(start_d="2010-01-01",end_d="2013-12-31",end_d_backtest="2020-12-31",asset_name="AMZN",sign_level=0.95,bt_sign_level=0.05,garch_zero_order=F, show_graph=F,period_evt = 253)`

### 2. Bitcoin

- `back_testing(start_d="2014-01-01",end_d="2017-12-31",end_d_backtest="2021-12-31",asset_name="BTC-USD",sign_level=0.95,bt_sign_level=0.05,garch_zero_order=F, show_graph=F,period_evt = 365)`

### 3. Tesla

- `back_testing(start_d="2010-01-01",end_d="2013-12-31",end_d_backtest="2020-12-31",asset_name="TSLA",sign_level=0.95,bt_sign_level=0.05,garch_zero_order=F, show_graph=F,period_evt = 253)`

### 4. Dow Jones Index

- `back_testing(start_d="2000-01-01",end_d="2003-12-31",end_d_backtest="2013-12-31",asset_name="^DJI",sign_level=0.95,bt_sign_level=0.05,garch_zero_order=F,show_graph=F,period_evt = 253)`

### 5. FTSE Index

- `back_testing(start_d="2012-01-01",end_d="2015-12-31",end_d_backtest="2021-12-31",asset_name="^FTAS",sign_level=0.95,bt_sign_level=0.05,garch_zero_order=F,show_graph=F,period_evt = 253)`

### 6. Nasdaq Index

- `back_testing(start_d="2014-01-01",end_d="2017-12-31",end_d_backtest="2021-12-31",asset_name="^IXIC",sign_level=0.95,bt_sign_level=0.05,garch_zero_order=F,show_graph=F,period_evt = 253)`

όπου στο όρισμα *sign\_level* μπορεί να επιλεγεί μια εκ των τιμών 0.95, 0.99, 0.995 και 0.999, ανάλογα με το επίπεδο σημαντικότητας του υπολογιζόμενου  $VaR$ .

# ΠΑΡΑΡΤΗΜΑ Ι

```
#install.packages("ismev")
#install.packages("xts")
#install.packages("quantmod")
#install.packages("tseries")
#install.packages("BETS")
#install.packages("ggplot2")
#install.packages("eva")
#install.packages("POT")
#install.packages("moments")
#install.packages("tea")
#install.packages("EnvStats")

#Απαραίτητες βιβλιοθήκες για τον κώδικα
library("ismev")
library("xts")
library("quantmod")
library("tseries")
library("forecast")
library("BETS")
library("ggplot2")
library("eva")
library("POT")
library("moments")
library("tea")
library("EnvStats")

#Η συνάρτηση predict_VaR εκτιμά το VaR τη χρονική στιγμή t+1 μιας χρηματοοικονομικής χρονοσειράς,
#χρησιμοποιώντας τα εξής μοντέλα:
#1) Τάση+ARMA(p,q)+GARCH(p',q')+EVT
#2) Τάση+ARMA(p,q)+GARCH(p',q')+Normal Distribution
#3) Τα δεδομένα να ακολουθούν κανονική κατανομή  $N(\mu, \sigma^2)$ 
#4) Δειγματικό ποσοστιαίο σημείο
#Αναφέρεται ότι για την περίπτωση του μοντέλου (1) υπολογίζεται και το ανάλογο TVaR
#Τα δεδομένα λαμβάνονται από την ιστοσελίδα «https://finance.yahoo.com». Τα ορίσματα της συνάρτησης είναι τα εξής:
#start_date: Έτος, μήνας και ημέρα που τα ξεκινάει η χρηματοοικονομική σειρά. Η μορφή θα πρέπει να είναι: "yyyy-mm-dd".
Παράδειγμα "2010-01-01"
#end_date: Έτος, μήνας και ημέρα που τελειώνει η χρηματοοικονομική σειρά. Η μορφή θα πρέπει να είναι: "yyyy-mm-dd".
Παράδειγμα "2015-12-31"
#asset_name: Όνομα μετοχής, δείκτη, κρυπτονομίσματος όπως αυτό περιγράφεται από την ιστοσελίδα
«https://finance.yahoo.com».
#Η μορφή που πρέπει να δίνεται το όνομα είναι: "name". Παράδειγμα, για τον δείκτη S&P500 πρέπει να δοθεί η τιμή
"^GSPC"
#sign_level: Ορίζεται το επίπεδο σημαντικότητας για το οποίο υπολογίζεται το η στάθμη απόδοσης.
#Η τιμή αυτή είναι προεπιλεγμένη στο 95%, σε περίπτωση που ο χρήστης δεν χρησιμοποιεί το όρισμα αυτό.
#arma_order: Δισδιάστατο διάνυσμα όπου εκφράζει την μέγιστη τάξη του μοντέλου ARMA ανάμεσα στα οποία ο αλγόριθμος
θα προσαρμόσει στα δεδομένα. Έχει προεπιλεγμένη την τιμή arma_order=c(2,2).
#garch_zero: Λαμβάνει τις λογικές TRUE ή FALSE, με προεπιλεγμένη τιμή FALSE. Στην περίπτωση όπου η μεταβλητή
garch_zero = TRUE,
#γίνεται ένας έλεγχος μεταξύ των GARCH μοντέλων τάξης p,q={0,1,2} που μπορούν να εφαρμοστούν στα δεδομένα
χρησιμοποιώντας το κριτήριο AIC.
#Στην περίπτωση όπου garch_zero = FALSE, τότε γίνεται έλεγχος μεταξύ των GARCH μοντέλων τάξης p,q={1,2}. Δηλαδή,
όταν garch_zero = FALSE
#εφαρμόζεται πάντα ένα μοντέλο GARCH στα δεδομένα, σε αντίθεση με την περίπτωση garch_zero = TRUE.
```

```

#graph: Λαμβάνει τις λογικές TRUE ή FALSE, με προεπιλεγμένη τιμή FALSE. Αν graph = TRUE, σαν αποτέλεσμα δίνεται μια
σειρά
#από διαγνωστικά γραφήματα που αφορούν τη χρονοσειρά των δεδομένων και την εφαρμογή της θεωρίας ακραίων τιμών
#threshold_val: Ορίζει το κατώφλι εκείνο για το οποίο εφαρμόζεται η μέθοδος POT και λαμβάνει πραγματικές τιμές.
#Δεν μπορεί να χρησιμοποιηθεί ταυτόχρονα με τη μεταβλητή n_ordered_p. Έχει προεπιλεγμένη τιμή threshold_val = Inf,
#για την περίπτωση που δεν χρειάζεται να χρησιμοποιηθεί
#n_ordered_p: Ορίζει το πλήθος των διατεταγμένων σημείων του δείγματος που χρησιμοποιούνται για την εφαρμογή της
μεθόδου POT.
#Δεν χρησιμοποιείται ταυτόχρονα με τη μεταβλητή threshold_val. Έχει προεπιλεγμένη τιμή n_ordered_p = 0 που
αντιστοιχεί στο σύνολο
#των διατεταγμένων σημείων του δείγματος που βρίσκονται πάνω από το 90% ποσοστιαίο σημείο.
#period: Εκφράζει τον αριθμό των ημερών ανά περίοδο. Έχει προεπιλεγμένη τιμή 253 διότι τόσες ημέρες κατά μέσο όρο
είναι ανοικτά τα χρηματιστήρια. Στην περίπτωση όπου τα δεδομένα είναι ετήσια πρέπει να τίθεται η τιμή 365. Γενικότερα η
μεταβλητή αυτή εκφράζει το πλήθος των δεδομένων σε ένα BLOCK και χρησιμοποιείται για τον υπολογισμό της στάθμης
απόδοσης.

#Παράδειγμα 1: predict_VaR(start_date = "2012-01-01",end_date = "2014-12-31",asset_name = "AAPL",garch_zero =
FALSE)
#Παράδειγμα 2: predict_VaR(start_date = "2012-01-01",end_date = "2014-12-31",asset_name = "AAPL",garch_zero =
TRUE, threshold_val = 0.4)
#Παράδειγμα 3: predict_VaR(start_date = "2012-01-01",end_date = "2014-12-31",asset_name = "AAPL",garch_zero =
TRUE,graph = TRUE, n_ordered_p= 100)
#Παράδειγμα 4: predict_VaR(start_date = "2012-01-01",end_date = "2014-12-31",asset_name = "AAPL",sign_level =
0.995,garch_zero = TRUE,graph = TRUE, n_ordered_p= 100)
#Παράδειγμα 5: predict_VaR(start_date = "2015-01-01",end_date = "2017-12-31",asset_name = "BTC-USD",sign_level =
0.995,garch_zero = TRUE,graph = TRUE, period=365)

predict_VaR<-
function(start_date,end_date,asset_name,sign_level=0.95,arma_order=c(2,2),garch_zero=FALSE,graph=FALSE,thresho
ld_val=Inf,n_ordered_p=0,period=253){
  defaultW <- getOption("warn")
  options(warn = -1)
  #####
  #Εισαγωγή δεδομένων μετοχής, χρηματιστηριακού δείκτη ή κρυπτονομίσματος από την ιστοσελίδα Yahoo-finance
  from.dat= as.Date(start_date,format = "%Y-%m-%d")
  to.dat= as.Date(end_date,format = "%Y-%m-%d")
  data<-getSymbols(asset_name,from=from.dat,to=to.dat,periodicity="daily",auto.assign=F)

  #Γραφική παράσταση των τιμών
  if (graph==T)
  {
    dev.new()
    chartSeries(data)
  }

  #Εισαγωγή των αποδόσεων σε νέα μεταβλητή. Το μέϊον χρησιμοποιείται γιατί στη συνέχεια μας ενδιαφέρει η μελέτη των
αρνητικών αποδόσεων (Adjusted values)
  data_re=-dailyReturn(data[,6],type="log")
  #Γραφική παράσταση της χρονοσειράς των αποδόσεων, της αυτοσυσχέτισης και έλεγχος Ljung-Box για την ανεξαρτησία των
δεδομένων

  if (graph==T)
  {
    dev.new()
    par(mfrow=c(3,1))
    plot(index(data_re),ts(data_re),type="l",xlab="Time",ylab="Daily negative log return",main="Γράφημα
χρονοσειράς\ντου αρνητικού λογάριθμου των αποδόσεων")
    plot(acf(data_re,plot=FALSE),main="Γράφημα αυτοσυσχετίσεων χρονοσειράς\ντου αρνητικού λογάριθμου των
αποδόσεων")
    plot(acf(data_re^2,plot=FALSE),main="Γράφημα αυτοσυσχετίσεων του τετραγώνου της χρονοσειράς\ντου αρνητικού
λογάριθμου των αποδόσεων")
    cat("=====\n")
    cat("Ljung-Box test για την χρονοσειρά του αρνητικού λογαρίθμου των αποδόσεων: \n")
    print(Box.test(data_re,type="Ljung-Box", lag = 15))
    cat("Ljung-Box test για την χρονοσειρά του τετραγώνου του αρνητικού λογαρίθμου των αποδόσεων: \n")
    print(Box.test(data_re^2,type="Ljung-Box", lag = 15))
    par(mfrow=c(1,1))
  }
}

```

```

#Μετατροπή χρονοσειράς σε αντικείμενο ts και ορισμός μεταβλητής pred που αποθηκεύει την πρόβλεψη του μοντέλου για
την επόμενη ημέρα
data_re_ts<-as.data.frame(data_re)
data_re_ts<-ts(data_re,start =
c(as.numeric(format(from.dat,format="%Y")),as.numeric(format(from.dat,format="%m")),as.numeric(format(from.dat,
format="%d"))),frequency = period)
pred<-0

#Εύρεση τάσης στα δεδομένα χρησιμοποιώντας γραμμική παλινδρόμηση.
#Σημειώνεται ότι στην περίπτωση όπου υπάρχει θετική συσχέτιση για lag=1 προτιμάται η χρήση γενικευμένων γραμμικών
μοντέλων ώστε να εισαχθεί ο βαθμός συσχέτισης στον υπολογισμό. Η περίπτωση αυτή δεν λαμβάνεται υπόψη στον κάτωθι
κώδικα.

#Επίσης γίνεται ο υπολογισμός της πρόβλεψης για την επόμενη ημέρα
Time<-1:length(data_re_ts)
if (summary(lm(data_re_ts~Time))$coefficients[8]<0.05)
{
  pred<-sum(lm(data_re_ts~Time)$coef*c(1,length(data_re_ts)+1))
  data_re_ts<-ts(resid(lm(data_re_ts~Time)),start=
c(as.numeric(format(from.dat,format="%Y")),as.numeric(format(from.dat,format="%m")),as.numeric(format(from.dat,
format="%d"))),frequency = period)
  trend<-"Yes"
} else
{
  trend<-"No"
}

#Εύρεση και υπολογισμός βέλτιστου ARMA μοντέλου
arima_model<-auto.arima(data_re_ts, seasonal = FALSE,max.p = arma_order[1],max.q = arma_order[2])
arima_model<-Arima(data_re_ts,order = arimaorder(arima_model))

if (graph==T)
{
  cat("=====\n")
  cat("Τάξη αρχικού μοντέλου ARMA:\n",arimaorder(arima_model),"\n")
}

#Υπολογισμός t-test για τη σημαντικότητα των μεταβλητών που έχουν εισαχθεί στο ARMA μοντέλο.
#Σε περίπτωση που κάποια μεταβλητή δεν χρειάζεται να εισαχθεί στο μοντέλο αφαιρείται
#και γίνεται επαναπροσδιορισμός του μοντέλου
t_val_arima<-t_test(arima_model,alpha =0.05 )$t
t_val_arima<-t_val_arima[-length(t_val_arima)]
crit_val_arima<-t_test(arima_model,alpha =0.05 )$Crit.Values[1]

while (sum(t_val_arima>crit_val_arima)<sum(arimaorder(arima_model)))
{
  if (which(t_val_arima==min(t_val_arima))<=arimaorder(arima_model)[1])
  {
    arima_model<-Arima(data_re_ts,order = arimaorder(arima_model)-c(1,0,0))
    t_val_arima<-t_test(arima_model,alpha =0.05 )$t
    t_val_arima<-t_val_arima[-length(t_val_arima)]
    crit_val_arima<-t_test(arima_model,alpha =0.05 )$Crit.Values[1]
  } else
  {
    arima_model<-Arima(data_re_ts,order = arimaorder(arima_model)-c(0,0,1))
    t_val_arima<-t_test(arima_model,alpha =0.05 )$t
    t_val_arima<-t_val_arima[-length(t_val_arima)]
    crit_val_arima<-t_test(arima_model,alpha =0.05 )$Crit.Values[1]
  }
}

if (graph==T)
{
  cat("=====\n")
  cat("Τάξη τελικού μοντέλου ARMA:\n",arimaorder(arima_model),"\n")
}

#Υπολογισμός πρόβλεψης επόμενης ημέρας. Επίσης γίνεται υπολογισμός της κανονικοποιημένης χρονοσειράς των
καταλοίπων με μέση τιμή μηδέν στην περίπτωση που εφαρμόζεται ένα μοντέλο ARMA στα δεδομένα
if (sum(arimaorder(arima_model))!=0)
{
  pred<-pred+stats::predict(arima_model,n.ahead=1)$pred[1]
}

```

```

mean_garch<-mean(arima_model$res)
garch_ts<-arima_model$res-mean_garch
} else
{
mean_garch<-mean(data_re_ts)
garch_ts<-data_re_ts-mean_garch
}

#Γραφική παράσταση της χρονοσειράς των καταλοίπων, της αυτοσυσχέτισης και έλεγχος Ljung-Box για την ανεξαρτησία
των δεδομένων
if (graph==T & sum(arimaorder(arima_model))!=0)
{
dev.new()
par(mfrow=c(3,1))
plot(garch_ts,ylab="Daily negative log returns",main="Γραφική παράσταση της χρονοσειράς καταλοίπων\n αφού έχει
εφαρμοστεί ένα μοντέλο ARMA")
plot(acf(garch_ts,plot=FALSE),main="Γράφημα αυτοσυσχετίσεων της χρονοσειράς καταλοίπων\nαφού έχει εφαρμοστεί
ένα μοντέλο ARMA")
cat("=====\n")
cat("Ljung-Box test για την χρονοσειρά των καταλοίπων αφού έχει\nεφαρμοστεί ένα μοντέλο ARMA \n")
print(Box.test(garch_ts,type="Ljung-Box",lag = 15))
cat("=====\n")
plot(acf(garch_ts^2,plot=FALSE),main="Γράφημα αυτοσυσχετίσεων του τετραγώνου της χρονοσειράς καταλοίπων")
cat("Ljung-Box test για την χρονοσειρά του τετραγώνου των καταλοίπων αφού έχει αφαιρεθεί τάση,\n και έχει εφαρμοστεί
ένα μοντέλο ARMA\n")
print(Box.test(garch_ts^2,type="Ljung-Box",lag = 15))
par(mfrow=c(1,1))
}

#Εύρεση βέλτιστου μοντέλου GARCH βάση κριτηρίου Akaike
aic=Inf
if(garch_zero==TRUE)
{
g_order<-0:2
} else {
g_order<-1:2
}
p_garch<-q_arch<-0
for (i in g_order)
{
for (j in g_order)
{
if(i!=0 | j!=0)
{
if (aic>AIC(garch(garch_ts, order = c(i,j), trace = FALSE)))
{
aic=AIC(garch(garch_ts, order = c(i,j), trace = FALSE))
p_garch=i
q_arch=j
best_garch<-garch(garch_ts, order = c(p_garch,q_arch), trace = FALSE)
}
}
}
}
}

if (graph==T)
{
cat("=====\n")
cat("Τάξη αρχικού μοντέλου GARCH(p,q):\n",c(p_garch,q_arch),"\n")
}

#Εισαγωγή των p_value απο το t-test, χωρίς να λαμβάνεται υπόψη η τιμή για τον σταθερό όρο a0. Η μεταβλητή αυτή
(p_val) χρησιμοποιείται για τον προσδιορισμό του βέλτιστου μοντέλου
#λαμβάνοντας υπόψιν το t-test
p_val<-summary(best_garch)$coef[-1,4]

#Σε περίπτωση που υπάρχουν NA στον έλεγχο t-test, τότε μειώνουμε την τάξη του μοντέλου
if(garch_zero==TRUE){
while (sum(which(is.na(p_val)))!=0)
{
p_garch<-p_garch-1
}
}

```

```

best_garch<-garch(garch_ts, order = c(p_garch,q_arch), trace = FALSE)
p_val<-summary(best_garch)$coef[-1,4]
}
} else {
while (sum(which(is.na(p_val)))!=0 & p_garch!=1)
{
p_garch<-p_garch-1
best_garch<-garch(garch_ts, order = c(p_garch,q_arch), trace = FALSE)
p_val<-summary(best_garch)$coef[-1,4]
}
}

#Έλεγχος συντελεστών βέλτιστου επιλεγμένου μοντέλου βάση t-test. Ο έλεγχος γίνεται μετρώντας τα True στο διάνυσμα
που προκύπτει μέσα από την συνθήκη που υπάρχει στο while
#Σε περίπτωση που ο αριθμός αυτός είναι μικρότερος από το άθροισμα των μεταβλητών που εισάγονται στο μοντέλο,
υπολογίζεται το αμέσως μικρότερο αφαιρώντας την τιμή με το μεγαλύτερο p-value
#Διαφορετικά, δίνεται στη μεταβλητή best_garch το μοντέλο που βρέθηκε παραπάνω
if(garch_zero==TRUE)
{
while (sum(p_val<0.05)<p_garch+q_arch)
{
#Εάν η παράμετρος αφορά το arch μοντέλο, τότε η σειρά κατάταξης της στο διάνυσμα των παραμέτρων θα είναι
μικρότερη ή το πολύ ίση
#με την τάξη q_arch του επιλεγμένου arch μοντέλου
if (which(p_val==max(p_val))<=q_arch)
{
q_arch<-q_arch-1
if(p_garch+q_arch!=0)
{
best_garch<-garch(garch_ts, order = c(p_garch,q_arch), trace = FALSE)
p_val<-summary(best_garch)$coef[-1,4]
while (sum(which(is.na(p_val)))!=0)
{
p_garch<-p_garch-1
if(p_garch+q_arch!=0)
{
best_garch<-garch(garch_ts, order = c(p_garch,q_arch), trace = FALSE)
p_val<-summary(best_garch)$coef[-1,4]
} else
{
break
}
}
} else
{
break
}
} else
{
p_garch<-p_garch-1
if(p_garch+q_arch!=0)
{
best_garch<-garch(garch_ts, order = c(p_garch,q_arch), trace = FALSE)
p_val<-summary(best_garch)$coef[-1,4]
while (sum(which(is.na(p_val)))!=0)
{
p_garch<-p_garch-1
if(p_garch+q_arch!=0)
{
best_garch<-garch(garch_ts, order = c(p_garch,q_arch), trace = FALSE)
p_val<-summary(best_garch)$coef[-1,4]
} else
{
break
}
}
} else
{
break
}
}
}
}
}
}
}

```



```

}
} else {
while (sum(p_val<0.05)<p_garch+q_arch & p_garch+q_arch!=2)
{
#Εάν η παράμετρος αφορά το arch μοντέλο, τότε η σειρά κατάταξης της στο διάγραμμα των παραμέτρων θα είναι
μικρότερη ή το πολύ ίση
#με την τάξη q_arch του επιλεγμένου arch μοντέλου
if (which(p_val==max(p_val))<=q_arch)
{
q_arch<-q_arch-1
best_garch<-garch(garch_ts, order = c(p_garch,q_arch), trace = FALSE)
p_val<-summary(best_garch)$coef[-1,4]
while (sum(which(is.na(p_val)))!=0 & p_garch+q_arch!=2)
{
p_garch<-p_garch-1
best_garch<-garch(garch_ts, order = c(p_garch,q_arch), trace = FALSE)
p_val<-summary(best_garch)$coef[-1,4]
}
} else
{
p_garch<-p_garch-1
best_garch<-garch(garch_ts, order = c(p_garch,q_arch), trace = FALSE)
p_val<-summary(best_garch)$coef[-1,4]
while (sum(which(is.na(p_val)))!=0 & p_garch+q_arch!=2)
{
p_garch<-p_garch-1
best_garch<-garch(garch_ts, order = c(p_garch,q_arch), trace = FALSE)
p_val<-summary(best_garch)$coef[-1,4]
}
}
}
}
}

if (graph==T)
{
cat("=====\n")
cat("Τάξη τελικού μοντέλου GARCH(p,q):\n",c(p_garch,q_arch),"\n")
}

#Γραφική παράσταση χρονοσειράς και υπό συνθήκη διασποράς σύμφωνα με το αποτελέσματα της προσαρμογής του
μοντέλου GARCH
if (graph==T & garch_zero==F)
{
dev.new()
plot(ts(garch_ts[!is.na(best_garch$res) & !is.nan(best_garch$res)],start =
c(as.numeric(format(from.dat,format="%Y")),as.numeric(format(from.dat,format="%m")),as.numeric(format(from.dat,
format="%d"))),frequency = period,ylab="Residual time series",col=3,main="Γραφική παράσταση χρονοσειράς για την
σποία εφαρμόζεται ένα μοντέλο GARCH\n σε συνδυασμό με την εκτιμώμενη (υπό συνθήκη) διασπορά")
lines(ts(2*best_garch$fitted.values[!is.na(best_garch$res) & !is.nan(best_garch$res)],1],start =
c(as.numeric(format(from.dat,format="%Y")),as.numeric(format(from.dat,format="%m")),as.numeric(format(from.dat,
format="%d"))),frequency = period,type="l",col=5)
lines(ts(-2*best_garch$fitted.values[!is.na(best_garch$res) & !is.nan(best_garch$res)],1],start =
c(as.numeric(format(from.dat,format="%Y")),as.numeric(format(from.dat,format="%m")),as.numeric(format(from.dat,
format="%d"))),frequency = period,type="l",col=5)

#Γραφική παράσταση διασποράς
dev.new()
plot(ts(best_garch$fitted.values[!is.na(best_garch$res) & !is.nan(best_garch$res)],1],start =
c(as.numeric(format(from.dat,format="%Y")),as.numeric(format(from.dat,format="%m")),as.numeric(format(from.dat,
format="%d"))),frequency = period,ylab="Estimated Variance",type="l",col=5,main = "Γραφική παράσταση εκτιμώμενης
(υπό συνθήκη) διασποράς")

#Γράφημα που περιέχει τα δυο παραπάνω σε ένα
dev.new()
par(mfrow=c(2,1))
plot(ts(garch_ts[!is.na(best_garch$res) & !is.nan(best_garch$res)],start =
c(as.numeric(format(from.dat,format="%Y")),as.numeric(format(from.dat,format="%m")),as.numeric(format(from.dat,
format="%d"))),frequency = period,ylab="Residual time series",col=3,main="Γραφική παράσταση χρονοσειράς για την
σποία εφαρμόζεται ένα μοντέλο GARCH\n σε συνδυασμό με την εκτιμώμενη (υπό συνθήκη) διασπορά")

```

```

lines(ts(2*best_garch$fitted.values[!is.na(best_garch$res) & !is.nan(best_garch$res)],1],start =
c(as.numeric(format(from.dat,format="%Y")),as.numeric(format(from.dat,format="%m")),as.numeric(format(from.dat,
format="%d"))),frequency = period,type="l",col=5)
lines(ts(-2*best_garch$fitted.values[!is.na(best_garch$res) & !is.nan(best_garch$res)],1],start =
c(as.numeric(format(from.dat,format="%Y")),as.numeric(format(from.dat,format="%m")),as.numeric(format(from.dat,
format="%d"))),frequency = period,type="l",col=5)
plot(ts(best_garch$fitted.values[!is.na(best_garch$res) & !is.nan(best_garch$res)],1],start =
c(as.numeric(format(from.dat,format="%Y")),as.numeric(format(from.dat,format="%m")),as.numeric(format(from.dat,
format="%d"))),frequency = period,ylab="Estimated Variance",type="l",col=5,main = "Γραφική παράσταση εκτιμώμενης
(υπό συνθήκη) διασποράς")
par(mfrow=c(1,1))
}

```

```

#Έλεγχος ανεξαρτησίας χρονοσειράς σφαλμάτων του μοντέλου GARCH και οι γραφικές παραστάσεις για τις αυτοσυσχετίσεις
#των σφαλμάτων και των τετραγώνων αυτών
if (graph==T & garch_zero==F)
{
if (sum(arimaorder(arima_model))!=0)
{
dev.new()
par(mfrow=c(2,1))
plot(acf(best_garch$res[!is.na(best_garch$res) & !is.nan(best_garch$res)],plot=FALSE),main="Γράφημα
αυτοσυσχετίσεων χρονοσειράς καταλοίπων\ναφού έχει εφαρμοστεί ένα μοντέλο ARMA-GARCH")
cat("=====\n")
cat("Ljung-Box test για την χρονοσειρά καταλοίπων μετά την εξάλειψη τάσης\η και αφού έχει εφαρμοστεί ένα μοντέλο
ARMA-GARCH \η")
print(Box.test(best_garch$res[!is.na(best_garch$res) & !is.nan(best_garch$res)],type="Ljung-Box",lag = 15))
plot(acf(best_garch$res[!is.na(best_garch$res) & !is.nan(best_garch$res)]^2,plot=FALSE),main="Γράφημα
αυτοσυσχετίσεων χρονοσειράς του τετραγώνου των καταλοίπων\ναφού έχει εφαρμοστεί ένα μοντέλο ARMA-GARCH")
cat("=====\n")
cat("Ljung-Box test για την χρονοσειρά του τετραγώνου των καταλοίπων μετά την εξάλειψη τάσης\ηκαι αφού έχει
εφαρμοστεί ένα μοντέλο ARMA-GARCH: \η")
print(Box.test(best_garch$res[!is.na(best_garch$res) & !is.nan(best_garch$res)]^2,type="Ljung-Box",lag = 15))
par(mfrow=c(1,1))
} else{
dev.new()
par(mfrow=c(2,1))
plot(acf(best_garch$res[!is.na(best_garch$res) & !is.nan(best_garch$res)],plot=FALSE),main="Γράφημα
αυτοσυσχετίσεων χρονοσειράς καταλοίπων\ναφού έχει εφαρμοστεί ένα μοντέλο GARCH")
cat("=====\n")
cat("Ljung-Box test για την χρονοσειρά καταλοίπων μετά την εξάλειψη τάσης\η και αφού έχει εφαρμοστεί ένα μοντέλο
GARCH \η")
print(Box.test(best_garch$res[!is.na(best_garch$res) & !is.nan(best_garch$res)],type="Ljung-Box",lag = 15))
plot(acf(best_garch$res[!is.na(best_garch$res) & !is.nan(best_garch$res)]^2,plot=FALSE),main="Γράφημα
αυτοσυσχετίσεων χρονοσειράς του τετραγώνου\η των καταλοίπων αφού έχει εφαρμοστεί ένα μοντέλο GARCH")
cat("=====\n")
cat("Ljung-Box test για την χρονοσειρά του τετραγώνου των καταλοίπων μετά την εξάλειψη τάσης\ηκαι αφού έχει
εφαρμοστεί ένα μοντέλο GARCH: \η")
print(Box.test(best_garch$res[!is.na(best_garch$res) & !is.nan(best_garch$res)]^2,type="Ljung-Box",lag = 15))
par(mfrow=c(1,1))
}
}

```

```

#Έλεγχος κανονικότητας της χρονοσειράς των σφαλμάτων μέσω QQplot, στατιστικών tests
dev.new()
stats::qqnorm(best_garch$res[!is.na(best_garch$res) & !is.nan(best_garch$res)],col=2)
qqline(best_garch$res[!is.na(best_garch$res) & !is.nan(best_garch$res)],col=3)
cat("=====\n")
cat("Έλεγχος κανονικότητας χρονοσειράς σφαλμάτων, χρησιμοποιώντας τον έλεγχο Shapiro-Wilk:\η")
print(shapiro.test(best_garch$res[!is.na(best_garch$res) & !is.nan(best_garch$res)]))
cat("=====\n")
cat("Έλεγχος κανονικότητας χρονοσειράς σφαλμάτων, χρησιμοποιώντας τον έλεγχο Kolmogorov-Smirnov:\η")
print(ks.test(best_garch$res[!is.na(best_garch$res) & !is.nan(best_garch$res)],y="pnorm"))
cat("=====\n")

#Ιστόγραμμα και σύγκριση συνάρτησης κατανομής με την κανονική. Επίσης σχεδιάζεται και η δεξιά ουρά της κατανομής
dev.new()
par(mfrow=c(2,1))
hist(best_garch$res[!is.na(best_garch$res) & !is.nan(best_garch$res)],breaks="scott",probability = TRUE,freq =
FALSE,xlab="Residuals",ylim = c(0,0.8),main="Ιστόγραμμα χρονοσειράς καταλοίπων\ηκαι σύγκριση με την κανονική
κατανομή")
lines(density(best_garch$res[!is.na(best_garch$res) & !is.nan(best_garch$res)],adjust=3),col=3)

```

```

lines(seq(min(best_garch$res[!is.na(best_garch$res) & !is.nan(best_garch$res)]),
max(best_garch$res[!is.na(best_garch$res) & !is.nan(best_garch$res)]), by=.2),
dnorm(seq(min(best_garch$res[!is.na(best_garch$res) & !is.nan(best_garch$res)]),
max(best_garch$res[!is.na(best_garch$res) & !is.nan(best_garch$res)]),
by=.2),mean(best_garch$res[!is.na(best_garch$res) & !is.nan(best_garch$res)]),
sd(best_garch$res[!is.na(best_garch$res) & !is.nan(best_garch$res)])), col="blue")
hist(best_garch$res[!is.na(best_garch$res) & !is.nan(best_garch$res)],breaks="scott",probability = TRUE,freq =
FALSE,xlim = c(4,7),xlab="Residuals",ylim = c(0,0.01),main="Ιστόγραμμα χρονοσειράς καταλοίπων και σύγκριση με\λητην
κανονική κατανομή δίνοντας έμφαση στην ουρά της κατανομής")
lines(density(best_garch$res[!is.na(best_garch$res) & !is.nan(best_garch$res)],adjust=3),col=3)
lines(seq(min(best_garch$res[!is.na(best_garch$res) & !is.nan(best_garch$res)]),
max(best_garch$res[!is.na(best_garch$res) & !is.nan(best_garch$res)]), by=.2),
dnorm(seq(min(best_garch$res[!is.na(best_garch$res) & !is.nan(best_garch$res)]),
max(best_garch$res[!is.na(best_garch$res) & !is.nan(best_garch$res)]),
by=.2),mean(best_garch$res[!is.na(best_garch$res) & !is.nan(best_garch$res)]),
sd(best_garch$res[!is.na(best_garch$res) & !is.nan(best_garch$res)])), col="blue")
par(mfrow=c(1,1))
}

#Εκτίμηση επόμενης ημέρας (one step ahead prediction) διασποράς (standard deviation). Η συνάρτηση rev
χρησιμοποιείται για να πάρουμε τους όρους με τη σωστή
#σειρά που απαιτείται για τον υπολογισμό. Επίσης γίνεται υπολογισμός των αντίστοιχων ποσοτήτων θεωρώντας ότι η
κατανομή των σφαλμάτων είναι η κανονική
if(p_garch+q_arch!=0)
{
#Πλήθος όρων χρονοσειράς σφαλμάτων
n_garch_res<-length(best_garch$res)

if (p_garch!=0)
{
#Υπολογισμός των X(t) του μοντέλου GARCH, όπου X(t)=σ(t)e(t)
garch_pred<-rev(best_garch$fitted.values[-(1:(n_garch_res-q_arch)),1]*best_garch$res[-(1:(n_garch_res-q_arch))])
#Υπολογισμός της εκτιμώμενης, από το μοντέλο GARCH, διασποράς σ(t+1)
standar_dev_pred<-sqrt(sum(c(best_garch$coef)*c(1,garch_pred^2,rev(best_garch$fitted.values[-(1:(n_garch_res-
p_garch)),1]^2))))
} else {
#Υπολογισμός των X(t) του μοντέλου GARCH, όπου X(t)=σ(t)e(t)
garch_pred<-rev(best_garch$fitted.values[-(1:(n_garch_res-q_arch)),1]*best_garch$res[-(1:(n_garch_res-q_arch))])
#Υπολογισμός της εκτιμώμενης, από το μοντέλο GARCH, διασποράς σ(t+1)
standar_dev_pred<-sqrt(sum(c(best_garch$coef)*c(1,garch_pred^2)))
}
}

#Σε περίπτωση που δεν βρεθεί κάποιο μοντέλο GARCH, τότε στη μεταβλητή best_garch δίνεται η χρονοσειρά που
προέκυψε μετά την εξάλειψη της τάσης και την εφαρμογή του μοντέλου ARMA
if (p_garch+q_arch==0)
{
best_garch<-garch_ts
}

#Υπολογισμός κύρτωσης και λοξότητας της κατανομής των καταλοίπων του μοντέλου GARCH
if(p_garch+q_arch==0)
{
garch_kurt=best_garch
garch_skew=best_garch
} else {
garch_kurt=best_garch$res[!is.na(best_garch$res) & !is.nan(best_garch$res)]
garch_skew=best_garch$res[!is.na(best_garch$res) & !is.nan(best_garch$res)]
}

if (kurtosis(garch_kurt)>3)
{
kurt<-"Leptokurtic: Fat tails"
} else if (kurtosis(garch_kurt)<3) {
kurt<-"Platykurtic: Thin tails"
} else {
kurt<-"Mesokurtic: Normal distribution"
}

if (skewness(garch_skew)>0)
{

```

```

skew<-"Positive Skew: Fat right tail"
} else if (skewness(garch_skew)<0) {
skew<-"Negative Skew: Fat left tail"
} else{
skew<-"Symmetrical Distribution"
}

#####

#Εφαρμογή EVT στην (κανονικοποιημένη) χρονοσειρά των σφαλαμάτων που υπολογίστηκε από την εφαρμογή του μοντέλου
GARCH
if(p_garch+q_arch!=0)
{
gpd_data_mean<-mean(best_garch$res[!is.na(best_garch$res) & !is.nan(best_garch$res)])
gpd_data_sd<-sd(best_garch$res[!is.na(best_garch$res) & !is.nan(best_garch$res)])
gpd_data<-(best_garch$res[!is.na(best_garch$res) & !is.nan(best_garch$res)]-gpd_data_mean)/gpd_data_sd

} else{
gpd_data_mean<-mean(best_garch)
gpd_data_sd<-sd(best_garch)
gpd_data<-(best_garch-gpd_data_mean)/gpd_data_sd
}

#Πλήθος δείγματος δεδομένων
n<-length(gpd_data)

if (graph==T)
{
#Γραφική παράσταση προς εξέταση τιμών
dev.new()
par(mfrow=c(1,1))
barplot(as.numeric(gpd_data), main="Γραφική παράσταση χρονοσειράς καταλοίπων\η για τα οποία εφαρμόζεται η
EVT")

#Γραφική παράσταση μέσης υπερβάλλουσας συνάρτησης
dev.new()
mrlplot(gpd_data[gpd_data>-0.5],main="Γράφημα μέσης υπερβάλλουσας συνάρτησης")

#Γραφικές παραστάσεις παραμέτρων σ, ξ για την σωστή επιλογή του κατωφλιού
dev.new()
gpd.fitrange(gpd_data,-1.5,2,nint = 20)
}

#Εκτίμηση παραμέτρων
if (threshold_val==Inf)
{
if (n_ordered_p==0)
{
n_ordered_p=as.integer(n*0.1) #10% των ακραίων παρατηρήσεων χρησιμοποιούνται για την προσαρμογή της GPD
}
gpd_fit<-gpdFit(gpd_data,nextremes=n_ordered_p, method = "mle", npp = period,plot=FALSE)
} else{
gpd_fit<-gpdFit(gpd_data,threshold = threshold_val, method = "mle", npp = period,plot=FALSE)
}

#QQplot για την περίπτωση της GPD
if (graph==T)
{
dev.new()
par(mfrow=c(1,1))
qqgpd(gpd_data,nextremes = gpd_fit$n.exceed,scale = gpd_fit$par.est$[1],shape = gpd_fit$par.est$[2])
}

#Υπολογισμός δ.ε. για την παράμετρο ξ χρησιμοποιώντας Profile Deviance Function
if (graph==T)
{
dev.new()
par(mfrow=c(1,1))
gpdProfShape(gpd_fit)
}

```

```

#Έλεγχος Cramer-von Mises για το αν τα δεδομένα, πάνω από το επιλεγμένο κατώφλι u, προέρχονται από μια GPD
κατανομή
cat("Έλεγχος Cramer-von Mises για το αν τα δεδομένα, πάνω από το επιλεγμένο κατώφλι u, προέρχονται από μια
κατανομή GPD:\n")
print(gpdCvm(gpd_data[gpd_data>=gpd_fit$threshold])
cat("===== \n")
}

#Εκτίμηση στάθμης απόδοσης. Η μεταβλητή period εκφράζει τον (μέσο) αριθμό ημερών που απαιτούνται
#ώστε μια Παρατήρηση να λάβει τιμή μεγαλύτερη από το κατώφλι u.
#Επιλέγουμε η μεταβλητή period να είναι περίπου το σύνολο του δείγματος μας. Το δείγμα χωρίζεται (συνήθως) σε
#blocks των 253 ημερών
return_level<-gpdRl(gpd_fit, period = 1/((1-sign_level)*gpd_fit$np), method = "profile",plot=FALSE)
return_level<-return_level$Estimate

#Υπολογισμός TVaR (ή expected shortfall)
sigma<-gpd_fit$par.sum[1,1]
xi<-gpd_fit$par.sum[2,1]
expected_sh<-return_level*(1/(1-xi)+(sigma-xi*gpd_fit$threshold)/((1-xi)*return_level))

#####

#Υπολογισμός VaR και TVaR για την μη-κανονικοποιημένη χρονοσειρά των σφαλμάτων. Επίσης, υπολογίζεται το VaR για
την περίπτωση που υποθέσουμε ότι τα σφάλματα
#ακολουθούν κανονική κατανομή με μ=0 και σ=1
return_VaR<-return_level*gpd_data_sd+gpd_data_mean
expected_sh_TVaR<-expected_sh*gpd_data_sd+gpd_data_mean
Normal_VaR<-qnorm(sign_level,0,1)

#Υπολογισμός VaR για τη χρονοσειρά των αποδόσεων, χρησιμοποιώντας τα μοντέλα
#Trend+ARMA+GARCH+EVT
#GARCH+Normal distribution
#Normal distribution
#Sample quantile
if(p_garch+q_arch!=0)
{
pred_VaR_evt<-pred+standar_dev_pred*return_VaR+mean_garch
pred_TVaR_evt<-pred+standar_dev_pred*expected_sh_TVaR+mean_garch
pred_VaR_garch_Normal<-pred+standar_dev_pred*Normal_VaR+mean_garch
pred_VaR_Normal<-qnorm(sign_level,mean = enorm(as.numeric(data_re),method = "mle")$parameters[1],sd =
enorm(as.numeric(data_re),method = "mle")$parameters[2])
pred_VaR_garch_quantile<-quantile(data_re,probs = sign_level)
} else{
pred_VaR_evt<-pred+return_VaR+mean_garch
pred_TVaR_evt<-pred+expected_sh_TVaR+mean_garch
pred_VaR_garch_Normal<-pred+Normal_VaR+mean_garch
pred_VaR_Normal<-qnorm(sign_level,mean = enorm(as.numeric(data_re),method = "mle")$parameters[1],sd =
enorm(as.numeric(data_re),method = "mle")$parameters[2])
pred_VaR_garch_quantile<-quantile(data_re,probs = sign_level)
}

#Λίστα με τα τελικά αποτελέσματα
results<-list("Volume of data"=n,"Trend"=trend,"Kurtosis"=kurt,"Kurtosis
result"=kurtosis(garch_kurt),"Skewness"=skew,"Skewness result"=skewness(garch_skew),"VaR prediction
(Trend+ARMA+GARCH+EVT)"=pred_VaR_evt,"TVaR prediction (Trend+ARMA+GARCH+EVT)"=pred_TVaR_evt,"VaR
prediction (Trend+ARMA+GARCH+Normal)"=pred_VaR_garch_Normal,"VaR prediction
(Normal)"=pred_VaR_Normal,"VaR prediction (Quantile)"=pred_VaR_garch_quantile,"GPD
estimates"=gpd_fit,"Confidence interval for ξ"= gpdProfShape(gpd_fit)$CI,"Volume of used points"=
gpd_fit$n.exceed,"threshold"=gpd_fit$threshold,"Garch order"=c(p_garch,q_arch),"Return_level χρονοσειράς
καταλοίπων GARCH"=return_level,"Arima order"=arimaorder(arima_model))
return(results)
options(warn = defaultW)
}

#####

#Εκ των υστέρων έλεγχος (Backtesting)
#Γίνεται σύγκριση των εκτιμώμενων τιμών με αυτές του έτους end_d+1
#back_testing(start_d="1970-01-01",end_d="1973-12-31",end_d_backtest="2020-12-31",asset_name="^GSPC")
#Η συνάρτηση back_testing ελέγχει την καταλληλότητα του μοντέλου GARCH+EVT χρησιμοποιώντας ιστορικά δεδομένα

```

```

#Συγκεκριμένα στα ορίσματα start_d και end_d δίνονται οι ημερομηνίες για το σύνολο των δεδομένων που θέλουμε να
λάβουμε υπόψη για τον προσδιορισμό
#των παραμέτρων του μοντέλου, ενώ στο όρισμα end_d_backtest δίνεται η ημερομηνία που θα τελειώσει το backtesting.
#Τα ορίσματα sign_level και bt_sign_level αναφέρονται στο επίπεδο εμπιστοσύνης για τον υπολογισμό VaR και το επίπεδο
εμπιστοσύνης για τον έλεγχο καταλληλότητας
#του μοντέλου χρησιμοποιώντας τη διωνυμική κατανομή.
#Ο έλεγχος του μοντέλου πραγματοποιείται ως εξής:
#αφού εκτιμηθούν οι παράμετροι του μοντέλου που χρησιμοποιούμε, στη συνέχεια γίνεται εκτίμηση της απόδοσης του
χρηματοοικονομικού στοιχείου για την επόμενη ημέρα,
#και η τιμή αυτή συγκρίνεται με την πραγματική. Σε περίπτωση που η πραγματική τιμή βρίσκεται κάτω από το εκτιμώμενο
VaR, τότε θεωρούμε επιτυχή την πρόβλεψη του μοντέλου
#Παράδειγμα: back_testing(start_d="2010-01-01",end_d="2013-12-31",end_d_backtest="2021-12-
31",asset_name="^GSPC",sign_level=0.95,bt_sign_level=0.05,garch_zero_order=F,show_graph=F)

```

```

back_testing<-
function(start_d,end_d,end_d_backtest,asset_name,sign_level=0.95,bt_sign_level=0.05,garch_zero_order=F,show_gra
ph=F, period_evt=253){

```

```

  from.dat= as.Date(start_d,format = "%Y-%m-%d")
  to.dat1= as.Date(end_d,format = "%Y-%m-%d")
  to.dat2= as.Date(end_d_backtest,format = "%Y-%m-%d")
  data1<-getSymbols(asset_name,from=from.dat,to=to.dat1,periodicity="daily",auto.assign=F)
  data2<-getSymbols(asset_name,from=from.dat,to=to.dat2,periodicity="daily",auto.assign=F)
  date_count1<-length(index(data1))
  date_count2<-length(index(data2))
  date<-date_count2-date_count1

```

```

  if (is.na(summary(is.na(getSymbols(Symbols = asset_name,from=from.dat,to=
to.dat1,periodicity="daily",auto.assign=F)))[,6][3])!=FALSE)
  {
    stop("Τερματισμός κώδικα. Τα δεδομένα περιέχουν NA τιμές")
  }

```

```

  #Μετράει πόσες φορές η πραγματική τιμή έλαβε τιμή μικρότερη από το VaR υπολογισμένο βάση
Trend+ARMA+GARCH+EVT
count_VaR_evt<-0

```

```

  #Μετράει πόσες φορές η πραγματική τιμή έλαβε τιμή μικρότερη από το TVaR υπολογισμένο βάση
Trend+ARMA+GARCH+EVT
count_TVaR_evt<-0

```

```

  #Μετράει πόσες φορές η πραγματική τιμή έλαβε τιμή μικρότερη από το VaR υπολογισμένο βάση
Trend+ARMA+GARCH+Normal
count_garch_Normal_VaR<-0

```

```

  #Μετράει πόσες φορές η πραγματική τιμή έλαβε τιμή μικρότερη από το VaR υπολογισμένο βάση της κανονικής κατανομής
count_Normal_VaR<-0

```

```

  #Μετράει πόσες φορές η πραγματική τιμή έλαβε τιμή μικρότερη από το VaR υπολογισμένο βάση δειγματικού ποσοστιαίου
σημείου
count_quantile_VaR<-0

```

```

  #Μετράει το πλήθος των ημερών που γίνεται το back-testing
data3<-getSymbols(asset_name,from=index(data2)[date_count1+1],to=to.dat2,periodicity="daily",auto.assign=F)
data3=-dailyReturn(data3[,6],type="log")
count_backtesting_days=length(data3)

```

```

  #Μετράει τις φορές που δεν έγινε υπολογισμός λόγω σφάλματος του κώδικα
code_err<-0

```

```

  #Υπολογισμός σωστών προβλέψεων
for (i in 1:date)
  {
    skip_to_next<-FALSE
    if(-dailyReturn(data2[,6],type="log")[date_count1+i]>0)
    {
      tryCatch(var_prediction<-predict_VaR(start_date = index(data2)[i],end_date = index(data2)[date_count1-
1+i],asset_name,sign_level,garch_zero=garch_zero_order,graph=show_graph, period = period_evt),
      error=function(err){
        code_err<-code_err+1
        skip_to_next<-TRUE})
      if(skip_to_next)
    }
  }

```

```

{
  next
}

if(var_prediction[7]>-dailyReturn(data2[,6],type="log")[date_count1+i])
{
  count_VaR_evt<-count_VaR_evt+1
}
if(var_prediction[8]>-dailyReturn(data2[,6],type="log")[date_count1+i])
{
  count_TVaR_evt<-count_TVaR_evt+1
}
if(var_prediction[9]>-dailyReturn(data2[,6],type="log")[date_count1+i])
{
  count_garch_Normal_VaR<-count_garch_Normal_VaR+1
}
if(var_prediction[10]>-dailyReturn(data2[,6],type="log")[date_count1+i])
{
  count_Normal_VaR<-count_Normal_VaR+1
}
if(var_prediction[11]>-dailyReturn(data2[,6],type="log")[date_count1+i])
{
  count_quantile_VaR<-count_quantile_VaR+1
}
}
cat("# υπολειπόμενων επαναλήψεων:",date-i,"\n")
}

cat("\nΣυνολικές αποτυχίες κώδικα:",code_err,"\n")
#Αφαίρεση τώσων ημερών (από τις συνολικές backtesting ημέρες), όσες οι φορές που δεν εκτελέστηκε ο κώδικας
count_backtesting_days<-count_backtesting_days-code_err

#Επιστρέφει το ποσοστό επιτυχίας
results<-list("% Correct VaR prediction Trend+ARMA+GARCH+EXT:
"=round((count_VaR_evt+length(data3[data3<0]))/count_backtesting_days,digits = 2),"% Correct TVaR prediction
Trend+ARMA+GARCH+EXT: " =round((count_TVaR_evt+length(data3[data3<0]))/count_backtesting_days,digits =
2),"% Correct VaR prediction Trend+ARMA+GARCH+Normal:
"=round((count_garch_Normal_VaR+length(data3[data3<0]))/count_backtesting_days,digits = 2),"% Correct VaR
prediction Normal distribution: " =round((count_Normal_VaR+length(data3[data3<0]))/count_backtesting_days,digits
= 2),"% Correct VaR prediction sample quantile:
"=round((count_quantile_VaR+length(data3[data3<0]))/count_backtesting_days,digits = 2),"Total backtested days:
"=count_backtesting_days)

#Διωνυμικός έλεγχος για τον έλεγχο καταλληλότητας των μεθόδων
test_resultl=binom.test(count_VaR_evt+length(data3[data3<0]),count_backtesting_days,sign_level)
if (test_resultl$p.value < bt_sign_level)
{
  cat("\nΑπορρίπτεται το μοντέλο για την περίπτωση των VaR προβλέψεων βάση Trend+ARMA+GARCH+EVT:\n")
  print(test_resultl)
} else {
  cat("\nΑποδεκτό μοντέλο για την περίπτωση των VaR προβλέψεων βάση Trend+ARMA+GARCH+EVT:\n")
  print(test_resultl)
}

test_resultl=binom.test(count_TVaR_evt+length(data3[data3<0]),count_backtesting_days,sign_level)
if (test_resultl$p.value < bt_sign_level)
{
  cat("\nΑπορρίπτεται το μοντέλο για την περίπτωση των TVaR προβλέψεων βάση Trend+ARMA+GARCH+EVT:\n")
  print(test_resultl)
} else {
  cat("\nΑποδεκτό μοντέλο για την περίπτωση των TVaR προβλέψεων βάση Trend+ARMA+GARCH+EVT:\n")
  print(test_resultl)
}

test_resultl=binom.test(count_garch_Normal_VaR+length(data3[data3<0]),count_backtesting_days,sign_level)
if ( test_resultl$p.value < bt_sign_level)
{
  cat("\nΑπορρίπτεται το μοντέλο για την περίπτωση των VaR προβλέψεων βάση Trend+ARMA+GARCH+Normal:\n")
  print(test_resultl)
} else {

```

```

cat("\nΑποδεκτό μοντέλο για την περίπτωση των VaR προβλέψεων βάση Trend+ARMA+GARCH+Normal:\n")
print(test_resutl)
}

test_resutl=binom.test(count_Normal_VaR+length(data3[data3<0]),count_backtesting_days,sign_level)
if ( test_resutl$p.value < bt_sign_level)
{
cat("\nΑπορρίπτεται το μοντέλο για την περίπτωση των VaR προβλέψεων βάση κανονικής κατανομής:\n")
print(test_resutl)
} else {
cat("\nΑποδεκτό μοντέλο για την περίπτωση των VaR προβλέψεων βάση κανονικής κατανομής:\n")
print(test_resutl)
}

test_resutl=binom.test(count_quantile_VaR+length(data3[data3<0]),count_backtesting_days,sign_level)
if ( test_resutl$p.value < bt_sign_level)
{
cat("\nΑπορρίπτεται το μοντέλο για την περίπτωση των VaR προβλέψεων βάση του δειγματικού ποσοστιμορίου:\n")
print(test_resutl)
} else {
cat("\nΑποδεκτό μοντέλο για την περίπτωση των VaR προβλέψεων βάση του δειγματικού ποσοστιμορίου:\n")
print(test_resutl)
}

return(results)
}

```



# ΒΙΒΛΙΟΓΡΑΦΙΑ

## Ελληνική

Βρόντος, Ι. (2006). Εφαρμογές Στατιστικών Μοντέλων στα Χρηματοοικονομικά, Οικονομικό Πανεπιστήμιο Αθηνών, Αθήνα.

Μπούτσικας, Μ. (2008). Σημειώσεις διαλέξεων στη «Θεωρία ακραίων τιμών», Πανεπιστήμιο Πειραιώς, Πειραιάς.

Μπούτσικας, Μ. (2019-2020). Εισαγωγή στην Ανάλυση Χρονοσειρών, Πανεπιστήμιο Πειραιώς, Πειραιάς.

## Ξενόγλωσση

Beguin, J. M., Gouriéroux, C. and Monfort, A., (1980). Identification of a Mixed Autoregressive-Moving Average Process : The Corner Method, In Time Series, O. D. Anderson, Ed., Amsterdam, North-Holland, pp. 423-436.

Beirlant, J., Goegebeur, Y., Segers, J., Teugels, J., De Waal, D. and Ferro, C. (2004). Statistics of Extremes Theory and Applications, John Wiley & Sons Ltd, England.

Bollerslev, T. Ray, Y. C. and Kenneth, F. K. (1992). ARCH Modeling in Finance, Journal of Econometrics, 52, 5-59, North-Holland.

Box, G. E. P., Jenkins, G., Reinsel, G. C. and Ljung, G. M. (2016). Time Series Analysis: Forecasting and Control, 4<sup>th</sup> ed., John Wiley & Sons, Inc., Hoboken, New Jersey.

Coles, S. (2001). An Introduction to Statistical Modeling of Extreme Values, Springer, London.

Cowpertwait, P. S. P. and Metcalfe, A. V. (2009). Introductory Time Series with R, Springer, New York.

Embrechts, P., Kluppelberg, C. and Mikosch, T. (1997). Modelling Extremal Events for Insurance and Finance, Springer, New York.

Engle, R. (1982). Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation, Econometrica, Volume 50, Issue 4, 987-1008.

- Francq, C. and Zakoian, J. M. (2019). *GARCH Models Structure, Statistical Inference and Financial Applications*, 2<sup>nd</sup> ed., John Wiley & Sons Ltd., United Kingdom.
- Gooijer, J. G. and Heuts, R. M. J. (1981). The Corner Method: An Investigation of an Order Discrimination Procedure for General ARMA Processes, *Journal of the Operational Research Society*, 32:11, 1039-1042.
- Haan, L. and Ferreira, A (2006). *Extreme Value Theory: An Introduction*, Springer, New York.
- Longin, F. (2017). *Extreme Events in Finance: A Handbook of Extreme Value Theory and its Applications*, John Wiley & Sons, Inc., Hoboken, New Jersey.
- McNeil, A. J. and Frey, R. (2000). Estimation of Tail-Related Risk Measures for Heteroscedastic Financial Time Series: An Extreme Value Approach, *Journal of Empirical Finance* 7 (2000) 271–300.
- Smith, R. L. (1985). Maximum likelihood estimation in a class of non-regular cases, *Biometrika*, 72, 1, pp. 67-90.
- Tsay, R.S. (2002). *Analysis of Financial Time Series in Probability and Statistics*, John Wiley & Sons, Inc., New York.
- Xekalaki, E. and Degiannakis, S. (2010). *ARCH Models for Financial Applications*, John Wiley & Sons Ltd., Great Britain.
- Xiangxia, Z. and Wenlei, G. (2009). A New Method to Choose the Threshold in the POT Model, *The 1st International Conference on Information Science and Engineering*.





