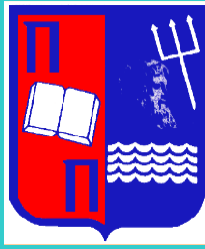


# ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΕΙΡΑΙΩΣ

Σχολή Χρηματοοικονομικής και Στατιστικής



Τμήμα Στατιστικής  
και Ασφαλιστικής Επιστήμης

ΜΕΤΑΠΤΥΧΙΑΚΟ ΠΡΟΓΡΑΜΜΑ ΣΠΟΥΔΩΝ  
ΣΤΗΝ ΕΦΑΡΜΟΣΜΕΝΗ ΣΤΑΤΙΣΤΙΚΗ

Μελέτη ακραίων παρατηρήσεων μέσω της  
μεθόδου υπερβάσεων κατωφλιού και  
εφαρμογές στον κίνδυνο αγοράς

Ηλίας Γεώργιος Π. Τρανάκος

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 *Εφαρμοσμένη Στατιστική*

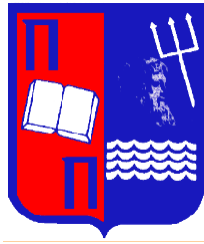
Πειραιάς

Οκτώβριος 2025



# ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΕΙΡΑΙΩΣ

Σχολή Χρηματοοικονομικής και Στατιστικής



Τμήμα Στατιστικής  
και Ασφαλιστικής Επιστήμης

ΜΕΤΑΠΤΥΧΙΑΚΟ ΠΡΟΓΡΑΜΜΑ ΣΠΟΥΔΩΝ  
ΣΤΗΝ ΕΦΑΡΜΟΣΜΕΝΗ ΣΤΑΤΙΣΤΙΚΗ

**Μελέτη ακραίων παρατηρήσεων μέσω της  
μεθόδου υπερβάσεων κατωφλιού και  
εφαρμογές στον κίνδυνο αγοράς**

**Ηλίας Γεώργιος Π. Τρανάκος**

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 *Εφαρμοσμένη Στατιστική*

Πειραιάς

Οκτώβριος 2025

Η παρούσα Διπλωματική Εργασία εγκρίθηκε ομόφωνα από την Τριμελή Εξεταστική Επιτροπή που ορίστηκε από τη ΓΣΕΣ του Τμήματος Στατιστικής και Ασφαλιστικής Επιστήμης του Πανεπιστημίου Πειραιώς στην υπ' αριθμ. 22<sup>η</sup> / 7-Ιουλίου 2021 συνεδρίασή του σύμφωνα με τον Εσωτερικό Κανονισμό Λειτουργίας του Προγράμματος Μεταπτυχιακών Σπουδών στην Εφαρμοσμένη Στατιστική

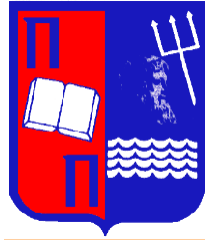
Τα μέλη της επιτροπής ήταν:

1. Μιχαήλ Μπούτσικας, Αναπληρωτής Καθηγητής (Επιβλέπων)
2. Γεώργιος Ψαράκος, Καθηγητής
3. Γεώργιος Τζαβελάς, Αναπληρωτής Καθηγητής

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

**UNIVERSITY OF PIRAEUS**

**School of Finance and Statistics**



**Department of Statistics and Insurance Science**

**POSTGRADUATE PROGRAM IN  
APPLIED STATISTICS**

**Modelling Extreme observations via Peaks  
over Threshold method with applications in  
Market Risk**

By

**Ilias Georgios P. Tranakos**

MSc Dissertation

submitted to the Department of Statistics and Insurance Science of the University of Piraeus in partial fulfilment of the requirements for the degree of Master of Science in Applied Statistics

**Piraeus, Greece  
October 2025**



## **Ευχαριστίες**

Σε αυτό το σημείο θα ήθελα να εκφράσω ένα πολύ μεγάλο και ειλικρινές ευχαριστώ στον επιβλέπων καθηγητή μου και σπουδαίο άνθρωπο, Κύριο Μπούτσικα Μιχαήλ για όλη του την υπομονή, την καθοδήγηση και τις πολύτιμες συμβουλές του.

Ένα πολύ μεγάλο ευχαριστώ επίσης θα ήθελα να εκφράσω και στην οικογένεια μου που ήταν πάντα και παντού δίπλα μου.



## Περίληψη

Η εργασία επικεντρώνεται στη Θεωρία Ακραίων Τιμών (Extreme Value Theory - EVT) και στην εφαρμογή της σε χρηματοοικονομικά δεδομένα. Αρχικά παρουσιάζονται οι βασικές μεθοδολογίες, όπως η μέθοδος μεγίστων ανά περίοδο (Block Maxima) και η μέθοδος υπερβάσεων κατωφλίου (Peaks Over Threshold ή απλά POT). Αναλύεται η θεωρητική βάση τους, οι κατανομές που χρησιμοποιούνται, πιο συγκεκριμένα οι κατανομές Gumbel, Frechet, Weibull καθώς και η Γενικευμένη Κατανομή Ακραίων Τιμών (Generalized Extreme Value Distribution ή απλά GEV) και η Γενικευμένη Κατανομή Pareto (Generalized Pareto Distribution ή απλά GPD). Επίσης παρουσιάζονται και τεχνικές εκτίμησης παραμέτρων.

Στη συνέχεια, παρουσιάζονται τα βασικά μέτρα κινδύνου, η Αξία σε Κίνδυνο (Value at Risk ή απλά VaR) και το Αναμενόμενο Έλλειμα (Expected Shortfall ή απλά ES), τα οποία χρησιμοποιούνται για την ποσοτικοποίηση του επενδυτικού κινδύνου. Γίνεται επίσης αναφορά στη διαδικασία εκ των υστέρων ελέγχου (backtesting), ώστε να ελεγχθεί η ακρίβεια και η αξιοπιστία των εκτιμήσεων σε ιστορικά δεδομένα.

Στόχος της διπλωματικής είναι η σύγκριση εμπειρικά των τριών μεθόδων (Block Maxima, POT, Διασποράς-Συνδιασποράς) με χρήση ιστορικών δεδομένων. Η ανάλυση στοχεύει στο να προσδιορίσει ποια μέθοδος είναι αποτελεσματικότερη για την εκτίμηση των ακραίων κινδύνων και να προσφέρει συμπεράσματα σχετικά με την καταλληλότητα κάθε προσέγγισης σε συνθήκες έντονης μεταβλητότητας.

Τέλος όλα τα παραπάνω εφαρμόζονται σε πραγματικά δεδομένα μέσω του στατιστικού πακέτου R. Συγκρίνουμε τις τρεις μεθόδους και εξάγουμε συμπεράσματα για την αποτελεσματικότητα τους για τις μετοχές της Amazon, Apple, Netflix αλλά και τις τιμές του Bitcoin.

**Λέξεις κλειδιά:** ακραίες παρατηρήσεις, υπερβάσεις κατωφλίου, κίνδυνος αγοράς, αξία σε κίνδυνο, αναμενόμενο έλλειμα, γενικευμένη Παρέτο κατανομή, εκ των υστέρων έλεγχος, μετοχές, κρυπτονομίσματα, μέγιστες τιμές περιόδων

## Abstract

This master thesis focuses on Extreme Value Theory (EVT) and its application to finance. Initially, the basic methodologies are presented, such as the Block Maxima and the Peaks Over Threshold method. Their theoretical basis, the distributions used (Gumbel, Frechet, Weibull, GEV, GPD) and the parameter estimation techniques are analyzed.

Next, the most common risk measures, Value at Risk (VaR) and Expected Shortfall (ES), which are used to quantify investment risk, are presented. Reference is also made to backtesting, in order to check the accuracy and reliability of the estimates using historical data.

The aim of this thesis is to empirically compare three methods: Block Maxima, POT and Variance-Covariance using stock prices. The analysis aims to determine which method is more effective for estimating extreme risks and to offer conclusions regarding the suitability of each approach under certain conditions.

Finally, the above are applied to real data employing the statistical package R. We compare the above-mentioned methods and draw conclusions about their effectiveness for Amazon, Apple, Netflix stocks as well as Bitcoin prices.

**Key words:** extreme values, peaks over threshold, market risk, value at risk, expected shortfall, generalized Pareto distribution, backtesting, shares, cryptocurrencies, block maxima

# Περιεχόμενα

## 1 Θεωρία Ακραίων Τιμών

### 1.1. Μέθοδος Block Maxima

1.1.1 Διαίρεση των δεδομένων

1.1.2 Ανάλυση ακραίων τιμών

1.1.3 Πρόβλεψη ακραίων τιμών

### 1.2. Η μέθοδος Peaks Over Threshold (POT)

1.2.1 Μοντελοποίηση των ακραίων τιμών

1.2.2 Καθορισμός του κατωφλίου (threshold)

1.2.3 Εκτίμηση παραμέτρων της GPD

1.2.4. Πρόβλεψη ακραίων τιμών

### 1.3. Η μέθοδος Variance-Covariance

1.3.1. Υπολογισμός του μέσου, των αποκλίσεων και των των διαφορών

1.3.2. Ανάλυση των ακραίων τιμών:

## 2. Μέτρα κινδύνου Value At Risk και Expected Shortfall (VaR και ES)

### 2.1 Value at Risk (VaR)

### 2.2. Expected Shortfall (ES)

### 2.3. Backtesting

## 3. Εφαρμογή

### 3.1. Εκτίμηση VaR και ES επενδύσεων σε μετοχές (μετοχή NFLX)

3.1.1. Καταγραφή και επεξεργασία δεδομένων

3.1.2 Παραμετρική Μέθοδος

3.1.3 Απαραμετρική Μέθοδος

3.1.4 Μέθοδος POT

3.1.5. Σύγκριση μεθόδων

3.1.6. Εκ των Υστέρων έλεγχος

3.1.7. Συμπεράσματα

### 3.2. Εφαρμογή με βάση τις τιμές της μετοχής της APPLE

3.2.1 Εκτιμήσεις των μέτρων κινδύνου

3.2.2. Συμπεράσματα

### 3.3. Εφαρμογή με βάση τις τιμές της μετοχής της AMAZON

3.3.1 Εκτιμήσεις των μέτρων κινδύνου

3.3.2. Συμπεράσματα

3.4. Εφαρμογή με βάση τις τιμές του BITCOIN

3.4.1 Εκτιμήσεις των μέτρων κινδύνου

3.4.2. Συμπεράσματα

3.5. Συμπεράσματα

## **ΚΕΦΑΛΑΙΟ 1.**

### **Η θεωρία ακραίων τιμών**

Η θεωρία ακραίων τιμών, επίσης γνωστή ως Extreme Value Theory (EVT), αποτελεί ένα πλαίσιο ανάλυσης που επικεντρώνεται στη μελέτη των ακραίων παρατηρήσεων σε ένα σύνολο δεδομένων. Αυτή η θεωρία αναζητά πώς και γιατί εμφανίζονται αυτές οι ακραίες τιμές, καθώς και πώς επηρεάζουν τη στατιστική ανάλυση και την ερμηνεία των δεδομένων.

Οι ακραίες τιμές αναφέρονται σε παρατηρήσεις που ξεχωρίζουν από την υπόλοιπη μάζα των παρατηρήσεων στο σύνολο δεδομένων, είτε λόγω της μεγάλης τους απόκλισης από τη μέση τιμή του συνόλου είτε λόγω της συχνότητας εμφάνισής τους σε σχέση με το υπόλοιπο δείγμα. Οι ακραίες τιμές μπορεί να οφείλονται σε διάφορους παράγοντες, όπως λάθη μέτρησης, ανωμαλίες στα δεδομένα ή ακόμα και πραγματικά ασυνήθιστα γεγονότα.

Η ανάλυση των ακραίων τιμών είναι σημαντική γιατί μπορεί να επηρεάσει σημαντικά τα στατιστικά μέτρα κεντρικής τάσης και διασποράς, όπως η μέση τιμή και η τυπική απόκλιση. Επίσης, η παρουσία ακραίων τιμών μπορεί να επηρεάσει την αξιοπιστία των στατιστικών εκτιμήσεων και των συμπερασμάτων που προκύπτουν από την ανάλυση δεδομένων.

Υπάρχουν διάφορες μέθοδοι για τη διαχείριση των ακραίων τιμών, συμπεριλαμβανομένων της αφαίρεσής τους από το σύνολο δεδομένων, της αντικατάστασής τους με πιο τυπικές τιμές ή της εφαρμογής μεθόδων που είναι ανθεκτικές στις ακραίες τιμές κατά την ανάλυση αλλά και η προσπάθεια μοντελοποίησης τους.

Συνολικά, η θεωρία ακραίων τιμών αποτελεί ένα σημαντικό εργαλείο στη στατιστική ανάλυση δεδομένων, καθώς βοηθά στην αναγνώριση, τη διαχείριση και την ερμηνεία των ακραίων παρατηρήσεων που μπορεί να επηρεάσουν τα αποτελέσματα μιας μελέτης.

Σε όσα θεωρητικά αποτελέσματα ακολουθούν χρησιμοποιήθηκαν ως βάση κατά κύριο λόγο τα συγγράμματα των Beirlant et al. (2004), Reiss and Thomas (2007), Rocco (2013), Georgescu (2009), Boutsikas (2020), Hull (2015), Dowd (2002), Danielsson (2011), Roncalli (2020).

## 1.1. Μέθοδος μέγιστων τιμών περιόδων (Block Maxima)

Η μέθοδος "block maxima" (μέγιστες τιμές περιόδων) είναι μια τεχνική που χρησιμοποιείται στη στατιστική και στον τομέα της ανάλυσης ακραίων τιμών (extreme value analysis) για τη μοντελοποίηση και την πρόβλεψη ακραίων γεγονότων σε μεγάλα δεδομένα. Αυτή η μέθοδος συχνά χρησιμοποιείται σε περιπτώσεις όπου οι ακραίες τιμές των δεδομένων αποτελούν το αντικείμενο ενδιαφέροντος, όπως για παράδειγμα στη μελέτη των ακραίων καιρικών φαινομένων, στη χρηματοοικονομική επιστήμη για τη μοντελοποίηση ακραίων αποτιμήσεων ή και στον ασφαλιστικό κλάδο για την πρόβλεψη (π.χ. βλ. Reiss and Thomas (2007)).

Έστω οι παρατηρήσεις  $X_1, X_2, \dots, X_m \sim F$  (μία άγνωστη κατανομή). Έχουμε την δυνατότητα να θεωρήσουμε ότι οι  $X_1, X_2, \dots, X_m, m = kn$  διαχωρίζονται σε  $k$  τον αριθμό υποσύνολα (blocks) από  $n$  παρατηρήσεις το κάθε σύνολο.

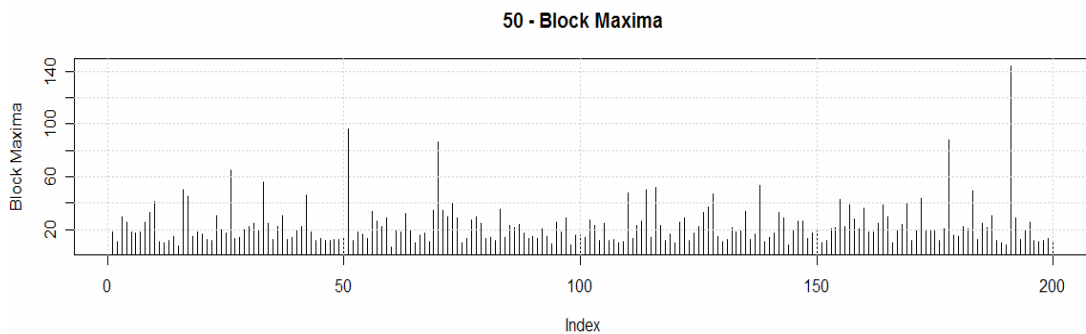
$$X_1, X_2, \dots, X_n | X_{n+1}, X_{n+2}, \dots, X_{n+n} \dots | X_{(k-1)n+1}, X_{(k-1)n+2}, \dots, X_{(k-1)n+n}$$

### 1.1.1. Διαίρεση των δεδομένων

Τα δεδομένα χρονοσειρών ή διανυσμάτων δειγμάτων διαχωρίζονται σε ομάδες (μπλοκ), συνήθως μη επικαλυπτόμενα. Σε κάθε ομάδα, επιλέγεται η μέγιστη τιμή (το μέγιστο) ως αντιπροσωπευτική τιμή για εκείνη την ομάδα.

Συμβολίζουμε με  $Y_1, Y_2, \dots, Y_k$ , τις μέγιστες τιμές για κάθε ένα από τα παραπάνω,  $k$  υποσύνολα, τις οποίες και καλούμε block maxima (μέγιστα περιόδων ή ομάδων).

Για να οπτικοποιήσουμε τα blocks, ένα τυχαίο παράδειγμα διαγράμματος με  $n = 50, k = 200$  έχει την μορφή που φαίνεται στο ακόλουθο σχήμα.



Σχήμα 1.1

### 1.1.2. Ανάλυση ακραίων τιμών

Οι τιμές των μέγιστων από κάθε ομάδα (μπλοκ) ακολουθούν μια κατανομή ακραίων τιμών.

Από το θεώρημα *Fisher-Tippett* (π.χ. βλ. Beirlant et al. (2004)) σύμφωνα με το οποίο, υποθέτοντας μία σειρά  $n$  ισόνομων και ανεξάρτητων τυχαίων μεταβλητών, εάν οι κατάλληλα κανονικοποιημένες μέγιστες τιμές αυτών των μεταβλητών συγκλίνουν σε μια μη τετριμμένη κατανομή καθώς το  $n$  τείνει στο άπειρο, τότε η κατανομή αυτή ανήκει σε μία από τις τρεις κατηγορίες παρακάτω. Συγκεκριμένα, μπορούμε να ισχυριστούμε τα ακόλουθα:

Τα (κανονικοποιημένα) block maxima ακολουθούν μία από τις ακόλουθες κατανομές

$\Lambda((x - d)/c)$  (κατανομή Gumbel)

$\Phi_a((x - d)/c)$  (κατανομή Frechet)

$\Psi_a((x - d)/c)$  (κατανομή Weibull)

ανάλογα με το αν  $F \in MDA(\Lambda)$  ή  $MDA(\Phi_a)$  ή  $MDA(\Psi_a)$  αντίστοιχα, δηλαδή ανάλογα με το αν η κατανομή των αρχικών παρατηρήσεων ανήκει στην περιοχή έλξης (MDA) της Gumbel, Frechet ή Weibull αντίστοιχα. Οι ορισμοί των κατανομών αυτών δίνονται παρακάτω.

**Κατανομή Gumbel.** Η συνάρτηση πυκνότητας πιθανότητας της κατανομής Gumbel είναι

$$f(x) = \frac{1}{\beta} \exp\left(-\left(\frac{x - \mu}{\beta} + \exp\left(-\frac{x - \mu}{\beta}\right)\right)\right), x \in \mathbb{R}$$

όπου με  $\mu$  συμβολίζουμε την παράμετρο θέσης και με  $\beta$  συμβολίζουμε την παράμετρο κλίμακας.

**Κατανομή Frechet.** Η συνάρτηση πυκνότητας πιθανότητας είναι

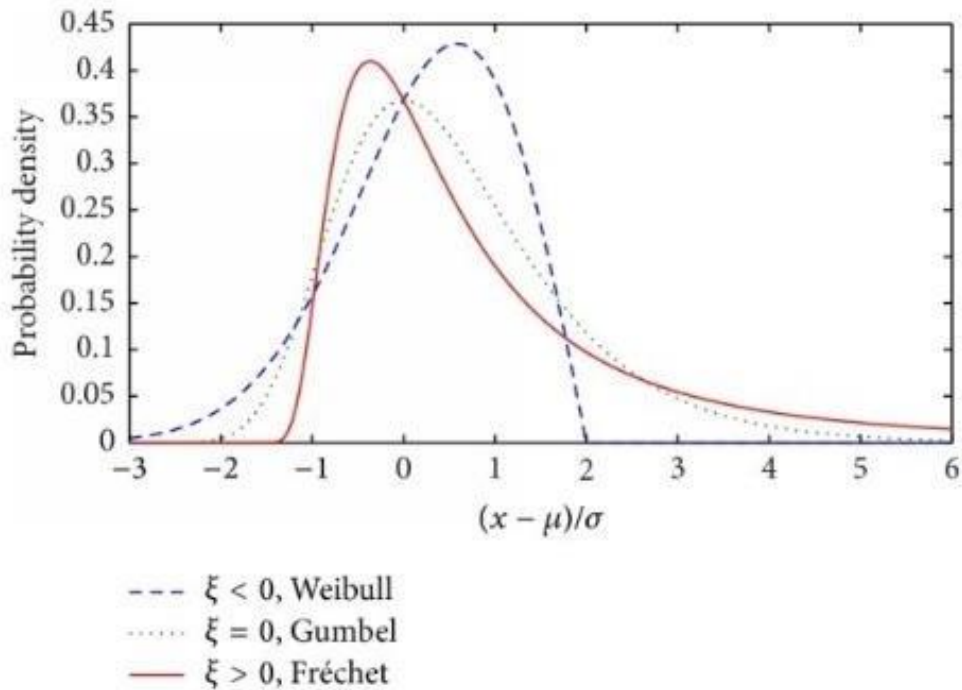
$$f(x) = \frac{\alpha}{\sigma} \left(\frac{x - \mu}{\sigma}\right)^{-\alpha-1} \exp\left(-\left(\frac{x - \mu}{\sigma}\right)^{-\alpha}\right), x > \mu$$

όπου με  $\mu$  συμβολίζουμε την παράμετρο θέσης, με  $\sigma$  συμβολίζουμε την παράμετρο κλίμακας και με  $\alpha$  συμβολίζουμε την παράμετρο σχήματος.

**Κατανομή Weibul.** Ορίζεται ως η κατανομή με σ.π.π.

$$f(x) = \frac{\alpha}{\beta} \left(\frac{x - \mu}{\beta}\right)^{\alpha-1} \exp\left(-\left(\frac{x - \mu}{\beta}\right)^\alpha\right), x < \mu$$

όπου με  $\mu$  συμβολίζουμε την παράμετρο θέσης, με  $\beta$  την παράμετρο κλίμακας και με  $\alpha$  την παράμετρο σχήματος.



Σχήμα 1.2

Συμβολίζουμε με  $x_F$ , το δεξί άκρο της κατανομής  $F$  των  $X_i$ . Αν  $F \in MDA(\lambda)$  τότε  $x_F < \infty$  ή  $x_F = \infty$  με λεπτή ή μέτρια βαριά ουρά (η ουρά της ίδιας της  $\lambda$  θα μειώνεται εκθετικά). Αν  $F \in MDA(\Phi_\alpha)$  τότε  $x_F = \infty$  με «βαριά ουρά». Μάλιστα,

$$1 - F(x) = x^{-\alpha}L(x),$$

όπου η  $L \in \mathcal{R}_0$  είναι βραδείας κύμανσης. Ως συνάρτηση βραδείας κύμανσης ορίζουμε μία συνάρτηση η οποία καθώς τείνει στα άκρα του πεδίου ορισμού της (συνήθως στο άπειρο) "κυμαίνεται" αργά, δηλαδή δεν αυξάνεται ή μειώνεται γρήγορα σε σχέση με μία δύναμη του  $x$ . Πιο συγκεκριμένα, μία συνάρτηση  $L(x)$  ορίζεται ως βραδείας κύμανσης (slowly varying) αν ισχύει η εξής ιδιότητα:

$$\lim_{x \rightarrow \infty} \frac{L(cx)}{L(x)} = 1 \text{ για κάθε σταθερά } c > 0$$

Αυτό σημαίνει πως καθώς το  $x$  τείνει στο άπειρο, η συνάρτηση  $L(x)$  δεν αλλάζει

σημαντικά όταν πολλαπλασιάσουμε το  $x$  με μία σταθερά  $c$ .

Αν  $F \in MDA(\Psi_a)$  τότε  $x_F < \infty$  με ουρά τέτοια ώστε να ικανοποιεί την

$$1 - F(x_F - x^{-1}) = x^{-a}L(x).$$

Τώρα, θα πρέπει να προσδιορίσουμε σε ποια από τις τρεις κατανομές ακροτάτων ταιριάζουν καλύτερα τα block maxima  $Y_1, Y_2, \dots, Y_k$  που έχουμε πάρει. Ανάλογα δηλαδή το MDA που ανήκουν τα  $X_i$ .

Εδώ αξίζει να σημειώσουμε πως η παραπάνω διαδικασία χρησιμοποιείται στην περίπτωση κατά την οποία παρέχεται βεβαιότητα για τον τύπο της οριακής κατανομής ( $L$ ,  $\Phi$  ή  $\Psi$ ) των  $X_i$ . Εναλλακτικά, θα πρέπει να χρησιμοποιηθούν κάποιες μέθοδοι για την επιλογή της κατάλληλου τύπου της οριακής κατανομής των ακροτάτων.

Γενικότερα, σε συνέχεια του παραπάνω, μία εσφαλμένη χρήση κάποιου τύπου οριακής κατανομής μπορεί πολύ εύκολα να οδηγήσει σε τελείως λανθασμένα συμπεράσματα για τα εκάστοτε δεδομένα μας.

Για τον λόγο αυτό η ασφαλέστερη οδός που ακολουθείται τις περισσότερες φορές είναι οι τρεις παραπάνω οικογένειες κατανομών να ενοποιούνται σε μία γενικότερη οικογένεια την οποία καλούμε γενικευμένη κατανομή ακραίων τιμών (Generalized Extreme Value, GEV).

Αν το  $n$  είναι αρκετά μεγάλο, μπορούμε να θεωρήσουμε ότι τα block maxima  $Y_1, Y_2, \dots, Y_k \sim GEV$ , και να εκτιμήσουμε τις παραμέτρους αυτής της κατανομής και με αυτόν τον τρόπο δεν χρειάζεται να επιλέξουμε μία από τις τρεις περιπτώσεις οριακών κατανομών. Ακολουθεί ο ορισμός της κατανομής GEV.

**Γενικευμένη κατανομή ακραίων τιμών (GEV).** Ορίζεται ως η κατανομή με σ.π.π.

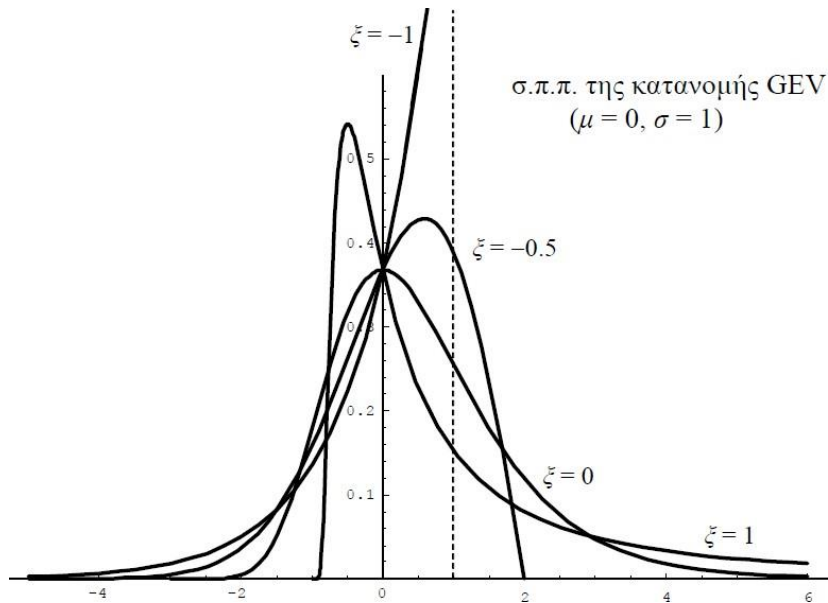
$$F(x) = \begin{cases} \exp\left(-\left(1 + \xi \frac{x - \mu}{\sigma}\right)^{\frac{-1}{\xi}}\right), & \text{αν } \xi \neq 0 \\ \exp\left(-\exp\left(-\frac{x - \mu}{\sigma}\right)\right), & \text{αν } \xi = 0 \end{cases}$$

όπου με  $\mu$  συμβολίζουμε την παράμετρο θέσης, με  $\sigma$  την παράμετρο κλίμακας (όπου  $\sigma > 0$ ), με  $\xi$  την παράμετρο σχήματος. Πιο συγκεκριμένα για την παράμετρο σχήματος:

Όταν  $\xi > 0$  η GEV είναι η κατανομή Freshet (βαριά ουρά)

Όταν  $\xi < 0$  η GEV είναι η κατανομή Weibull

Όταν  $\xi = 0$  η GEV είναι η κατανομή Gumbel (ελαφριά ουρά)



Σχήμα 1.3

Στο διάγραμμα της συνάρτησης πυκνότητας πιθανότητας της GEV παρατηρούμε σε σύγκριση και με το παραπάνω διάγραμμα πως η GEV προσαρμόζεται ανάλογα το  $\xi$  στις κατανομές Weibull , Gumbel και Frechet.

### 1.1.3. Πρόβλεψη ακραίων τιμών

Στην συνέχεια θα προχωρήσουμε στην εκτίμηση για τις παραμέτρους σχήματος, θέσης και κλίμακας μέσω μεθόδων εκτίμησης όπως για η μέθοδος της μέγιστης πιθανοφάνειας (method of maximum likelihood) , η μέθοδος των ροπών και η μέθοδος των ποσοστημορίων (π.χ. βλ. Danielsson (2011)). Ακολουθεί μια σύντομη αναφορά στις γνωστές αυτές μεθόδους όσον αφορά τις εκτιμήσεις των παραμέτρων της GEV.

**Μέθοδος Μέγιστης Πιθανοφάνειας.** Για να εκτιμήσουμε τις παραμέτρους της GEV θα πρέπει αρχικά να υπολογίσουμε την λογαριθμική συνάρτηση της και εν συνεχεία να την μεγιστοποιήσουμε ως προς κάθε παράμετρο ξεχωριστά ώστε να πάρουμε τους εκτιμητές που επιθυμούμε. Η λογαριθμική συνάρτηση της ορίζεται ως:

$$l(\mu, \sigma, \xi) = -n \log \sigma - \sum_{i=1}^n \left[ \left(1 + \xi \frac{x_i - \mu}{\sigma}\right)^{-\frac{1}{\xi}} + \left(1 + \xi \frac{x_i - \mu}{\sigma}\right)^{-1 - \frac{1}{\xi}} \log \left(1 + \xi \frac{x_i - \mu}{\sigma}\right) \right]$$

Τώρα, για να γίνει η μεγιστοποίηση της παραπάνω ως προς κάθε παράμετρο ξεχωριστά μπορούν να χρησιμοποιηθούν διάφορες αριθμητικοί μέθοδοι όπως η μέθοδος Newton-Raphson ή η μέθοδος Nelder-Mead μέσω κάποιου υπολογιστικού πακέτου όπως η R ή η Python.

**Μέθοδος των Ροπών.** Για να γίνει ο υπολογισμός των παραμέτρων της GEV μέσω της μεθόδου των ροπών θα πρέπει να εξισώσουμε τις θεωρητικές ροπές με τις δειγματικές. Αποδεικνύεται ότι η μέση τιμή της GEV είναι

$$E(X) = \mu + \frac{\sigma}{\xi} (\Gamma(1 - \xi) - 1)$$

ενώ η διασπορά είναι

$$Var(X) = \sigma^2 (\Gamma(1 - 2\xi) - (\Gamma(1 - \xi) - 1)^2)$$

Όπου  $\Gamma(X)$  εννοούμε την συνάρτηση Γάμμα. Τώρα θα εξισώσουμε τις παραπάνω με την δειγματική μέση τιμή και την δειγματική διασπορά αντίστοιχα. Για την μέση τιμή

$$\bar{x} = E(X)$$

$$\bar{x} = \mu + \frac{\sigma}{\xi} (\Gamma(1 - \xi) - 1)$$

ενώ για την διασπορά

$$Var(X) = s^2$$

$$s^2 = \sigma^2 (\Gamma(1 - 2\xi) - (\Gamma(1 - \xi) - 1)^2)$$

Τώρα λύνοντας ως προς τις παραμέτρους της GEV παίρνουμε εκτιμήτριες ροπών. Όμως για να γίνει αυτό, λόγω του ότι οι παραπάνω συναρτήσεις είναι μη γραμμικές θα χρειαστεί ξανά να επικαλεστούμε κάποια αριθμητική μέθοδο όπως η μέθοδος Newton-Raphson ή η μέθοδος Nelder-Mead μέσω κάποιου υπολογιστικού πακέτου όπως η R ή η Python.

**Η μέθοδος των ποσοστημορίων.** Αρχικά θα πρέπει να υπολογίσουμε τα δειγματικά ποσοστημόρια  $Q_p$  για διάφορες τιμές του  $p$  από τα δεδομένα. Θα εξισώσουμε τα θεωρητικά με τα δειγματικά ποσοστημόρια με σκοπό να

λύσουμε εκ νέου ως προς τις παραμέτρους της GEV για να πάρουμε τις εκτιμήσεις τους.

$$Q(p) = \mu + \sigma(-\log(1-p))^{\frac{1}{\xi} - 1}$$

Για να λύσουμε ως προς τις παραμέτρους της GEV ξανά θα χρειαστεί να επικαλεστούμε κάποια αριθμητική μέθοδο όπως η μέθοδος Newton-Raphson ή η μέθοδος Nelder-Mead μέσω κάποιου υπολογιστικού πακέτου όπως η R ή η Python.

Αξίζει να σημειωθεί πως για την περίπτωση της Gumbell κατανομής ( $\xi = 0$ ) η παραπάνω εξίσωση γίνεται

$$Q(p) = \mu - \sigma \log(-\log(p))$$

Να σημειώσουμε πως θα μπορούσε να γίνει εκτίμηση και με την χρήση διαφόρων κατάλληλων γραφημάτων (παραλλαγές των γραφημάτων Q-Q) αλλά προφανώς μία οπτική εκτίμηση δεν μπορεί να έχει την ακρίβεια μίας αναλυτικής μαθηματικής εκτίμησης.

Με βάση τώρα αυτές τις εκτιμημένες παραμέτρους από κάποια μέθοδο εκτίμησης, είναι δυνατή η πρόβλεψη της πιθανότητας εμφάνισης ακραίων τιμών σε μελλοντικά δεδομένα.

Συνοψίζοντας και τα τρία είδη κατανομών ακροτάτων αποδεικνύεται το ακόλουθο αποτέλεσμα.

**Θεώρημα Fisher-Tippett.** Έστω  $X_1, X_2, \dots$  μία ακολουθία από ανεξάρτητες και ισόνομες τυχαίες μεταβλητές (iid περίπτωση). Αν υπάρχουν ακολουθίες για τις οποίες  $c_n > 0$ ,  $d_n \in R$  και μία μη εκφυλισμένη κατανομή  $G$  ώστε

$$\frac{M_n - d_n}{c_n} = \frac{\max\{X_1, \dots, X_n\} - d_n}{c_n} \rightarrow_d G$$

τότε η  $G$  θα είναι της μορφής

$$G(x; \mu, \sigma, \xi) = e^{-\left(1 + \xi \frac{x - \mu}{\sigma}\right)^{\frac{1}{\xi}}}, \quad x: 1 + \xi \frac{x - \mu}{\sigma} > 0$$

όπου  $\mu, \xi \in R, \sigma > 0$ .

Τώρα συνεχίζοντας με την μέθοδο Block Maxima από θεώρημα Fisher-Tippett τα Block Maxima  $Y_1, Y_2, \dots, Y_k \sim GEV$  άρα

$$P\left(\frac{Y_i - d_n}{c_n} \leq x\right) = P\left(\frac{M_i^{(i)} - d_n}{c_n} \leq x\right) \approx G(x; \mu_0, \sigma_0, \xi)$$

Θέτοντας τώρα  $y = c_n x + d_n$  παρατηρούμε πως (π.χ. βλ. Rocco (2013))

$$\begin{aligned} P(Y_i \leq y) &\approx G\left(\frac{y - d_n}{c_n}; \mu_0, \sigma_0, \xi\right) = \exp\left(-\left(1 + \frac{\xi(y - d_n - c_n \mu_0)}{c_n \sigma_0}\right)^{\frac{1}{\xi}}\right) \\ &= G(y; \mu_n, \sigma_n, \xi). \end{aligned}$$

Και έτσι έχουμε την δυνατότητα να θεωρήσουμε πως και τα μη κανονικοποιημένα block-maxima  $Y_1, Y_2, \dots, Y_k$  ακολουθούν και αυτά προσεγγιστικά την GEV και αποφεύγουμε με αυτόν τον τρόπο και τον προσδιορισμό των  $d_n, c_n$ .

Για να μελετήσουμε τώρα την κατανομή της μεγαλύτερης παρατήρησης  $M_n = \max\{X_1, \dots, X_n\}$  θα πρέπει να εργασθούμε ως εξής:

- Χωρίζουμε τα  $Y_1, \dots, Y_k$  σε  $k$  ομάδες, τα blocks, και κρατάμε τις μεγαλύτερες παρατηρήσεις,  $Y_i = M_n^{(i)}, i = 1, \dots, k$  για κάθε ομάδα (block maxima). Ενδεικτικά,

$$Y_1 = \max\{X_1, X_2, \dots, X_n\},$$

$$Y_2 = \max\{X_{n+1}, X_{n+2}, \dots, X_{2n}\},$$

, ...,

$$Y_k = \max\{X_{(k-1)n+1}, X_{(k-1)n+2}, \dots, X_{(k-1)n+n}\}$$

Εφόσον το  $n$  είναι ικανοποιητικά μεγάλο, τα block maxima

$$Y_1, \dots, Y_k \sim GEV(\mu_n, \sigma_n, \xi)$$

- Στην συνέχεια εκτιμούμε τις παραμέτρους  $\mu_n, \sigma_n, \xi$  της GEV. Να σημειώσουμε πως όταν έχουμε ένα ακριβές σύνολο δεδομένων το  $n$  το θεωρούμε ως σταθερό και έτσι στην συνέχεια θα γράφουμε τις παραμέτρους  $\mu_n, \sigma_n, \xi$  χωρίς τον δείκτη  $n$ .

### Στάθμη απόδοσης $z_p$ ή return level

Στάθμη απόδοσης καλείται το κατώφλι  $z_p$  το οποίο δεν υπερβαίνει καμία παρατήρηση  $X_i$  για μία χρονική περίοδο (block) με την πιθανότητα  $1 - p$  και έτσι δεν υπερβαίνει και το αντίστοιχο μέγιστο εντός block, το  $M_n$ .

Άρα όταν το μέγεθος του block θα τείνει στο άπειρο τότε ,

$$P(M_n \leq z_p) = G(z_p) = 1 - p$$

άρα

$$z_p = G^{-1}(1 - p) = \begin{cases} \mu - \frac{\sigma}{\xi} (1 - (-\ln(1 - p))^{-\xi}), & \xi \neq 0 \\ \mu - \sigma \ln(-\ln(1 - p)), & \xi = 0 \end{cases}$$

### Εκτίμηση της στάθμης απόδοσης

Ισχύει ότι  $z_p = g(\mu, \sigma, \xi)$  , άρα

$$\hat{z}_p = g(\hat{\mu}, \hat{\sigma}, \hat{\xi}) = \begin{cases} \hat{\mu} - \frac{\hat{\sigma}}{\hat{\xi}} (1 - \exp(\xi w_p)), & \xi \neq 0 \\ \hat{\mu} + \hat{\sigma} w_p, & \xi = 0 \end{cases}$$

Για το τέλος θα πρέπει να βεβαιωθούμε πως τα block maxima προσαρμόζονται σε ικανοποιητικό βαθμό σε κάποια κατανομή GEV έτσι ώστε να μπορούμε να κάνουμε ακριβείς εκτιμήσεις που να έχουν νόημα, άρα θα πρέπει να γίνει κάποιος έλεγχος καλής προσαρμογής των δεδομένων μας στην GEV.

Ένας τέτοιος κλασικός και εύκολος σχετικά τρόπος ελέγχου είναι τα Q-Q plots ή τα P-P plots των block maxima ως προς την  $GEV(\hat{\mu}, \hat{\sigma}, \hat{\xi})$  .

Η μέθοδος "block maxima" είναι αποτελεσματική για τη μοντελοποίηση ακραίων τιμών σε σύνθετα συστήματα, καθώς επιτρέπει τη μείωση της διάστασης των δεδομένων, καθιστώντας την υπολογιστικά πιο εφικτή.

### 1.2. Η μέθοδος των Υπερβάσεων Κατωφλιού ή Peaks Over Threshold (POT)

Η μέθοδος "Peaks Over Threshold" (POT) είναι μια τεχνική που χρησιμοποιείται στη στατιστική για την ανάλυση ακραίων τιμών σε σύνολα δεδομένων. Αυτή η μέθοδος επικεντρώνεται στις τιμές που υπερβαίνουν ένα καθορισμένο κατώφλι (threshold), και στη συνέχεια εκτιμά την κατανομή των

ακραίων τιμών πάνω από αυτό το κατώφλι. Ας δούμε μια σύντομη ανάλυση της μεθόδου Peaks Over Threshold. (π.χ. βλ. Beirlant et al. (2004))

### 1.2.1. Μοντελοποίηση των ακραίων τιμών:

Έστω  $X_1, \dots, X_n \sim F$  για την οποία θέλουμε να αναλύσουμε την δεξιά της ουρά. Θα πρέπει να στηριχτούμε στην κατανομή υπέρβασης μίας τυχαίας μεταβλητής  $X_i$  πάνω από ένα κατώφλι  $u$  δεδομένου πως η  $X_i$  έχει ξεπεράσει το κατώφλι  $u$ .

$$F_u(y) = P(X - u < -y | X > u) = 1 - \frac{1 - F(u + y)}{1 - F(u)}, \quad y > 0$$

Έχοντας επιλέξει ως ακραίες τιμές αυτές που υπερβαίνουν το  $u$ , τώρα μπορούμε να χρησιμοποιήσουμε ένα κατάλληλο μοντέλο για να περιγράψουμε την κατανομή τους. Η δημοφιλέστερη μοντελοποίηση είναι η Γενικευμένη Κατανομή Pareto (Generalized Pareto Distribution - GPD).

### Οριακή κατανομή των POT

Έστω ανεξάρτητες και ισόνομες τυχαίες μεταβλητές  $X_1, \dots, X_n \sim F$  και έστω ότι η  $M_n = \max\{X_1, \dots, X_n\}$   $\xrightarrow{cn} GEV$ . Θα υποθέσουμε πως το  $n$  είναι αρκετά μεγάλο έτσι ώστε

$$P(M_n \leq x) \approx G(x; \mu, \sigma, \xi), \quad M_n = \max\{X_1, \dots, X_n\}$$

όπου  $G$  είναι η συνάρτηση κατανομής της GEV. Έτσι για πολύ μεγάλο  $u$  θα έχουμε ότι

$$F_u(y) = P(X - u \leq y | X > u) \approx H(y; \sigma, \xi) = 1 - \left(1 + \frac{\xi y}{\sigma}\right)^{-\frac{1}{\xi}}$$

για  $y > 0$  με  $1 + \frac{\xi y}{\sigma} > 0$  και  $\tilde{\sigma} = \sigma + \xi(-\mu + u)$ .

Η  $H(\cdot; \tilde{\sigma}, \xi)$  καλείται γενικευμένη κατανομή Pareto (GPD) με μέση τιμή

$$\frac{\tilde{\sigma}}{1 - \xi}$$

Αξίζει να παρατηρήσουμε ότι:

- Όταν  $\xi < 0$  τότε η GPD έχει άνω όριο το  $u - \frac{\tilde{\sigma}}{\xi}$

- Όταν  $\xi > 0$  τότε η GPD εκτείνεται έως το άπειρο από δεξιά

- Όταν  $\xi = 0$  τότε η GPD ταυτίζεται με την εκθετική κατανομή με παράμ.  $\lambda = \frac{1}{\tilde{\sigma}}$

Άρα οι υπερβάσεις ενός μεγάλου κατωφλιού  $u$  των  $X_i$  θα ακολουθούν προσεγγιστικά μία Γενικευμένη κατανομή Pareto με παραμέτρους  $\tilde{\sigma}$  και  $\xi$ .

### 1.2.2. Καθορισμός του κατωφλιού (threshold):

Το κατώφλι, το οποίο αντιπροσωπεύει την τιμή πάνω από την οποία θεωρούμε μια παρατήρηση ως ακραία. Αυτό το κατώφλι μπορεί να επιλεγεί με βάση την εμπειρία ή με τη χρήση στατιστικών μεθόδων, όπως ο υπολογισμός του ποσοστού των παρατηρήσεων που υπερβαίνουν διάφορα κατώφλια.

Επιλογή των ακραίων τιμών: Αφού έχει καθοριστεί το κατώφλι, επιλέγουμε όλες τις παρατηρήσεις που υπερβαίνουν αυτό το κατώφλι. Αυτές οι παρατηρήσεις θεωρούνται ως "ακραίες".

Στην πράξη τώρα εάν το  $u$  δεν οριστεί αρκετά υψηλό διατρέχουμε τον κίνδυνο η  $F_u$  να μην προσεγγίζεται σε ικανοποιητικό βαθμό από την γενικευμένη Pareto. Στην αντίπερα όχθη όμως αν πάρουμε το  $u$  πολύ μεγάλο τότε διατρέχουμε τον κίνδυνο οι υπερβάσεις των  $X_i$  που θα πάρουμε να μην είναι αρκετές και έτσι να μην μπορούμε να κάνουμε ασφαλείς εκτιμήσεις. Θα πρέπει να πάρουμε το μικρότερο δυνατό κατώφλι έτσι ώστε να ισχύει ότι  $F_u \sim GPD$ .

Μία συνήθης μέθοδος επιλογής του κατωφλιού  $u$  είναι η ακόλουθη: Κάνουμε εκτίμηση των παραμέτρων  $\tilde{\sigma}, \xi$  της GPD για διάφορες τιμές του  $u$ . Αν ισχύει ότι  $F_u \sim GPD$  θα πρέπει η εκτίμηση του  $\xi$  να μην επηρεάζεται από το  $u$  ενώ ταυτόχρονα η εκτίμηση του  $\tilde{\sigma} = \sigma + \xi(u - \mu)$  να μεταβάλλεται γραμμικά ως προς το  $u$ . Μπορούμε λοιπόν έτσι να επιλέξουμε το μικρότερο  $u$  που όλα τα παραπάνω ικανοποιούνται.

### 1.2.3. Εκτίμηση παραμέτρων της GPD

Η επόμενη διαδικασία είναι η εκτίμηση των παραμέτρων του μοντέλου που έχουμε επιλέξει. Στην περίπτωση της GPD, θα πρέπει να εκτιμήσουμε τη μορφή και την κλίμακα της κατανομής. Αυτό γίνεται και πάλι μέσω της μεθόδου μέγιστης πιθανοφάνειας όπως έγινε και στην περίπτωση εκτίμησης των παραμέτρων της GEV που αναφέρθηκε παραπάνω.

#### 1.2.4. Πρόβλεψη ακραίων τιμών

Με τις εκτιμημένες παραμέτρους του μοντέλου μας, μπορούμε να κάνουμε προβλέψεις για το πόσο συχνά αναμένουμε να εμφανίζονται ακραίες τιμές στο μέλλον. Η μέθοδος "Peaks Over Threshold" είναι χρήσιμη για τη μοντελοποίηση και την πρόβλεψη ακραίων γεγονότων σε ποικίλους τομείς, όπως οικονομικά καιρικά φαινόμενα, ασφαλιστικές αξιολογήσεις κινδύνου, καθώς και στη μελέτη του αντίκτυπου της κλιματικής αλλαγής στο περιβάλλον. Επιπλέον, η μέθοδος POT μπορεί να προσαρμοστεί σε διάφορα σενάρια ανάλυσης, επιλέγοντας διαφορετικά μοντέλα και κατώφλια ανάλογα με την εφαρμογή. (π.χ. βλ. Rocco (2013))

Γενικά, η μέθοδος Peaks Over Threshold είναι ένα ισχυρό εργαλείο για τη μοντελοποίηση και την πρόβλεψη ακραίων τιμών, και έχει ευρεία εφαρμογή σε πολλούς τομείς της επιστήμης και της βιομηχανίας.

### 1.3. Η μέθοδος Διασπορών-Συνδιασπορών ή Variance-Covariance

Η μέθοδος **Διασπορών-Συνδιασπορών** είναι μια τεχνική που χρησιμοποιείται και για την εκτίμηση ποσοστημορίων μιας κατανομής. Αυτή η μέθοδος συνήθως χρησιμοποιείται όταν οι ακραίες τιμές προκύπτουν από μια κανονική κατανομή (και συνεπώς δεν χρησιμοποιείται το θεώρημα Fisher-Tippett). Ας περιγράψουμε συνοπτικά τη μέθοδο **Διασπορών-Συνδιασπορών** για τη θεωρία ακραίων τιμών παρακάτω. Τα πρώτα βήματα είναι ο υπολογισμός του μέσου (ή της μέσης τιμής) και της διακύμανσης της μεταβλητής. Αυτό μπορεί να γίνει με τη χρήση των κατάλληλων στατιστικών συναρτήσεων του δειγματικού μέσου και δειγματικής διασποράς. Μπορούμε να χρησιμοποιήσουμε διάφορα. Μια συνήθης προσέγγιση είναι η χρήση της κατανομής Gaussian ή της κατανομής της Student-t για να προσαρμόσουμε ένα μοντέλο στις ακραίες τιμές.

#### 1.3.1. Ανάλυση των ακραίων τιμών

Η μέθοδος **Διασπορών-Συνδιασπορών** είναι χρήσιμη για την ανάλυση ακραίων τιμών όταν οι τιμές ακολουθούν κανονική κατανομή ή κοντά σε αυτήν. Ωστόσο, πρέπει να ληφθεί υπόψη ότι η ακρίβειά της μεθόδου μπορεί να μειωθεί αν τα δεδομένα δεν ακολουθούν αυτήν την υπόθεση. Επίσης, η μέθοδος Variance-Covariance μπορεί να αποδειχθεί λιγότερο αποτελεσματική όταν οι

ακραίες τιμές είναι σπάνιες ή όταν η κατανομή των δεδομένων δεν είναι συμμετρική.

Σε γενικές γραμμές, η μέθοδος **Διασπορών-Συνδιασπορών** αποτελεί ένα απλό αλλά δυνατό εργαλείο για την ανάλυση ακραίων τιμών, που μπορεί να χρησιμοποιηθεί σε ποικίλες εφαρμογές από την οικονομία και τα χρηματοοικονομικά μέχρι τη μετεωρολογία και την ασφάλιση. Ωστόσο, πρέπει να εφαρμόζεται με προσοχή και να συνυπολογίζονται οι περιορισμοί και οι υποθέσεις που κάθε φορά εμπεριέχει.

## ΚΕΦΑΛΑΙΟ 2.

### Μέτρα κινδύνου: Αξία σε Κίνδυνο ή Value At Risk και Αναμενόμενο έλλειμα ή Expected Shortfall (VaR και ES)

Τόσο το Value at Risk (VaR) όσο και το Expected Shortfall (ES) είναι μέτρα κινδύνου που χρησιμοποιούνται κυρίως στον χώρο της οικονομικής και χρηματοοικονομικής ανάλυσης για την αξιολόγηση του κινδύνου των επενδύσεων σε διάφορους τομείς ή των διαφόρων χρηματοοικονομικών προϊόντων. Ας αναλύσουμε παρακάτω σύντομα τι αντιπροσωπεύει καθένα από αυτά. (π.χ. βλ. Boutsikas (2020), Rocco (2013))

#### 2.1 Αξία σε Κίνδυνο (VaR).

Το Value at Risk (Αξία σε Κίνδυνο ή Μέγιστη Δυνητική Ζημία) είναι μια ποσότητα που προσδιορίζει το μέγιστο ποσοστό απώλειας που μπορεί να προκύψει σε ένα χρηματοοικονομικό χαρτοφυλάκιο (ή σε ένα επενδυτικό προϊόν) με συγκεκριμένο επίπεδο εμπιστοσύνης και για ένα συγκεκριμένο χρονικό ορίζοντα. Για παράδειγμα, μπορεί να λέμε ότι το VaR του χαρτοφυλακίου μας είναι 1 εκατομμύριο δολάρια με επίπεδο εμπιστοσύνης 95% για μια ημερήσια περίοδο. Αυτό σημαίνει ότι υπάρχει μόνο μία πιθανότητα  $p = 5\%$  να χάσουμε περισσότερα από  $L = 1$  εκατομμύριο δολάρια μέσα σε μια ημέρα. (π.χ. βλ. Boutsikas, (2020))

Πολλές φορές το VaR, εκτός από τους μετόχους και τις ρυθμιστικές αρχές, παρουσιάζεται και σε υποψήφιους επενδυτές ή αγοραστές μίας εταιρείας με σκοπό την ενίσχυση της εικόνας μίας εταιρείας και την απόδειξη πως η εκάστοτε εταιρεία κινείται σε ασφαλή πλαίσια. Ειδικότερα το VaR ορίζεται ως ακολούθως:

**Ορισμός.** Η αξία σε Κίνδυνο (Value at Risk) μιας ζημίας  $Z$  σε επίπεδο εμπιστοσύνης  $p \in (0,1)$  είναι ίση με

$$\text{VaR}_p(L) = \inf \{x \in \mathbb{R}: F_L(x) \geq p\}$$

όπου  $F_L : \mathbb{R} \rightarrow [0,1]$  η συνάρτηση κατανομής της ζημίας  $L$  (π.χ. βλ. Roncalli (2020))

Η παραπάνω σχέση ορίζεται στην θεωρία των κατανομών και ως η γενικευμένη αντίστροφη συνάρτηση κατανομής της ζημιάς  $L$  στο  $p$  ( $F^{-1}(p)$ )  $p$ -ποσοστημόριο της  $F_L$  (συμβολίζεται και με  $q_p(F_L)$ ).

Στην συγκεκριμένη εργασία, όπου χρειαστεί θα δουλεύουμε με συνεχείς κατανομές ζημιών (με αντιστρέψιμη συνάρτηση κατανομής) όπου και θα ισχύει:

$$\text{VaR}_p(L) = \inf \{x \in \mathbb{R}: F_L(x) \geq p\} = \inf \{x \in \mathbb{R}: x \geq F_L^{-1}(p)\} = F_L^{-1}(p)$$

Για τις συνεχείς κατανομές ζημιών ισχύει ισοδύναμα ότι

$$P(L \leq \text{VaR}_p(L)) = P(L \leq F_L^{-1}(p)) = F_L(F_L^{-1}(p)) = p$$

άρα έτσι το  $\text{VaR}_p(L)$  θα παριστάνει το κατώφλι/τιμή το οποίο ξεπερνά την ζημία  $L$  με πιθανότητα  $1 - p$ .

## 2.2. Αναμενόμενο Έλλειμμα (ES):

Το Expected Shortfall ή αλλιώς το Αναμενόμενο Έλλειμμα είναι μια εναλλακτική μέτρηση του κινδύνου που εστιάζει στο μέσο όρο των απωλειών πέρα από το VaR. Το ES λοιπόν ορίζεται και μαθηματικά ως η παρακάτω δεσμευμένη μέση τιμή (π.χ. βλ. Roncalli (2020)):

$$\text{ES}_p(L) = E(L | L \geq \text{VaR}_p(L))$$

Στην ουσία, αναφέρεται στην μέση τιμή των απωλειών σε περιπτώσεις που οι απώλειες υπερβαίνουν το καθορισμένο επίπεδο VaR. Για παράδειγμα, αν το VaR είναι 1 εκατομμύριο δολάρια με επίπεδο εμπιστοσύνης  $p = 95\%$ , το ES ύψους 1.5 εκατομμυρίων δολαρίων εκφράζει τον μέσο όρο των απωλειών όταν οι απώλειες υπερβαίνουν αυτό το VaR. Με άλλα λόγια είναι η προσδοκόμενη ζημία δοθέντος πως εκείνη βρίσκεται στο χείριστο  $100(1 - p)\%$  των περιπτώσεων.

Αν η τυχαία μεταβλητή  $L$  (ζημιά) είναι συνεχής τότε το Αναμενόμενο Έλλειμμα σε επίπεδο εμπιστοσύνης  $p$  δίνεται από τους παρακάτω τύπους:

$$\text{ES}_p(L) = \frac{1}{1-p} \int_{\text{VaR}_p(L)}^{\infty} x f_L(x) dx = \frac{1}{1-p} \int_p^1 \text{VaR}_y(L) dy, \text{ για } p \in (0,1)$$

Παράλληλα, ξανά αν η  $L$  συνεχής τ.μ. τότε

$$\text{ES}_p(cL + d) = c\text{ES}_p(L) + d, d \in \mathbb{R}, \quad c > 0.$$

Συνοψίζοντας τελικώς, το VaR αναφέρεται στο μέγιστο ποσοστό απώλειας σε ένα συγκεκριμένο επίπεδο εμπιστοσύνης, ενώ το ES αναφέρεται στον μέσο όρο των απωλειών πέρα από αυτό το όριο. Ταυτόχρονα και τα δύο μέτρα κινδύνου είναι σημαντικά για τη διαχείριση του ρίσκου σε χρηματοοικονομικά χαρτοφυλάκια και επενδύσεις.

Εκτίμηση των μέτρων κινδύνου VaR και ES

Έχουν προταθεί αρκετές μέθοδοι για την εκτίμηση των VaR και ES. Στο επόμενο κεφάλαιο θα χρησιμοποιήσουμε τις τρεις βασικότερες από αυτές που είναι

A. Η μέθοδος Διασπορών-Συνδιασπορών θεωρώντας ότι τα δεδομένα προέρχονται από κανονική κατανομή.

B. Η Εμπειρική (απαραμετρική) μέθοδος

Γ. Η εκτίμηση μέσω της θεωρίας POT.

Θα παρουσιάσουμε εν συντομία παρακάτω αυτές τις μεθόδους.

### A. Εκτιμήσεις μέσω της κανονικής κατανομής (Παραμετρική μέθοδος).

Ας θεωρήσουμε ότι τα δεδομένα μας είναι οι τ.μ.

$$W_1, W_2, \dots, W_m.$$

Για να εφαρμόσουμε την παραμετρική μέθοδο εκτίμησης του VaR και ES κάνουμε την υπόθεση ότι οι παρατηρήσεις αυτές προέρχονται από την  $N(\mu, \sigma^2)$  και εκτιμούμε τις παραμέτρους

$$\hat{\mu} = \bar{W} = \frac{1}{m} \sum_{i=1}^m W_i, \quad \hat{\sigma}^2 = \frac{1}{m-1} \sum_{i=1}^m (W_i - \bar{W})^2$$

Είναι εύκολο να αποδειχθεί ότι σε αυτή την περίπτωση θα είναι

$$\text{VaR}_p(W) = \mu + \sigma \Phi^{-1}(p), \quad \text{ES}_p(W) = \mu + \sigma \frac{\varphi(\Phi^{-1}(p))}{1-p}$$

Συνεπώς εκτιμούμε

$$\hat{\text{VaR}}_p(W) = \hat{\mu} + \hat{\sigma} \Phi^{-1}(p), \quad \hat{\text{ES}}_p(W) = \hat{\mu} + \hat{\sigma} \frac{\varphi(\Phi^{-1}(p))}{1-p}$$

όπου  $\Phi, \varphi$  η συνάρτηση κατανομής και πυκνότητας πιθανότητας της τυπικής κανονικής. (π.χ. βλ. Boutsikas (2020))

## Β. Απαραμετρικές εκτιμήσεις

Σε αυτή την περίπτωση, δεν θεωρούμε κάποια συγκεκριμένη κατανομή για τα αν.ισ. (IID)  $W_1, W_2, \dots, W_m$ . Εκτιμούμε το  $\text{VaR}_p(W)$  (που είναι το θεωρητικό  $p$ -ποσοστημόριο της κατανομής των  $W_i$ ) από το δειγματικό  $p$ -ποσοστημόριο (καλείται και εμπειρικό VaR)

$$\hat{\text{VaR}}_p(L) = W_{[mp]:m}$$

όπου  $W_{1:m} \leq W_{2:m} \leq \dots \leq W_{m:m}$  είναι οι διατεταγμένες παρατηρήσεις  $W_1, W_2, \dots, W_m$  και  $[x]$  είναι ο μικρότερος ακέραιος που είναι  $\geq x$  (οροφή, ceiling).

Όμοια, το αναμενόμενο έλλειμα

$$\text{ES}_p(W) = E(L|L \geq \text{VaR}_p(W))$$

μπορεί να εκτιμηθεί από το δειγματικό μέσο των παρατηρήσεων που είναι  $\geq \hat{\text{VaR}}_p(W) = W_{[mp]:m}$  (καλείται εμπειρικό ES),

$$\text{ES}_p(W) = \frac{W_{[mp]:m} + \dots + W_{m:m}}{m - [mp] + 1}$$

Το πλεονέκτημα είναι ότι τα παραπάνω μπορούν να εφαρμοστούν χωρίς να γίνει κάποια υπόθεση για την κατανομή των παρατηρήσεων αλλά γενικά λαμβάνονται εκτιμήσεις με μεγαλύτερη διασπορά σε σχέση με την παραμετρική μέθοδο. Επιπλέον η μέθοδος αυτή είναι αναποτελεσματική για  $p$  κοντά στο 1 διότι τότε

$$\hat{\text{VaR}}_p(W) = \max W_i, \quad \hat{\text{ES}}_p(W) = \max W_i.$$

## Γ. Η εκτίμηση μέσω της θεωρίας POT.

Συμβολίζουμε με  $Z_1, Z_2, \dots, Z_\kappa$  τις υπερβάσεις πάνω από ένα κατώφλι  $u$ , δηλαδή τις

$$W_1 - u, W_2 - u, \dots, W_m - u$$

που είναι μεγαλύτερες του 0. Σύμφωνα με την θεωρία ακραίων τιμών αυτές οι  $\kappa$  υπερβάσεις θα ακολουθούν προσεγγιστικά την γενικευμένη Pareto κατανομή. Δηλαδή,

$$Z_1, Z_2, \dots, Z_\kappa \sim F_u \approx \text{GPD}(y; \sigma, \xi) = 1 - \left(1 + \frac{\xi y}{\sigma}\right)^{-\frac{1}{\xi}},$$

για κάποιες άγνωστες παραμέτρους  $\xi, \sigma$ . Η σημαντικότερη είναι παράμετρος είναι η παράμετρος σχήματος  $\xi$  (extremal index). Για  $\xi > 0$  η κατανομή των αρχικών παρατηρήσεων  $W_1, W_2, \dots, W_m$  θα έχει βαριά δεξιά ουρά. Μάλιστα, όσο μεγαλύτερο είναι το  $\xi$ , τόσο βαρύτερη θα είναι η δεξιά ουρά. Μπορούμε τώρα να χρησιμοποιήσουμε τις  $\kappa$  υπερβάσεις για να εκτιμήσουμε τις δύο αυτές παραμέτρους. Θα χρησιμοποιήσουμε τις εκτιμήτριες μέγιστης πιθανοφάνειας. Ως γνωστό, ο λογάριθμος της πιθανοφάνειας των  $Z_1, Z_2, \dots, Z_\kappa$  είναι:

$$\ell(\sigma_u, \xi) = -\kappa \ln \sigma_u - (1 + \frac{1}{\xi}) \sum_{i=1}^{\kappa} \ln (1 + \frac{\xi Z_i}{\sigma_u}) + \sum_{i=1}^{\kappa} \ln I(1 + \xi Z_i / \sigma_u > 0)$$

Η παραπάνω  $\ell(\sigma_u, \xi)$  δεν μεγιστοποιείται ως προς  $\sigma, \xi$  αναλυτικά (μέσω κλειστού τύπου). Δεδομένου όμως ενός τ.δ.  $z_1, \dots, z_\kappa$  των  $Z_1, Z_2, \dots, Z_\kappa$ , μπορούμε να βρούμε τις  $\hat{\sigma}_u, \hat{\xi}$  που την μεγιστοποιούν χρησιμοποιώντας αριθμητικές μεθόδους βελτιστοποίησης. Μπορούμε τώρα να εκτιμήσουμε το VaR από το  $p$  ποσοστημόριο της GPD και συγκεκριμένα (για  $\text{VaR}_p(X) > u$ )

$$\hat{\text{VaR}}_p(X) = u + \frac{\hat{\sigma}_u}{\hat{\xi}} \left( \left( \frac{\kappa}{m(1-p)} \right)^{\hat{\xi}} - 1 \right)$$

Για το Expected Shortfall (όταν  $\xi < 1$  και  $\text{VaR}_p(X) > u$ ) θα είναι

$$\text{ES}_p(X) = E(X|X > \text{VaR}_p(X)) = \frac{1}{1-p} \int_p^1 \text{VaR}_y(X) dy = \frac{\sigma_u - \xi u}{1-\xi} + \frac{\text{VaR}_p(X)}{1-\xi}.$$

### 2.3. Εκ των υστέρων έλεγχος ή Backtesting

Το backtesting είναι η διαδικασία ελέγχου μιας στρατηγικής ή ενός μοντέλου σε ιστορικά δεδομένα για να διαπιστωθεί η αποτελεσματικότητα του μοντέλου. Στο πλαίσιο της ανάλυσης ακραίων τιμών, ο στόχος του backtesting είναι να αξιολογήσει πώς μια στρατηγική ή ένα μοντέλο αντιδρά σε διάφορες συνθήκες ακραίων τιμών.

Η διαδικασία του backtesting συνήθως περιλαμβάνει τα ακόλουθα βήματα:

1. *Ορισμός της στρατηγικής ή του μοντέλου:* Αυτό περιλαμβάνει την περιγραφή των κανόνων και των παραμέτρων της στρατηγικής ή του μοντέλου που θα δοκιμαστεί.
2. *Εφαρμογή στα ιστορικά δεδομένα:* Η στρατηγική ή το μοντέλο εφαρμόζεται σε ιστορικά δεδομένα.

3. *Αξιολόγηση των αποτελεσμάτων*: Ανάλυση των αποτελεσμάτων για να κατανοήσουμε πώς αντέδρασε η στρατηγική ή το μοντέλο σε διάφορες συνθήκες ακραίων τιμών. Αυτό μπορεί να περιλαμβάνει την ανάλυση των απωλειών, των κερδών, της ακρίβειας κ.λπ.

Ανάλογα με τα αποτελέσματα του backtesting, μπορεί να απαιτηθεί προσαρμογή της στρατηγικής ή του μοντέλου για βελτίωση των επιδόσεών τους.

Στην περίπτωση της ανάλυσης ακραίων τιμών, το backtesting μπορεί να βοηθήσει στην αξιολόγηση του πόσο καλά μια στρατηγική αντιμετωπίζει τις ακραίες καταστάσεις της αγοράς και εάν οδηγεί σε αποτελέσματα που είναι αποδεκτά και κερδοφόρα σε αυτές τις συνθήκες.

## ΚΕΦΑΛΑΙΟ 3

### Εφαρμογή

#### 3.1. Εκτίμηση VaR και ES επενδύσεων σε μετοχές (μετοχή NFLX)

Στο κεφάλαιο αυτό θα υλοποιηθεί η διαδικασία εκτίμησης των μέτρων κινδύνου που παρουσιάστηκε στα προηγούμενα κεφάλαια. Επίσης θα γίνει εκ των υστέρων έλεγχος (backtesting) με σκοπό να εξεταστεί κατά πόσο η εκτίμηση των μέτρων κινδύνου μέσω της θεωρίας ακραίων τιμών και ιδιαίτερα της μεθόδου POT είναι αποτελεσματικότερη από την παραμετρική μέθοδο (μέσω της κανονικής κατανομής) και την εμπειρική μέθοδο (μέσω των δειγματικών ποσοστημορίων).

Η εφαρμογή αρχικά θα πραγματοποιηθεί με βάση τις ημερήσιες χρηματιστηριακές τιμές κλεισίματος της μετοχής NFLX, έστω  $S_1, S_2, \dots, S_n$ .

##### 3.1.1. Καταγραφή και επεξεργασία δεδομένων

Ως πρώτο βήμα φορτώνουμε τα δεδομένα μας μέσω του R (βιβλιοθήκη quantmod).

```
library("quantmod")
library(ismev)
from.dat <- as.Date("2014-01-01")
to.dat <- as.Date("2024-12-31")
ticker <- "NFLX"
data <- getSymbols(ticker, from = from.dat, to = to.dat,
                    periodicity = "daily", auto.assign = FALSE)
```

Λαμβάνουμε παρατηρήσεις από 1/1/2014 έως 31/12/2024. Οι 10 πρώτες και οι 10 τελευταίες γραμμές του πίνακα δεδομένων data είναι

```
S=data$NFLX.Adjusted
n0=length(S)
length(S)
data[1:10];
data[(n0-9):n0]
```

```
> data[1:10]
```

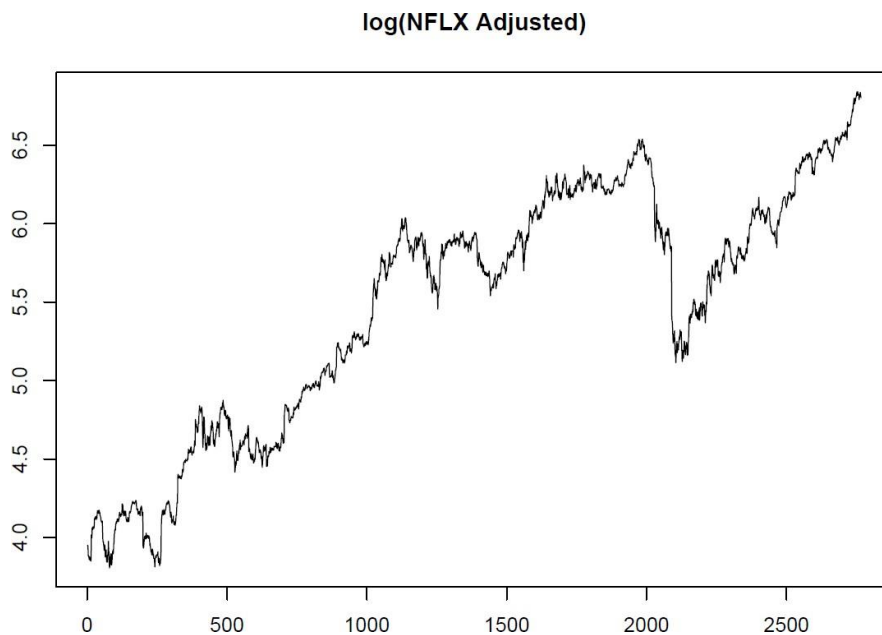
	NFLX.Open	NFLX.High	NFLX.Low	NFLX.Close	NFLX.Volume	NFLX.Adjusted
2014-01-02	52.40143	52.51143	51.54286	51.83143	12325600	51.83143
2014-01-03	52.00000	52.49571	51.84286	51.87143	10817100	51.87143
2014-01-06	51.89000	52.04428	50.47572	51.36714	15501500	51.36714
2014-01-07	49.68428	49.69857	48.15286	48.50000	36167600	48.50000
2014-01-08	48.10429	49.42571	48.07429	48.71286	20001100	48.71286
2014-01-09	48.82429	49.14000	47.85714	48.15000	17007200	48.15000
2014-01-10	48.29572	48.37714	47.21857	47.44857	16237900	47.44857
2014-01-13	47.30429	49.81000	47.30429	48.11572	27237000	48.11572
2014-01-14	48.77714	48.80000	47.51000	48.28000	19740000	48.28000
2014-01-15	47.85714	48.36714	45.58143	47.19714	40432000	47.19714

```
> data[(n0-9):n0]
```

	NFLX.Open	NFLX.High	NFLX.Low	NFLX.Close	NFLX.Volume	NFLX.Adjusted
2024-12-16	923.94	928.94	919.50	921.08	2674900	921.08
2024-12-17	921.54	928.45	916.42	919.13	2459500	919.13
2024-12-18	919.13	922.44	886.56	889.55	2938200	889.55
2024-12-19	908.18	912.62	899.00	902.04	2815500	902.04
2024-12-20	893.21	919.92	890.49	909.05	4530500	909.05
2024-12-23	913.42	915.00	899.10	911.45	2339400	911.45
2024-12-24	915.00	935.85	911.70	932.12	2320300	932.12
2024-12-26	928.40	930.49	915.30	924.14	2340300	924.14
2024-12-27	916.01	918.13	894.50	907.55	3226200	907.55
2024-12-30	894.51	908.23	889.71	900.43	2203000	900.43

Μπορούμε να δούμε στο ακόλουθο γράφημα την εξέλιξη της τιμής της μετοχής NFLX.

```
plot(log(data$NFLX.Adjusted))
```



Σχήμα 3.1

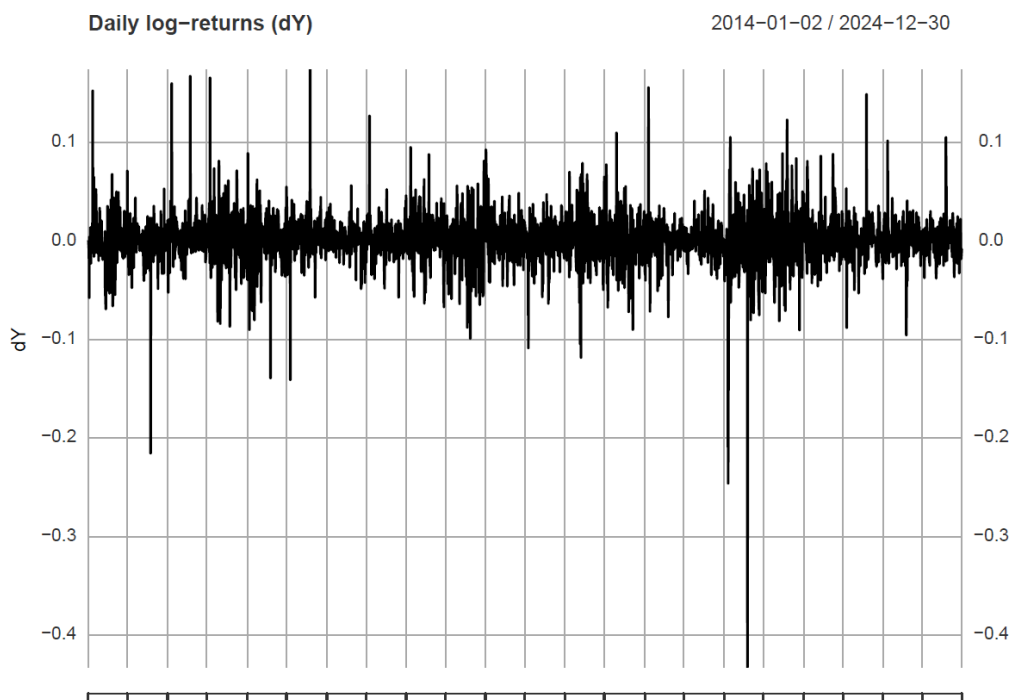
Σύμφωνα με το υπόδειγμα Black and Scholes θα πρέπει οι πρώτες διαφορές των λογαρίθμων των παραπάνω τιμών  $S_1, S_2, \dots, S_n$  να είναι ανεξάρτητες τ.μ. που ακολουθούν την κανονική κατανομή με κάποιες παραμέτρους οι οποίες μπορεί να μεταβάλλονται στον χρόνο. Μετασχηματίσουμε λοιπόν τα δεδομένα λαμβάνοντας

$$Y_i = \ln S_i, \quad i = 1, 2, \dots, n,$$

$$\delta Y_i = -(Y_{i+1} - Y_i), \quad i = 1, 2, \dots, n - 1$$

Παραπάνω λαμβάνουμε τις ημερήσιες ζημίες  $\delta Y_i$  από την κατοχή μιας μετοχής NFLX (ενώ οι διαφορές  $Y_{i+1} - Y_i$  εκφράζουν κέρδη) έτσι ώστε στη συνέχεια να μελετήσουμε την δεξιά ουρά των ζημιών. Το γράφημα των πρώτων διαφορών θα είναι της μορφής:

```
Y=log(S) ; dY=-diff(Y)  
plot(dY) ;
```



Σχήμα 3.2

Στη συνέχεια, θα κρατήσουμε 2000 από αυτές τις παρατηρήσεις.

```
n=3000;start=1000;  
X=rep(0,n)  
for(i in 1:n){X[i]=100*dY[[start+i]]}
```

Οι παρατηρήσεις αυτές φαίνεται ότι μάλλον δεν είναι στάσιμες (π.χ. από μια συγκεκριμένη κατανομή με σταθερές παραμέτρους) διότι η διασπορά τους δεν φαίνεται να είναι σταθερή στον χρόνο (υπάρχουν περίοδοι με μεγαλύτερη ή μικρότερη μεταβλητότητα). Παρόλα αυτά, στη συνέχεια για ευκολία θα θεωρήσουμε ότι οι παράμετροι της κατανομής τους μεταβάλλονται σχετικά αργά, ώστε να μπορούμε να θεωρήσουμε ότι π.χ. για παράθυρα διάρκειας μερικών μηνών παραμένουν περίπου σταθερές. Με αυτή την παραδοχή στην συνέχεια μπορούμε να εφαρμόζουμε την μέθοδο POT αλλά και την παραμετρική μέθοδο μέσω της κανονικής κατανομής σε κυλιόμενα χρονικά παράθυρα συγκεκριμένου, σχετικά μικρού, μήκους

Κάποια περιγραφικά στοιχεία είναι τα ακόλουθα:

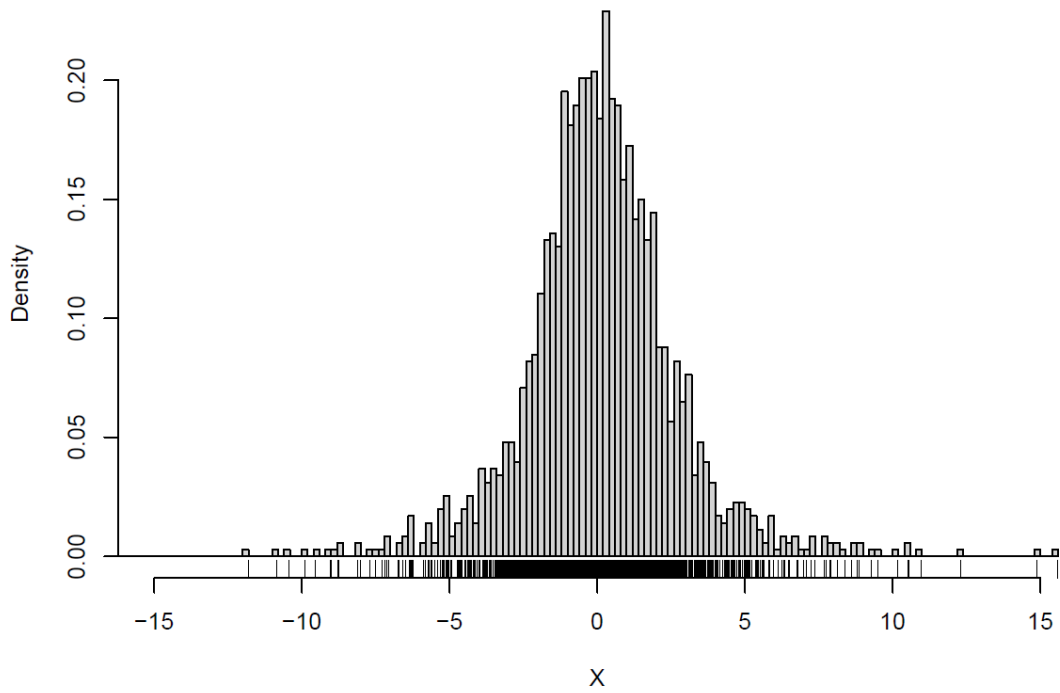
```
summary(X)
```

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
-43.25784	-1.21475	0.08039	0.08894	1.44996	15.57580

Επίσης ένα ιστόγραμμα τους και ένα QQ plot ως προς την κανονική είναι της μορφής:

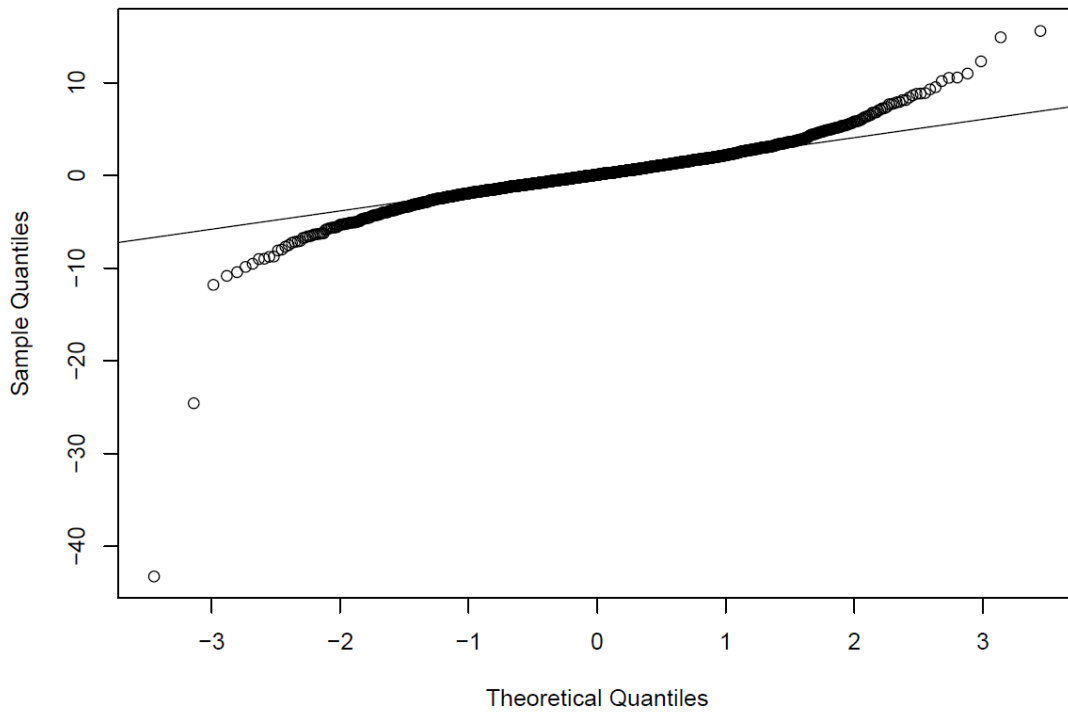
```
hist(X,400,prob=TRUE,xlim=c(-15,15));rug(X)  
qqnorm(X);qqline(X)
```

Histogram of X (returns in %)



Σχήμα 3.3

QQ-plot of X



Σχήμα 3.4

### 3.1.2 παραμετρική Μέθοδος

Θα φορτώσουμε το πακέτο `ismen` μέσω του οποίου θα εφαρμοστεί η μέθοδος POT, όχι στο σύνολο των παραπάνω δεδομένων, αλλά σε χρονικά παράθυρα  $m$  διαδοχικών παρατηρήσεων (ώστε να θεωρηθεί προσεγγιστικά ότι ισχύει στασιμότητα). Σε κάθε τέτοιο παράθυρο ξεκινάμε από την  $X_i$  και βασιζόμαστε στις  $X_i, X_{i+1}, \dots, X_{i+m-1}$  και αυτό για κάθε  $i = 1, 2, \dots, n - m$ . Για απλότητα ας συμβολίσουμε τις  $X_i, X_{i+1}, \dots, X_{i+m-1}$  με

$$W_1, W_2, \dots, W_m$$

Για να εφαρμόσουμε την παραμετρική μέθοδο εκτίμησης του VaR και ES θεωρούμε ότι οι παραπάνω  $m$  παρατηρήσεις προέρχονται από την  $N(\mu, \sigma^2)$  και εκτιμώντας τις παραμέτρους

$$\hat{\mu} = \bar{W} = \frac{1}{m} \sum_{i=1}^m W_i, \quad \hat{\sigma}^2 = \frac{1}{m-1} \sum_{i=1}^m (W_i - \bar{W})^2$$

εκτιμούμε

$$\hat{\text{VaR}}_p(W) = \hat{\mu} + \hat{\sigma} \Phi^{-1}(p), \quad \hat{\text{ES}}_p(W) = \hat{\mu} + \hat{\sigma} \frac{\varphi(\Phi^{-1}(p))}{1-p}$$

όπου  $\Phi, \varphi$  η συνάρτηση κατανομής και πυκνότητας πιθανότητας της τυπικής κανονικής.

Μέσω R, για κάθε παράθυρο  $i \in \{1, 2, \dots, n - m + 1\}$  θέτουμε,

```
W=X[i:(i+m-1)]
```

```
VaR2[i]=mean(W)+sd(W)*qnorm(p);
```

```
ES2[i]=mean(W)+sd(W)/(1-p)*dnorm(qnorm(p));
```

### 3.1.3 Απαραμετρική Μέθοδος

Επίσης μπορούμε να εκτιμήσουμε τα μέτρα κινδύνου απαραμετρικά, χρησιμοποιώντας το εμπειρικό VaR και το εμπειρικό ES. Σε αυτή την περίπτωση, δεν θεωρούμε κάποια συγκεκριμένη κατανομή για τα αν.ισ. (IID)  $W_1, W_2, \dots, W_m$ . Εκτιμούμε το  $\text{VaR}_p(W)$  από το δειγματικό  $p$ -ποσοστημόριο

$$\hat{\text{VaR}}_p(L) = W_{[mp]:m} \quad (\text{καλείται και εμπειρικό VaR})$$

όπου  $W_{1:m} \leq W_{2:m} \leq \dots \leq W_{m:m}$  είναι οι διατεταγμένες παρατηρήσεις  $W_1, W_2, \dots, W_m$  και  $[x]$  είναι ο μικρότερος ακέραιος που είναι  $\geq x$  (οροφή, ceiling).

Όμοια, το αναμενόμενο έλλειμα  $ES_p(W)$  μπορεί να εκτιμηθεί από το δειγματικό μέσο των παρατηρήσεων που είναι  $\geq \hat{VaR}_p(W) = W_{[mp]:m}$  (καλείται εμπειρικό ES),

$$ES_p(W) = \frac{W_{[mp]:m} + \dots + W_{m:m}}{m - [mp] + 1}$$

Μέσω R, και πάλι για κάθε παράθυρο  $i \in \{1, 2, \dots, n - m + 1\}$  θέτουμε,

```
W=X[i:(i+m-1)]
VaR3[i]=quantile(W,p)
mean(W[W>VaR3[i]])
```

### 3.1.4 Μέθοδος POT

Ως τρίτη μέθοδο θα εφαρμόσουμε την μέθοδο POT θέτουμε ως κατώφλι  $u$  το δειγματικό 0.7 ποσοστημόριο, δηλαδή κρατάμε ως υπερβάσεις τις  $\kappa = 0.3m$  μεγαλύτερες παρατηρήσεις από τις  $W_1, W_2, \dots, W_m$ . Συμβολίζουμε με  $Z_1, Z_2, \dots, Z_\kappa$  τις υπερβάσεις αυτές, δηλαδή κρατάμε τις  $W_1 - u, W_2 - u, \dots, W_m - u$ , που είναι μεγαλύτερες του 0. Μπορούμε τώρα να χρησιμοποιήσουμε τις  $\kappa$  υπερβάσεις για να εκτιμήσουμε τις δύο αυτές παραμέτρους. Θα χρησιμοποιήσουμε τις εκτιμήτριες μέγιστης πιθανοφάνειας. Ως γνωστό, ο λογάριθμος της πιθανοφάνειας των  $Z_1, Z_2, \dots, Z_\kappa$  είναι:

$$\ell(\sigma_u, \xi) = -\kappa \ln \sigma_u - \left(1 + \frac{1}{\xi}\right) \sum_{i=1}^{\kappa} \ln \left(1 + \frac{\xi Z_i}{\sigma_u}\right) + \sum_{i=1}^{\kappa} \ln I\left(1 + \xi Z_i / \sigma_u > 0\right)$$

Η παραπάνω  $\ell(\sigma_u, \xi)$  δεν μεγιστοποιείται ως προς  $\sigma, \xi$  αναλυτικά (μέσω κλειστού τύπου). Δεδομένου όμως ενός τ.δ.  $z_1, \dots, z_\kappa$  των  $Z_1, Z_2, \dots, Z_\kappa$ , μπορούμε να βρούμε τις  $\hat{\sigma}_u, \hat{\xi}$  που την μεγιστοποιούν χρησιμοποιώντας αριθμητικές μεθόδους βελτιστοποίησης. Όπως έχουμε ήδη αναφέρει, η αριθμητική εκτίμηση των παραμέτρων αυτών γίνεται με την χρήση του πακέτου ismев:

```
library("ismev")
```

Για κάθε παράθυρο  $i \in \{1, 2, \dots, n - m + 1\}$  θέτουμε, σύμφωνα με τα παραπάνω,

```
W=X[i:(i+m-1)]
u<-quantile(W,0.7)
```

και εκτιμούμε τις παραμέτρους της GPD μέσω της εντολής gpd.fit:

```
pot=gpd.fit(W,u);
sigma=pot$mle[1];xi=pot$mle[2];
```

Σημειώνεται ότι το κατώφλι  $u$  επιλέγεται ως το 70% δειγματικό ποσοστημόριο του δείγματος (διατεταγμένη παρατήρηση  $0.7m = 140$  από τις 200, αφήνοντας 60 υπερβάσεις σε κάθε παράθυρο από τις οποίες εκτιμώνται οι παράμετροι της GPD).

Μπορούμε τώρα να εκτιμήσουμε το VaR από το  $p$  ποσοστημόριο της GPD και συγκεκριμένα (για  $\text{VaR}_p(X) > u$ )

$$\hat{\text{VaR}}_p(X) = u + \frac{\hat{\sigma}_u}{\xi} \left( \left( \frac{\kappa}{m(1-p)} \right)^\xi - 1 \right)$$

Μέσω του R:

```
k1=pot$ncx;
```

```
VaR[i]=u+sigma2/xi2* ((k1/(m*(1-p)))^xi2-1);
```

Για το Expected Shortfall (όταν  $\xi < 1$  και  $\text{VaR}_p(X) > u$ ) θα είναι

$$\text{ES}_p(X) = E(X|X > \text{VaR}_p(X)) = \frac{1}{1-p} \int_p^1 \text{VaR}_y(X) dy = \frac{\sigma_u - \xi u}{1-\xi} + \frac{\text{VaR}_p(X)}{1-\xi}$$

και συνεπώς μέσω R:

```
ES1[i]=(sigma-xi*u+VaR[i])/(1-xi)
```

Συνοψίζοντας σε έναν ενιαίο κώδικα, οι παραπάνω εκτιμήσεις των VaR, ES επαναλαμβάνονται για  $i \in \{1, 2, \dots, n - m + 1\}$  : (οι εντολές sink δίνονται ώστε τα αποτελέσματα να μην εξάγονται στην οθόνη σε κάθε επανάληψη, κάτι που δεν χρειάζεται αφού αποθηκεύουμε τις εκτιμήσεις σε κάθε παράθυρο στις λίστες VaR, VaR2, VaR3 και ES1, ES2, ES3). Επίσης αρχικά επιλέγουμε επίπεδο εμπιστοσύνης για το VaR, ES το  $p = 99.5\%$  ενώ τα επικαλυπτόμενα παράθυρα εφαρμογής του μοντέλου POT επιλέγουμε να αποτελούνται από  $m = 300$  παρατηρήσεις.

```
m=300;
```

```
library("ismev")
```

```
VaR=rep(0,n-m);VaR2=rep(0,n-m);VaR3=rep(0,n-m);
```

```
ES3=rep(0,n-m);ES2=rep(0,n-m);ES1=rep(0,n-m);
```

```
xii=rep(0,n-m);
```

```

sink(file="nul");
p=0.995;
for(i in 1:(n-m))
{
W=X[i:(i+m-1)]
u<-quantile(W,0.7)
pot=gpd.fit(W,u);
sigma=pot$mle[1];xi=pot$mle[2];
xii[i]=xi;
k1=pot$nexc;
VaR[i]=u+sigma/xi*((k1/(m*(1-p)))^xi-1);
VaR2[i]=mean(W)+sd(W)*qnorm(p);
VaR3[i]=quantile(W,p)[[1]]

ES1[i]=(sigma-xi*u+VaR[i])/(1-xi)
ES2[i]=mean(W)+sd(W)/(1-p)*dnorm(qnorm(p));
ES3[i]=mean(W[W>VaR3[i]])
}
sink()

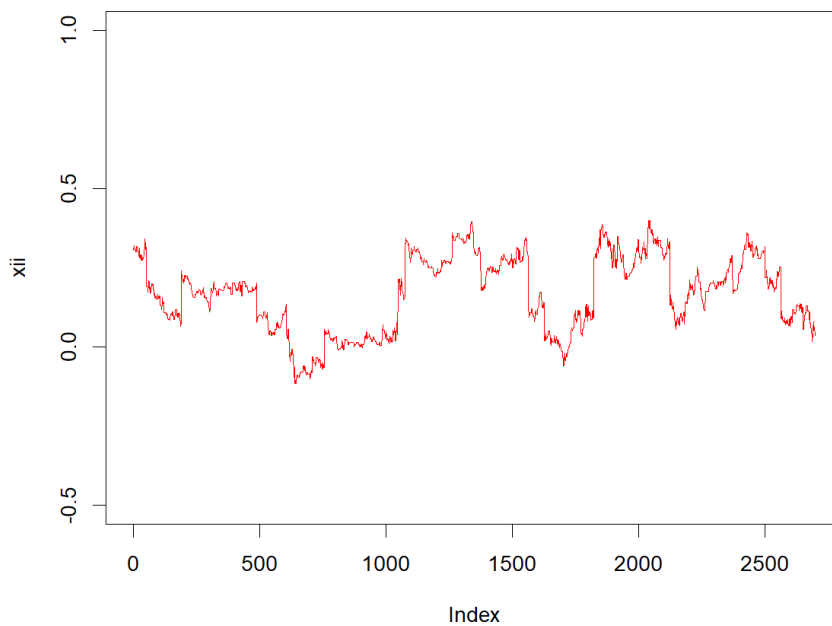
```

Στη συνέχεια κατασκευάζουμε ένα γράφημα με τις εκτιμήσεις του  $\xi$ :

```

plot(xii,type="l",col="red",ylim=c(-0.5,1))

```



Σχήμα 3.5

Παρατηρούμε ότι η εκτίμηση της παραμέτρου σχήματος  $\xi$  της GPD μεταβάλλεται στο χρονικό διάστημα που εξετάζουμε την τιμή της μετοχής NFLX. Αυτό ίσως ήταν αναμενόμενο και από το διάγραμμα των  $\delta Y_i = -(Y_{i+1} - Y_i)$ ,  $i = 1, 2, \dots, n - 1$ , όπου διαφαίνονταν κάποια χρονικά διαστήματα με περισσότερες και κάποια με λιγότερες ακραίες παρατηρήσεις. Σε κάποιες περιόδους λοιπόν παρατηρούμε ότι οι πρώτες διαφορές των λογαρίθμων των τιμών φαίνεται να προέρχονται από μια κατανομή με βαριά ουρά.

### 3.1.5. Σύγκριση μεθόδων

Στον παραπάνω κώδικα, καταχωρούμε στην λίστα VaR τις εκτιμήσεις της αξίας σε κίνδυνο με την μέθοδο POT για την  $m + 1$  ημέρα (με βάση τις παρατηρήσεις 1, 2, ...,  $m$ ), για την  $m + 2$  ημέρα (με βάση τις παρατηρήσεις 2, 3, ...,  $m + 1$ ) κ.ο.κ. για την  $n$  ημέρα (με βάση τις παρατηρήσεις  $n - m + 1$ , ...,  $n$ ) ενώ αντίστοιχα στις λίστες VaR2, VaR3 καταχωρούμε τις εκτιμήσεις της αξίας σε κίνδυνο μέσω της κανονικής κατανομής και μέσω της απαραμετρικής μεθόδου (εμπειρικό VaR).

Οι 20 πρώτες εκτιμήσεις μετά τον χρόνο  $m + 1$  (δηλ. οι  $X_{m+1}, \dots, X_{m+20}$ ) μαζί με τις αντίστοιχες εκτιμήσεις του VaR με τις τρεις παραπάνω μεθόδους και για  $p=0.995$ .

**XX=X[(m+1):n]**

```
> head(results[["0.995"]][xx, 20)
 [1] 0.9243776 1.4876093 -1.9685783 -0.8544626 2.6142638 -0.7243460 1.3770643
 [8] -0.6638593 0.7330471 0.5462866 -1.2904844 4.4803140 0.7037684 -4.5644522
[15] 1.4355895 -1.7240883 -1.8505045 0.3502900 0.5483969 2.8750305
```

```
> print(table_VaR, digits = 6)
      t1      t2      t3      t4      t5      t6      t7      t8      t9
[1,] 8.47757 8.47757 8.47757 8.47757 8.47757 8.46025 8.48635 8.48635 8.49179
      t10     t11     t12     t13     t14     t15     t16     t17     t18
[1,] 8.47904 8.47904 8.48227 8.46576 8.46576 8.45621 8.46871 8.47274 8.47274
      t19     t20
[1,] 8.47274 8.47274
```

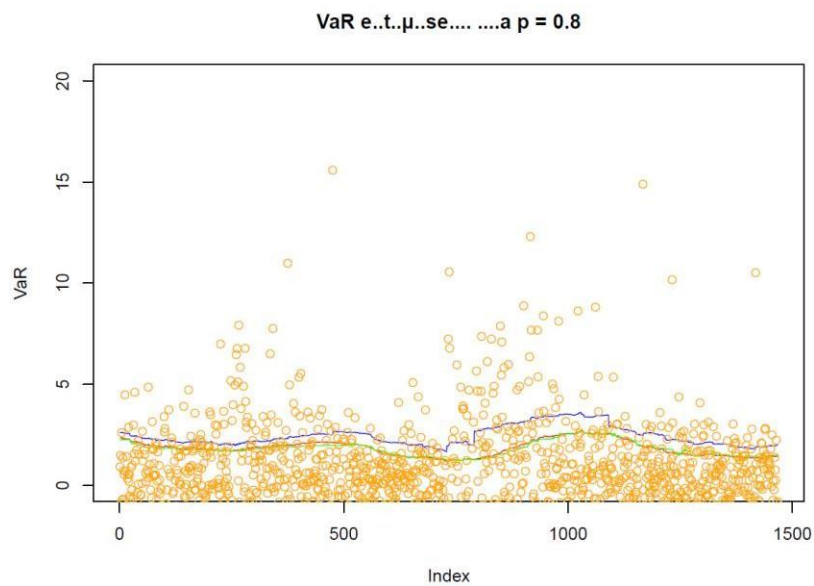
```
> print(table_VaR2, digits = 6)
      t1      t2      t3      t4      t5      t6      t7      t8      t9
[1,] 7.58548 7.58529 7.5929 7.59079 7.5907 7.60921 7.58131 7.58871 7.54254
      t10     t11     t12     t13     t14     t15     t16     t17     t18
[1,] 7.53417 7.53523 7.52169 7.56025 7.56397 7.57492 7.56864 7.55811 7.55754
      t19     t20
[1,] 7.55891 7.55474
```

```

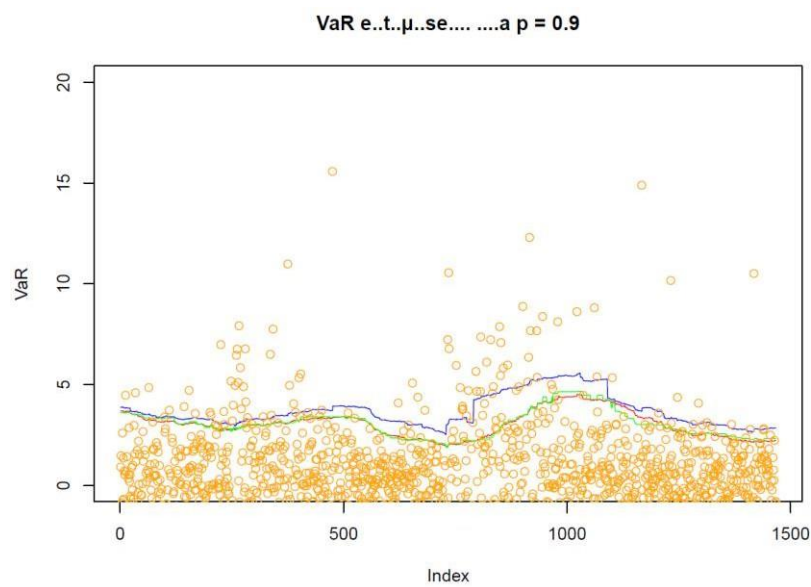
> print(table_VaR3, digits = 6)
      t1      t2      t3      t4      t5      t6      t7      t8      t9
[1,] 9.03731 9.03731 9.03731 9.03731 9.03731 9.03731 9.03731 9.03731 9.03731
      t10     t11     t12     t13     t14     t15     t16     t17     t18
[1,] 9.03731 9.03731 9.03731 9.03731 9.03731 9.03731 9.03731 9.03731 9.03731
      t19     t20
[1,] 9.03731 9.03731

```

Για μια καλύτερη εικόνα των παραπάνω εκτιμήσεων κατασκευάζουμε σε ένα κοινό γράφημα τις παρατηρήσεις  $X_{m+1}, \dots, X_n$  μαζί με τις εκτιμήσεις του VaR για τις ίδιες ημέρες ( $m + 1, \dots, n$ ) και για διάφορα  $p$ .

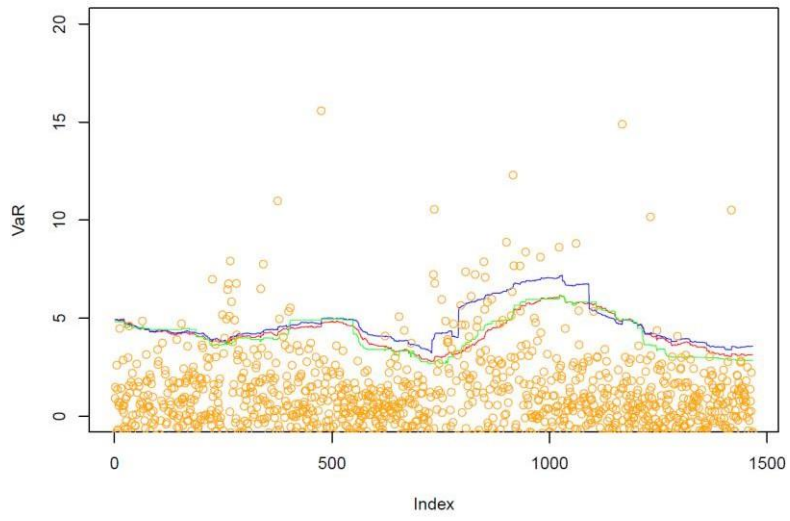


Σχήμα 3.6



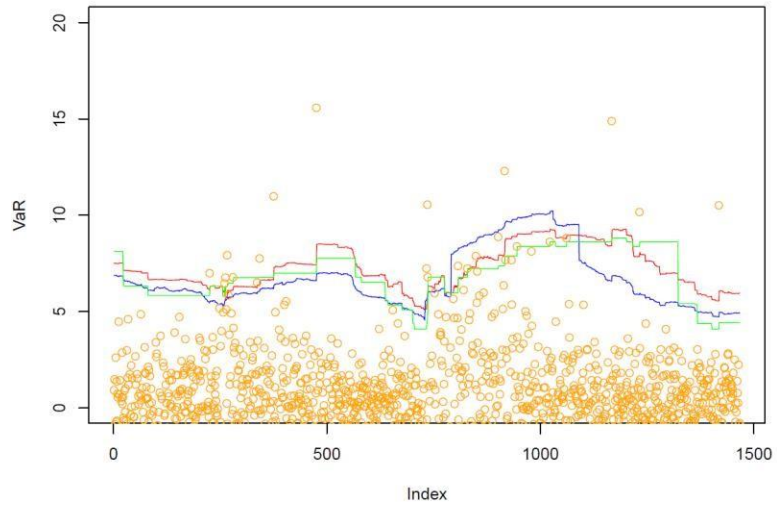
Σχήμα 3.7

VaR e..t..μ..se..... a p = 0.95



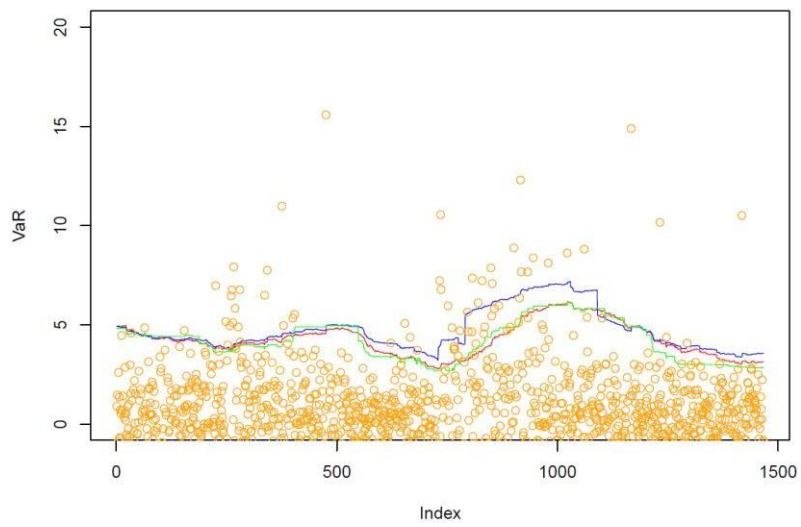
Σχήμα 3.8

VaR e..t..μ..se..... a p = 0.99



Σχήμα 3.9

VaR e..t..μ..se..... a p = 0.95



Σχήμα 3.10

Παρατηρούμε ότι η εκτίμηση του VaR μέσω της μεθόδου POT (κόκκινες τιμές) είναι τις περισσότερες φορές μεγαλύτερη από την παραμετρική εκτίμηση μέσω κανονικής (μπλέ τιμές) και την απαραμετρική εκτίμηση (πράσινες τιμές).

### 3.1.6. Εκ των Υστέρων έλεγχος

Στην συνέχεια θα πραγματοποιηθεί εκ των υστέρων έλεγχος (**back-testing**) ώστε να εξεταστεί κατά πόσο η εκτίμηση των μέτρων κινδύνου μέσω της θεωρίας ακραίων τιμών (μέθοδος POT) υπερτερεί έναντι άλλων μεθόδων (παραμετρική μέσω κανονικής κατανομής, απαραμετρική μέσω δειγματικών ποσοστημορίων).

Συγκεκριμένα, όπως περιγράψαμε παραπάνω, από κάθε παράθυρο  $m$  παρατηρήσεων  $X_i, X_{i+1}, \dots, X_{i+m-1}$  εκτιμάται το VaR και το ES με τις 3 μεθόδους. Ο εκ των υστέρων έλεγχος συνίσταται στη σύγκριση της εκτιμώμενης τιμής του VaR για την επόμενη ( $i + m$ ) ημέρα με την πραγματική τιμή,  $X_{i+m}$ . Αυτό επαναλαμβάνεται για  $i = 1, 2, \dots, n - m$ .

Αν η διαδικασία εκτίμησης του VaR είναι ικανοποιητική, αναμένεται ότι το δειγματικό ποσοστό των  $X_{i+m}$  που δεν υπερβαίνουν το εκάστοτε εκτιμημένο  $VaR_p$  (για την  $i + m$  ημέρα) να είναι ίσο με  $p$ . Εάν αυτό το δειγματικό ποσοστό των μη υπερβάσεων είναι σημαντικά μεγαλύτερο του  $1 - p$  (π.χ. είναι 99.9% αντί  $p = 99.5\%$ ) τότε γίνεται υπερεκτίμηση του  $VaR$  με συνέπεια π.χ. να κρατείται μεγαλύτερο κεφάλαιο ασφαλείας από ότι χρειάζεται. Στην αντίθετη περίπτωση, αν το δειγματικό αυτό ποσοστό είναι μικρότερο του  $p$  (π.χ. είναι 98% αντί  $p = 99.5\%$ ) τότε γίνεται υποεκτίμηση του  $VaR$  με συνέπεια να γίνεται έκθεση σε μεγαλύτερο κίνδυνο από τον εκτιμώμενο.

Προκειμένου να ελέγξουμε την υπόθεση:

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

καταγράφουμε το πλήθος των μη υπερβάσεων του εκτιμημένου VaR, έστω  $v$ . Υπό την  $H_0$  το  $v$  θα ακολουθεί Διωνυμική κατανομή με πιθανότητα επιτυχίας  $p$ . Άρα το αντίστοιχο  $p - value$  του αμφίπλευρου ελέγχου ( $H_0: \text{το ποσοστό των μη υπερβάσεων} \neq p$ ) θα είναι

$$p - value = \min\{P(B \geq v|H_0), P(B \leq v|H_0)\}$$

όπου  $B \sim$  Διωνυμική ( $n - m, p$ ).

Λαμβάνουμε τα ποσοστά  $v/(n - m)$  καθώς και τα αντίστοιχα  $p$ -values για κάθε μία από τις τρεις μεθόδους μέσω του κώδικα

Συνοψίζουμε τα παραπάνω αποτελέσματα του back-testing, στον ακόλουθο πίνακα, όπου έχουμε συμπεριλάβει και τα αποτελέσματα για διάφορες τιμές του επιπέδου εμπιστοσύνης.

Model	p = 0.995	p = 0.99	p = 0.95	p = 0.9	p = 0.8
<b>POT</b>	<b>0.99182</b> <b>(0.0922108)</b>	0.98569 (0.112554)	0.95433 (0.509556)	<b>0.91070</b> <b>(0.177449)</b>	0.81050 (0.327568)
<b>Normal</b>	0.98978 (0.00900261)	<b>0.98841</b> <b>(0.510211)</b>	0.96592 (0.00397152)	0.94001 (6.08596e-08)	0.87389 (1.00962e-13)
<b>Empiric</b>	0.98978 (0.00900261)	0.98500 (0.0643338)	<b>0.95160</b> <b>(0.857274)</b>	0.91411 (0.0742762)	<b>0.80845</b> <b>(0.433532)</b>

Πίνακας 3.1

Τέλος για την σύγκριση της αποτελεσματικότητας των εκτιμήσεων του ES μέσω των τριών μεθόδων, υπολογίζουμε την μέση (απόλυτη) διαφορά της εκτίμησης του ES από την πραγματική τιμή (για το σύνολο των ημερών  $m + 1, \dots, n$ ), δεδομένου ότι είχαμε υπέρβαση του VaR (για την εκτίμηση του VaR χρησιμοποιούμε την μέθοδο POT που από τους προηγούμενους πίνακες φαίνεται να είναι ακριβέστερη). Δηλαδή για κάθε μέθοδο υπολογίσουμε το μέσο απόλυτο σφάλμα

$$MAE = \frac{1}{n - m} \sum_{\substack{i=m+1 \\ i: X_i > VaR}}^n |ES_i - X_i|$$

Παρατηρούμε και εδώ ότι η εκτίμηση του ES μέσω της μεθόδου POT φαίνεται να είναι ακριβέστερη (οδηγεί σε μικρότερο απόλυτο σφάλμα).

Συνοψίζουμε τα αποτελέσματα του back-testing για το ES, στον ακόλουθο πίνακα, όπου και πάλι έχουμε συμπεριλάβει και τα αποτελέσματα για διάφορες τιμές του επιπέδου εμπιστοσύνης.

Model	p = 0.995	p = 0.990	p = 0.950	p = 0.900	p = 0.800
<b>POT</b>	8.66760	7.28147	4.28101	<b>3.05225</b>	1.84530
<b>Normal</b>	<b>7.40708</b>	<b>6.69416</b>	4.74674	3.70857	2.45143
<b>Empiric</b>	8.38663	6.89275	<b>4.21695</b>	3.07051	<b>1.82697</b>

Πίνακας 3.2

### 3.1.7. Συμπεράσματα (για την μετοχή NFLX)

Όπως διαφαίνεται και από τους εκ των υστέρων ελέγχους (back-testing), το πρώτο μοντέλο (POT), εκτιμά γενικά αποτελεσματικότερα την αξία σε κίνδυνο για τις διάφορες τιμές του  $p$  σε σχέση με τα υπόλοιπα μοντέλα. Συγκεκριμένα, στη μετοχή **Netflix (NFLX)** η εκτίμηση του VaR μέσω POT υπερτερεί, καθώς τα ποσοστά επιτυχιών που καταγράφηκαν βρίσκονται πιο κοντά στις θεωρητικές πιθανότητες, ενώ τα αντίστοιχα  $p$ -values υποδηλώνουν ότι η μέθοδος είναι στατιστικά αποδεκτή στις περισσότερες περιπτώσεις.

Το παραμετρικό μοντέλο που βασίζεται στην υπόθεση ότι οι λογαριθμικές αποδόσεις ακολουθούν κανονική κατανομή φαίνεται ότι οδηγεί σε **υποεκτίμηση του κινδύνου για υψηλές τιμές του  $p$**  και υπερεκτίμηση για σχετικά χαμηλές τιμές του  $p$ . Αυτό δικαιολογείται από τη διαπίστωση ότι οι παρατηρήσεις φαίνεται να προέρχονται από κατανομή με βαρύτερη ουρά σε σχέση με αυτή της κανονικής, με αποτέλεσμα η χρήση της να είναι συστηματικά ανακριβής.

Στο τρίτο, απαραμετρικό μοντέλο παρατηρούμε ότι η μέθοδος αυτή δίνει καλύτερα αποτελέσματα σε σχέση με το παραμετρικό μοντέλο, ωστόσο **δεν είναι ικανοποιητική για πολύ υψηλές τιμές του  $p$** , καθώς η εμπειρική εκτίμηση του VaR επηρεάζεται δυσανάλογα από τις μεγαλύτερες παρατηρήσεις σε κάθε περίοδο. Παρά το γεγονός ότι σε μεσαίες τιμές εμπιστοσύνης (π.χ. 0.9, 0.8) η μέθοδος προσεγγίζει αρκετά ικανοποιητικά, δεν μπορεί να θεωρηθεί αξιόπιστη για την αποτίμηση ακραίων κινδύνων.

Συνολικά, η μέθοδος POT αναδείχθηκε ως η **πιο συνεπής και ακριβής** στην εκτίμηση τόσο του VaR όσο και του ES, επιβεβαιώνοντας τη θεωρητική της καταλληλότητα για δεδομένα με βαριές ουρές όπως οι αποδόσεις της μετοχής Netflix.

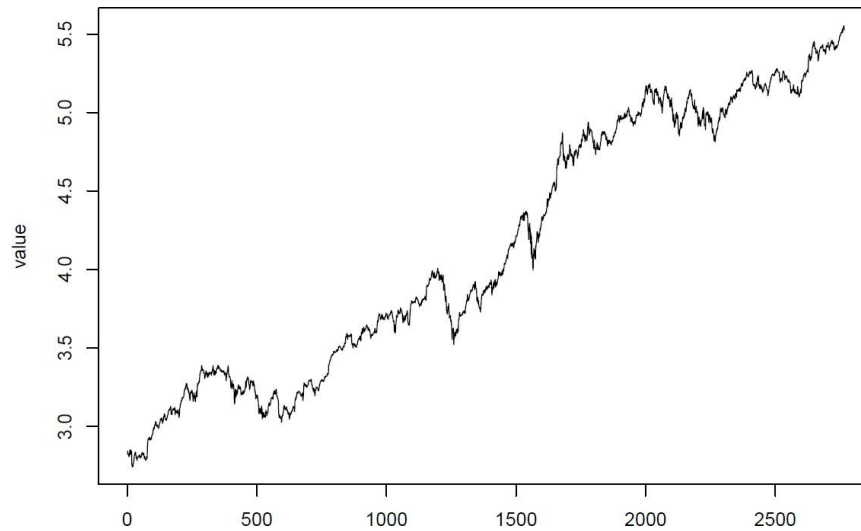
Παρακάτω θα κάνουμε την ίδια δουλειά για τις τιμές του κρυπτονομίσματος Bitcoin και τις μετοχές της Apple και της Amazon προκειμένου να εξετάσουμε αν υπάρχει κάποια διαφοροποίηση στα παραπάνω συμπεράσματα.

Σε όσα ακολουθούν θα λάβουμε παρατηρήσεις και πάλι από 1/1/2014 έως 31/12/2024.

## 3.2. Εφαρμογή με βάση τις τιμές της μετοχής της APPLE

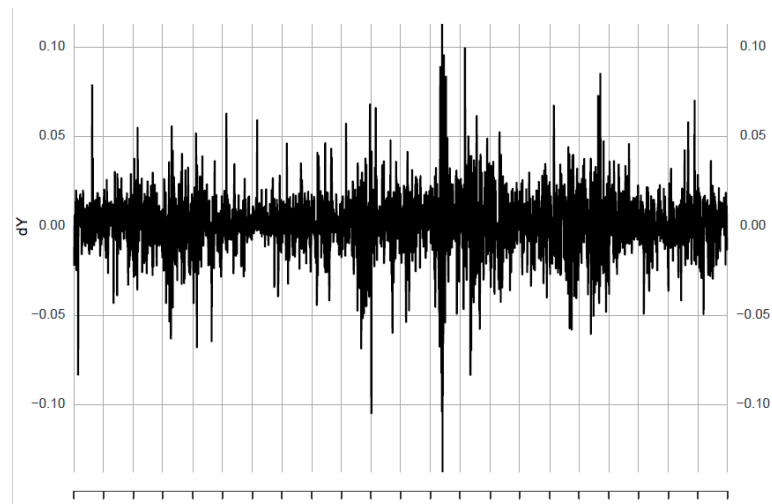
### 3.2.1 Εκτιμήσεις των μέτρων κινδύνου

Μπορούμε να δούμε στο ακόλουθο γράφημα την εξέλιξη της τιμής της μετοχής APPLE.



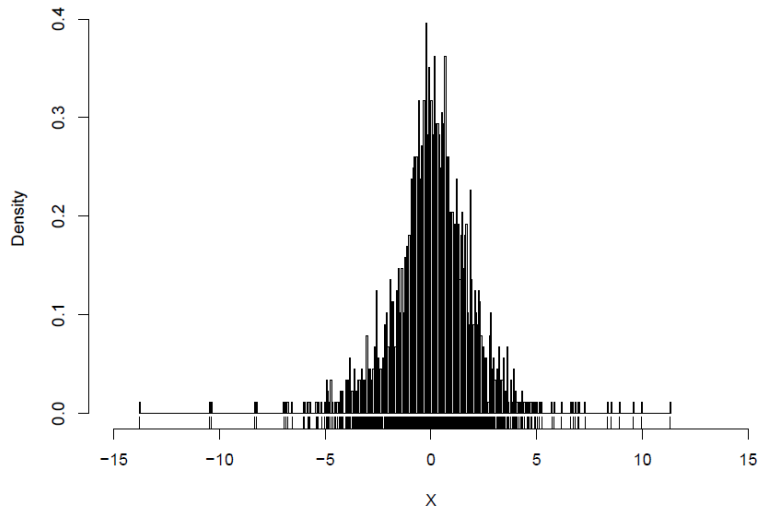
Σχήμα 3.11

Το γράφημα των πρώτων διαφορών θα είναι της μορφής



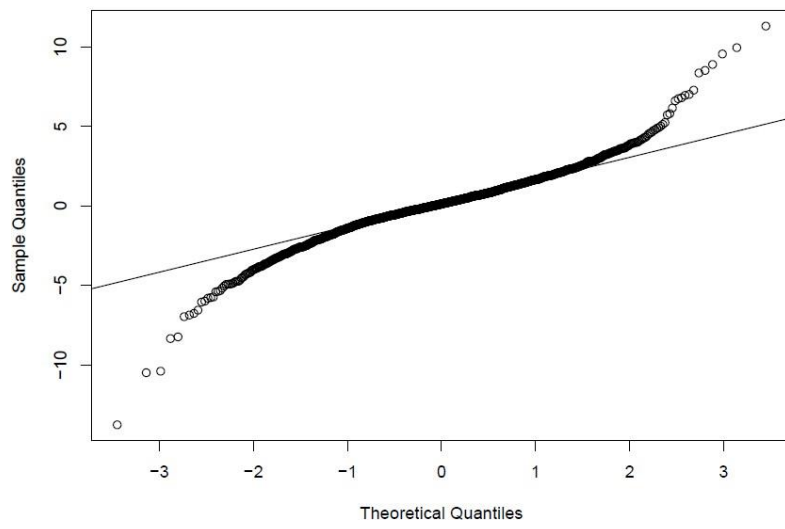
Σχήμα 3.12

Επίσης ένα ιστόγραμμα τους και ένα QQ plot ως προς την κανονική είναι της μορφής



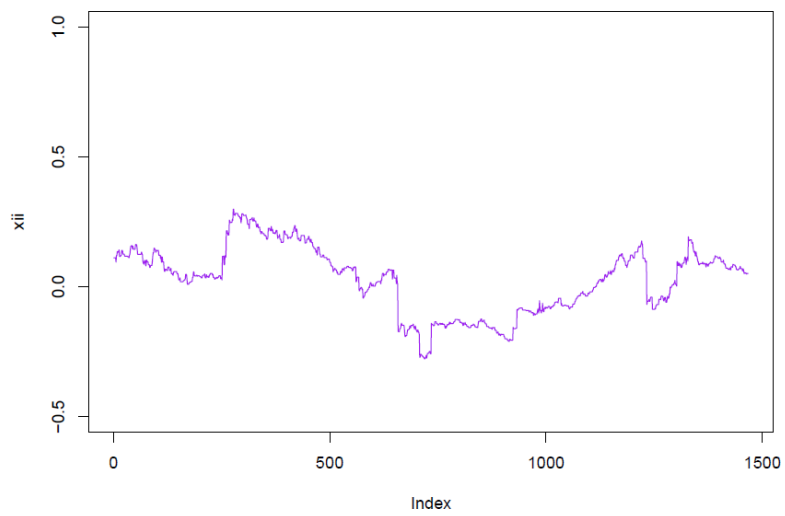
Σχήμα 3.13

QQ-plot of X



Σχήμα 3.14

Για την εκτίμηση του  $\xi$

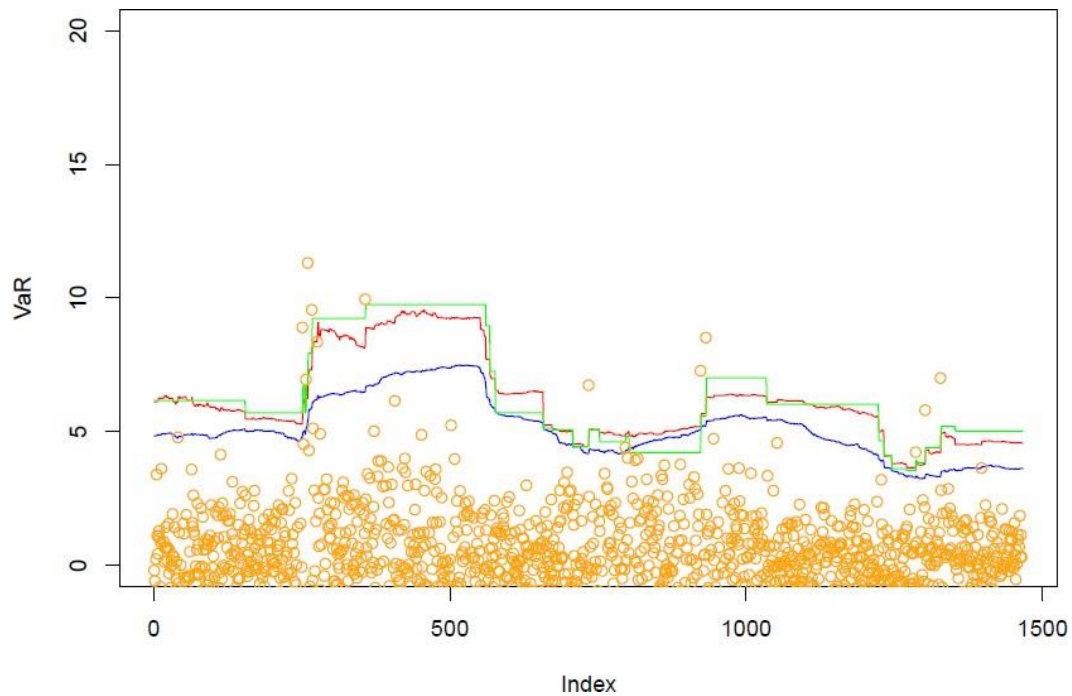


Σχήμα 3.15

Τώρα για κάθε διαφορετική τιμή του  $p$  ακολουθούν τα αντίστοιχα διαγράμματα:

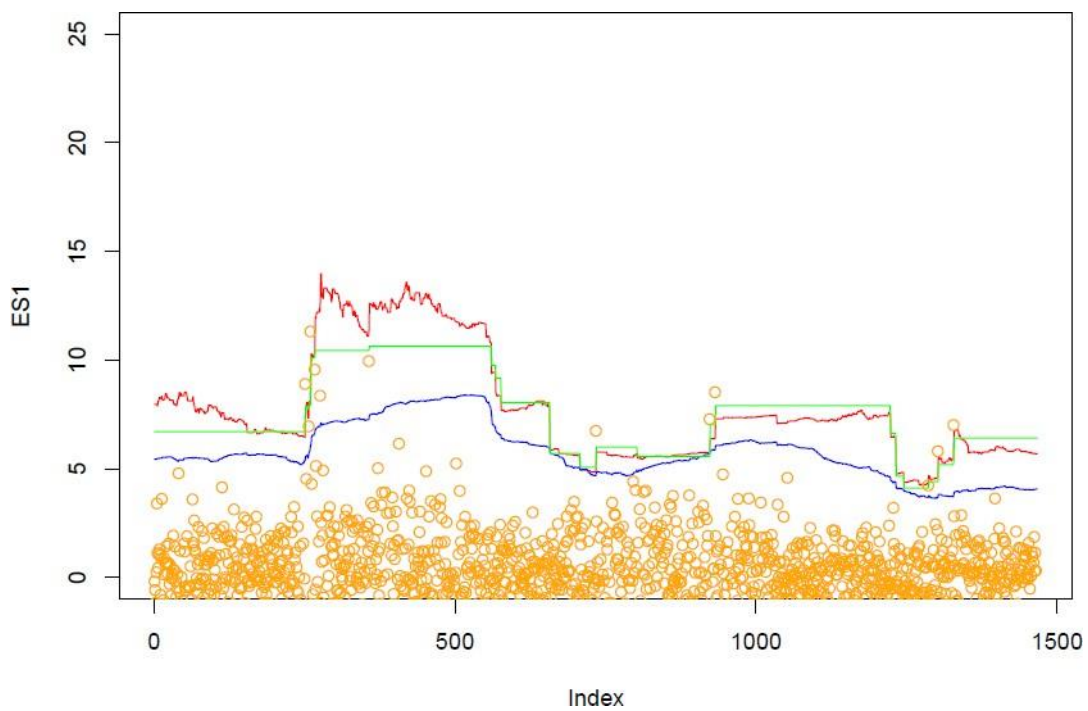
**Για  $p = 0.995$**

Για το VaR



Σχήμα 3.16

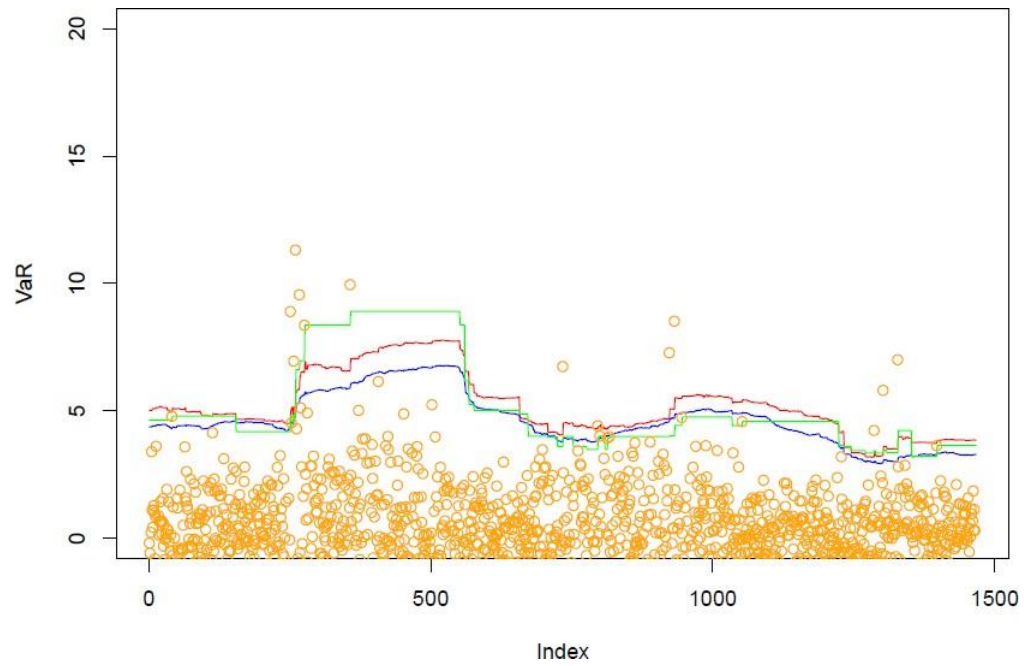
Για το ES



Σχήμα 3.17

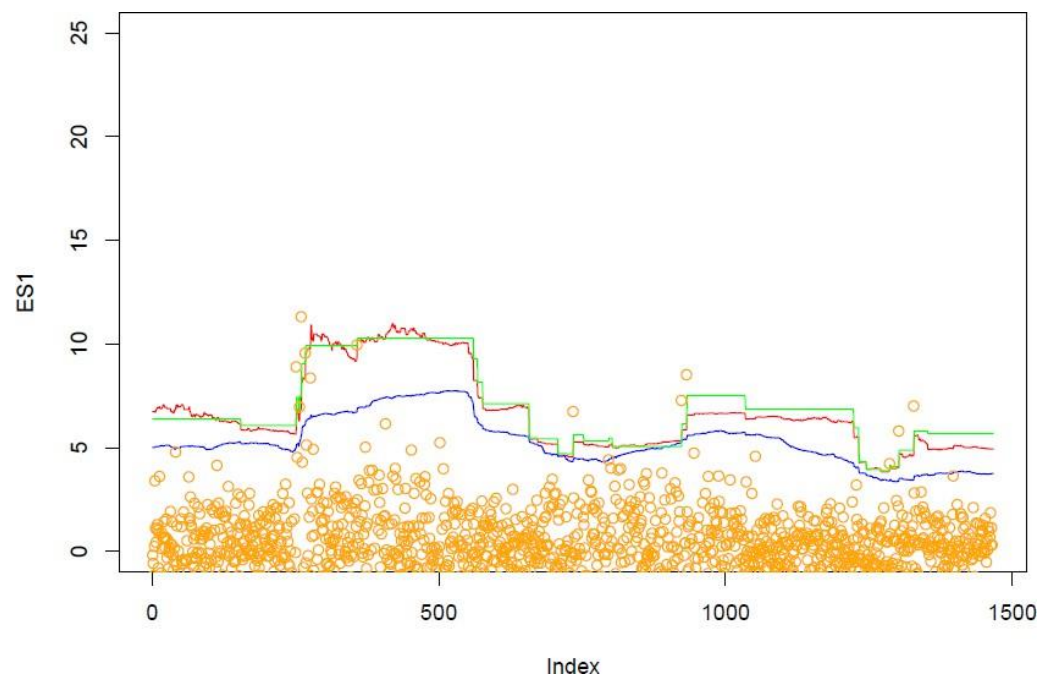
**Για  $p = 0.99$**

Για το VaR



Σχήμα 3.18

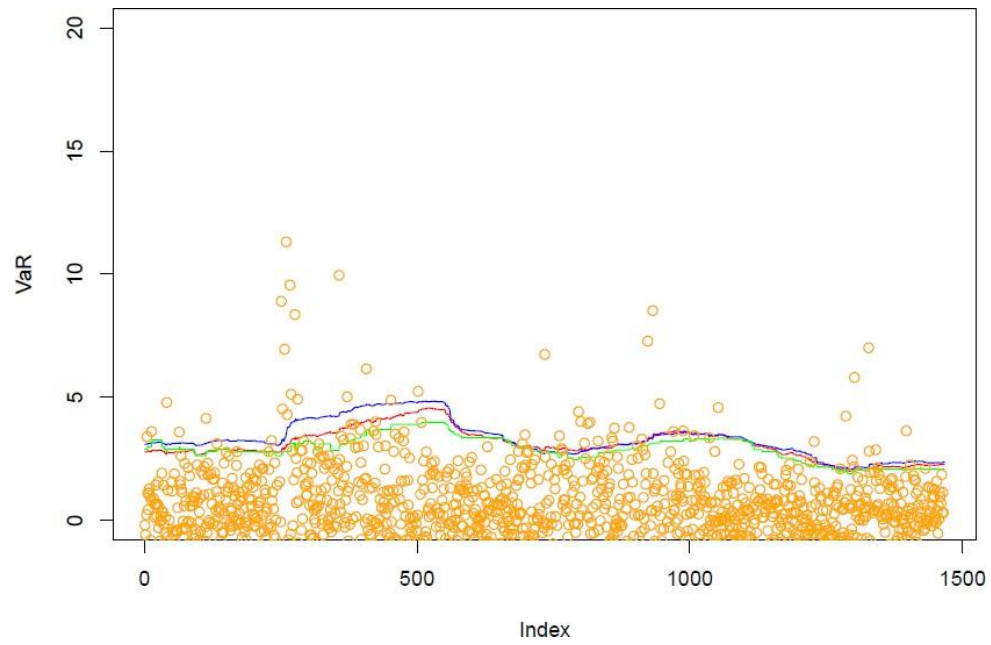
Για το ES



Σχήμα 3.19

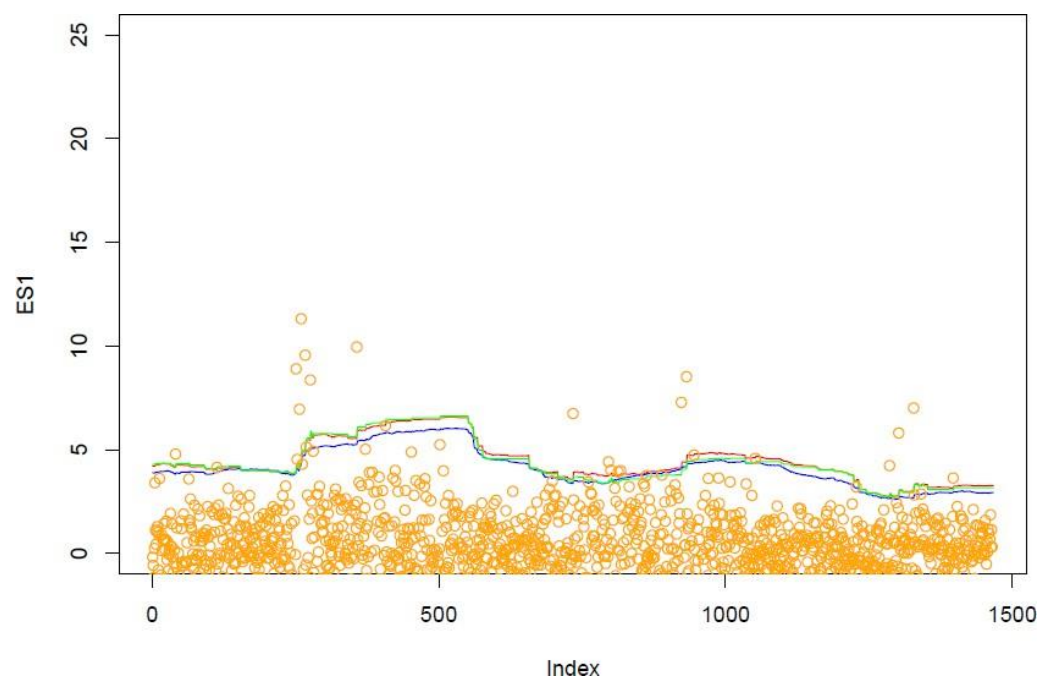
**Για  $p = 0.95$**

Για το VaR



Σχήμα 3.20

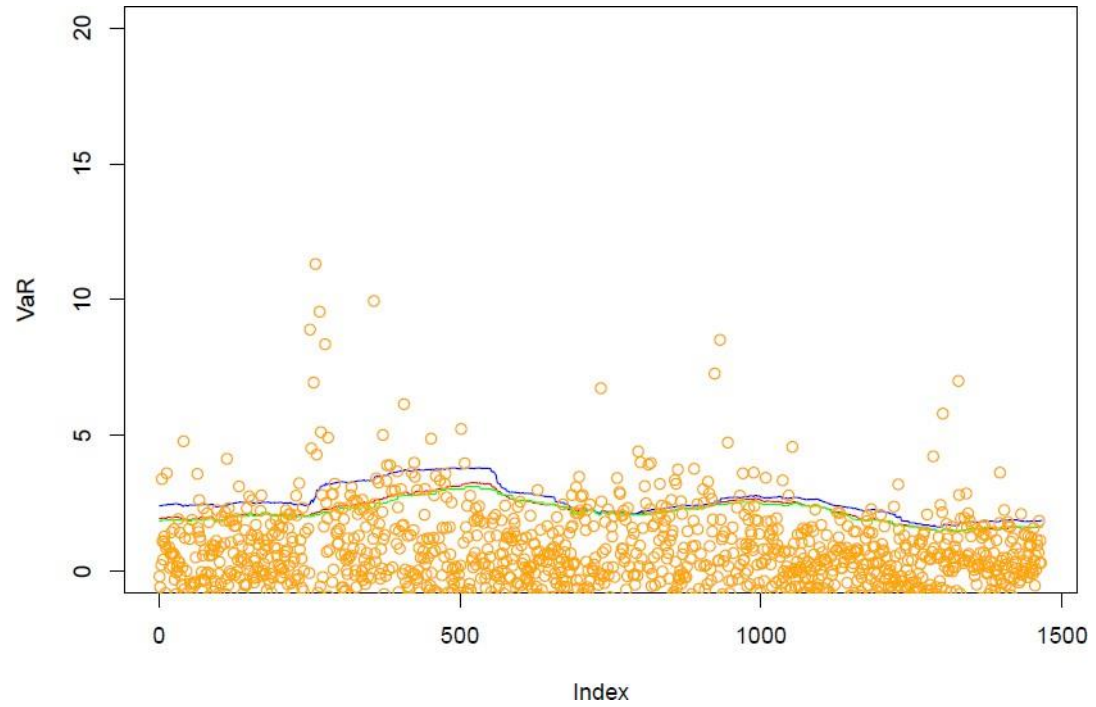
Για το ES



Σχήμα 3.21

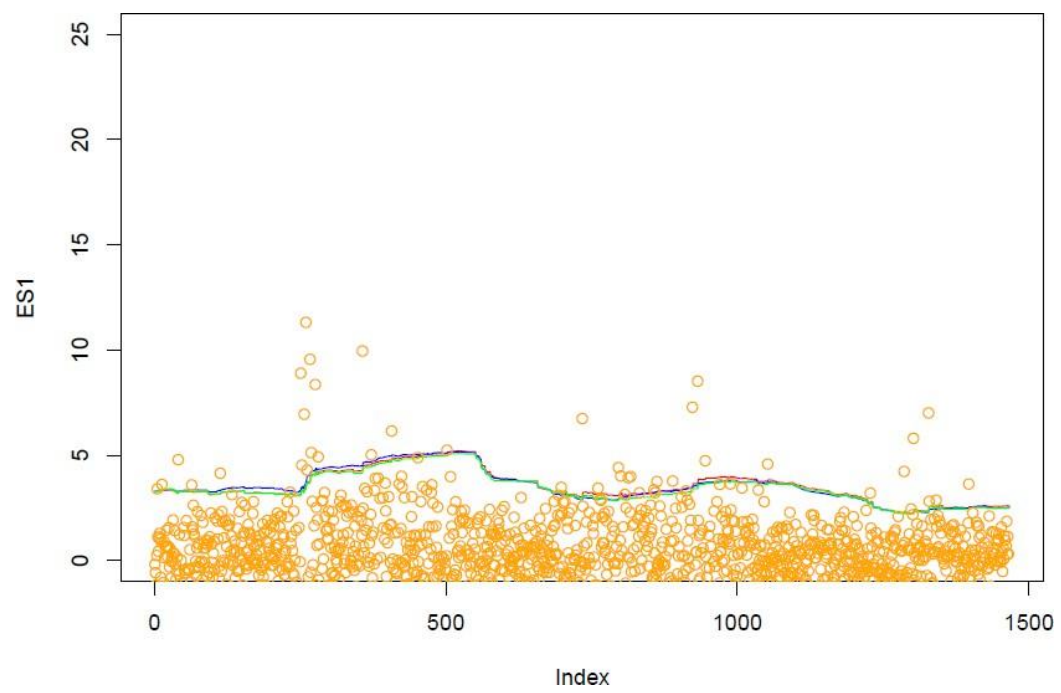
**Για  $p = 0.9$**

Για το VaR



Σχήμα 3.22

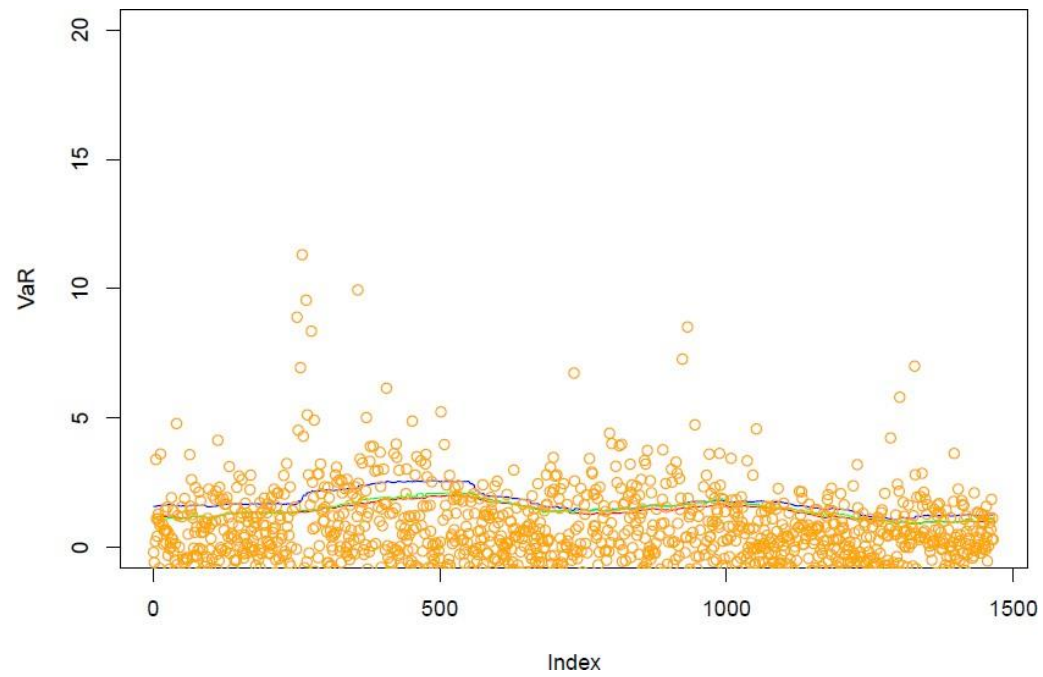
Για το ES



Σχήμα 3.23

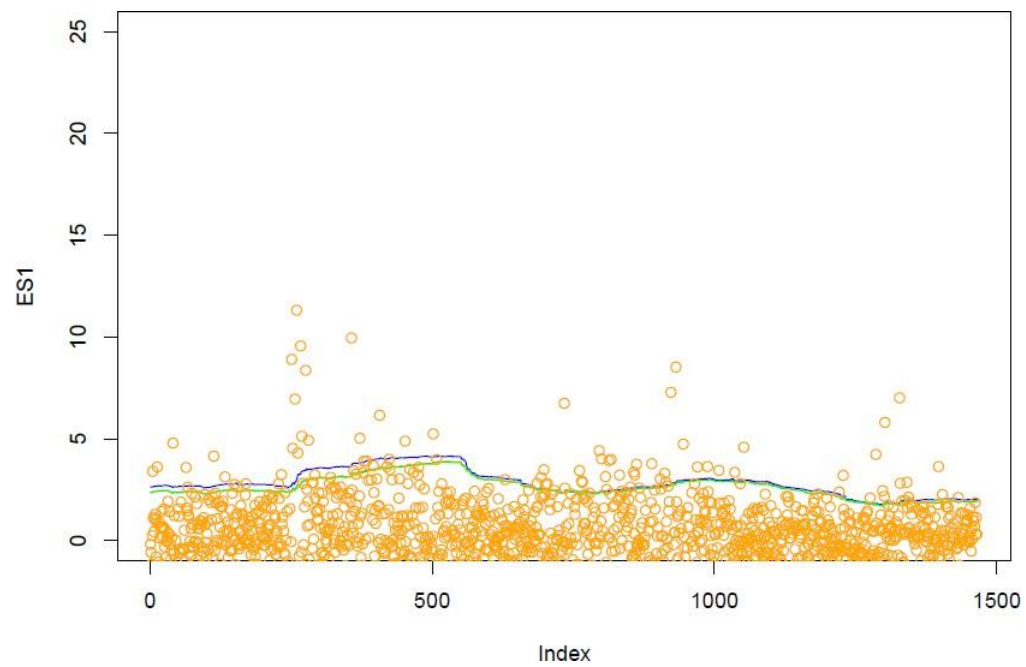
**Για  $p = 0.8$**

Για το VaR



Σχήμα 3.24

Για το ES



Σχήμα 3.25

Συνοπτικά τώρα κατασκευάζουμε τους αντίστοιχους πίνακες:

Model	p = 0.995	p = 0.990	p = 0.950	p = 0.900	p = 0.800
<b>POT</b>	0.99182 (0.09221)	<b>0.99114</b> <b>(0.79258)</b>	0.96115 (0.05486)	<b>0.89843</b> <b>(0.82776)</b>	0.79141 (0.41453)
<b>Normal</b>	0.99046 (0.02339)	0.98909 (0.69264)	0.96455 (0.00984)	0.93320 (1.05e-05)	0.83776 (0.000253)
<b>Empiric</b>	<b>0.99250</b> <b>(0.18888)</b>	0.98637 (0.18632)	<b>0.94956</b> <b>(0.90467)</b>	0.89162 (0.27680)	<b>0.79686</b> <b>(0.76896)</b>

Πίνακας 3.3

Model	p = 0.995	p = 0.990	p = 0.950	p = 0.900	p = 0.800
<b>POT</b>	6.17270	5.20117	3.10702	2.25478	<b>1.42149</b>
<b>Normal</b>	<b>5.12936</b>	<b>4.64265</b>	3.31315	2.60440	1.74616
<b>Empiric</b>	6.31080	5.15732	<b>2.91736</b>	<b>2.20998</b>	1.47957

Πίνακας 3.4

### 3.2.2. Συμπεράσματα (για την μετοχή AAPL)

Όπως φαίνεται και από τους εκ των υστέρων ελέγχους (back-testing), τα μηχανικά μοντέλα αποδίδουν καλύτερα στη μετοχή της Apple (AAPL). Συνολικά, το παραμετρικό/εμπειρικό (Empiric) επιτυγχάνει την πιο σταθερή βαθμονόμηση: τα ποσοστά επιτυχιών (hit rates) βρίσκονται πολύ κοντά στις θεωρητικές πιθανότητες σε όλα σχεδόν τα επίπεδα  $p$ , με  $p$ -values που δεν υποδεικνύουν σημαντικές αποκλίσεις. Το POT επίσης αποδίδει ικανοποιητικά—συχνά κοντά στο target  $p$  και με  $p$ -values που σπάνια οδηγούν σε απόρριψη—ενώ τείνει να είναι ελαφρώς πιο «συντηρητικό» γύρω στο 95%.

Το παραμετρικό μοντέλο Κανονικής κατανομής παρουσιάζει συστηματικές αστοχίες στη βαθμονόμηση: για μεσαία επίπεδα εμπιστοσύνης ( $p = 0.95, 0.90, 0.80$ ) εμφανίζει σημαντικές αποκλίσεις (χαμηλά  $p$ -values), γεγονός που συνεπάγεται κακή προσαρμογή στην ουρά. Στο άκρο ( $p = 0.995$ ) επίσης απορρίπτεται στατιστικά. Η εικόνα αυτή είναι συμβατή με το ότι οι

λογαριθμικές αποδόσεις της AAPL έχουν βαρύτερες ουρές από την Κανονική, οπότε το Normal υποεκτιμά ή/και παραμορφώνει τον κίνδυνο ανάλογα με το επίπεδο  $p$ .

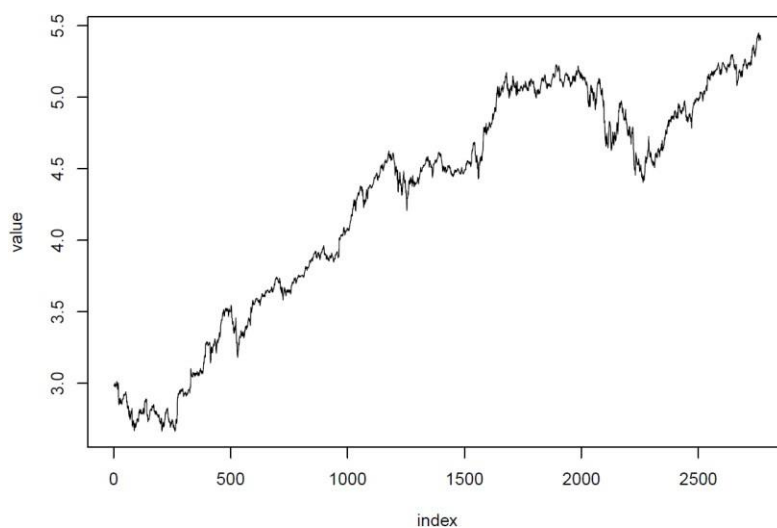
Για τα πολύ υψηλά quantiles, η εμπειρική προσέγγιση εμφανίζεται ιδιαίτερα ανταγωνιστική: στο 99.5% αποδίδει υψηλότερο μέσο VaR από τα υπόλοιπα (δηλαδή πιο αυστηρή κάλυψη) με hit rate πολύ κοντά στο θεωρητικό  $p$ . Το POT ακολουθεί από κοντά και παραμένει στατιστικά αποδεκτό στις περισσότερες περιπτώσεις, στοιχείο που συνάδει με τη θεωρητική του καταλληλότητα για βαριές ουρές. Σε μεσαία quantiles (0.95–0.80), τόσο το Empiric όσο και το POT διατηρούν ικανοποιητική βαθμονόμηση, με το Empiric να υπερέχει οριακά σε σταθερότητα, ενώ το Normal αποτυγχάνει.

Συνολικά, για την AAPL προκύπτει ότι η εμπειρική μέθοδος και η POT είναι οι προτεινόμενες επιλογές: Η εμπειρική ως προεπιλογή λόγω σταθερής βαθμονόμησης σε όλο το εύρος  $p$ , και η POT ως ισχυρή εναλλακτική/διασταύρωση ειδικά όταν το ενδιαφέρον εστιάζει σε ακραίους κινδύνους και tail behavior. Η παραμετρική μέθοδος (μέσω κανονικής) δεν συστήνεται για παραγωγική χρήση στα συγκεκριμένα δεδομένα, καθώς οι έλεγχοι βαθμονόμησης καταδεικνύουν επίμονες αποκλίσεις.

### 3.3. Εφαρμογή με βάση τις τιμές της μετοχής της AMAZON

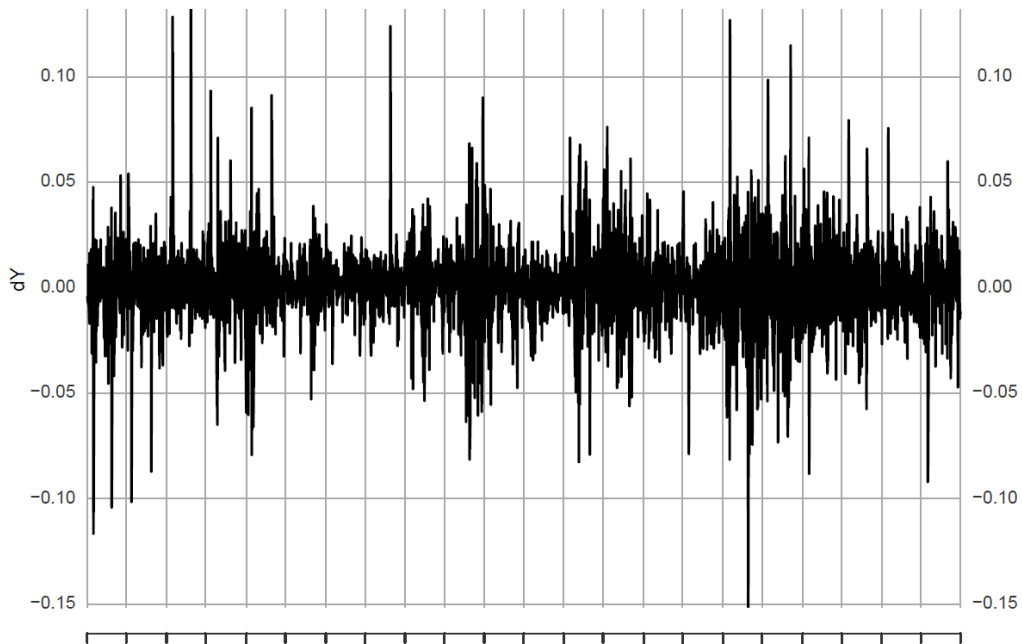
#### 3.3.1 Εκτιμήσεις των μέτρων κινδύνου

Μπορούμε να δούμε στο ακόλουθο γράφημα την εξέλιξη της τιμής της μετοχής AMAZON.



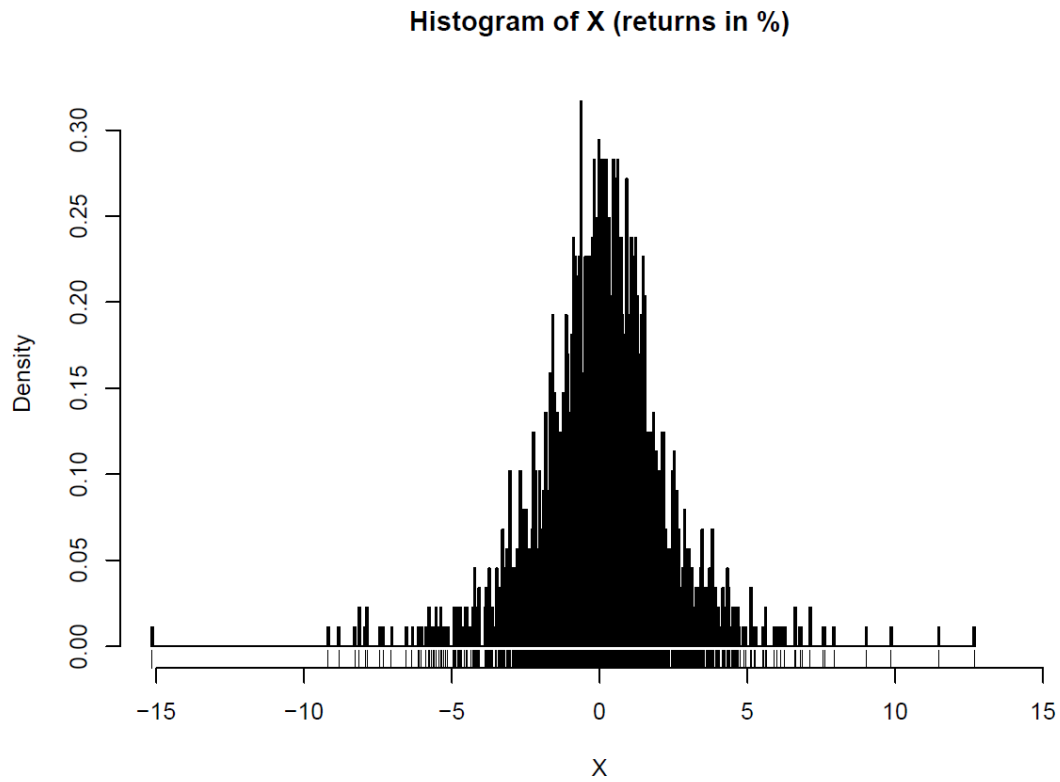
Σχήμα 3.26

Το γράφημα των πρώτων διαφορών θα είναι της μορφής

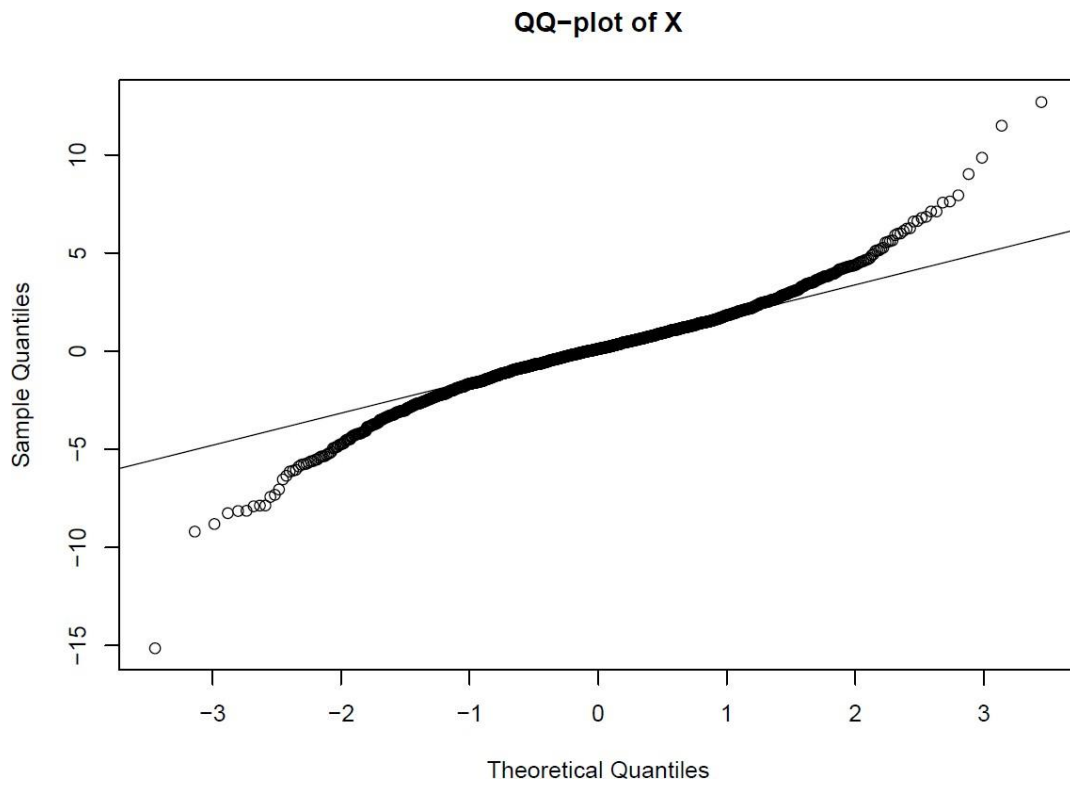


Σχήμα 3.27

Επίσης ένα ιστόγραμμα τους και ένα QQ plot ως προς την κανονική είναι της μορφής

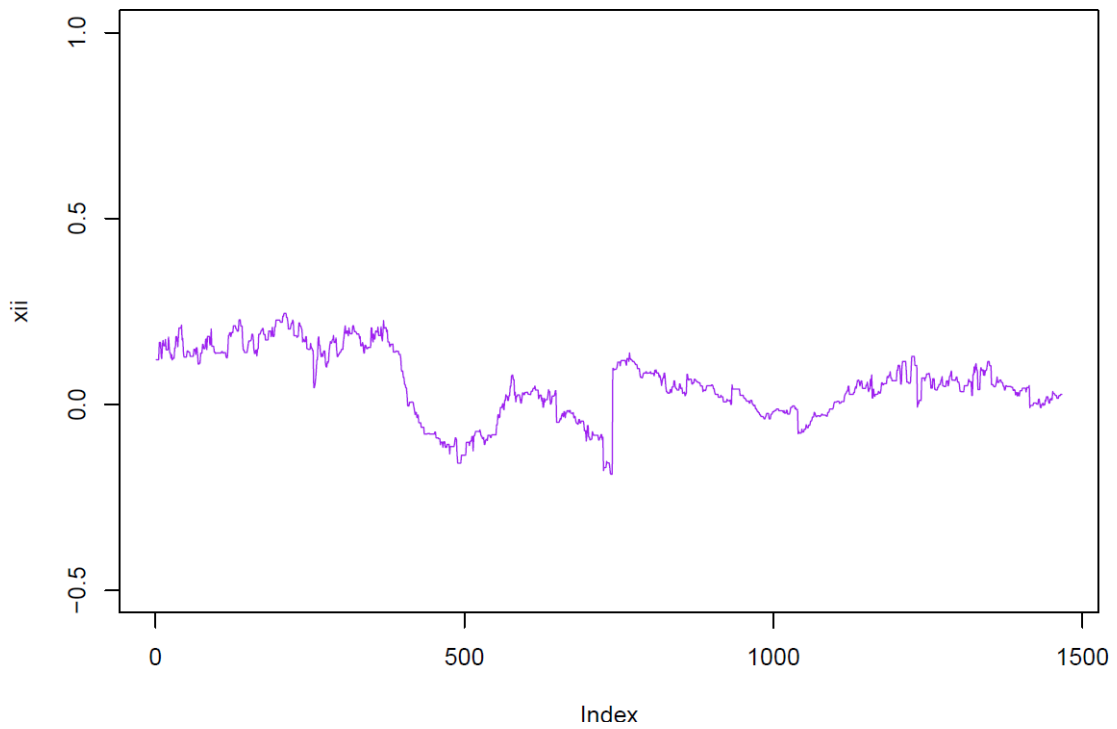


Σχήμα 3.28



Σχήμα 3.29

Για την εκτίμηση του  $\xi$

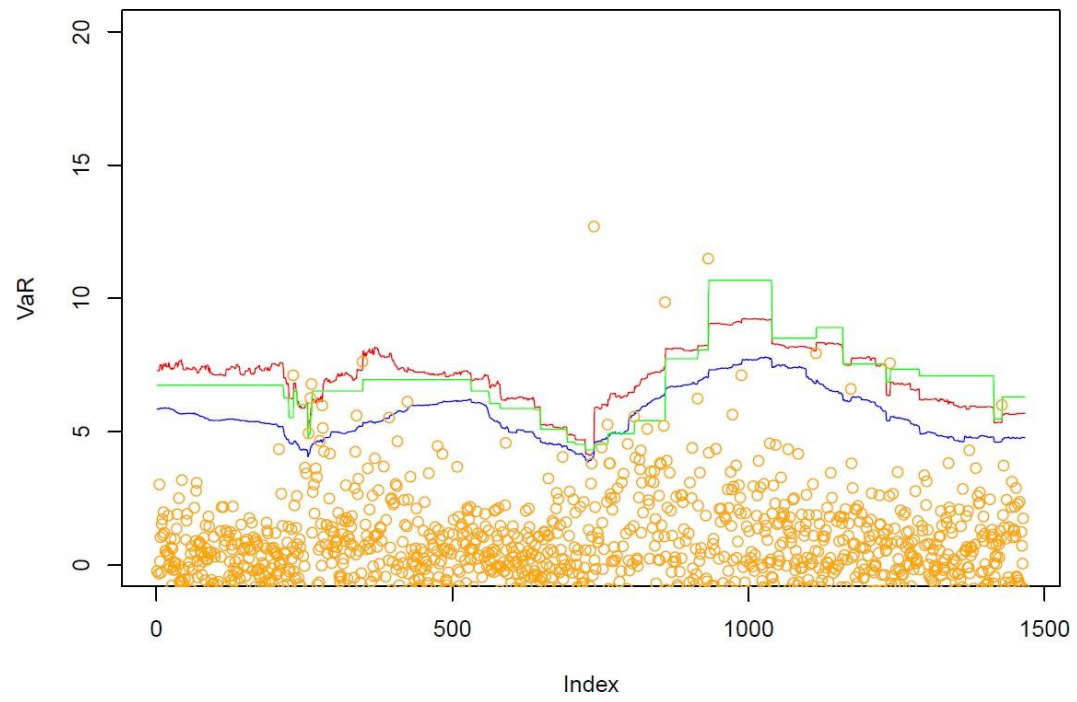


Σχήμα 3.30

Τώρα για κάθε διαφορετική τιμή του  $p$  ακολουθούν τα αντίστοιχα διαγράμματα:

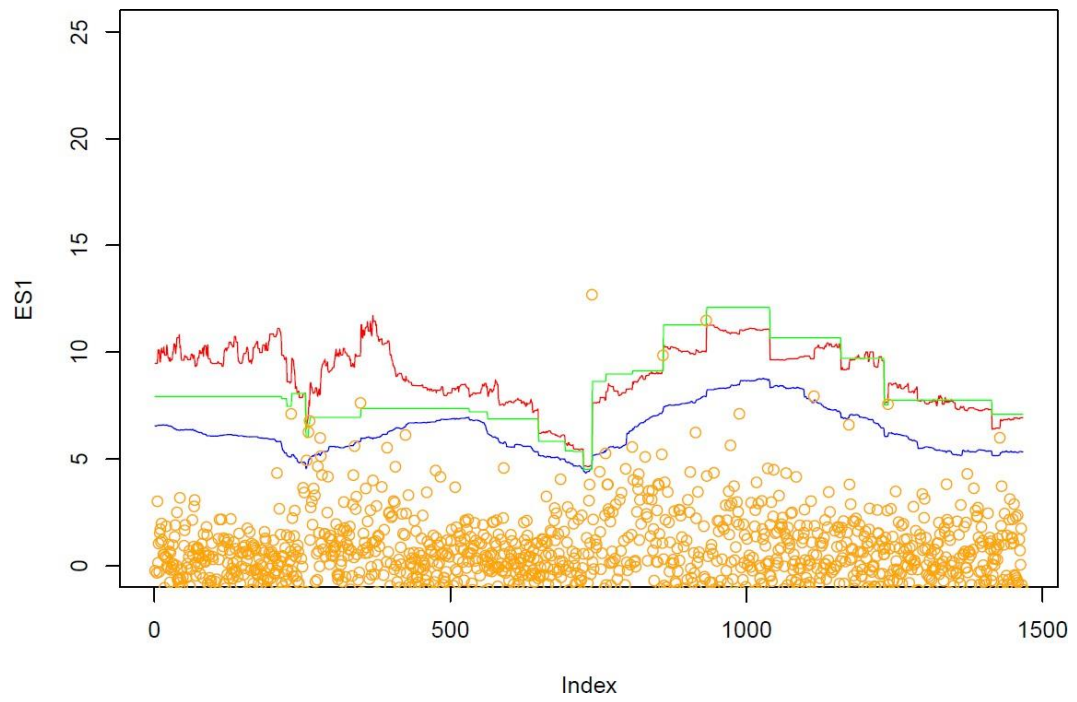
**Για  $p = 0.995$**

Για το VaR



Σχήμα 3.31

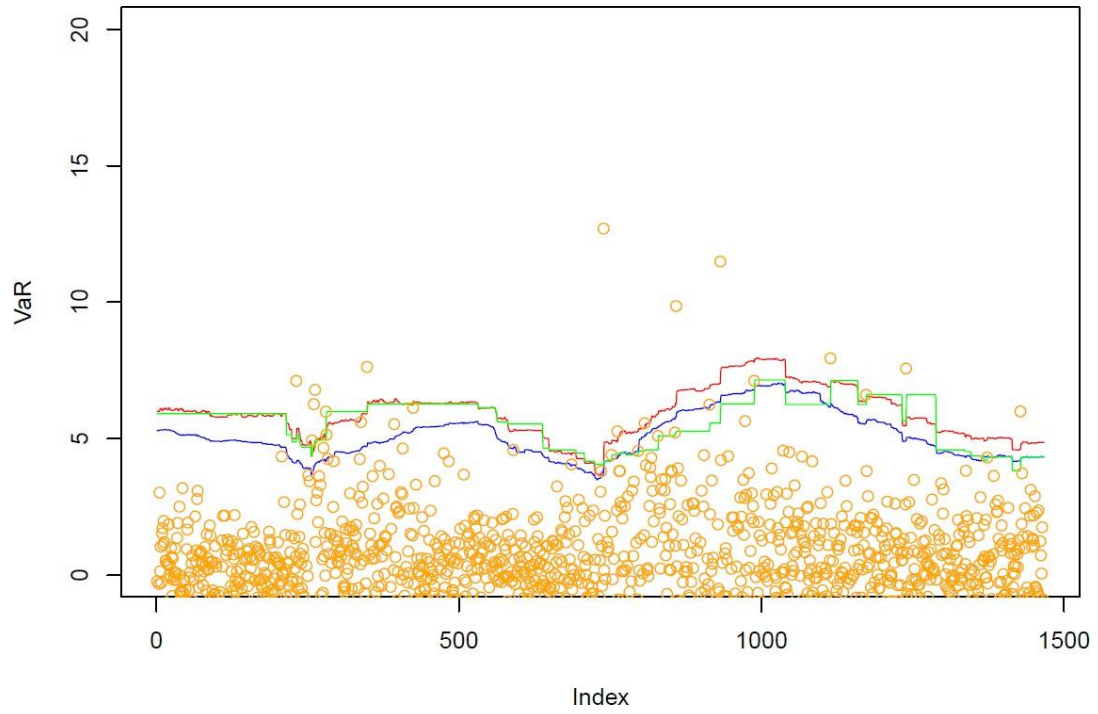
Για το ES



Σχήμα 3.32

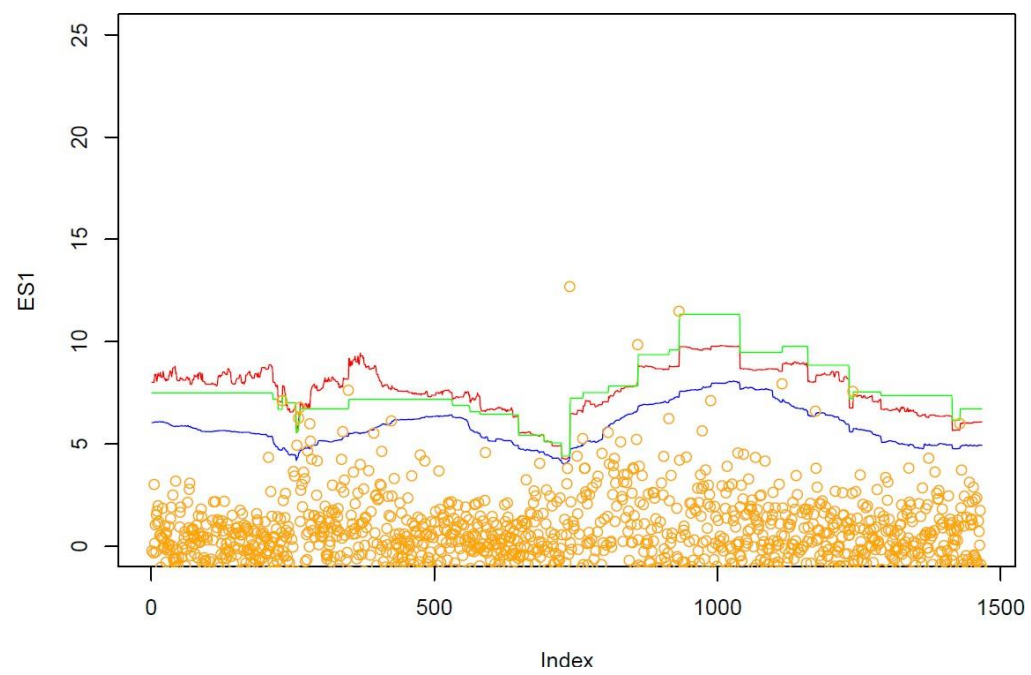
**Για  $p = 0.99$**

Για το VaR



Σχήμα 3.33

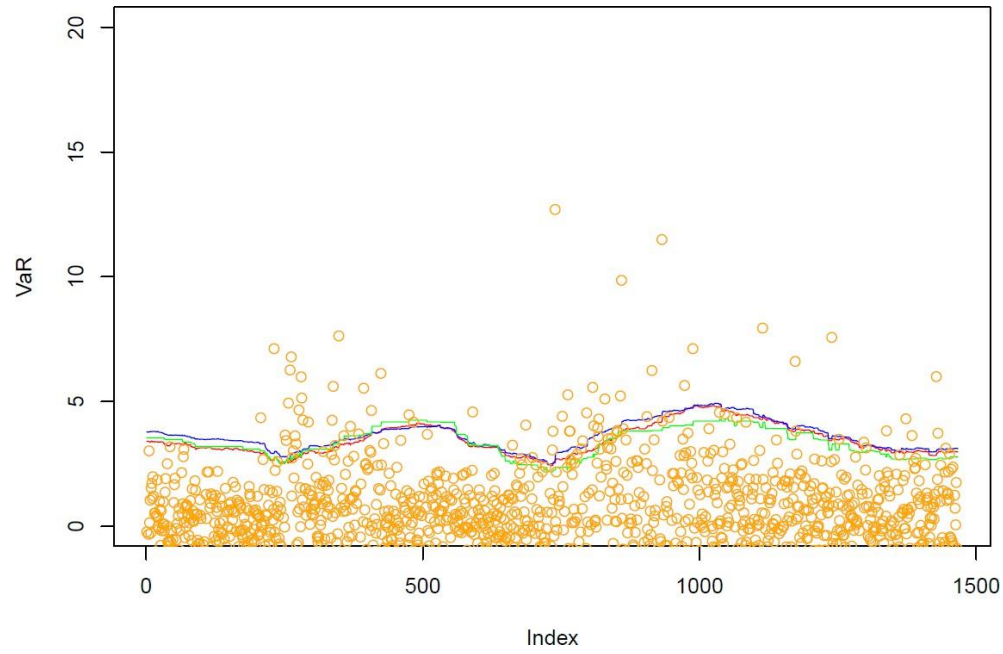
Για το ES



Σχήμα 3.34

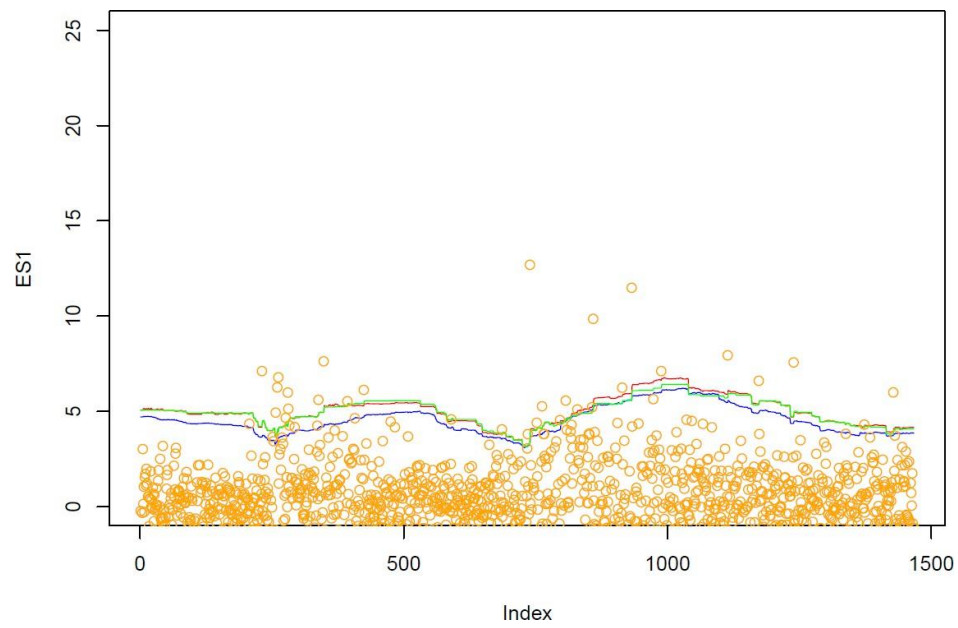
**Για  $p = 0.95$**

Για το VaR



Σχήμα 3.35

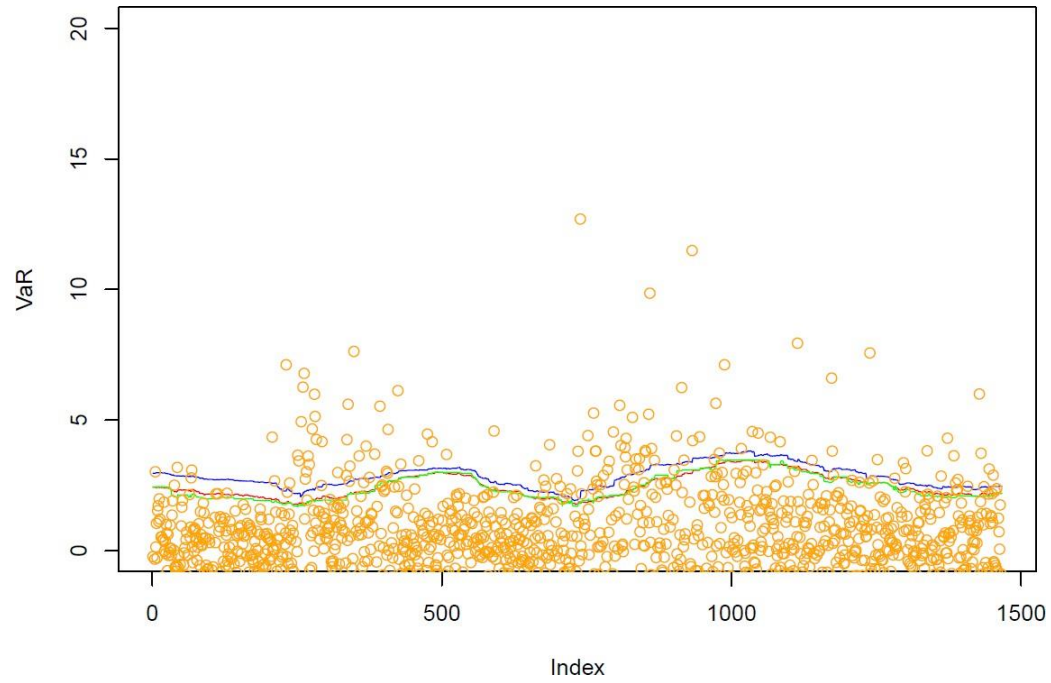
Για το ES



Σχήμα 3.36

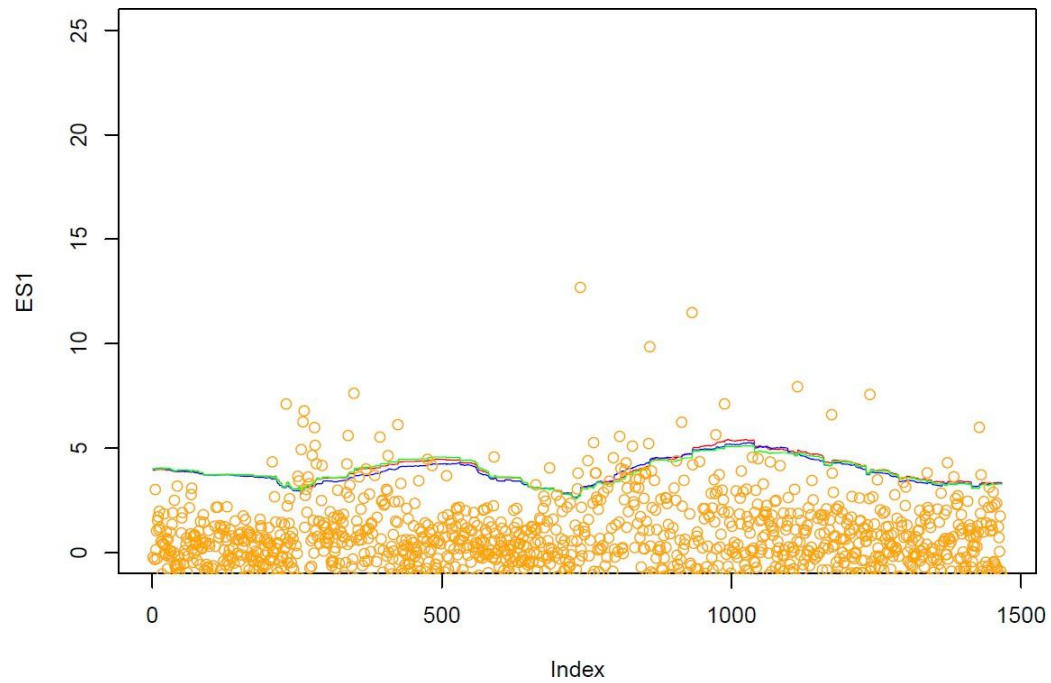
**Για  $p = 0.9$**

Για το VaR



Σχήμα 3.27

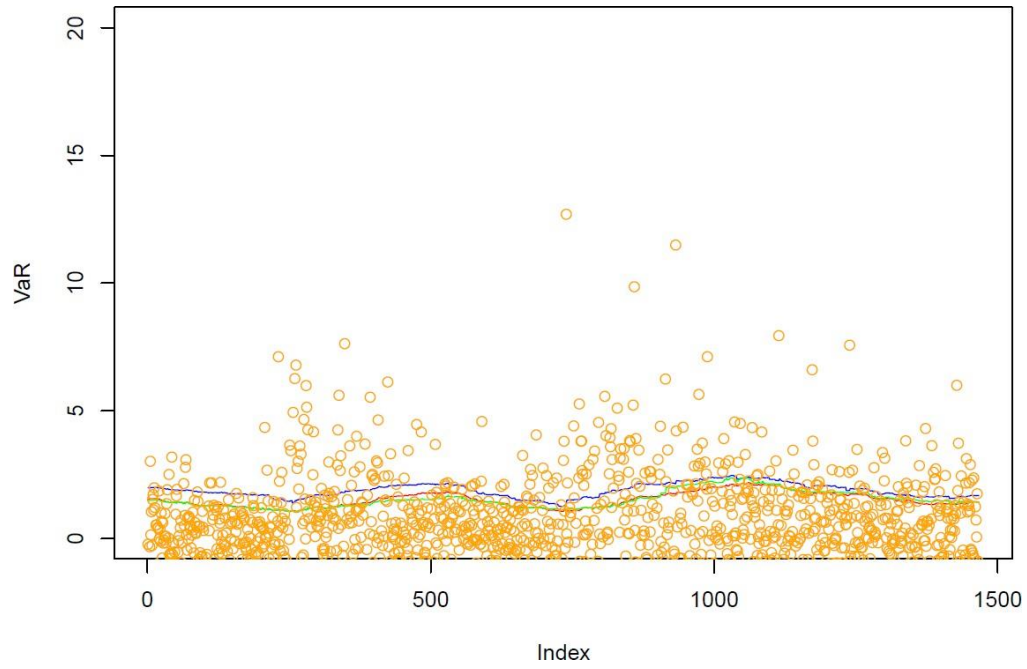
Για το ES



Σχήμα 3.28

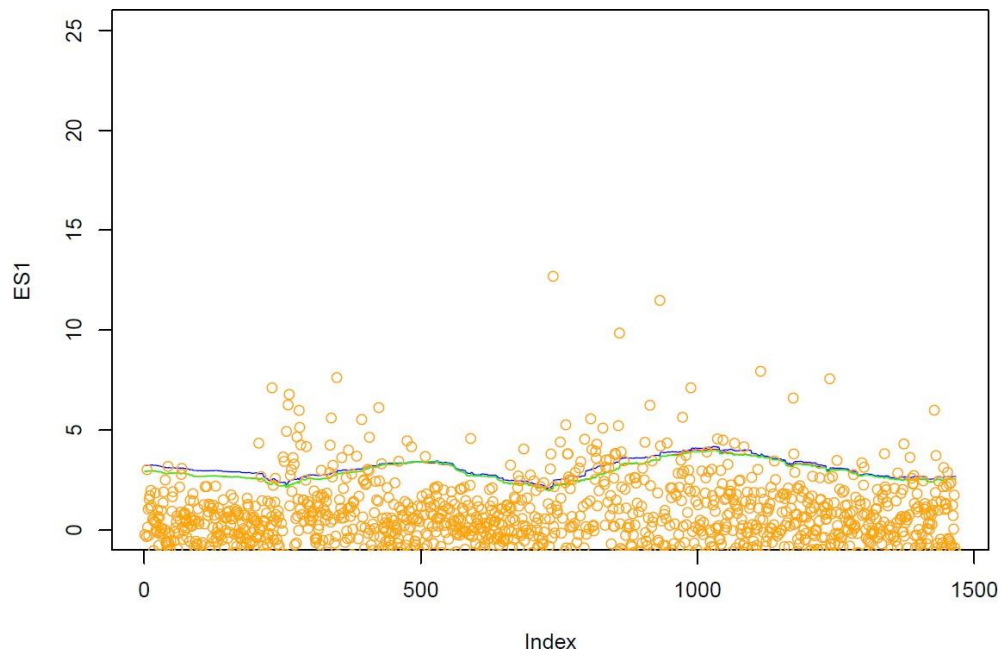
**Για  $p = 0.8$**

Για το VaR



Σχήμα 3.29

Για το ES



Σχήμα 3.30

Συνοπτικά τώρα κατασκευάζοντας τους αντίστοιχους πίνακες:

Model	p = 0.995	p = 0.99	p = 0.95	p = 0.9	p = 0.8
<b>POT</b>	<b>0.99387</b> <b>(0.459193)</b>	<b>0.98909</b> <b>(0.692642)</b>	<b>0.95092</b> <b>(0.952197)</b>	<b>0.90048</b> <b>(1.0)</b>	0.80982 (0.360857)
<b>Normal</b>	0.98705 (0.0002212)	0.98296 (0.0117621)	0.95842 (0.150254)	0.91888 (0.0147715)	0.85072 (5.68e-07)
<b>Empiric</b>	0.99182 (0.0922108)	0.98500 (0.0643338)	0.94138 (0.133889)	0.89707 (0.695338)	<b>0.80709</b> <b>(0.514009)</b>

Πίνακας 3.5

Model	p = 0.995	p = 0.990	p = 0.950	p = 0.900	p = 0.800
<b>POT</b>	7.07110	5.93153	3.47216	2.47918	<b>1.51854</b>
<b>Normal</b>	<b>5.67365</b>	<b>5.12927</b>	3.64221	2.84947	1.88952
<b>Empiric</b>	6.98233	5.64256	<b>3.37972</b>	<b>2.45216</b>	1.53242

Πίνακας 3.6

### 3.3.2. Συμπεράσματα

Όπως διαπιστώνεται από τα αποτελέσματα των εκ των υστέρων ελέγχων (back-testing), το **μοντέλο POT** εμφανίζεται πιο συνεπές στην εκτίμηση του VaR για όλες σχεδόν τις τιμές του p. Στις περισσότερες περιπτώσεις, τα ποσοστά υπερβάσεων (hit rates) βρίσκονται πολύ κοντά στις θεωρητικές πιθανότητες και τα αντίστοιχα p-values υποδηλώνουν ότι η μέθοδος είναι στατιστικά αποδεκτή. Για παράδειγμα, για p = 0.995, 0.99 και 0.95 το POT παρουσιάζει αποτελέσματα απόλυτα ευθυγραμμισμένα με τις αναμενόμενες τιμές, με υψηλή στατιστική αποδοχή.

Αντίθετα, το **παραμετρικό μοντέλο** που βασίζεται στην κανονική κατανομή εμφανίζει συστηματικές αποκλίσεις. Παρατηρείται ότι στις υψηλές τιμές εμπιστοσύνης (π.995 και 0.99) το Normal τείνει να **υποεκτιμά τον κίνδυνο**, με αποτέλεσμα να δίνει περισσότερες υπερβάσεις και χαμηλά p-values (π.χ. 0.00022 για p=0.995). Στις μεσαίες και χαμηλότερες τιμές εμπιστοσύνης (π.90

και 0.80) το Normal καταλήγει σε **υπερεκτίμηση** του κινδύνου, καθώς τα ποσοστά υπερβάσεων ξεφεύγουν σημαντικά από τα θεωρητικά και τα p-values απορρίπτουν τη στατιστική συμβατότητα.

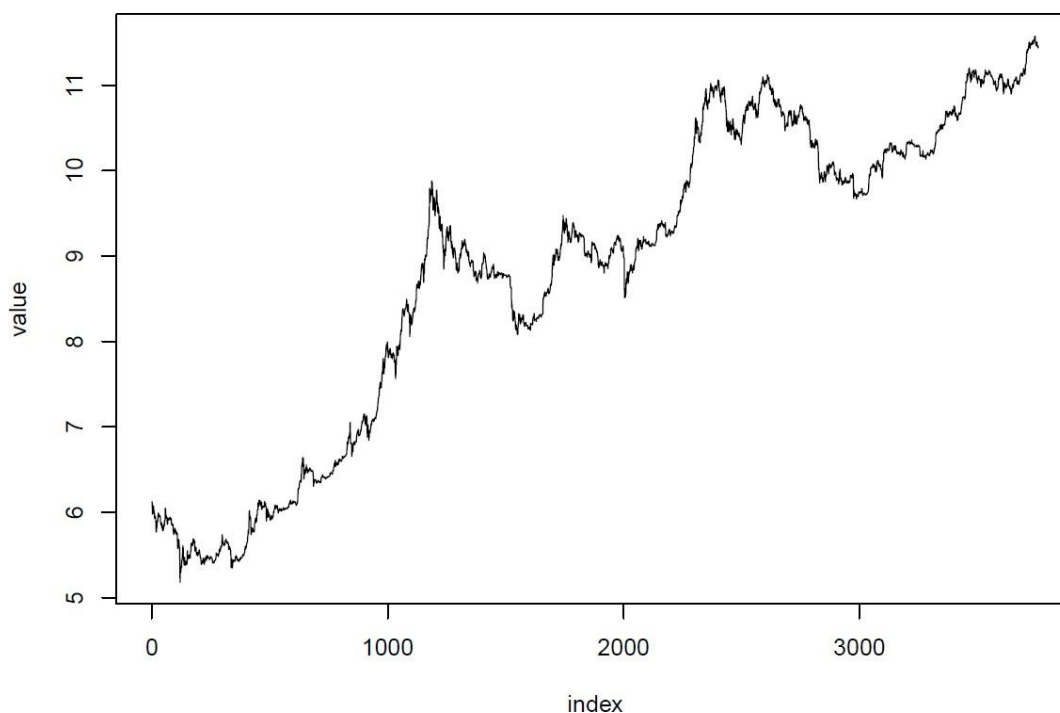
Το **εμπειρικό μοντέλο** (Empiric) αποδίδει καλύτερα από το κανονικό στις περισσότερες περιπτώσεις, ιδιαίτερα για ενδιάμεσες τιμές του  $p$ . Παρόλα αυτά, παρουσιάζει αστάθεια στις υψηλές τιμές εμπιστοσύνης, όπου επηρεάζεται δυσανάλογα από ακραίες παρατηρήσεις και οι εκτιμήσεις του VaR αποκλίνουν (π.χ. για  $p=0.95$  έχουμε hit rate 0.94138 έναντι θεωρητικού 0.95, με οριακή στατιστική αποδοχή).

Συνολικά, για τη μετοχή **Amazon (AMZN)**, η μέθοδος **POT** αναδείχθηκε η πιο αξιόπιστη και ακριβής τόσο στην εκτίμηση του VaR όσο και του ES. Η απόδοσή της επιβεβαιώνει τη θεωρητική της καταλληλότητα για δεδομένα με βαριές ουρές, όπως είναι οι αποδόσεις της AMZN, σε αντίθεση με το κανονικό μοντέλο που απέτυχε να αποτυπώσει σωστά τον κίνδυνο.

### 3.4. Εφαρμογή με βάση τις τιμές του BITCOIN

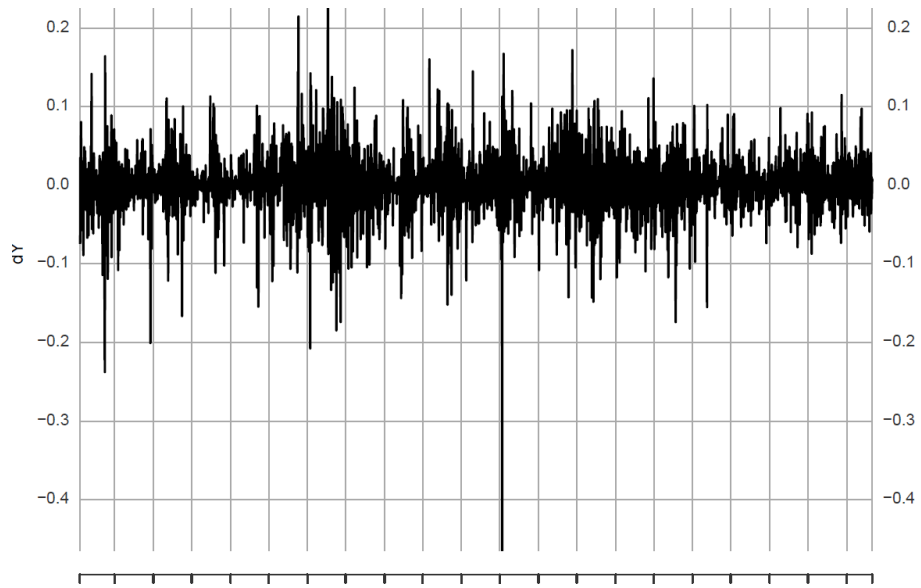
#### 3.4.1 Εκτιμήσεις των μέτρων κινδύνου

Μπορούμε να δούμε στο ακόλουθο γράφημα την εξέλιξη της τιμής του BITCOIN.



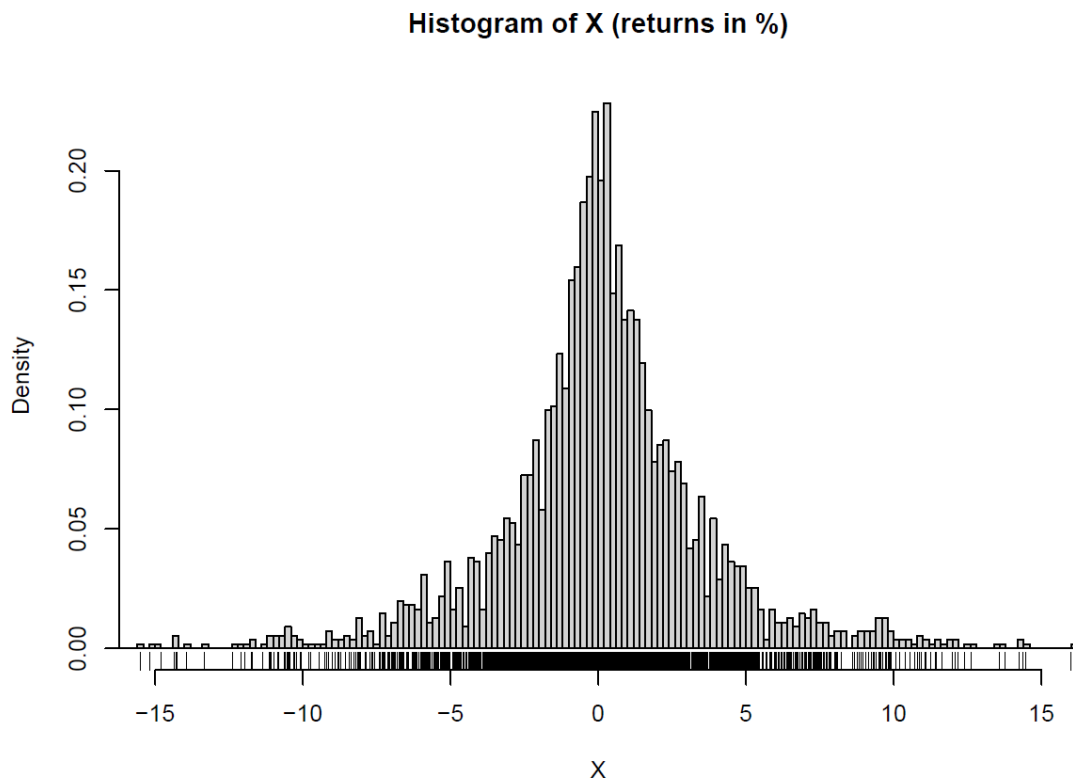
Σχήμα 3.31

Το γράφημα των πρώτων διαφορών θα είναι της μορφής

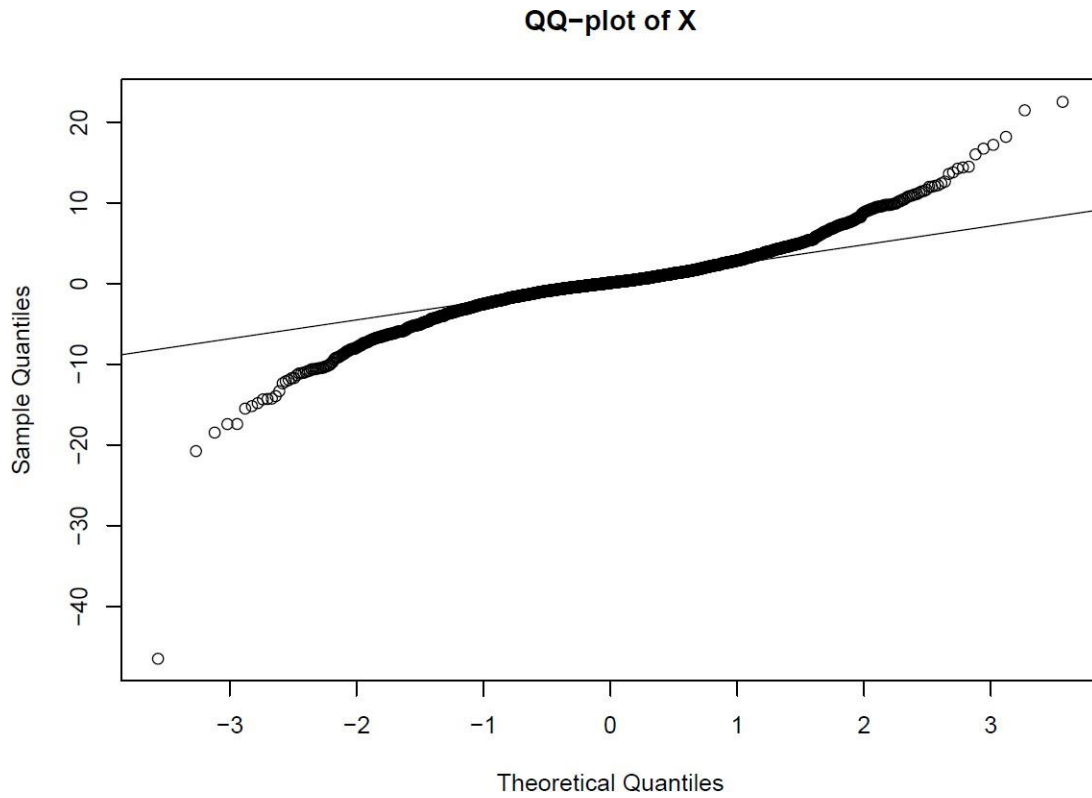


Σχήμα 3.32

Επίσης ένα ιστόγραμμα τους και ένα QQ plot ως προς την κανονική είναι της μορφής

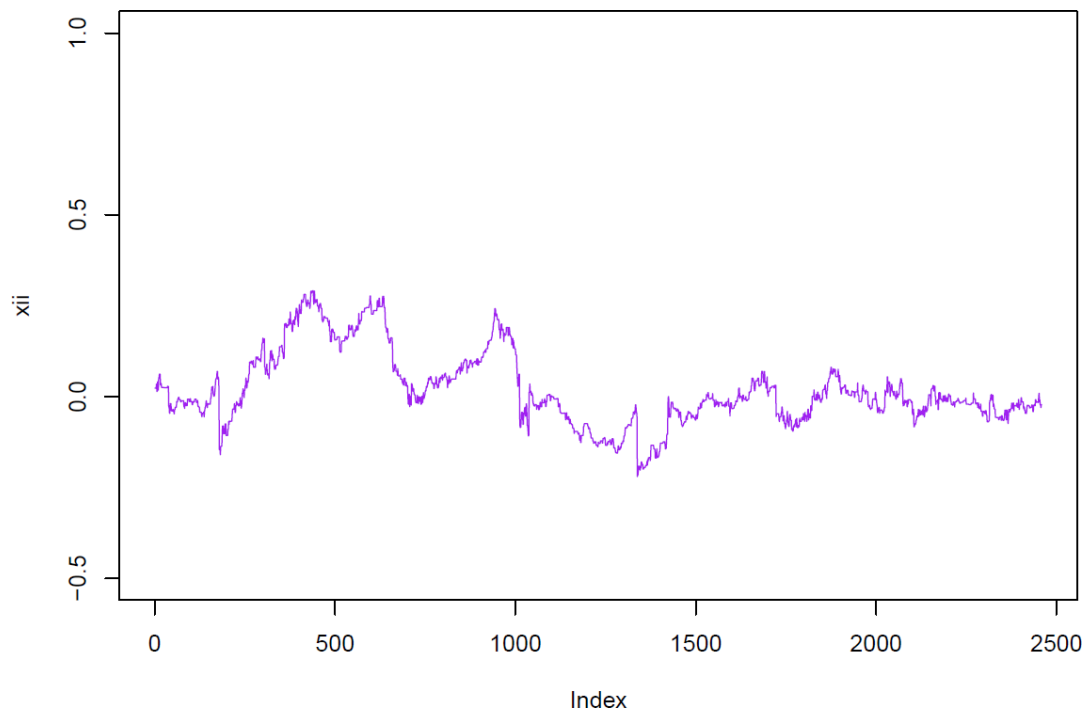


Σχήμα 3.33



Σχήμα 3.34

Για την εκτίμηση του  $\xi$

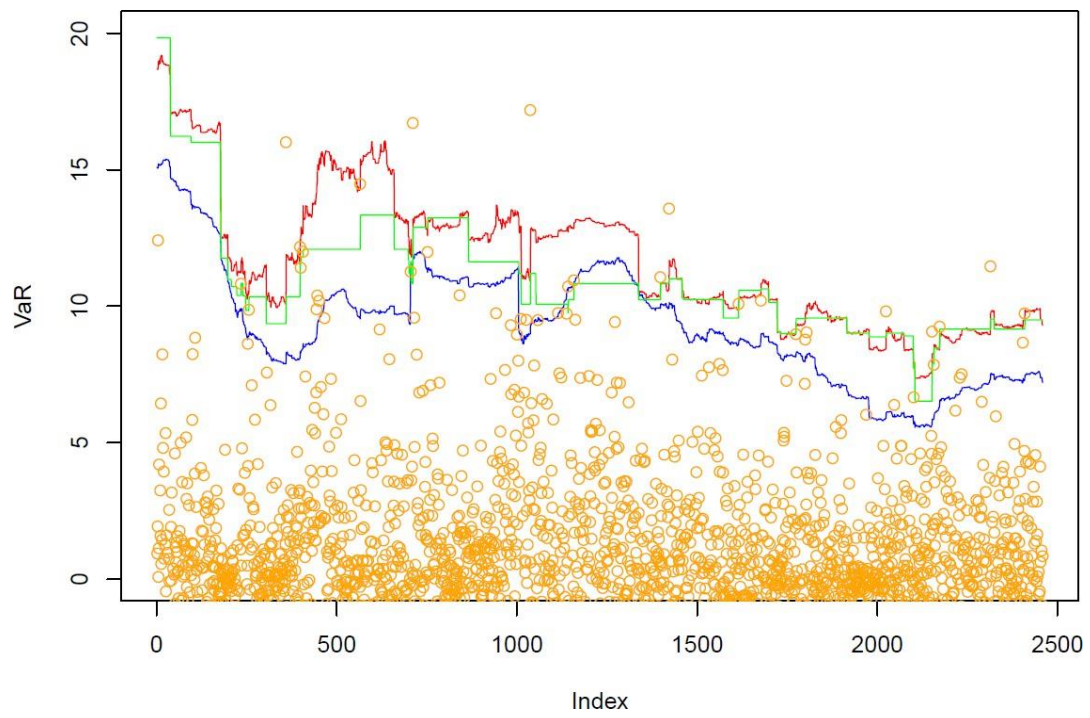


Σχήμα 3.35

Τώρα για κάθε διαφορετική τιμή του  $p$  ακολουθούν τα αντίστοιχα διαγράμματα:

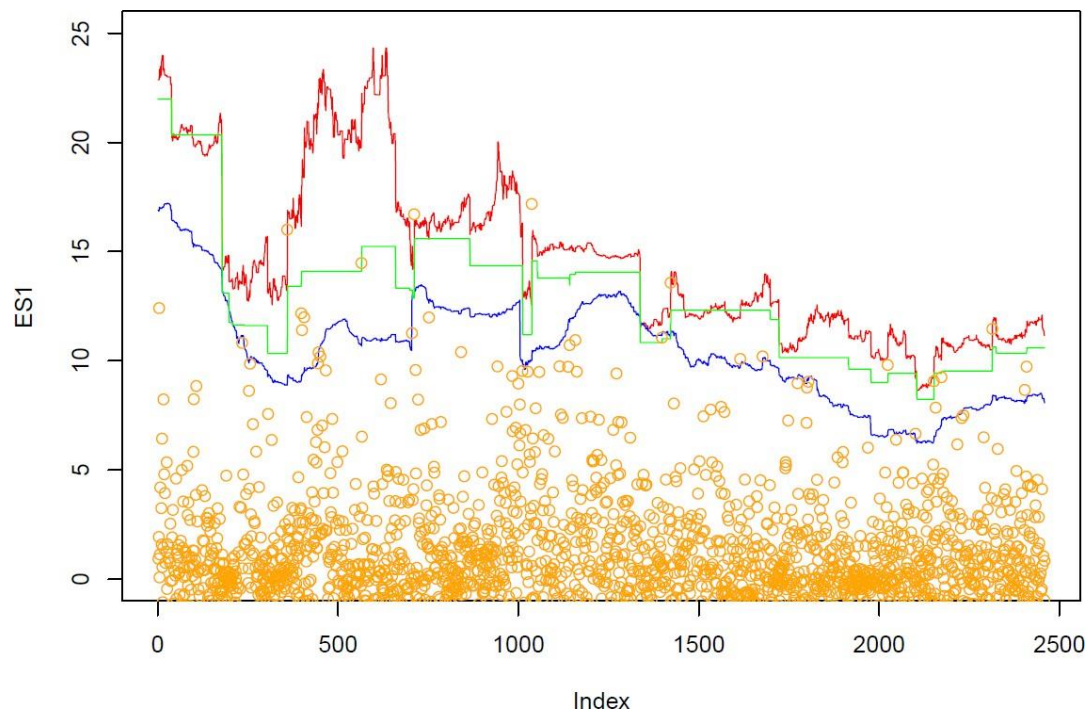
**Για  $p = 0.995$**

Για το VaR



Σχήμα 3.36

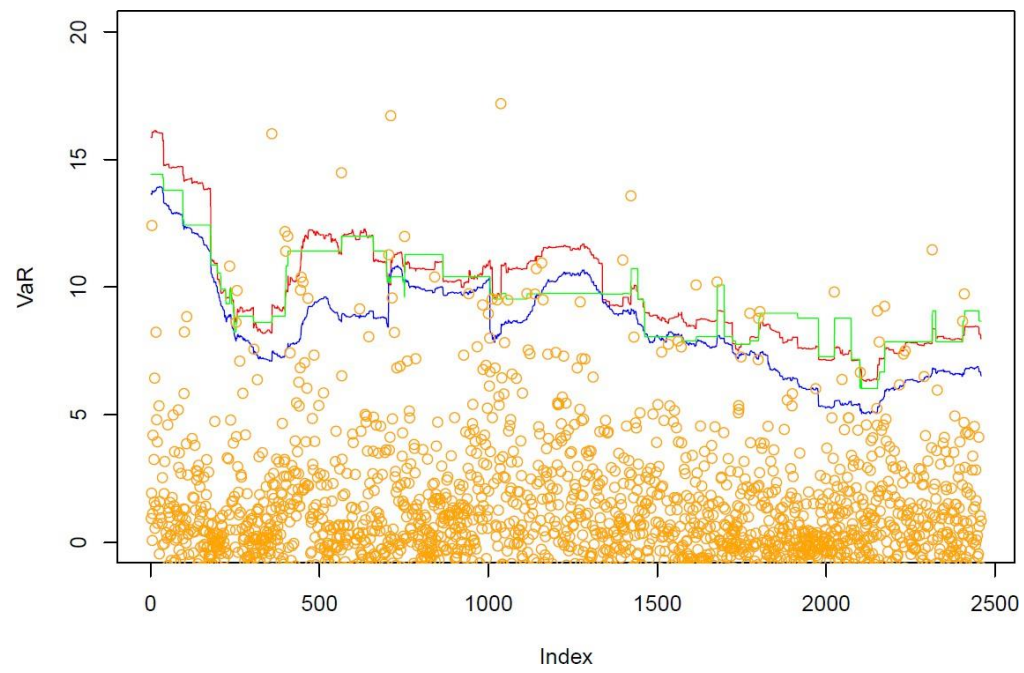
Για το ES



Σχήμα 3.37

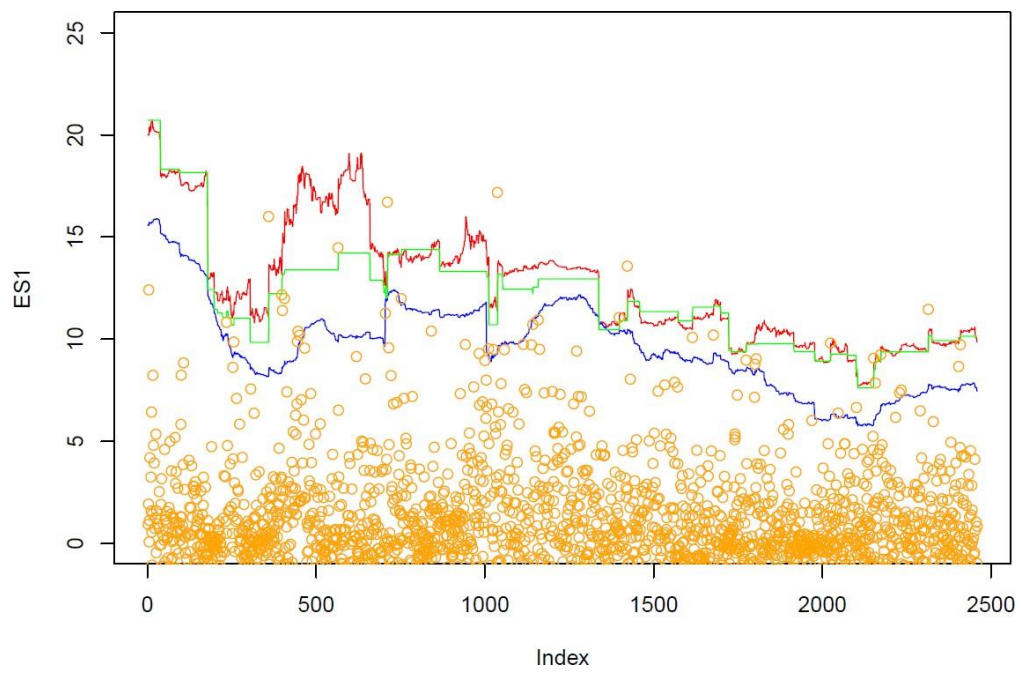
**Για  $p = 0.99$**

Για το VaR



Σχήμα 3.38

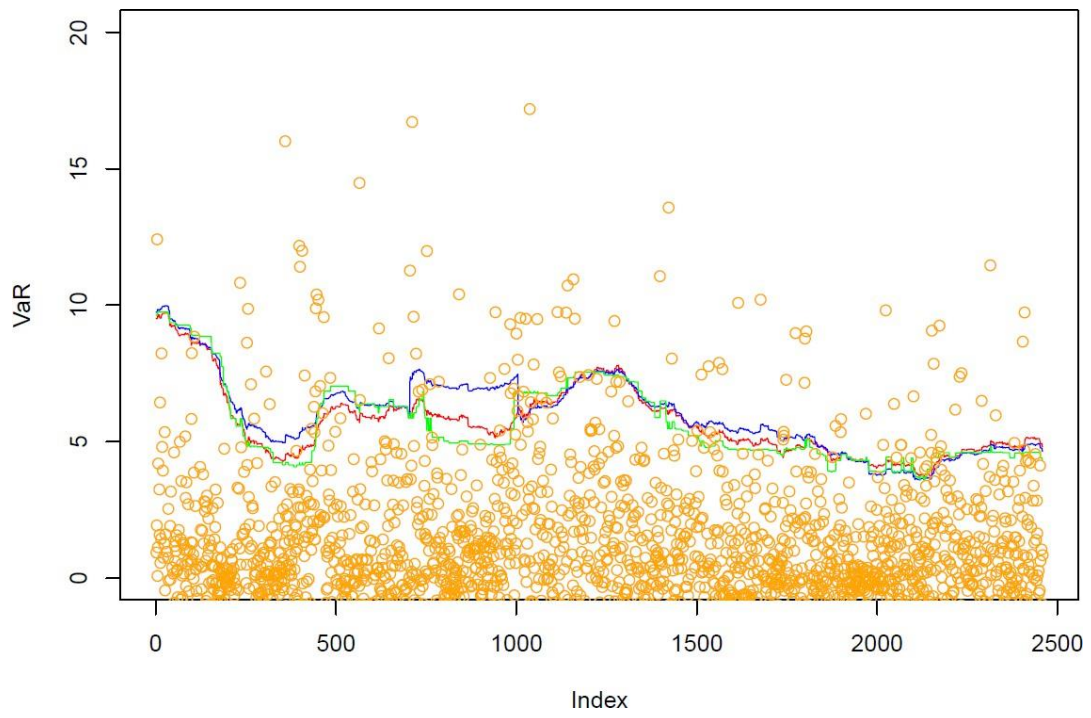
Για το ES



Σχήμα 3.39

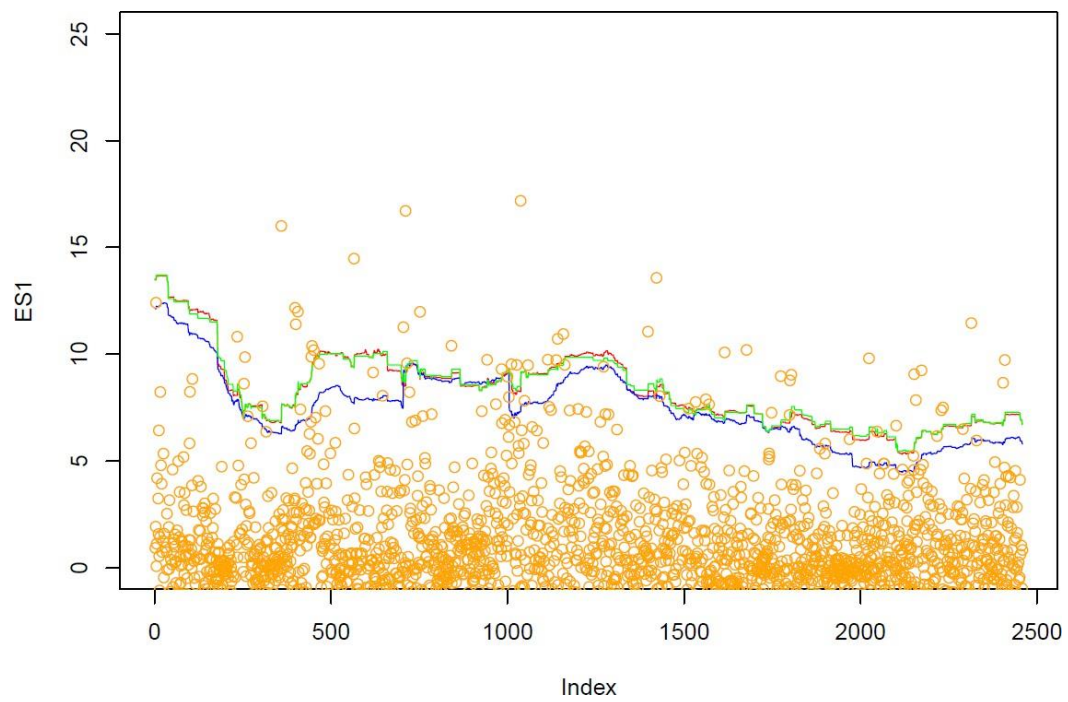
**Για  $p = 0.95$**

Για το VaR



Σχήμα 3.40

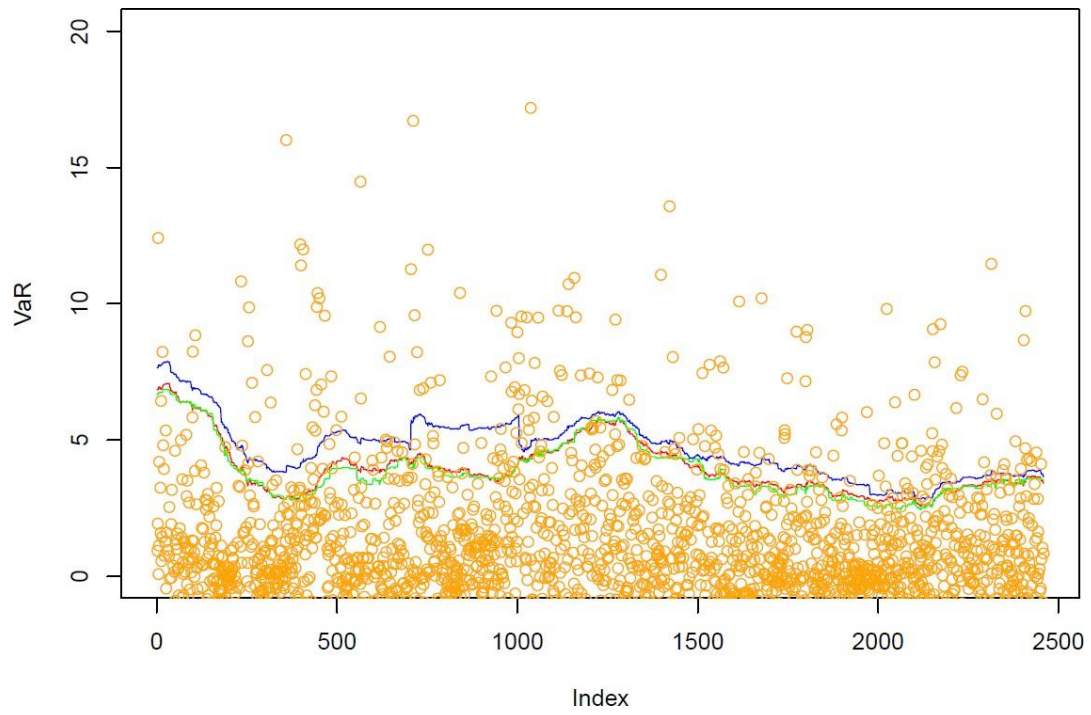
Για το ES



Σχήμα 3.40

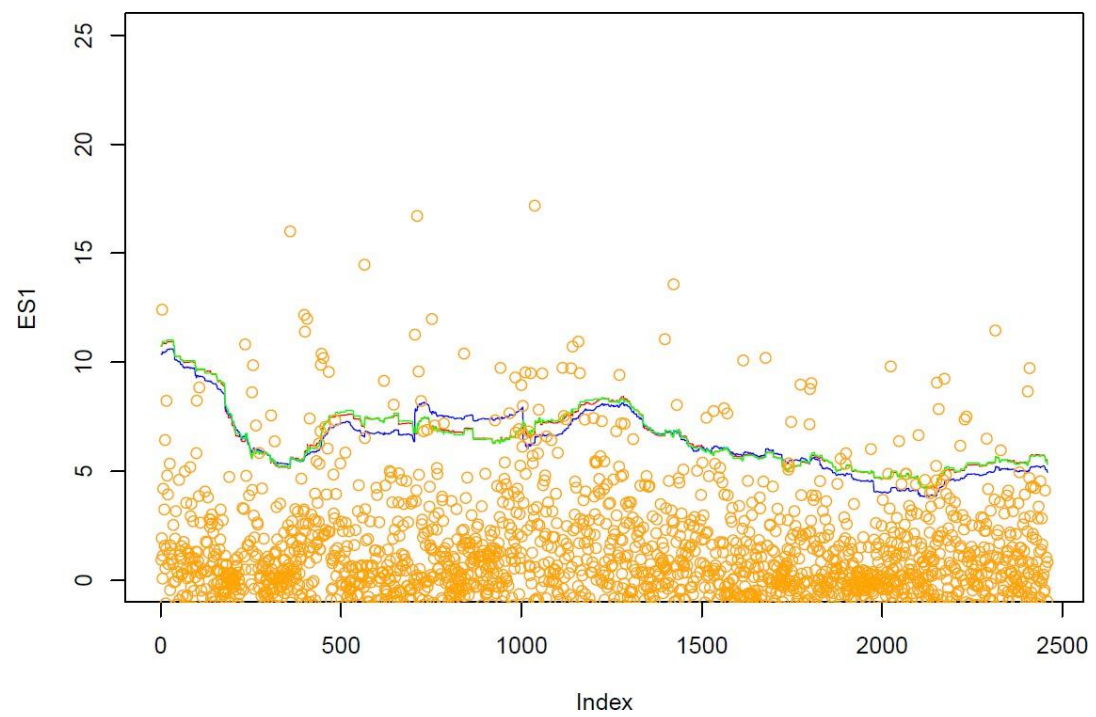
**Για  $p = 0.9$**

Για το VaR



Σχήμα 3.41

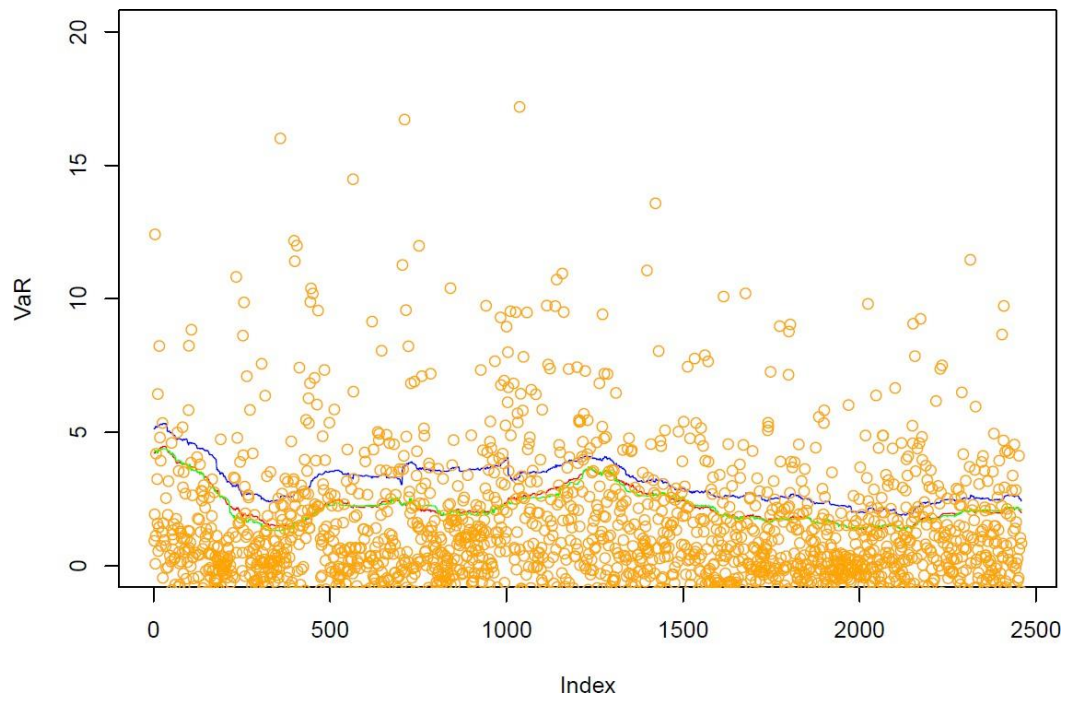
Για το ES



Σχήμα 3.42

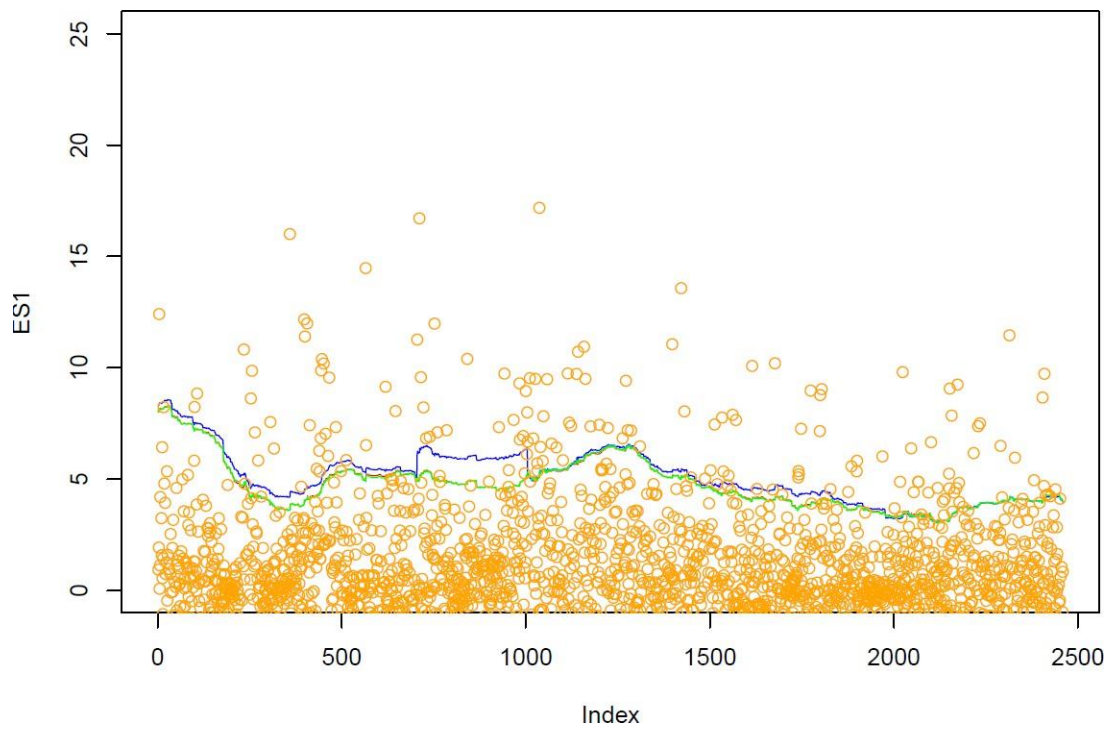
**Για  $p = 0.8$**

Για το VaR



Σχήμα 3.43

Για το ES



Σχήμα 3.44

Συνοπτικά τώρα κατασκευάζοντας τους αντίστοιχους πίνακες:

Model	p = 0.995	p = 0.99	p = 0.95	p = 0.9	p = 0.8
<b>POT</b>	<b>0.99431</b> <b>(0.566662)</b>	<b>0.98902</b> <b>(0.611299)</b>	<b>0.95120</b> <b>(0.817153)</b>	0.90606 (0.329731)	0.80643 (0.434627)
<b>Normal</b>	0.98495 (9.0e-09)	0.98007 (1.0e-05)	0.95608 (0.179466)	0.93127 (7.0e-08)	0.86783 (1.3e-18)
<b>Empiric</b>	0.99227 (0.0617973)	0.98780 (0.263955)	0.95364 (0.431621)	<b>0.90240</b> <b>(0.711683)</b>	<b>0.80439</b> <b>(0.59664)</b>

Πίνακας 3.7

Model	p = 0.995	p = 0.99	p = 0.95	p = 0.9	p = 0.8
<b>POT</b>	11.7066	9.84141	5.70399	3.97949	2.27587
<b>Normal</b>	<b>9.3306</b>	<b>8.43632</b>	5.99346	4.69118	3.11422
<b>Empiric</b>	10.9508	9.62036	<b>5.64403</b>	<b>3.89053</b>	<b>2.23849</b>

Πίνακας 3.8

### 3.4.2. Συμπεράσματα

Όπως φαίνεται από τους εκ των υστέρων ελέγχους (back-testing), το **μοντέλο POT** παρέχει τη σταθερότερη και πιο αξιόπιστη εκτίμηση της αξίας σε κίνδυνο (VaR) για το Bitcoin σε όλες τις πιθανότητες εμπιστοσύνης. Τα ποσοστά επιτυχιών που καταγράφηκαν βρίσκονται κοντά στις θεωρητικές τιμές (ιδιαίτερα για  $p = 0.995$ ,  $0.99$  και  $0.95$ ), ενώ τα p-values δείχνουν ότι η μέθοδος είναι στατιστικά αποδεκτή και δεν απορρίπτεται στις περισσότερες περιπτώσεις. Αυτό επιβεβαιώνει ότι η μέθοδος POT είναι πιο κατάλληλη για δεδομένα με βαριές ουρές, όπως αυτά του Bitcoin.

Το **παραμετρικό μοντέλο Normal** αποδείχθηκε ανακριβές, καθώς οδηγεί σε **υποεκτίμηση του κινδύνου** στις πολύ υψηλές πιθανότητες ( $p = 0.995$ ,  $0.99$ ) με πολύ χαμηλά p-values, και σε **υπερεκτίμηση** στις χαμηλότερες πιθανότητες (π.χ.  $p = 0.8$  και  $0.9$ ). Το αποτέλεσμα αυτό είναι αναμενόμενο, καθώς οι

αποδόσεις του Bitcoin έχουν σαφώς βαρύτερες ουρές από την κανονική κατανομή, κάτι που καθιστά τη συγκεκριμένη υπόθεση ακατάλληλη.

Η **εμπειρική (empiric) μέθοδος** έχει σε αρκετές περιπτώσεις καλύτερη απόδοση σε σχέση με το κανονικό μοντέλο και καταγράφει ποσοστά επιτυχιών που βρίσκονται κοντά στις θεωρητικές τιμές. Ωστόσο, παραμένει λιγότερο σταθερή από την POT, καθώς εξαρτάται υπερβολικά από ακραίες παρατηρήσεις σε κάθε χρονικό παράθυρο. Στις ενδιάμεσες πιθανότητες (0.8, 0.9) η συμπεριφορά της είναι ικανοποιητική, αλλά όχι πλήρως αξιόπιστη.

**Συνολικά**, η μέθοδος **POT** αναδείχθηκε ως η πιο συνεπής και ακριβής στην εκτίμηση τόσο του VaR όσο και του ES για το Bitcoin. Η καταλληλότητά της σε δεδομένα με έντονες διακυμάνσεις και βαριές ουρές (όπως αυτά των κρυπτονομισμάτων) επιβεβαιώνεται, καθιστώντας την προτιμητέα για την εκτίμηση του κινδύνου σε σχέση με τα εναλλακτικά μοντέλα.

### 3.5. Συμπεράσματα

#### **POT (Θεωρία ακραίων τιμών):**

Αναδείχθηκε η **σταθερότερη και πιο αξιόπιστη μέθοδος** και για τις τέσσερις περιπτώσεις.

Δουλεύει εξαιρετικά για δεδομένα με βαριές ουρές (Netflix, Amazon, Bitcoin) αλλά και για «ηπιότερες» μετοχές όπως η Apple.

Είναι η μόνη μέθοδος που αντέχει σε **ακραίες πιθανότητες εμπιστοσύνης** ( $p = 0.995, 0.99$ ).

#### **Κανονική Κατανομή (Παραμετρική):**

Απέτυχε συστηματικά, ιδίως για Netflix, Amazon και Bitcoin.

Υποεκτιμά σε ακραίες περιπτώσεις και υπερεκτιμά σε χαμηλότερες.

Μπορεί να δώσει μια πρώτη προσέγγιση, αλλά δεν είναι αξιόπιστη.

#### **Εμπειρική/Απαραμετρική (Empiric):**

Δίνει καλύτερα αποτελέσματα από την κανονική προσέγγιση στις μεσαίες πιθανότητες (0.8–0.9).

Ωστόσο, **ασταθής σε ακραία ποσοστημόρια** (0.995, 0.99), άρα δεν είναι κατάλληλη για διαχείριση κινδύνου σε υψηλό επίπεδο εμπιστοσύνης.

## Τελικό Συμπέρασμα

Για όλες τις παραπάνω περιπτώσεις (**Netflix, Apple, Amazon, Bitcoin**), η **POT** αναδείχθηκε η πιο αξιόπιστη μέθοδος για εκτίμηση Value-at-Risk και Expected Shortfall.

Η **κανονική** κατανομή δεν είναι κατάλληλη, ειδικά σε δεδομένα με βαριές ουρές (όπως οι μετοχές τεχνολογίας και το Bitcoin).

Η **εμπειρική/απαραμετρική** μπορεί να χρησιμοποιηθεί συμπληρωματικά σε μεσαίες πιθανότητες, αλλά δεν επαρκεί για ακραίους κινδύνους.

Συνεπώς, ως τελικό συμπέρασμα, και με βάση τις περιπτώσεις, η POT φαίνεται να είναι η ενδεδειγμένη μέθοδος και για μετοχές και για κρυπτονομίσματα, επιβεβαιώνοντας τη θεωρητική της υπεροχή σε δεδομένα με υψηλή μεταβλητότητα και βαριές ουρές.

## Βιβλιογραφία

- Beirlant, J., Goegebeur, Y., Segers, J., & Teugels, J. (2004). *Statistics of Extremes: Theory and Applications*. Wiley.
- Boutsikas, M. (2020). *Διαχείριση κινδύνων. Πανεπιστημιακές Σημειώσεις*. Πανεπιστήμιο Πειραιά.
- Danielsson, J. (2011). *Financial Risk Forecasting: The Theory and Practice of Forecasting Market Risk with Implementation in R and Matlab*. Wiley.
- Dowd, K. (2002). *Measuring Market Risk*. Wiley.
- Georgescu, V. (2009). *Measuring Risk with Extreme Value Theory*. Master's Thesis, Πανεπιστήμιο Humboldt του Βερολίνου.
- Hull, J. C. (2015). *Risk Management and Financial Institutions*. Wiley.
- Reiss, R.-D., & Thomas, M. (2007). *Statistical Analysis of Extreme Values: with Applications to Insurance, Finance, Hydrology*. Birkhäuser
- Rocco, M. (2013). Extreme Value Theory for Finance: A Survey. *Journal of Economic Surveys*, 28, pp. 82-108.
- Roncalli, T. (2020). *Handbook of Financial Risk Management*. Chapman and Hall/CRC



