

**ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΕΙΡΑΙΩΣ**



**ΤΜΗΜΑ ΣΤΑΤΙΣΤΙΚΗΣ  
ΚΑΙ ΑΣΦΑΛΙΣΤΙΚΗΣ ΕΠΙΣΤΗΜΗΣ**

**ΜΕΤΑΠΤΥΧΙΑΚΟ ΠΡΟΓΡΑΜΜΑ ΣΠΟΥΔΩΝ  
ΣΤΗΝ ΕΦΑΡΜΟΣΜΕΝΗ ΣΤΑΤΙΣΤΙΚΗ**

**ΜΟΝΤΕΛΑ ΕΚΤΙΜΗΣΗΣ ΤΟΥ ΔΕΙΚΤΗ  
ΑΚΡΑΙΩΝ ΤΙΜΩΝ ΚΑΙ  
ΑΞΙΟΛΟΓΗΣΗ ΤΟΥΣ**

**ΓΑΡΑΝΤΖΙΩΤΗΣ ΓΕΩΡΓΙΟΣ**

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 2017



**ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΕΙΡΑΙΩΣ**



**ΤΜΗΜΑ ΣΤΑΤΙΣΤΙΚΗΣ  
ΚΑΙ ΑΣΦΑΛΙΣΤΙΚΗΣ ΕΠΙΣΤΗΜΗΣ**

**ΜΕΤΑΠΤΥΧΙΑΚΟ ΠΡΟΓΡΑΜΜΑ ΣΠΟΥΔΩΝ  
ΣΤΗΝ ΕΦΑΡΜΟΣΜΕΝΗ ΣΤΑΤΙΣΤΙΚΗ**

**ΜΟΝΤΕΛΑ ΕΚΤΙΜΗΣΗΣ ΤΟΥ ΔΕΙΚΤΗ  
ΑΚΡΑΙΩΝ ΤΙΜΩΝ ΚΑΙ  
ΑΞΙΟΛΟΓΗΣΗ ΤΟΥΣ**

**ΓΑΡΑΝΤΖΙΩΤΗΣ ΓΕΩΡΓΙΟΣ**

*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 2017

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' αριθμόν ..... συνεδρίασή του, σύμφωνα με τον Εσωτερικό Κανονισμό Λειτουργίας του Προγράμματος Μεταπτυχιακών Σπουδών στην Εφαρμοσμένη Στατιστική.

Τα μέλη της Επιτροπής ήταν:

- Επίκουρος ΚΑΘΗΓΗΤΗΣ ΜΠΟΥΤΣΙΚΑΣ ΜΙΧΑΗΛ (επιβλέπων)
- Επίκουρος ΚΑΘΗΓΗΤΗΣ ΤΖΑΒΕΛΛΑΣ ΓΕΩΡΓΙΟΣ
- Επίκουρος ΚΑΘΗΓΗΤΗΣ ΨΑΡΡΑΚΟΣ ΓΕΩΡΓΙΟΣ

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**



**DEPARTMENT OF STATISTICS  
AND INSURANCE SCIENCE**

**POSTGRADUATE PROGRAM IN  
APPLIED STATISTICS**

**A review of methods for tail index estimation**

By

**GARANTZIOTIS GEORGIOS**

MSc Dissertation

Submitted to the Department of Statistics and Insurance Science of the University of Piraeus as part of the requirements for obtaining the MSc in Applied Statistics.

Piraeus  
October 2017

# Περίληψη

Η θεωρία ακραίων τιμών αποτελεί το πιθανοθεωρητικό πλαίσιο για την μελέτη μοντέλων, στα οποία εμφανίζονται τιμές που θεωρούνται ακραίες (π.χ. πολύ μεγάλες αποζημιώσεις, απότομες αυξομειώσεις τιμών χρηματιστηριακών προϊόντων, ακραίες τιμές περιβαλλοντολογικών δεικτών, ακραία γεωλογικά ή καιρικά φαινόμενα).

Μια από τις κύριες επιδιώξεις της θεωρίας αυτής είναι ο προσδιορισμός της μορφής της δεξιάς (ή αριστερής) ουράς της κατανομής των υπό μελέτη παρατηρήσεων, ώστε να είναι δυνατή η μακροπρόθεσμη πρόβλεψη εμφάνισης ακραίων παρατηρήσεων. Η μορφή της δεξιάς ουράς της κατανομής χαρακτηρίζεται από την παράμετρο  $\xi$  (tail index), η τιμή της οποίας προσδιορίζει μια περισσότερο ή λιγότερο βαριά ουρά. Σκοπός της συγκεκριμένης εργασίας είναι η επισκόπηση των βασικών μεθόδων εκτίμησης της συγκεκριμένης παραμέτρου (εκτιμητές Hill, Negative Hill, Pickand, Ροπών, Λόγου Ροπών, Peng, W, PWM και Μεγίστης Πιθανοφάνειας) και η σύγκριση της αποτελεσματικότητας των μεθόδων αυτών, είτε αναλυτικά είτε κυρίως μέσω Monte Carlo προσομοίωσης, με χρήση διάφορων κατανομών (για τον σκοπό αυτό θα χρησιμοποιηθεί το λογισμικό R).



# Abstract

The extreme value theory consists of the probabilistic theoretical framework for the study of models in which sample values are considered to be extreme (e.g. very large claims, large stock price fluctuations, extreme values of environmental measurements, extreme geological or weather phenomena).

One of the main aims of this theory is the determination of the shape of the right (or left) tail of the data distribution, so that long-term prediction of extreme observations can be efficiently achieved. The shape of the right tail of a distribution is characterized by the tail index  $\xi$ , whose value identifies a more or less heavy tail. The purpose of this dissertation is to present a review of well known methods for estimating this parameter. In particular we are interested in Hill's Estimator, Negative Hill's Estimator, Pickand's Estimator, MLE Estimator, Moments Estimator, Moments Ratio Estimator, Peng Estimator, W Estimator and PWM Estimator. We compare the effectiveness of these estimators either analytically or, mainly, through Monte Carlo simulation (for this purpose R software will be used).





# Ευχαριστίες

Θα ήθελα να ευχαριστήσω τους γονείς μου, Ιωάννη και Παναγιώτα, την θεία μου Ελένη τον θείο μου Νίκο, τις γιαγιάδες μου, Ελένη και Σταματία καθώς και τη θεία μου Θεοδώρα για όλη την αμέριστη υποστήριξη και αγάπη που μου έδειξαν όλα αυτά τα χρόνια, καθώς και τους παππούδες μου Γιώργο και Σωτήρη που δεν είναι πια μαζί μας.

Επίσης θα ήθελα να ευχαριστήσω την Μαρία μου, τον Γιάννη τον Αντώνη και τον Ιάκωβο που στάθηκαν δίπλα μου και χωρίς την βοήθειά τους ο κώδικας ακόμα θα έτρεχε. Τέλος ένα μεγάλο ευχαριστώ στον καθηγητή μου κ. Μπούτσικα, για την καθοδήγηση τις συμβουλές, τις προτάσεις και τις διορθώσεις του στην παρούσα διπλωματική εργασία.



# ΠΕΡΙΕΧΟΜΕΝΑ

---

Περίληψη.....	i
Abstract .....	iii
Ευχαριστίες .....	v
Περιεχόμενα .....	vii
Κατάλογος σχημάτων.....	ix
Κατάλογος Πινάκων.....	x
Κατάλογος Συντομογραφιών .....	xi
1 Εισαγωγή.....	1
2 Θεωρία Ακραίων Τιμών.....	3
2.1 Βάσεις της Θεωρίας Ακραίων Τιμών (Extreme Value Theory).....	4
2.1.1 Μέθοδος Block Maxima .....	4
2.1.2 Μέθοδος Peak Over Threshold (POT).....	8
3 Εκτίμηση Δείκτη Ακραίων Τιμών .....	13
3.1 Ορισμός Παραμέτρου .....	13
3.2 Διατεταγμενες παρατηρησεις (Order Statistics).....	13
3.3 Εκτιμητής Pickand.....	14
3.4 Εκτιμητής Hill .....	15
3.4.1 Κλασσικός Εκτιμητής Hill .....	15
3.4.2 Εκτιμητής Negative Hill.....	16
3.4.3 Εκτιμητής Shifted Hill .....	17
3.5 Εκτίμηση με Χρήση Ροπών.....	17
3.5.1 Εκτιμητής Ροπών.....	17
3.5.2 Εκτιμητής Λόγου Ροπών .....	18
3.5.3 Εκτιμητές Peng και W .....	18
3.6 Εκτιμητής Ροπών Σταθμισμένης Πιθανότητας .....	19
3.7 Εκτιμητής Μεγίστης Πιθανοφάνειας.....	19
4 Αποτελέσματα Προσομοίωσης .....	21
4.1 Μεθοδολογία .....	21
4.2 Εκθετική Κατανομή.....	24
4.3 Κανονική Κατανομή.....	28
4.4 Κατανομή Γάμμα.....	31
4.5 Κατανομή Weibull.....	36

4.6	Κατανομή Βήτα .....	43
4.7	Ομοιόμορφη Κατανομή .....	50
4.8	Κατανομή Burr .....	53
4.9	Κατανομή Fréchet .....	64
5	Συμπεράσματα .....	77
6	Βιβλιογραφία .....	80
7	Παράρτημα Ι .....	83
8	Παράρτημα ΙΙ .....	141

## ΚΑΤΑΛΟΓΟΣ ΣΧΗΜΑΤΩΝ

---

Εικόνα 4.2.1 Διαγράμματα Καλής Προσαρμογής Στην Gen Για Την Εκθετική ( $\Lambda = 1$ ) .....	25
Εικόνα 4.2.2 Διαγράμματα Για Την Επιλογή Του Βελτιστού U Στην Περίπτωση Της Εκθετικής ( $\Lambda=1$ ) .....	25
Εικόνα 4.2.3 Διαγράμματα Καλής Προσαρμογής Στην Gpd Για Την Εκθετική ( $\Lambda = 1$ ) .....	26
Εικόνα 4.3.1 Διαγράμματα Καλής Προσαρμογής Στην Gen Για Την Κανονική (10, 1) .....	28
Εικόνα 4.3.2 Διαγράμματα Για Την Επιλογή Του Βελτιστού U Στην Περίπτωση Της Κανονικής (10, 1) .....	29
Εικόνα 4.3.3 Διαγράμματα Καλής Προσαρμογής Στην Gpd Για Την Κανονική (10, 1) .....	29
Εικόνα 4.4.1 Διαγράμματα Καλής Προσαρμογής Στην Gen Για Την Γάμμα (1, 0.5) .....	31
Εικόνα 4.4.2 Διαγράμματα Για Την Επιλογή Του Βελτιστού U Στην Περίπτωση Της Γάμμα (1, 0.5) .....	32
Εικόνα 4.4.3 Διαγράμματα Καλής Προσαρμογής Στην Gpd Για Την Γάμμα (1, 0.5) .....	32
Εικόνα 4.4.4 Διαγράμματα Καλής Προσαρμογής Στην Gen Για Την Γάμμα(1, 1.5) .....	34
Εικόνα 4.4.5 Διαγράμματα Για Την Επιλογή Του Βελτιστού U Στην Περίπτωση Της Γάμμα (1, 1.5) .....	35
Εικόνα 4.4.6 Διαγράμματα Καλής Προσαρμογής Στην Gpd Για Την Γάμμα (1, 1.5) .....	35
Εικόνα 4.5.1 Διαγράμματα Καλής Προσαρμογής Στην Gen Για Την Weibull (1, 0.5) .....	37
Εικόνα 4.5.2 Διαγράμματα Για Την Επιλογή Του Βελτιστού U Στην Περίπτωση Της Weibull (1, 0.5) .....	38
Εικόνα 4.5.3 Διαγράμματα Καλής Προσαρμογής Στην Gpd Για Την Weibull (1, 0.5) .....	38
Εικόνα 4.5.4 Διαγράμματα Καλής Προσαρμογής Στην Gen Για Την Weibull (1, 1.5) .....	40
Εικόνα 4.5.5 Διαγράμματα Για Την Επιλογή Του Βελτιστού U Στην Περίπτωση Της Weibull (1, 1.5) .....	41
Εικόνα 4.5.6 Διαγράμματα Καλής Προσαρμογής Στην Gpd Για Την Weibull (1, 1.5) .....	41
Εικόνα 4.6.1 Διαγράμματα Καλής Προσαρμογής Στην Gen Για Την Βήτα (0.5, 4) .....	43
Εικόνα 4.6.2 Διαγράμματα Για Την Επιλογή Του Βελτιστού U Στην Περίπτωση Της Βήτα (0.5, 4) .....	44
Εικόνα 4.6.3 Διαγράμματα Καλής Προσαρμογής Στην Gpd Για Την Βήτα (0.5, 4) .....	45
Εικόνα 4.6.4 Διαγράμματα Καλής Προσαρμογής Στην Gen Για Την Βήτα (5, 1) .....	47
Εικόνα 4.6.5 Διαγράμματα Για Την Επιλογή Του Βελτιστού U Στην Περίπτωση Της Βήτα (5, 1) .....	47
Εικόνα 4.6.6 Διαγράμματα Καλής Προσαρμογής Στην Gpd Για Την Βήτα (5, 1) .....	48
Εικόνα 4.7.1 Διαγράμματα Καλής Προσαρμογής Στην Gen Για Την Ομοιομορφή (0, 1) .....	50
Εικόνα 4.7.2 Διαγράμματα Για Την Επιλογή Του Βελτιστού U Στην Περίπτωση Της Ομοιομορφής (0, 1) .....	51
Εικόνα 4.7.3 Διαγράμματα Καλής Προσαρμογής Στην Gpd Για Την Ομοιομορφή (0, 1) .....	52
Εικόνα 4.8.1 Διαγράμματα Καλής Προσαρμογής Στην Gen Για Την Burr (4, 1, 1) .....	54
Εικόνα 4.8.2 Διαγράμματα Για Την Επιλογή Του Βελτιστού U Στην Περίπτωση Της Burr (4, 1, 1) .....	55
Εικόνα 4.8.3 Διαγράμματα Καλής Προσαρμογής Στην Gpd Για Την Burr (4, 1, 1) .....	56
Εικόνα 4.8.4 Διαγράμματα Καλής Προσαρμογής Στην Gen Για Την Burr (1, 2, 1) .....	58
Εικόνα 4.8.5 Διαγράμματα Για Την Επιλογή Του Βελτιστού U Στην Περίπτωση Της Burr (1, 2, 1) .....	58
Εικόνα 4.8.6 Διαγράμματα Καλής Προσαρμογής Στην Gpd Για Την Burr (1, 2, 1) .....	59
Εικόνα 4.8.7 Διαγράμματα Καλής Προσαρμογής Στην Gen Για Την Burr (1, 1, 1) .....	61
Εικόνα 4.8.8 Διαγράμματα Για Την Επιλογή Του Βελτιστού U Στην Περίπτωση Της Burr (1, 1, 1) .....	62
Εικόνα 4.8.9 Διαγράμματα Καλής Προσαρμογής Στην Gpd Για Την Burr (1, 1, 1) .....	63
Εικόνα 4.9.1 Διαγράμματα Καλής Προσαρμογής Στην Gen Για Την Frechet (0, 1, 4) .....	65
Εικόνα 4.9.2 Διαγράμματα Για Την Επιλογή Του Βελτιστού U Στην Περίπτωση Της Frechet (0, 1, 4) .....	65
Εικόνα 4.9.3 Διαγράμματα Καλής Προσαρμογής Στην Gpd Για Την Frechet (0, 1, 4) .....	66
Εικόνα 4.9.4 Διαγράμματα Καλής Προσαρμογής Στην Gen Για Την Frechet (0, 1, 2) .....	68
Εικόνα 4.9.5 Διαγράμματα Για Την Επιλογή Του Βελτιστού U Στην Περίπτωση Της Frechet (0, 1, 2) .....	69
Εικόνα 4.9.6 Διαγράμματα Καλής Προσαρμογής Στην Gpd Για Την Frechet (0, 1, 2) .....	70
Εικόνα 4.9.7 Διαγράμματα Καλής Προσαρμογής Στην Gen Για Την Frechet (0, 2, 1) .....	72
Εικόνα 4.9.8 Διαγράμματα Για Την Επιλογή Του Βελτιστού U Στην Περίπτωση Της Frechet (0, 2, 1) .....	72
Εικόνα 4.9.9 Διαγράμματα Καλής Προσαρμογής Στην Gpd Για Την Frechet (0, 2, 1) .....	73

## ΚΑΤΑΛΟΓΟΣ ΠΙΝΑΚΩΝ

---

Πίνακας 4.1.1 Λίστα Των Κατανομών Με Τους Αντιστοιχούς Θεωρητικούς Δείκτες $\Xi$ .....	22
Πίνακας 4.1.2 Κατανομές Που Χρησιμοποιήθηκαν Και Η Αντιστοιχη Τιμή Του Δείκτη Ακρικών Τιμών Που Αναμενόμε.....	22
Πίνακας 4.1.3 Οι Εκτιμητές Που Εξετάσθηκαν Μεσω Προσομοίωσης Και Οι Κλειστοί Τύποι Που Τους Περιγράφουν .....	23
Πίνακας 4.2.1 Περιγραφικά Χαρακτηριστικά Της Κατανομής Των Προσομοιωμένων Εκτιμητών Για Την Εκθετική ( $\Lambda = 1$ ) .....	26
Πίνακας 4.3.1 Περιγραφικά Χαρακτηριστικά Της Κατανομής Των Προσομοιωμένων Εκτιμητών Για Την Κανονική (10, 1) .....	30
Πίνακας 4.4.1 Περιγραφικά Χαρακτηριστικά Της Κατανομής Των Προσομοιωμένων Εκτιμητών Για Την Γάμμα (1, 0,5) .....	33
Πίνακας 4.4.2 Περιγραφικά Χαρακτηριστικά Της Κατανομής Των Προσομοιωμένων Εκτιμητών Για Την Γάμμα (1, 1,5) .....	36
Πίνακας 4.5.1 Περιγραφικά Χαρακτηριστικά Της Κατανομής Των Προσομοιωμένων Εκτιμητών Για Την Weibull (1, 0,5) .....	39
Πίνακας 4.5.2 Περιγραφικά Χαρακτηριστικά Της Κατανομής Των Προσομοιωμένων Εκτιμητών Για Την Weibull (1, 1,5) .....	42
Πίνακας 4.6.1 Περιγραφικά Χαρακτηριστικά Της Κατανομής Των Προσομοιωμένων Εκτιμητών Για Την Βήτα(0,5 , 4).....	45
Πίνακας 4.6.2 Περιγραφικά Χαρακτηριστικά Της Κατανομής Των Προσομοιωμένων Εκτιμητών Για Την Βήτα (5, 1).....	49
Πίνακας 4.7.1 Περιγραφικά Χαρακτηριστικά Της Κατανομής Των Προσομοιωμένων Εκτιμητών Για Την Ομοιομορφή (0, 1).....	52
Πίνακας 4.8.1 Περιγραφικά Χαρακτηριστικά Της Κατανομής Των Προσομοιωμένων Εκτιμητών Για Την Burr (4, 1, 1).....	56
Πίνακας 4.8.2 Περιγραφικά Χαρακτηριστικά Της Κατανομής Των Προσομοιωμένων Εκτιμητών Για Την Burr (1, 2, 1).....	59
Πίνακας 4.8.3 Περιγραφικά Χαρακτηριστικά Της Κατανομής Των Προσομοιωμένων Εκτιμητών Για Την Burr (1, 1, 1).....	63
Πίνακας 4.9.2 Περιγραφικά Χαρακτηριστικά Της Κατανομής Των Προσομοιωμένων Εκτιμητών Για Την Frechet (0 , 1 ,4) .....	66
Πίνακας 4.9.3 Περιγραφικά Χαρακτηριστικά Της Κατανομής Των Προσομοιωμένων Εκτιμητών Για Την Frechet (0 , 1 , 2) .....	70
Πίνακας 4.9.4 Περιγραφικά Χαρακτηριστικά Της Κατανομής Των Προσομοιωμένων Εκτιμητών Για Την Frechet (0 , 2 , 1) .....	74

## ΚΑΤΑΛΟΓΟΣ ΣΥΝΤΟΜΟΓΡΑΦΙΩΝ

---

iid	Ανεξάρτητες και ισόνομες τυχαίες μεταβλητές (independent and identical distributed)
BM	Block Maxima
POT	Peak over Threshold
σ.κ.	Συνάρτηση κατανομής
σ.π.	Συνάρτηση πιθανότητας
σ.π.π.	Συνάρτηση πυκνότητας πιθανότητας
δ.ε.	Διάστημα Εμπιστοσύνης
MLE	Εκτιμητής Μέγιστης Πιθανοφάνειας (Maximum Likelihood Estimator)
PWM	Εκτιμητής Ροπών Σταθμισμένης Πιθανότητας (Probability Weighted Moment Estimator)
τ.μ.	Τυχαία μεταβλητή
GPD	Generalized Pareto Distribution
GEV	Generalized Extreme Value Distribution



# 1 ΕΙΣΑΓΩΓΗ

---

Η θεωρία των ακραίων τιμών αποτελεί ξεχωριστό κλάδο της στατιστικής με στόχο την ανάπτυξη θεωρητικών και στατιστικών μοντέλων για την μελέτη προβλημάτων που προκύπτουν από την εμφάνιση ακραίων παρατηρήσεων, είτε πολύ μεγάλων είτε πολύ μικρών. Δημιουργήθηκε με σκοπό να καλύψει την ανεπάρκεια της κλασσικής θεωρίας πιθανοτήτων να μετρήσει, μελετήσει και εκτιμήσει τον κίνδυνο που προκύπτει από την πραγματοποίηση ακραίων συμβάντων. Σύμφωνα με την θεωρία αυτή, οι ιδιότητες των κατανομών ακροτάτων, καθώς και η υπέρβαση του εκάστοτε προκαθορισμένου κατωφλιού, καθορίζονται από την δεξιά και αριστερή ουρά της κατανομής των ανεξάρτητων και ισόνομων (iid) τυχαίων μεταβλητών  $X_1, X_2, \dots, X_n$ .

Ιστορικά, η συστηματική ανάπτυξη της θεωρίας ξεκίνησε το 1922 από τον von Bortkiewicz που ασχολήθηκε με την κατανομή κύμανσης τυχαίων δειγμάτων προερχόμενα από την κανονική κατανομή, ενώ την επόμενη κιόλας χρονιά οι R. von Mises & E.L. Dodd ασχολήθηκαν με την εκτίμηση της αναμενόμενης τιμής και της διαμέσου της κατανομής του Bortkiewicz. Οι Fisher & Tippett (1928), με την από κοινού εργασία τους «*Limiting forms of the frequency distribution of the largest and smallest member of a sample*» αναπτύσσουν το πρώτο θεώρημα της θεωρίας ακραίων τιμών, στο οποίο διατύπωσαν τις τρεις οριακές κατανομές που δύναται να ακολουθεί η μέγιστη παρατήρηση ενός δείγματος iid (ανεξάρτητων και ισόνομα κατανεμημένων), επεκτείνοντας την έρευνα του M. Fréchet, (1927), ο οποίος είχε προσδιορίσει μία πιθανή οριακή κατανομή. Αργότερα, το 1936 ο R. von Mises παρουσίασε κάποιες απλές αλλά και χρήσιμες συνθήκες για την ασθενή σύγκλιση διατεταγμένων τυχαίων μεταβλητών για κάθε μία εκ των τριών οριακών κατανομών, με τον B. Gnedenko (1943) να δίνει μια αυστηρή περιγραφή των συνθηκών αυτών. Το έργο του B. Gnedenko αποτέλεσε αντικείμενο μελέτης επιστημόνων για πολλά χρόνια αργότερα, όπως των Mejlzer (1949), Juncosa (1949), Marcus & Pinsky (1969) και de Haan (1970). Οι L. de Haan, Balkema & J. Pickand's αναπτύσσουν το δεύτερο θεώρημα της θεωρίας ακραίων τιμών, σύμφωνα με το οποίο, η ασυμπτωτική ουρά της κατανομής των τυχαίων μεταβλητών μπορεί να είναι γνωστή ακόμη και στο ενδεχόμενο άγνοιας της πραγματικής κατανομής των παρατηρήσεων. Από τη δεκαετία του '80 και έπειτα, οι μέθοδοι της εφαρμογής της θεωρίας ακραίων τιμών έχουν γίνει πολύ πιο διαδεδομένες.

Οι επικρατέστερες προσεγγίσεις στην θεωρία ακραίων τιμών είναι δύο, οι οποίες εξαρτώνται από την διαδικασία καταγραφής των δεδομένων. Η πρώτη προσέγγιση είναι το πρώτο θεώρημα που αναφέρθηκε προηγουμένως, το οποίο βασίζεται στην μελέτη της μέγιστης παρατήρησης του δείγματος. Σύμφωνα με αυτή, τα δεδομένα διακρίνονται σε ομάδες εκ των οποίων επιλέγεται η μεγαλύτερη παρατήρηση ανά ομάδα (Block Maxima Method) και εφαρμόζεται όταν είναι διαθέσιμα τα πλήρη δεδομένα του υπό μελέτη φαινομένου ή τουλάχιστον οι μέγιστες παρατηρήσεις ανά χρονικές περιόδους. Η δεύτερη μέθοδος στην οποία έγινε αναφορά, έχει ως στόχο την μελέτη της ουράς της κατανομής των παρατηρήσεων βάσει των υπερβάσεων άνω (κάτω) από κάποιο υψηλό (χαμηλό)

κατώφλι  $c_p$  ή  $u$  (Peaks Over Threshold Method) και εφαρμόζεται κυρίως όταν είναι διαθέσιμες παρατηρήσεις άνω (κάτω) από κάποια τιμή.

Μέχρι το ξεκίνημα της δεκαετίας του '50, η θεωρία ακραίων τιμών βρήκε μεγάλη άνθιση σε θεωρητικό επίπεδο, ενώ, από τα τέλη της δεκαετίας του '30 και έπειτα εφαρμόστηκε κατά κόρον σε διάφορους τομείς (Kotz & Nadarajah, 2000). Οι εκπομπές ραδιενέργειας (E.J. Gumbel (1937)), η αντοχή υλικών (Weibull (1939), Frenkel & Kontorova (1943)), η ανάλυση πλημμύρων (E.J. Gumbel (1941), Rantz & Riggs (1949)), η σεισμολογία (Nordquist (1945)), οι βροχοπτώσεις (Potter (1949)), η μηχανική (E.J. Gumbel (1958), Castillo (1988)), τα χρηματοοικονομικά (Koedijk et al (1990), Loretan & Phillips (1994), Diebold (1998)), οι ασφαλίσεις (Beirlant, Teugels & Vynekier (1996)) κ.α. είναι μόνο λίγοι από τους τομείς που στηρίζονται στην θεωρία αυτή για την εξέλιξη τους. Από τα μέσα του 20<sup>ου</sup> αιώνα μέχρι και σήμερα έχουν δημοσιευθεί χιλιάδες εργασίες αναφορικά με την θεωρία ακραίων τιμών και των εφαρμογών της, γεγονός που αποδεικνύει την χρησιμότητα και αναγκαιότητα της τόσο στον επιστημονικό κλάδο όσο και στην καθημερινότητα.

## 2 ΘΕΩΡΙΑ ΑΚΡΑΙΩΝ ΤΙΜΩΝ

---

Μια πρώτη προσέγγιση, χρησιμοποιώντας την κλασσική θεωρία πιθανοτήτων, είναι να βρεθεί το σημείο κατανομής, το οποίο καμία παρατήρηση δεν θα υπερβαίνει, με πιθανότητα  $1 - p$ . Το σημείο αυτό ονομάζεται κατώφλι και συμβολίζεται με  $c_p$ .

Έστω,  $X_1, X_2, \dots, X_n \sim F_\theta$  με  $X_1, X_2, \dots, X_n$  iid,  $n \geq 2$  και  $\theta$  οι άγνωστες παράμετροι από τα δεδομένα.

$$\begin{aligned} P(\max\{X_1, \dots, X_n\} < c_p) &= 1 - p \\ \Leftrightarrow P(X_1 < c_p, \dots, X_n < c_p) &= 1 - p \\ \Leftrightarrow P(X_1 < c_p) \dots P(X_n < c_p) &= 1 - p \\ \Leftrightarrow F_\theta(c_p)^n &= 1 - p \\ \Leftrightarrow c_p &= F_\theta^{-1}\left((1 - p)^{\frac{1}{n}}\right) \end{aligned}$$

Οπότε, η εκτίμηση του κατωφλιού βρίσκεται από την σχέση,

$$\hat{c}_p = F_\theta^{-1}\left((1 - p)^{\frac{1}{n}}\right)$$

όπου,  $\hat{\theta}$ : οι εκτιμήσεις των παραμέτρων, υποθέτοντας ότι η  $F$  αντιστρέφεται

Η προσέγγιση του ζητήματος, όπως αναπτύχθηκε, παρουσιάζει τα εξής μειονεκτήματα:

- Η επιλογή της  $F_\theta$  είναι αυθαίρετη, επομένως δεν είναι σίγουρο ότι η συμπεριφορά της δεξιάς ουράς των δεδομένων ταυτίζεται με αυτήν της πραγματικής ουράς. Ωστόσο, ακόμη και να έχει επιλεγεί σωστά η  $F_\theta$ , υπάρχει ο κίνδυνος ελάχιστες αποκλίσεις των εκτιμήσεων  $\hat{\theta}$  από τις πραγματικές τιμές  $\theta$ , να παρουσιάσουν πολύ μεγάλες αποκλίσεις στα αποτελέσματα για τη μορφή των άκρων της κατανομής  $F_\theta$ .
- Οι εκτιμήσεις των παραμέτρων  $\hat{\theta}$  εξάγονται, κυρίως, από το σύνολο των παρατηρήσεων και όχι από τις ακραίες παρατηρήσεις, οι οποίες αποτελούν την δεξιά ουρά της πραγματικής κατανομής των δεδομένων.
- Δυσκολία εκτίμησης της  $F_\theta$  λόγω περικομμένων δεδομένων.

Όλες οι αδυναμίες που αναφέρθηκαν, μπορούν να παρακαμφθούν με αξιοποίηση της θεωρίας ακραίων τιμών, καθώς δίνεται η δυνατότητα εστίασης της ανάλυσης στο ζητούμενο του προβλήματος, που είναι η συμπεριφορά του μοντέλου σε εξαιρετικές περιπτώσεις και όχι στο σύνολο της κατανομής.

## 2.1 ΒΑΣΕΙΣ ΤΗΣ ΘΕΩΡΙΑΣ ΑΚΡΑΙΩΝ ΤΙΜΩΝ (EXTREME VALUE THEORY)

Το βασικό στοιχείο της θεωρίας ακραίων τιμών είναι η περιγραφή μιας κατανομής, η οποία θα μελετά την πιθανότητα εμφάνισης μιας τιμής πιο ακραίας από αυτές που παρατηρούνται κατά το σύννηθες. Βάσει αυτής της ιδέας, υπάρχουν δύο κατανομές, μία κατανομή «γονέας» (parent distribution), η οποία περιγράφει κυρίως μια κεντρική τάση, και μια κατανομή γόνου (child distribution), η οποία εστιάζει στις παρατηρήσεις της ουράς της κατανομής γονέα.

Στη συνέχεια πρέπει αντιμετωπίζουμε δύο ερωτήματα. Πρώτον, ποιες παρατηρήσεις μπορούν να θεωρηθούν ακραίες και με ποιόν τρόπο μπορεί να χαρακτηριστούν ως τέτοιες και, δεύτερον, ποια είναι η κατανομή γονέας, από την οποία και εξαρτώνται οι παρατηρήσεις αυτές.

Μέχρι σήμερα, στην βιβλιογραφία υπάρχουν δύο μέθοδοι προσδιορισμού των ακραίων τιμών, η μέθοδος της μέγιστης παρατήρησης (Block Maxima Method) και η μέθοδος των παρατηρήσεων πέρα από ένα υψηλό κατώφλι (Peaks Over Threshold Method). Όσον αφορά την κατανομή γονέα, δεν είναι γνωστή ούτε και δύναται να εκτιμηθεί, επομένως γίνεται χρήση, είτε της οριακής κατανομής της μέγιστης παρατήρησης ανά περίοδο, αν υπάρχει, είτε της οριακής κατανομής των παρατηρήσεων αυτών που βρίσκονται πέρα από ένα υψηλό κατώφλι, αν υπάρχει.

Ωστόσο, η χρήση της θεωρίας αυτής έχει τα εξής μειονεκτήματα:

- αναγκαστική εφαρμογή ασυμπτωτικών υποθέσεων λόγω έλλειψης επαρκούς όγκου δεδομένων,
- δημιουργία μοντέλων υπό ιδανικές συνθήκες, οι οποίες στην πράξη, μπορεί να μην είναι ούτε ακριβείς ούτε λογικές,
- ενδεχόμενη απώλεια πληροφορίας κατά την χρήση των μοντέλων, και
- πιθανή αδυναμία σύγκλισης.

### 2.1.1 Μέθοδος Block Maxima

Στην μέθοδο αυτή, τα διαθέσιμα δεδομένα χωρίζονται σε ισομεγέθη διαστήματα (blocks) με  $n$  παρατηρήσεις το καθένα. Σε κάθε ένα εξ αυτών επιλέγεται η μέγιστη τιμή  $M_n$ , η οποία αποτελεί αντικείμενο μελέτης και αναλύεται παρακάτω. Η επιλογή του μεγέθους των blocks είναι πολύ σημαντική για την ορθή μοντελοποίηση. Κατά την επιλογή τους θα πρέπει να διατηρείται ισορροπία μεταξύ μεροληψίας και διακύμανσης, καθώς, μικρά blocks οδηγούν σε σχηματισμό «φτωχού» μοντέλου λόγω απουσίας αμεροληψίας στις εκτιμήσεις, ενώ μεγάλα blocks οδηγούν σε υψηλή διακύμανση των εκτιμήσεων λόγω του μικρού αριθμού maxima που προκύπτουν. Πέραν αυτού, η επιλογή των blocks θα πρέπει να γίνει με τέτοιο τρόπο ώστε να εξασφαλίζεται ισονομία των  $M_n$  που προέρχονται από διαφορετικά blocks.

### 2.1.1.1 Διατύπωση Μοντέλου

Η μέθοδος αξιοποιεί τις μέγιστες παρατηρήσεις, όπως αυτές παρατηρούνται σε ένα καθορισμένο (χρονικό) διάστημα (block), από το οποίο λαμβάνουμε  $n$  παρατηρήσεις. Επομένως, το βασικό στοιχείο της μεθόδου είναι οι παρατηρήσεις  $M_n$ , οι οποίες ορίζονται ως ακολούθως:

$$M_n = \max\{X_1, X_2, \dots, X_n\}, n \geq 2$$

όπου,  $X_1, X_2, \dots, X_n$  iid με κοινή σ.κ.  $F$ .

Χρησιμοποιώντας την κλασική στατιστική, η ακριβής σ.κ. της  $M_n$  θα δινόταν από τον τύπο:

$$P(M_n \leq x) = P(\max\{X_1, X_2, \dots, X_n\} \leq x) = P(X_1 \leq x, \dots, X_n \leq x) = F(x)^n, x \in R$$

Δεδομένου, όμως, ότι η  $F$  δεν είναι γνωστή θα πρέπει να μελετηθεί η συμπεριφορά της  $M_n$  όταν  $n \rightarrow \infty$ . Εφόσον οι ακραίες παρατηρήσεις λαμβάνουν χώρα κοντά στο δεξί άκρο του στηρίγματος της κατανομής, διαισθητικά αναμένουμε, ότι η ασυμπτωτική συμπεριφορά της  $M_n$  σχετίζεται με τη συμπεριφορά της δεξιάς ουράς της  $F$  κοντά στο δεξί άκρο της. Οπότε, αν συμβολίσουμε με  $x_F$  το δεξί άκρο στηρίγματος της  $F$ , τέτοιο ώστε:

$$x_F = \sup\{x \in R: F(x) < 1\} \leq \infty$$

τότε,

- αν  $x_F < \infty$  τότε,  $P(M_n \leq x) = F(x)^n \rightarrow \begin{cases} 0 & , x < x_F \\ 1 & , x \geq x_F \end{cases}, x \in R$
- αν  $x_F = \infty$  τότε,  $P(M_n \leq x) = F(x)^n \rightarrow 0, x \in R$

Επειδή η ακολουθία των τ.μ.  $M_n$   $n = 1, 2, \dots$  είναι μη φθίνουσα, προκύπτει ότι

$$M_n \rightarrow x_F \text{ με πιθανότητα } 1.$$

Σε μία προσπάθεια να βρεθεί μια μη εκφυλισμένη οριακή κατανομή είναι απαραίτητη η γραμμική αναμόρφωση της μεταβλητής  $M_n$ . Αντίστοιχα, με το Κεντρικό Οριακό Θεώρημα (Κ.Ο.Θ.), θα πρέπει να βρεθούν ακολουθίες πραγματικών αριθμών  $c_n > 0$ ,  $d_n$  τέτοιες ώστε

$$\frac{M_n - d_n}{c_n} \rightarrow_d G$$

όπου,  $G$  μια μη εκφυλισμένη οριακή κατανομή για  $n \rightarrow \infty$ .

Είναι πιθανό, όμως, να μην είναι πάντοτε δυνατή η κανονικοποίηση του δειγματικού μεγίστου, έτσι ώστε να ισχύει η παραπάνω σύγκλιση. Ωστόσο, όταν γίνει ορθή επιλογή των ακολουθιών δύναται να σταθεροποιηθεί η θέση και η διασπορά, παρακάμπτοντας έτσι τις δυσκολίες που διέπουν την  $M_n$ .

Έχουν αποδειχθεί τόσο οι ιδιότητες όσο και η μορφή των μη εκφυλισμένων κατανομών  $G$ , όταν υπάρχουν τα κατάλληλα  $c_n$  και  $d_n$ . Η βασική τους ιδιότητα είναι ότι τα κανονικοποιημένα μέγιστα ανεξάρτητων τ.μ. από την  $G$  θα ακολουθούν και πάλι την  $G$ , δηλαδή είναι max-stable, δημιουργώντας έτσι ένα σύνολο όλων των μη εκφυλισμένων κατανομών που μπορούν να προκύψουν ως όριο κάποιου κανονικοποιημένου μεγίστου (MD). Ενώ, όσον αφορά την μορφή που μπορούν να έχουν, αυτή περιορίζεται σε έναν μόλις από τρεις τύπους κατανομών, οι οποίοι προσδιορίζονται από το Θεώρημα Fisher-Tippett.

### 2.1.1.2 Θεώρημα Fisher-Tippett

Το θεώρημα των Fisher-Tippett είναι ένα γενικό αποτέλεσμα της θεωρίας των ακραίων τιμών, το οποίο αφορά στην ασυμπτωτική κατανομή των ακραίων διατεταγμένων παρατηρήσεων. Έστω  $X_1, X_2, \dots, X_n$  μια ακολουθία από iid τ.μ. Αν υπάρχουν ακολουθίες  $c_n > 0$  και  $d_n \in R$  και μια μη εκφυλισμένη κατανομή  $G$ , τέτοια, ώστε:

$$\frac{M_n - d_n}{c_n} = \frac{\max\{X_1, X_2, \dots, X_n\} - d_n}{c_n} \rightarrow_d G$$

τότε, η  $G$  θα είναι του ίδιου τύπου<sup>1</sup> με μία από τις τρεις ακόλουθες κατανομές

$$\text{Gumbel:} \quad \Lambda(x) = e^{-e^{-x}}, \quad x \in R \quad (\text{τύπου I})$$

$$\text{Fréchet:} \quad \Phi_\alpha(x) = \begin{cases} 0 & , x \leq 0 \\ e^{-x^{-\alpha}} & , x > 0 \end{cases}, \quad \alpha > 0 \quad (\text{τύπου II})$$

$$\text{(Reversed) Weibull:} \quad \Psi_\alpha(x) = \begin{cases} e^{-(-x)^\alpha} & , x \leq 0 \\ 1 & , x > 0 \end{cases}, \quad \alpha > 0 \quad (\text{τύπου III})$$

Αποδεικνύεται ότι οι παραπάνω είναι οι μόνες (μη-εκφυλισμένες) λύσεις της συναρτησιακής εξίσωσης που προκύπτει από τον ορισμό των max-stable κατανομών. Οι κατανομές  $\Phi_\alpha$ ,  $\Psi_\alpha$  και  $\Lambda$  καλούνται τυπικές κατανομές ακροτάτων, ενώ κατανομές που είναι του ίδιου τύπου με αυτές καλούνται κατανομές ακροτάτων. Αν  $X_i \sim F = 1 - \bar{F}$ , τότε προκύπτουν τα παρακάτω

- Αν  $\bar{F}(x) = c(x)e^{-\int_z^x \frac{g(t)}{a(t)} dt}$  ( $c(x) \rightarrow c_0, g(x) \rightarrow 1, a'(x) = 0$ , για  $x \uparrow x_F$ )

$G = \Lambda$  (Gumbel) (πχ Gamma, Normal, Lognormal, Exponential  $\in$  MDA( $\Lambda$ ))

- Αν  $x_F = \infty$  και  $\bar{F}(x) = x^{-\alpha}L(x)$  τότε

<sup>1</sup> **Κατανομές Ιδίου Τύπου:** Δύο τ.μ.  $X, Y$  είναι ίδιου τύπου αν  $Y =_d cX + d$  για κάποια  $c, d$ . Αν υπάρχουν ακολουθίες  $c_n, d_n$  και  $G$  έτσι ώστε  $(M_n - d_n)/c_n \rightarrow_d G$  τότε αποδεικνύεται για  $\tilde{c}_n = c_n/c, \tilde{d}_n = d_n - dc_n/c$  για  $d \in R, c \neq 0$ ,

$$\frac{M_n - \tilde{d}_n}{\tilde{c}_n} = \frac{M_n - d_n + dc_n/c}{c_n/c} = c \frac{M_n - d_n}{c_n} + d \rightarrow_d G'$$

και η  $G'$  είναι ίδιου τύπου με την  $G$ . Παρότι οι ακολουθίες αυτές δεν είναι μοναδικές, αποδεικνύεται πως πάντοτε προκύπτουν οριακές κατανομές ίδιου τύπου ( $G'(x) = G(\frac{x-d}{c})$ ).

$$G = \Phi_\alpha \text{ (Frêchet)} \quad (\text{πχ Pareto, Cauchy, Loggamma} \in \text{MDA}(\Phi_\alpha))$$

( $L$ : συνάρτηση βραδείας κύμανσης, δηλαδή  $L(tx)/L(x) \rightarrow 1, \forall t > 0$ )

- Αν  $x_F < \infty$  και  $\bar{F}(x_F - x^{-1}) = x^{-\alpha}L(x)$  τότε

$$G = \Psi_\alpha \text{ (Reversed Weibull)} \quad (\text{πχ Uniform, Beta} \in \text{MDA}(\Psi_\alpha))$$

$$(\text{ισοδύναμα, } \bar{F}(y) = (x_F - y)^\alpha L\left(\frac{1}{x_F - y}\right))$$

Το παραπάνω θεώρημα έχει δύο μειονεκτήματα. Πρώτον, απαιτείται η εύρεση μιας τεχνικής για την επιλογή της κατάλληλης εκ των τριών κατανομής για τα εκάστοτε δεδομένα και δεύτερον, μια a priori αυθαίρετη επιλογή κατανομής μπορεί να οδηγήσει σε εσφαλμένα συμπεράσματα. Τα μειονεκτήματα αυτά, όμως, μπορούν να εξαλειφθούν μέσω της ενοποίησης των τριών τυπικών κατανομών ακροτάτων σε μία οικογένεια κατανομών, την GEV (Generalized Extreme Value Distribution).

### 2.1.1.3 Generalized Extreme Value Distribution

Έστω  $X_1, X_2, \dots$  μια ακολουθία από iid τ.μ. Αν υπάρχουν ακολουθίες  $c_n > 0, d_n \in R$  και μια εκφυλισμένη κατανομή  $G$  ώστε

$$\frac{M_n - d_n}{c_n} = \frac{\max\{X_1, X_2, \dots, X_n\} - d_n}{c_n} \rightarrow_d G$$

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,  $\mu, \xi \in R, \sigma > 0$ .

Το μοντέλο αυτό έχει τρεις παραμέτρους: την παράμετρο θέσης (location)  $\mu$ , την παράμετρο κλίμακας (scale)  $\sigma$  και την παράμετρο σχήματος (shape)  $\xi$ . Έτσι, για κατάλληλες τιμές των παραμέτρων  $\mu, \sigma$  και  $\xi$  οι τρεις τύποι κατανομών του θεωρήματος Fisher-Tippett παίρνουν την μορφή

- αν  $\xi \rightarrow 0, c = \sigma$  και  $\mu = d$ , τότε

$$\text{I. } G(x) = e^{-\left(1 + \xi \frac{x - \mu}{\sigma}\right)^{-\frac{1}{\xi}}} \xrightarrow{\xi \rightarrow 0} e^{-e \frac{x - \mu}{\sigma}} = \Lambda\left(\frac{x - d}{c}\right), \quad x \in R$$

- αν  $\xi > 0$  με  $\xi = \alpha^{-1}, c = \sigma/\alpha$  και  $\mu = c + d$ , τότε

$$\text{II. } G(x) = e^{-\left(1 + \xi \frac{x - \mu}{\sigma}\right)^{-\frac{1}{\xi}}} = e^{-\left(\frac{x - d}{c}\right)^{-\alpha}} = \Phi_\alpha\left(\frac{x - d}{c}\right),$$

$$\frac{x - d}{c} > 0 \Leftrightarrow 1 + \xi \frac{x - \mu}{\sigma} > 0$$

- αν  $\xi < 0$  με  $\xi = -\alpha^{-1}, c = \sigma/\alpha$  και  $\mu = -c + d$ , τότε

$$\text{III. } G(x) = e^{-\left(1+\xi\frac{x-\mu}{\sigma}\right)^{\frac{1}{\xi}}} = e^{-\left(\frac{x-d}{c}\right)^a} = \Psi_\alpha\left(\frac{x-d}{c}\right),$$

$$\frac{x-d}{c} < 0 \Leftrightarrow 1 + \xi \frac{x-\mu}{\sigma} < 0$$

Η GEV επιλύει τα προβλήματα που αντιμετώπιζε το θεώρημα Fisher-Tippett. Προσαρμόζοντας τα δεδομένα στην GEV και εκτιμώντας ορθά τις παραμέτρους είναι δυνατό, πλέον, να προσδιοριστεί η οριακή κατανομή των μεγίστων. Τέλος, με χρήση κατάλληλων εργαλείων, στα οποία θα γίνει αναφορά σε επόμενο κεφάλαιο, αξιολογείται η καλή προσαρμογή των δεδομένων στην GEV, γεγονός που επιβεβαιώνει την ορθότητα της απόφασης για την επιλογή της εκάστοτε κατανομής.

### 2.1.2 Μέθοδος Peak Over Threshold (POT)

Στις περιπτώσεις όπου έχουν καταγραφεί όλες τις παρατηρήσεις από τα δεδομένα για να συμπεριληφθούν στην ανάλυση και όχι μόνο τα μέγιστα ανά περιόδους, η μέθοδος Block Maxima παύει να είναι η βέλτιστη για την αντιμετώπιση του προβλήματος. Είναι προτιμότερο να αποφεύγεται η διαδικασία διαχωρισμού των δεδομένων σε blocks, ενώ συνίσταται να επιλέγονται ως maxima οι παρατηρήσεις εκείνες, που υπερβαίνουν ένα προκαθορισμένο κατώφλι, η τιμή του οποίου θα προσδιορίζεται κάθε φορά ανάλογα με τις ανάγκες του ερευνητή. Με αυτόν τον τρόπο επιτυγχάνεται η χρήση παρατηρήσεων από τα δεδομένα, τα οποία είναι χαρακτηριστικά της ουράς της κατανομής, αλλά η μέθοδος Block Maxima θα αγνοούσε, καθώς βρίσκονται σε block, στο οποίο υπάρχει κάποια άλλη ακραία παρατήρηση, διατηρώντας έτσι περισσότερη πληροφορία για το σχήμα της ουράς της κατανομής.

#### 2.1.2.1 Διατύπωση Μοντέλου

Έστω  $X_1, X_2, \dots, X_n$  μια ακολουθία από *iid* τ.μ. με από κοινού σ.κ.  $F$ . Ως ακραίες παρατηρήσεις χαρακτηρίζονται αυτές που υπερβαίνουν ένα κατώφλι  $u$  με σ.κ. της υπέρβασης  $X_i - u$  που δίνεται από την σχέση

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

Αν η κατανομή γονέας  $F$  ήταν γνωστή, τότε και η κατανομή του κατωφλιού  $F_u$  θα ήταν, επίσης, γνωστή. Δεδομένου, όμως, πως δεν υπάρχει αυτή η πληροφορία, επιδιώκεται να βρεθούν προσεγγίσεις, οι οποίες θα είναι ευρέως εφαρμόσιμες για υψηλές τιμές του κατωφλιού. Αντίστοιχα με την χρήση της GEV στην μέθοδο Block Maxima ως προσέγγιση της κατανομής των μεγίστων για μεγάλες ακολουθίες, όταν ο πληθυσμός γονέας είναι άγνωστος. Η ανάλυση μπορεί να περιλαμβάνει την προσαρμογή δύο κατανομών, μία για τον αριθμό των γεγονότων σε μία προκαθορισμένη περίοδο και μία για το μέγεθος των υπερβαινόντων σημείων. Μια συχνή υπόθεση για την πρώτη περίπτωση είναι η κατανομή Poisson, με την Generalized Pareto για τα μεγέθη. Μια



εκτίμηση της ουράς της κατανομής μπορεί να γίνει βάσει των θεωρημάτων των Pickand's(1975) και Balkema–de Haan(1974).

### 2.1.2.2 Generalized Pareto Distribution (GPD)

Έστω  $X_1, X_2, \dots$  μια ακολουθία από *iid* τ.μ. με από κοινού σ.κ.  $F$  για την οποία ισχύει ότι η κατανομή του κανονικοποιημένου μεγίστου συγκλίνει στην κατανομή GEV. Για  $n \rightarrow \infty$  και αν

$$P\left(\frac{M_n - d_n}{c_n} \leq x\right) \rightarrow_d G \approx G(x; \mu, \sigma, \xi)$$

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

Τότε, για μεγάλο  $u$  η συνάρτηση πιθανότητας της υπέρβασης  $X - u$  δεδομένου ότι  $X > u$  είναι κατά προσέγγιση η

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

$$\text{για } y > 0 \text{ και } \left(1 + \frac{\xi y}{\tilde{\sigma}}\right) > 0$$

όπου,  $X \sim F$  και  $\tilde{\sigma} = \sigma + \xi(u - \mu)$ .

Η οικογένεια κατανομών που προκύπτει καλείται Generalized Pareto Distribution. Επομένως, οι υπερβάσεις ενός υψηλού κατωφλιού των  $X_i$  ακολουθούν προσεγγιστικά μια GPD με παραμέτρους  $\tilde{\sigma}, \xi$ . Οι παράμετροι της GPD για τις υπερβάσεις του κατωφλιού προσδιορίζονται μοναδικά από τις αντίστοιχες παραμέτρους της GEV για τα block maxima. Επιλέγοντας διαφορετικό αλλά εξίσου μεγάλο  $n$ , οι παράμετροι της GEV επηρεάζονται, εν αντιθέσει με αυτούς της GPD. Πιο συγκεκριμένα, το  $\xi$  παραμένει σταθερό καθώς είναι ανεξάρτητο από το μέγεθος  $n$ , ενώ ο υπολογισμός του  $\tilde{\sigma}$  επηρεάζεται μόνον από το επίπεδο του κατωφλιού και όχι από τις αλλαγές στα  $\mu$  και  $\sigma$ , δεδομένου ότι η επιρροή του ενός αντισταθμίζει την επιρροή του άλλου.

Η δυαδικότητα μεταξύ των δύο αυτών οικογενειών κατανομών επιφέρει την κυριαρχία της παραμέτρου σχήματος  $\xi$  στον προσδιορισμό της ποιοτικής συμπεριφοράς της GPD, όπως ακριβώς και στην περίπτωση της GEV. Για τις διάφορες τιμές της παραμέτρου  $\xi$  έχει αποδειχθεί ότι

- για  $\xi < 0$ , η GDP έχει άνω όριο στηρίγματος το  $u - \tilde{\sigma}/\xi$ ,
- για  $\xi > 0$ , η GDP εκτείνεται δεξιά ως το άπειρο, ενώ,
- για  $\xi \rightarrow 0$ , η GDP εκφυλίζεται στην εκθετική κατανομή με παράμετρο  $1/\tilde{\sigma}$ .

### 2.1.2.3 Επιλογή Βέλτιστου Κατωφλιού $u$

Η GPD εμφανίζεται ως η κατανομή των υπερβάσεων πάνω από το  $u$  όταν το κατώφλι  $u$  είναι αρκετά μεγάλο, όπως και το πλήθος των τυχαίων μεταβλητών. Στην μέθοδο των Block Maxima θα πρέπει να επιλεγεί προσεκτικά το μέγεθος  $n$  των blocks, ώστε να υπάρχει καλή προσαρμογή στην GEV, αλλά και να προκύπτουν εκτιμήσεις με μικρή μεταβλητότητα. Αντίστοιχα, στη μέθοδο POT θα πρέπει να επιλεγεί εξίσου προσεκτικά το ύψος του κατωφλιού  $u$ , καθώς για πολύ μεγάλο  $u$  οι υπερβάσεις θα είναι λίγες με αποτέλεσμα το μοντέλο να παρουσιάζει μεγάλη διακύμανση, ενώ, για μικρό  $u$  η οριακή συνάρτηση κατανομής των υπερβάσεων δεν θα προσεγγίζει ικανοποιητικά την GPD με τις εκτιμήσεις να μην είναι αμερόληπτες.

Συνήθης πρακτική είναι η επιλογή του χαμηλότερου δυνατού κατωφλιού, τέτοιο ώστε το οριακό μοντέλο να παρέχει μια λογική προσέγγιση. Στην βιβλιογραφία έχουν αναπτυχθεί δύο μέθοδοι για τον σκοπό αυτό, με την πρώτη να αποτελεί μία εξερευνητική τεχνική που εφαρμόζεται πριν από την εκτίμηση του μοντέλου και την δεύτερη να αποτελεί αξιολόγηση της σταθερότητας των εκτιμήσεων των παραμέτρων, βασισμένη στην προσαρμογή του μοντέλου σε διάφορα κατώφλια.

Η πρώτη μέθοδος βασίζεται στον μέσο της GPD. Αν υπάρχει ελάχιστο κατώφλι  $u_0$  τέτοιο ώστε  $F_{u_0} \approx GPD$  τότε  $F_u \approx GPD \forall u > u_0$ , έτσι

$$e(u) = E(X - u | X > u) = \frac{\sigma_u}{1 - \xi} = \frac{\sigma_{u_0} + \xi u}{1 - \xi}$$

Επομένως, για  $u > u_0$ , η  $e(u)$  που εκφράζεται ως γραμμική συνάρτηση του κατωφλιού  $u$ , είναι ο μέσος των υπερβάσεων των  $X_i$  άνω του  $u$  και καλείται μέση υπερβάλλουσα συνάρτηση. Η εκτίμηση της  $e(u)$  προκύπτει από την δειγματική μέση τιμή των υπερβάσεων πάνω από το  $u$ ,

$$\hat{e}(u) = \frac{1}{n_u} \sum_{i=1}^{n_u} (X_i - u)$$

όπου,  $n_u$ : το πλήθος των  $X_i$  που υπερβαίνουν το  $u$ .

Στην πράξη, λοιπόν, κατασκευάζοντας το γράφημα των  $\{(u, \hat{e}(u)) : u < x_{max}\}$  επιλέγεται ως  $u_0$  το σημείο όπου η GPD δίνει μια καλή προσέγγιση της υπερβάλλουσας κατανομής, δηλαδή το σημείο στο οποίο το παραπάνω γράφημα γίνεται γραμμικό.

Η δεύτερη μέθοδος για την επιλογή κατωφλιού είναι η προσαρμογή της GPD σε ένα εύρος κατωφλιών και η αναζήτηση σταθερότητας των εκτιμήσεων των παραμέτρων. Αν υπάρχει ελάχιστο κατώφλι  $u_0$  τέτοιο ώστε  $F_{u_0} \approx GPD$  τότε,  $F_u \approx GPD \forall u > u_0$ . Οι παράμετροι σχήματος  $\xi$  για τις δύο αυτές κατανομές είναι ίδιες, ενώ οι παράμετροι κλίμακας  $\sigma_u$  μεταβάλλονται με το  $u$  εκτός αν  $\xi = 0$ , όπως φαίνεται ακολούθως:

$$\sigma_u = \sigma_{u_0} + \xi(u - u_0)$$

Ωστόσο, αν μεταβληθεί η παράμετρος κλίμακας σε  $\sigma^* = \sigma_u - \xi u$ , παρατηρείται ότι πάνω από το  $u_0$  οι παράμετροι  $\sigma^*$  και  $\xi$  είναι σταθερές. Επομένως, σε ένα διάγραμμα των  $\hat{\sigma}^*$ ,  $\hat{\xi}$  με το  $u$ , επιλέγεται ως  $u_0$  η μικρότερη τιμή του  $u$  για την οποία οι εκτιμήσεις των  $\sigma$  και  $\xi$  θα είναι σταθερές.



## 3 ΕΚΤΙΜΗΣΗ ΔΕΙΚΤΗ ΑΚΡΑΙΩΝ ΤΙΜΩΝ

---

### 3.1 ΟΡΙΣΜΟΣ ΠΑΡΑΜΕΤΡΟΥ

Η παράμετρος σχήματος της κατανομής των ακροτάτων αποτελεί βασικό σημείο στην ανάλυση κατανομών ακραίων τιμών. Δε χρειάζεται να γνωρίζουμε ή να εκτιμήσουμε την αρχική κατανομή των δεδομένων, καθώς η παράμετρος καθορίζει τις ανώτερες ροπές της συνάρτησης πυκνότητας πιθανότητας, καθώς και τη λοξότητα της μάζας πιθανότητας από περικεκομμένα δεδομένα στην κατανομή.

Όπως αναφέρθηκε προηγουμένως, μία παράμετρος  $\xi$  ίση με το μηδέν είναι αποτέλεσμα κατανομών με λεπτή ουρά της οικογένειας κατανομών Gumbel, που περιλαμβάνει την κανονική, την εκθετική, τη γάμμα και την λογαριθμοκανονική κατανομή. Αντίστοιχα, οι κατανομές με θετικές τιμές του  $\xi$  ανήκουν στο MDA της Fréchet περιλαμβάνοντας κατανομές με βαριά ουρά, όπως η Pareto, η Cauchy και η κατανομή Student's t. Ενώ τέλος, στις περιπτώσεις, όπου η παράμετρος  $\xi$ , βρίσκεται να είναι αρνητική η οικογένεια κατανομών που χαρακτηρίζει τα δεδομένα είναι η αντίστροφη Weibull, η οποία αποτελείται από κατανομές με μικρή ουρά και πεπερασμένο στήριγμα.

Σε όλες τις κατανομές, στις οποίες ισχύει η ανισότητα  $\xi \neq 0$ , η ανισότητα που υπάρχει ως προϋπόθεση στον τύπο της κατανομής  $(1 + \xi(x - \mu)/\sigma > 0)$  επιβάλλει έναν περιορισμό στη μάζα πιθανότητας στην δεξιά και αριστερή ουρά των κατανομών. Όταν μια αρκετά μεγάλη μάζα πιθανότητας υπάρχει σε μία από τις ουρές, η οποία υποδεικνύει μια μη αμελητέα πιθανότητα εμφάνισης ενός ακραίου γεγονότος προς την εκάστοτε κατεύθυνση, υπάρχει ένα απόλυτο μέγιστο (ελάχιστο) στην αντίθετη κατεύθυνση, πέρα από το οποίο οι τιμές του  $x$  έχουν μηδενική μάζα πιθανότητας.

Η κύρτωση της κατανομής Fréchet τείνει στο άπειρο σε περιπτώσεις, όπου ισχύει  $\xi \geq 0,25$  και όλες οι ροπές ανώτερης τάξης, συμπεριλαμβανομένου και της κύρτωσης, απειρίζονται σε περιπτώσεις όπου  $\xi \geq 0,33$ . Ακόμα και για πολύ μικρές θετικές τιμές του  $\xi$ , ο ρυθμός αύξησης της λοξότητας και της κύρτωσης έχει ως αποτέλεσμα τη συγκέντρωση της πυκνότητας πιθανότητας στη δεξιά ουρά της κατανομής Fréchet. Επομένως, καθώς το  $\xi$  αυξάνεται έχοντας θετικό πρόσημο τα σημεία, όπου τα  $x$  έχουν μηδενική πιθανότητα στην αντίθετη ουρά γίνονται πιο αυστηρά. Άρα, αναμένουμε να παρατηρήσουμε περισσότερες και πιο μεγάλες κατά μέγεθος ακραίες παρατηρήσεις.

### 3.2 ΔΙΑΤΕΤΑΓΜΕΝΕΣ ΠΑΡΑΤΗΡΗΣΕΙΣ (ORDER STATISTICS)

Το αντικείμενο των διατεταγμένων παρατηρήσεων (order statistics) είναι η διερεύνηση των ιδιοτήτων και των εφαρμογών ενός συνόλου από τυχαίες μεταβλητές και των συναρτήσεών τους, όταν η διάταξή τους έχει νόημα για την επίλυση του προβλήματος. Επομένως, αν είναι διαθέσιμες τυχαίες μεταβλητές  $\{X_i\}, i = 1, 2, \dots, n$  διατεταγμένες σε αύξουσα (ή φθίνουσα κατά περιπτώσεις) σειρά κατά το μέγεθος των τιμών τους, τότε αυτές μπορούν να γραφτούν με τον εξής τρόπο:

$$X_{1:n} \leq X_{2:n} \leq \dots \leq X_{n:n}$$

Σε αυτήν την περίπτωση, η τυχαία μεταβλητή  $X_{i:n}$  είναι η  $i$ -οστή κατά σειρά μεγέθους παρατήρηση σε ένα δείγμα μεγέθους  $n$ . Λόγω της ανισότητας, που έχει τεθεί ανάμεσα στις τυχαίες μεταβλητές, αυτές είναι αναγκαία εξαρτημένες μεταξύ τους, παρά το γεγονός, ότι οι αρχικές παρατηρήσεις θεωρούνταν ανεξάρτητες μεταξύ τους. Η συνάρτηση κατανομής μίας εκ των τυχαίων μεταβλητών  $X_{i:n}$  μπορεί να δειχθεί (Balakrishnan & Clifford Cohen, 1991), ότι βρίσκεται από τον ακόλουθο τύπο:

$$\begin{aligned} F_{i:n}(x) &= \Pr(X_{i:n} \leq x) \\ &= \Pr(\text{τουλάχιστον } i \text{ από τις } X_r \text{ είναι } \leq x) \\ &= \sum_{r=i}^n \binom{n}{r} \{F(x)\}^r \{1 - F(x)\}^{n-r} \end{aligned}$$

η οποία είναι η συνάρτηση πιθανότητας της ουράς της διωνυμικής κατανομής ξεκινώντας από την  $i$ -οστή μεταβλητή. Από την ίδια σχέση μπορούμε να βρούμε, ότι για τη μέγιστη παρατήρηση ισχύει:

$$F_{n:n}(x) = \{F(x)\}^n$$

Ενώ, αντίστοιχα για την ελάχιστη μεταβλητή  $X_{1:n}$ :

$$F_{1:n}(x) = 1 - \{1 - F(x)\}^n$$

Οι πιο συχνές συναρτήσεις διατεταγμένων δεδομένων είναι τα άκρα  $X_{1:n}$  και  $X_{n:n}$ , τα οποία είναι τα γνωστά μέγιστα και ελάχιστα ενός δείγματος, ενώ χρησιμοποιείται και το εύρος  $W = X_{n:n} - X_{1:n}$ , καθώς και η μέγιστη απόσταση από τον μέσο  $X_{n:n} - \bar{X}$ .

### 3.3 ΕΚΤΙΜΗΤΗΣ PICKAND

Η βασική ιδέα πίσω από αυτόν τον εκτιμητή, έγκειται στην εύρεση μιας ισοδύναμης της  $F \in MDA(H_\xi)$  συνάρτησης, η οποία να απλοποιεί την παράμετρο  $\xi$ . Για να επιτευχθεί αυτό, υιοθετείται μια ποσοστιαία μέθοδος εκτίμησης για τις διαφορές μεταξύ των άνω διατεταγμένων δεδομένων. Έτσι, λοιπόν, για  $X_1, X_2, \dots, X_n$  iid ακολουθία και  $\xi \in \mathbb{R}$ , ο Pickand's (1975) διατύπωσε την απλούστερη και πιο παλαιά εκτίμηση του  $\xi$ :

$$\hat{\xi}_{(k:n)}^{Pickand} = \frac{1}{\ln 2} \ln \left( \frac{X_{n-k:n} - X_{n-2k:n}}{X_{n-2k:n} - X_{n-4k:n}} \right)$$

Ιδιότητες

- (Ασθενής Συνέπεια) Αν  $k \rightarrow \infty, k/n \rightarrow 0$  για  $n \rightarrow \infty$ , τότε

$$\hat{\xi}^{Pickand} \xrightarrow{P} \xi$$

- (Ισχυρή Συνέπεια) Αν  $k/n \rightarrow 0, k/\ln \ln n \rightarrow \infty$  για  $n \rightarrow \infty$ , τότε

$$\hat{\xi}^{Pickand} \xrightarrow{a.s.} \xi$$

- (Ασυμπτωτική Κανονικότητα) Υπό συνθήκες για  $k$  και  $F$  (Dekkers & De Haan, 1989)

$$\sqrt{k}(\hat{\xi} - \xi) \xrightarrow{d} N(0, v(\xi)), \quad n \rightarrow \infty,$$

$$\text{όπου, } v(\xi) = \frac{\xi^2(2^{2\xi+1}+1)}{(2(2^\xi-1)\ln 2)^2}$$

Στον υπολογισμό του εκτιμητή γίνεται χρήση μιας ακολουθίας order statistics, η οποία αυξάνεται με το  $n$ . Εδώ, παρατηρείται στην ανάλυση το λεγόμενο Pickand's-plot:

$$\{(k, \hat{\xi}_{k:n}^{Pickand}) : k = 1, \dots, n\}$$

μέσω του οποίου γίνεται προσπάθεια εύρεσης του βέλτιστου  $k$ . Ωστόσο, δεν υπάρχει βέλτιστη λύση και η βέλτιστη τιμή για τον  $\hat{\xi}_{k:n}^{Pickand}$  επιλέγεται από μια  $k$ -περιοχή, στην οποία το παραπάνω διάγραμμα είναι σχεδόν οριζόντιο.

Τέλος, ο εκτιμητής αυτός αποτελεί βάση για πολλούς άλλους εκτιμητές, οι οποίοι αποτελούν είτε βελτιώσεις του είτε παραλλαγές του. Μια σειρά τέτοιων εκτιμητών συναντάται στο σύγγραμμα των Dekkers & De Haan.

## 3.4 ΕΚΤΙΜΗΤΗΣ HILL

### 3.4.1 Κλασικός Εκτιμητής Hill

Ένας από τους πιο βασικούς εκτιμητές της παραμέτρου  $\xi$  προτάθηκε από τον Hill (1975). Έστω  $X_1, \dots, X_n$  iid με  $F \in D(G_\xi)$  για  $\xi > 0$  τότε η παράμετρος εκτιμάται από την παρακάτω σχέση:

$$\hat{\xi}_{(k:n)}^{Hill} = \frac{1}{k} \sum_{i=n-k}^n \ln(X_{i:n}) - \ln(X_{n-k:n})$$

Έχουν προταθεί πολλές τροποποιήσεις του παραπάνω τύπου, γεγονός που φανερώνει ότι είναι ένας πολύ φυσικός εκτιμητής. Η αρχική προσέγγιση εκφράστηκε με χρήση της μεθόδου μεγίστης πιθανοφάνειας για την παράμετρο  $\alpha$  της κατανομής  $F$  (με  $\xi = \alpha^{-1} > 0$ ). Αργότερα προτάθηκαν και άλλες δημοφιλείς προσεγγίσεις, όπως η mean excess function, στην οποία ο εκτιμητής μπορεί να ερμηνευτεί ως η μέση εμπειρική υπερβάλλουσα συνάρτηση του  $\ln(X)$  υπολογισμένη στο κατώφλι  $u = \ln X_{k:n}$ .

Χρησιμοποιώντας την αρχική προσέγγιση του Hill παρατηρούνται οι παρακάτω ιδιότητες (Embrechts, et al., 1997) του εκτιμητή με  $\xi = \alpha^{-1} > 0$ :

- (Ασθενής Συνέπεια) Έστω ότι μια από τις ακόλουθες συνθήκες ικανοποιείται
  - $(X_n)$  είναι iid
  - $(X_n)$  είναι ασθενώς εξαρτημένη

- $(X_n)$  είναι μια γραμμική διαδικασία  
Αν  $k \rightarrow \infty, k/n \rightarrow 0$  για  $n \rightarrow \infty$ , τότε

$$\hat{a}^{Hill} \xrightarrow{P} a$$

- (Ισχυρή Συνέπεια) Αν  $k/n \rightarrow 0, k/\ln \ln n \rightarrow \infty$  για  $n \rightarrow \infty$  και  $(X_n)$  iid ακολουθία, τότε

$$\hat{a}^{Hill} \xrightarrow{a.s.} a$$

- (Ασυμπτωτική Κανονικότητα) Υπό συνθήκες για  $k$  και  $F$  και  $(X_n)$  iid ακολουθία,

$$\sqrt{k}(\hat{a}^{Hill} - a) \xrightarrow{d} N(0, a^2), n \rightarrow \infty$$

Παρά την οικουμενική αποδοχή του εκτιμητή Hill, παρουσιάζει κάποια μειονεκτήματα. Ο εκτιμητής περιορίζεται στην περίπτωση της Fréchet (όπου  $\xi > 0$ ), παρουσιάζει πολύ μεγάλη ευαισθησία όταν τα δεδομένα είναι εξαρτημένα, ενώ τέλος, ο ρυθμός σύγκλισης του αργός.

Όπως και στον εκτιμητή του Pickand's, μια ανάλυση βασισμένη στον εκτιμητή του Hill μπορεί να παρασταθεί γραφικά. Το Hill-plot, το οποίο αναπαρίσταται από τα σημεία:

$$\{(k, \hat{a}_{k:n}^{Hill}): k = 2, \dots, n\}$$

είναι το εργαλείο για την εύρεση του βέλτιστου  $k$ . Εξομαλύνοντας τα Hill-plots σε ένα συγκεκριμένο εύρος των  $k$ -τιμών μπορεί να εξαλειφθεί το πρόβλημα επιλογής του  $k$ .

### 3.4.2 Εκτιμητής Negative Hill

Ο Negative Hill εκτιμητής πρόκειται για μία από τις πιο γνωστές τροποποιήσεις του Hill εκτιμητή, ο οποίος προτάθηκε από τον Falk (1995) και χρησιμοποιείται στην περίπτωση όπου  $\xi < -1/2$ . Έτσι, για  $X_1, \dots, X_n$  iid με  $F \in D(G_\xi)$  και  $X_{n,n}$  το μεγαλύτερο order statistic ισχύει

$$\hat{\xi}_{(k:n)}^{N Hill} = \frac{1}{k} \sum_{i=1}^{k-1} \ln(X_{n:n} - X_{n-i:n}) - \ln(X_{n:n} - X_{n-k:n})$$

Ο εκτιμητής, πλέον, παρουσιάζει τα εξής πλεονεκτήματα:

- (Ασθενής Συνέπεια) Αν  $F \in D(G_\xi)$  με  $\xi < -\frac{1}{2}$  και  $k \rightarrow \infty, k/n \rightarrow 0, k^n/\log n \rightarrow \infty$ , για  $n \rightarrow \infty$ , τότε

$$\hat{\xi}^{N Hill} \xrightarrow{P} \xi$$

- (Ασυμπτωτική Κανονικότητα) Αν  $-1 < \xi < -\frac{1}{2}$ ,  $k \rightarrow \infty, k/n \rightarrow 0, k^n/\log n \rightarrow \infty$  για  $n \rightarrow \infty$  τότε ο  $\hat{\xi}^{N Hill}$  ακολουθεί μια κανονική κατανομή με



διακύμανση  $\xi^2$  και μέσο

$$\lambda \xi \int_0^1 s^\xi \Psi_{\xi, \rho}(s^{-1}) ds = \begin{cases} \frac{\lambda \xi}{\rho(1+\xi)(1-\rho)}, & \rho < 0 \\ \lambda, & \rho = 0 \end{cases}$$

Τέλος, αξίζει να αναφερθεί πως ο εκτιμητής Negative Hill διατηρεί το πλεονέκτημα του βασικού τύπου και δεν μεταβάλλεται από τις παραμέτρους θέσης και κλίμακας.

### 3.4.3 Εκτιμητής Shifted Hill

Ο εκτιμητής του Hill, όπως αναφέρθηκε, δεν μεταβάλλεται από την παράμετρο κλίμακας, δηλαδή, αν εφαρμοσθεί ένας πολλαπλασιαστικός παράγοντας στα δεδομένα, η εκτίμηση δεν θα επηρεαστεί. Ωστόσο, συνεχίζει να επηρεάζεται από την παράμετρο θέσης, δηλαδή αν εφαρμοσθεί ένας αθροιστικός όρος στα δεδομένα, η εκτίμηση θα παραμορφωθεί. Έτσι οι Inmaculada B. Aban\* and Mark M. Meerschaert (2001) διατύπωσαν μια τροποποίηση του βασικού τύπου, τον εκτιμητή Shifted Hill. Η ιδέα πίσω από αυτόν τον εκτιμητή είναι η χρήση του εκτιμητή μεγίστης πιθανοφάνειας βασισμένο στα  $r + 1$  ( $0 \leq r \leq n$ ) μεγαλύτερα order statistic τα οποία χρησιμοποιούνται για την εκτίμηση των παραμέτρων μια μετατοπισμένης Pareto κατανομής με τύπο:

$$F(x) = 1 - C(x - s)^{-\xi}, x > s + C^{1/\xi}$$

Αν ισχύει  $X_{r,n} > X_{r+1,n}$ , τότε ο εκτιμητής μεγίστης πιθανοφάνειας δίνεται από τον τύπο:

$$\hat{\xi} = \left[ r^{-1} \sum_{i=1}^r \{ \ln(X_{i:n} - \hat{s}) - \ln(X_{r+1:n} - \hat{s}) \} \right]^{-1}$$

όπου,  $\hat{C} = \frac{r}{n} (X_{r+1:n} - \hat{s})^{\hat{\xi}}$  και η εκτίμηση της μετατόπισης  $\hat{s}$  ικανοποιεί την σχέση:

$$\hat{\xi} (X_{r+1:n} - \hat{s})^{-1} = (\hat{\xi} + 1) r^{-1} \sum_{i=1}^r (X_{i:n} - \hat{s})^{-1}$$

## 3.5 ΕΚΤΙΜΗΣΗ ΜΕ ΧΡΗΣΗ ΡΟΠΩΝ

### 3.5.1 Εκτιμητής Ροπών

Ο Εκτιμητής Ροπών (Moment Estimator) προτάθηκε από τους Dekkers-Einmahl-de Haan το 1989. Πρόκειται για μία γενίκευση του εκτιμητή Hill, στην οποία ο κλασικός εκτιμητής εκτείνεται ώστε να καλύψει όλη την τάξη  $H_\xi$ , με  $\xi \in \mathbb{R}$ . Αναδιαμορφώνοντας τις συνθήκες ομαλής διακύμανσης της  $MDA(H_\xi)$ , για  $X_1, X_2, \dots, X_n$  ακολουθία από iid τ.μ. με από κοινού συνάρτηση κατανομής  $F \in D(H_\xi)$ ,  $\xi \in \mathbb{R}$  κατέληξαν στον εκτιμητή (Dekkers, et al., 1989)

$$\hat{\xi}^M = H_n^{(1)} + 1 - \frac{1}{2} \left( 1 - \frac{(H_n^{(1)})^2}{H_n^{(2)}} \right)^{-1}$$

όπου,

$$H_n^{(1)} = \frac{1}{k} \sum_{j=0}^{k-1} (\ln X_{n-j:n} - \ln X_{n-k:n}) \text{ και } H_n^{(2)} = \frac{1}{k} \sum_{j=0}^{k-1} (\ln X_{n-j:n} - \ln X_{n-k:n})^2$$

Επειδή οι  $H_n^{(1)}, H_n^{(2)}$  μπορούν να χαρακτηριστούν ως εμπειρικές ροπές, ο  $\hat{\xi}^M$  ονομάζεται και Εκτιμητής των Ροπών. Η ασθενής και ισχυρή συνέπεια, καθώς και η ασυμπτωτική κανονικότητα του εκτιμητή αποδεικνύονται από τους Dekkers-Einmahl-de Haan.

### 3.5.2 Εκτιμητής Λόγου Ροπών

Ο Εκτιμητής Λόγου Ροπών (Moment Ratio Estimator) προτάθηκε από τους Danielsson, Jansen και deVries. Βασισμένοι σε μια ασυμπτωτική δεύτερης τάξης επέκταση της συνάρτησης  $F$ , κατάφεραν να αφαιρέσουν την μεροληψία που παρουσιάζει ο εκτιμητής Hill. Έτσι, για  $\xi > 0$  και  $F \in D(H_\xi)$ :

$$\hat{\xi}^{MR} = \frac{1}{2} \frac{H_n^{(2)}}{H_n^{(1)}}$$

Οι (Danielsson, et al., 1996) απέδειξαν πως ο  $\hat{\xi}^{MR}$  παρουσιάζει χαμηλότερο ασυμπτωτικό τετράγωνο μεροληψίας από τον εκτιμητή Hill, για δεδομένο επίπεδο κατωφλιού, παρότι ο ρυθμός σύγκλισης παραμένει ο ίδιος.

### 3.5.3 Εκτιμητές Peng και W

Ένας εκτιμητής που σχετίζεται με τον εκτιμητή των ροπών  $\hat{\xi}^M$ , είναι ο εκτιμητής Peng, που προτάθηκε από τους Deheuvels, de Haan, Peng και Pereira, σχεδιάστηκε ώστε να μειώσει την μεροληψία του εκτιμητή των ροπών και δίνεται από την σχέση:

$$\hat{\xi}^L = \frac{H_n^{(2)}}{2H_n^{(1)}} + 1 - \frac{1}{2} \left( 1 - \frac{(H_n^{(1)})^2}{H_n^{(2)}} \right)^{-1}.$$

Ένας ακόμα εκτιμητής που προτάθηκε από τους ίδιους ερευνητές είναι ο εκτιμητής W:

$$\hat{\xi}^W = 1 - \frac{1}{2} \left( 1 - \frac{(L_n^{(2)})^2}{L_n^{(1)}} \right), \text{ όπου } L_n^{(j)} = \frac{1}{k} \sum_{i=0}^{k-1} (X_{n-j:n} - X_{(n-k):n})^j, j = 1, 2$$

Όπως αναφέρουν οι (Deheuvels, et al., 1997), ο εκτιμητής  $\hat{\xi}^L$  είναι συνεπής  $\forall \xi \in \mathbb{R}$  (υπό τις συνήθεις συνθήκες), ενώ ο  $\hat{\xi}^W$  είναι συνεπής μόνο για  $\xi < 1/2$ . Επιπλέον, υπό

κατάλληλες συνθήκες για τα  $F$  και  $k$ , ο  $\xi^L$  είναι ασυμπτωτικά κανονικός, ενώ, ο  $\xi^W$  είναι ασυμπτωτικά κανονικός όταν  $\xi < 1/4$ .

### 3.6 ΕΚΤΙΜΗΤΗΣ ΡΟΠΩΝ ΣΤΑΘΜΙΣΜΕΝΗΣ ΠΙΘΑΝΟΤΗΤΑΣ

Η μέθοδος των ροπών σταθμισμένης πιθανότητας (Probability Weighted Moments-PWM) αποτελεί μια γενίκευση της μεθόδου των ροπών της κατανομής πιθανότητας. Η βασική ιδέα εισήχθη από τους Landwehr, Matalas, Wallis και Greenwood, σύμφωνα με την οποία, οι ροπές αντιστοιχίζονται με τις εμπειρικές συναρτήσεις, όμοια με την κλασική μέθοδο των ροπών (Landwehr, et al., 1979), (Greenwood, et al., 1979). Ο Hosking εισήγαγε την έννοια των ροπών σταθμισμένης πιθανότητας στην θεωρία ακραίων τιμών (Hosking, 1985), ενώ σε συνεργασία με τον Wallis παρουσίασε τον εκτιμητή:

$$\hat{\xi}^{PWM} = \frac{P_n - 4Q_n}{P_n - 2Q_n} = 1 - \left( \frac{P_n}{2Q_n} - 1 \right)^{-1}$$

όπου,

$$P_n := \frac{1}{k} \sum_{i=0}^{k-1} (X_{n-i:n} - X_{n-k:n}) \text{ και } Q_n := \frac{1}{k} \sum_{i=0}^{k-1} \frac{i}{k} (X_{n-i:n} - X_{n-k:n})$$

Όπως αναφέρουν οι (Hosking & Wallis, 1987), ο εκτιμητής  $\hat{\xi}^{PWM}$  υπό τις συνήθειες συνθήκες είναι ασθενώς συνεπής για  $\xi < 1$ , κατανέμεται ασυμπτωτικά κανονικά για  $\xi < 1/2$ , ενώ τέλος, για  $1/2 < \xi < 1$  η σύγκλιση του στο  $\xi$  είναι πιο αργή απ' ό τι για  $\xi < 1/2$ .

### 3.7 ΕΚΤΙΜΗΤΗΣ ΜΕΓΙΣΤΗΣ ΠΙΘΑΝΟΦΑΝΕΙΑΣ

Οι κατανομές για τις οποίες ισχύει ότι  $F \in D(G_\xi)$  για κάποια  $\xi \in \mathbb{R}$  δεν μπορούν να παραμετροποιηθούν με τέτοιο τρόπο, ώστε να υπάρξει ένας πεπερασμένος αριθμός παραμέτρων, οπότε δεν δύναται να υπάρξει ένας απευθείας τύπος για τις εκτιμήτριες μεγίστης πιθανοφάνειας. Ωστόσο, αν για  $0 < x < (OV(-\xi))^{-1}$  και έστω  $f$  μία θετική μη φθίνουσα συνάρτηση έχουμε:

$$\lim_{t \uparrow c_p} \prod Pr \left( \frac{X-t}{f(t)} > x \mid X > t \right) = 1 - H_\xi(x) := (1 - \xi x)^{-1/\xi}.$$

Στην παραπάνω σχέση παρατηρούμε ότι οι μεγαλύτερες παρατηρήσεις (ξεχωρίζονται από τον περιορισμό  $X > t$ ) ακολουθούν ασυμπτωτικά μία GPD. Ωστόσο, επειδή η κλάση των GP κατανομών έχουν μόνο μία παράμετρο, την  $\xi$ , αν εφαρμοστεί η μέθοδος μεγίστης πιθανοφάνειας στις μεγαλύτερες παρατηρήσεις και ως μοντέλο επιλεγεί η GPD, μπορεί να εξαχθούν χρήσιμα συμπεράσματα για την παράμετρο  $\xi$ .

Επομένως, δοθέντος ενός δείγματος από block maxima  $Y_1, Y_2, \dots, Y_k \sim GPD$ , λογαριθμημένη συνάρτηση πιθανοφάνειας (log-likelihood) των  $Y_i$  θα είναι για  $\xi \neq 0$ :

$$l(\mu, \sigma, \xi) = \ln \prod_{i=1}^k G'(Y_i; \mu, \sigma, \xi) = \sum_{i=1}^k G'(Y_i; \mu, \sigma, \xi)$$

Το άθροισμα των  $\ln l(Y_i; \mu, \sigma, \xi)$  είναι ίσο με 0 αν τα  $\mu, \sigma, \xi$  ικανοποιούν την  $1 + \xi \frac{Y_i - \mu}{\sigma} > 0$  για όλα τα  $Y_i$  και αποκλίνει στο  $-\infty$  σε διαφορετική περίπτωση. Στην περίπτωση όπου ισχύει  $\xi = 0$ , έχουμε:

$$l(\mu, \sigma, 0) = -k \ln \sigma - \sum_{i=1}^k \frac{Y_i - \mu}{\sigma} - \sum_{i=1}^k e^{-\frac{Y_i - \mu}{\sigma}}$$

Έχοντας αυτά ως βάση, είναι δυνατή η εκτίμηση των παραμέτρων με τις τιμές  $\mu, \sigma, \xi$ , οι οποίες μεγιστοποιούν τις παραπάνω εξισώσεις. Για να βρεθούν τα μέγιστα των παραπάνω χρησιμοποιούνται αριθμητικές μέθοδοι βελτιστοποίησης (όπως η Newton-Raphson), καθώς δεν υπάρχουν αναλυτικές εκφράσεις για τους εκτιμητές βάσει των πιθανοφανειών.

## 4 ΑΠΟΤΕΛΕΣΜΑΤΑ ΠΡΟΣΟΜΟΙΩΣΗΣ

### 4.1 ΜΕΘΟΔΟΛΟΓΙΑ

Σ αυτή την ενότητα θα αναλυθούν ως προς την απόδοση και συγκριθούν μεταξύ τους, μέσω της προσομοίωσης Monte Carlo, οι εκτιμητές του δείκτη ακραίων τιμών που έχουν παρουσιαστεί στα προηγούμενα κεφάλαια. Ο στόχος είναι η παραγωγή δείγματος μεγέθους 100.000 παρατηρήσεων, ούτως ώστε να υπάρχει στη διάθεσή μας αρκετό πλήθος εκτιμήσεων για να υπολογιστούν περιγραφικά χαρακτηριστικά από την κατανομή του κάθε εκτιμητή για να υπάρξει δυνατότητα σύγκρισης.

Ο υπολογισμός κάθε τιμής από τον εκάστοτε εκτιμητή βασίζεται σε ένα τυχαίο σετ δεδομένων μεγέθους 5.000 παρατηρήσεων. Καθώς, όπως αναφέρθηκε στη θεωρία, το σχήμα της ουράς προσδιορίζει την τιμή της παραμέτρου  $\xi$  της κατανομής, χρησιμοποιήθηκαν αντίστοιχες κατανομές ανά περίπτωση. Για τις περιπτώσεις μηδενικής τιμής για την παράμετρο έχουν χρησιμοποιηθεί οι κατανομές της Γάμμα, Weibull, η κανονική και η εκθετική κατανομή, καθώς οι ουρές τους συγκλίνουν αρκετά γρήγορα στο μηδέν. Οι κατανομές με πιο βαριά ουρά που χρησιμοποιήθηκαν για τις θετικές περιπτώσεις της παραμέτρου ήταν οι Pareto, Frechét καθώς και Burr, η οποία όταν η τιμή της δεύτερης παραμέτρου σχήματος είναι ίση με 1 καλείται κατανομή Pareto Type II. Οι κατανομές με πεπερασμένο εύρος, τέλος, ήταν οι Βήτα και η Ομοιόμορφη, οι οποίες οδήγησαν σε δείγματα με αρνητική τιμή του δείκτη  $\xi$  για το σχήμα της ουράς της κατανομής τους.

Εφαρμόστηκαν οι μέθοδοι Block Maxima και Peaks Over Threshold, όπως αναφέρθηκαν στο προηγούμενο κεφάλαιο. Αυτή η παραμετρική προσέγγιση για την μοντελοποίηση των ακροτάτων βασίζεται στην υπόθεση ότι τα δεδομένα  $(X_1, X_2, \dots, X_N)$  σχηματίζουν ένα iid δείγμα GEV ή GPD οικογένειας κατανομών, γεγονός που θα εξετάζεται σε κάθε περίπτωση με τους γραφικούς ελέγχους προσαρμογής των δεδομένων στην αντίστοιχη οικογένεια κατανομών, τα διαγράμματα πιθανοτήτων και ποσοστημορίων (PP – plot και QQ – plot αντίστοιχα). Ταυτόχρονα εξετάζεται το διάγραμμα στάθμης απόδοσης και το ιστόγραμμα των δεδομένων, που θα επικυρώνουν τη μορφή της οικογένειας κατανομών και την ύπαρξη «βαριάς» ουράς στην κατανομή των δεδομένων.

Τα γραφήματα PP – plot και QQ – plot παρέχουν την πληροφορία ταύτισης των παρατηρούμενων και των θεωρητικών εκατοστημορίων (percentile) και ποσοστημορίων (quantile) αντίστοιχα. Το Γράφημα Στάθμης Απόδοσης (Return Level Plot) υπολογίζεται βάσει των  $\mu, \sigma, \xi$ , των παραμέτρων της GEV (ή GPD) που ακολουθούν τα Block Maxima και την τιμή  $z_p$  το κατώφλι, το οποίο εξ ορισμού είναι  $P(X_i > z_p) = p$ . Προϋποθέτοντας  $w = -\ln(-\ln(1-p)) = \Lambda^{-1}(1-p)$ , τότε

$$z_{p(w)} = \begin{cases} \mu - \frac{\sigma}{\xi}(1 - e^{\xi w}), & \xi \neq 0 \\ \mu = \sigma w & , \xi = 0 \end{cases}$$

και επομένως, όταν  $\xi > 0$  η  $z_p(w)$  είναι αύξουσα κυρτή συνάρτηση του  $w$ , όταν  $\xi < 0$  η  $z_p(w)$  είναι κοίλη συνάρτηση του  $w$ , ενώ όταν  $\xi = 0$  η  $z_p(w)$  είναι γραμμικά αύξουσα συνάρτηση του  $w$ .

Στην εφαρμογή της θεωρίας επιλέχθηκαν οι τιμές για τον δείκτη  $\xi \in \{-1, -0.25, 0, 0.25, 1\}$ , καθώς η παρουσία μεγάλων τιμών στα δεδομένα κάνει δύσκολη την προσαρμογή της κατανομής σε μια GEV ή GPD παρουσιάζοντας πρόβλημα στις εκτιμήσεις της παραμέτρου, όταν ισχύει  $\xi < -1,5$  ή  $\xi > 1,5$ . Η επιλογή των επιμέρους παραμέτρων των κατανομών σε αυτές τις περιπτώσεις έγινε βάσει της θεωρίας που περιγράφεται από τον (Beirlant, et al., 2004), όπως φαίνεται στον παρακάτω πίνακα, ενώ για την περίπτωση, όπου  $\xi = 0$  χρησιμοποιήθηκαν η κατανομή Γάμμα, η Εκθετική, η Κανονική και η Weibull, με αυθαίρετες παραμέτρους.

Πίνακας 4.1.1 Λίστα των κατανομών με τους αντίστοιχους θεωρητικού δείκτες  $\xi$

Κατανομή	$1 - F(x)$	Δείκτης ακραίων τιμών
<i>Burr</i> ( $\alpha, \gamma, \theta$ ) (Type XII)	$\left(\frac{\theta}{\theta - x^{-\gamma}}\right)^\alpha, x > 0; \alpha, \gamma, \theta > 0$	$\frac{1}{\alpha\gamma}$
<i>Fréchet</i> ( $\alpha$ )	$1 - \exp\left(-\frac{x^{-\mu-a}}{\sigma}\right), x > 0; a > 0$	$\frac{1}{a}$
Βήτα( $p, q$ )	$\int_{1-\frac{1}{x}}^1 \frac{\Gamma(p+q)}{\Gamma(p)\Gamma(q)} u^{p-1}(1-u)^{q-1} du$	$-\frac{1}{q}$
Ομοιόμορφη( $\alpha, \beta$ )	$\frac{1}{x}, x > 1$	-1

Η ομοιομορφία στην εφαρμογή διασφαλίστηκε με τη χρήση ενός συγκεκριμένου σετ παραγμένων δειγμάτων, τα οποία παράγονταν κάθε φορά με την τοποθέτηση μιας τιμής seed πριν την εκτέλεση κάθε αλγορίθμου (συγκεκριμένα της τιμής 2017, όπως φαίνεται από τον αλγόριθμο που εκτελέστηκε<sup>2</sup>).

Συνολικά οι κατανομές και οι αντίστοιχες παράμετροι, που θα εξεταστούν εμφανίζονται στον παρακάτω πίνακα.

Πίνακας 4.1.2 Κατανομές που χρησιμοποιήθηκαν και η αντίστοιχη τιμή του δείκτη ακραίων τιμών που αναμένουμε

Κατανομή	Δείκτης Ακραίων Τιμών $\xi$	Άλλες παράμετροι
Burr	0.25, 0.5, 1	$(\alpha, \gamma, \theta) = (4, 1, 1), (1, 2, 1), (1, 1, 1)$
Fréchet	0.25, 0.5, 1	$(\mu, \sigma, \alpha) = (0, 1, 4), (0, 1, 2), (0, 2, 1)$
Εκθετική	0	$\lambda = 1$
Κανονική	0	$\mu = 10, \sigma = 1$
Γάμμα	0	$(\alpha, \beta) = (1, 0.5), (1, 1.5)$

<sup>2</sup> Βλέπε Παράρτημα II

Κατανομή	Δείκτης Ακραίων Τιμών $\xi$	Άλλες παράμετροι
Weibull	0	$(\lambda, \tau) = (1, 0.5), (1, 1.5)$
Βήτα	-0.25, -1	$(\alpha, \beta) = (0.5, 4), (5, 1)$
Ομοιόμορφη	-1	$(\alpha, \beta) = (0, 1)$

Για τη μέθοδο Block Maxima χρησιμοποιήθηκε ο σχηματισμός blocks μεγέθους από 2 έως και 25 ανάλογα με τις ανάγκες της κατανομής και δημιουργήθηκαν δείγματα από 200 έως 2.500 παρατηρήσεις. Για τη μέθοδο Peaks Over Threshold ήταν αναγκαίος ο υπολογισμός του βέλτιστου κατωφλιού ανά περίπτωση μέσω κατάλληλων γραφημάτων (όπως το MRLP που παρουσιάζεται), επομένως η κάθε εκτίμηση με αυτήν τη μέθοδο επιλογής δεδομένων δε βασίζεται σε σταθερό πλήθος δεδομένων κάθε φορά. Πάνω σε αυτά τα επιλεγμένα maxima και peaks χρησιμοποιήθηκε ο εκτιμητής μεγίστης πιθανοφάνειας της γενικευμένης εκθετικής κατανομής και της γενικευμένης κατανομής Pareto. Ενώ η εφαρμογή των υπολοίπων εκτιμητών έγινε στο σύνολο των 5.000 παρατηρήσεων κάθε φορά ανά κατανομή. Συνοπτικά οι εκτιμητές που εξετάστηκαν μέσω προσομοίωσης είναι οι παρακάτω.

Πίνακας 4.1.3 Οι εκτιμητές που εξετάστηκαν μέσω προσομοίωσης και οι κλειστοί τύποι που τους περιγράφουν

Εκτιμητής	Τύπος
Εκτιμητής του Pickand ( $\xi \in \mathbb{R}$ )	$\hat{\xi}_{(k,n)}^{Pickand} = \frac{1}{\ln 2} \ln \left( \frac{X_{n-k,n} - X_{n-2k,n}}{X_{n-2k,n} - X_{n-4k,n}} \right)$
Εκτιμητής του Hill ( $\xi > 0$ )	$\hat{\xi}_{(k,n)}^{Hill} = \frac{1}{k} \sum_{i=1}^k \ln(X_{i,n}) - \ln(X_{k+1,n})$
Εκτιμητής Negative Hill ( $\xi < -1/2$ )	$\hat{\xi}_{(k,n)}^{N Hill} = \frac{1}{k} \sum_{i=1}^{k-1} \ln(X_{n,n} - X_{n-i,n}) - \ln(X_{n,n} - X_{n-k,n})$
Εκτιμητής των Ροπών ( $\xi \in \mathbb{R}$ )	$\hat{\xi}^M = H_n^{(1)} + 1 - \frac{1}{2} \left( 1 - \frac{(H_n^{(1)})^2}{H_n^{(2)}} \right)^{-1}$
Εκτιμητής Λόγου των Ροπών ( $\xi > 0$ )	$\hat{\xi}^{MR} = \frac{1}{2} \frac{H_n^{(2)}}{H_n^{(1)}}$
Εκτιμητής Peng ( $\xi \in \mathbb{R}$ )	$\hat{\xi}^L = \frac{H_n^{(2)}}{2H_n^{(1)}} + 1 - \frac{1}{2} \left( 1 - \frac{(H_n^{(1)})^2}{H_n^{(2)}} \right)^{-1}$
Εκτιμητής W ( $\xi < 1/2$ )	$\hat{\xi}^W = 1 - \frac{1}{2} \left( 1 - \frac{(L_n^{(2)})^2}{L_n^{(1)}} \right)$
Εκτιμητής Ροπών Σταθμισμένης Πιθανότητας ( $\xi < 1/2$ )	$\hat{\xi}^{PWM} = \frac{P_n - 4Q_n}{P_n - 2Q_n} = 1 - \left( \frac{P_n}{2Q_n} - 1 \right)^{-1}$

Στην περίπτωση του εκτιμητή μεγίστης πιθανοφάνειας δεν υπάρχει κλειστός τύπος για την εφαρμογή του. Αντιθέτως, χρησιμοποιείται το αποτέλεσμα από την αριθμητική λύση, που παρέχει η κατάλληλη εντολή του πακέτου fExtremes της R. Ενώ η εφαρμογή των εκτιμητών έγινε για τιμές της παραμέτρου  $\kappa$ , που φαίνεται στις συναρτήσεις των εκτιμητών, από 25 έως 500, για να μπορεί να γίνει σύγκριση και μεταξύ τους.

Η απόδοση κάθε εκτιμητή  $\xi$ , αξιολογήθηκε βάσει τριών βασικών περιγραφικών στοιχείων από την κατανομή που προέκυψε γι' αυτόν. Η μεροληψία (bias) της κατανομής, δηλαδή της απόστασης της μέσης τιμής της από την θεωρητική τιμή του δείκτη  $\xi$ , του τυπικού σφάλματος (s.e.), το οποίο δείχνει το εύρος τιμών της κατανομής και του μέσου τετραγωνικού σφάλματος (m.s.e.), που είναι οι μέσες τετραγωνικές αποστάσεις από την θεωρητική τιμή του δείκτη  $\xi$ , ενώ εξετάζεται και ο χρόνος (σε δευτερόλεπτα) που χρειάστηκε ο επεξεργαστής να ολοκληρώσει την παραγωγή του δείγματος κρατήθηκε μέσω της εντολής `proc.time()` του βασικού πακέτου της R.

## 4.2 ΕΚΘΕΤΙΚΗ ΚΑΤΑΝΟΜΗ

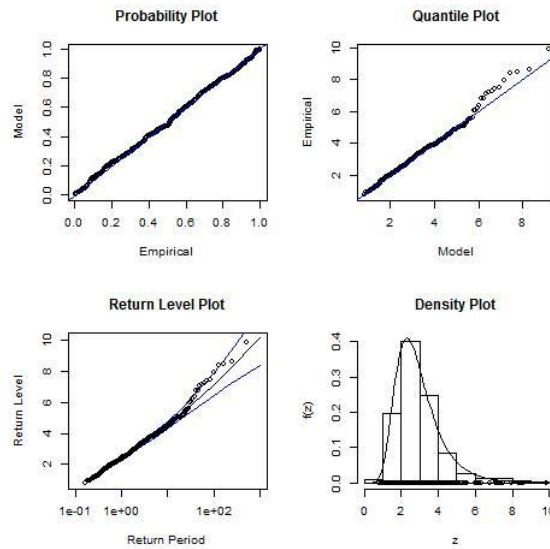
Αρχικά, ελέγχουμε την προσαρμογή των δεδομένων που παρήχθησαν με την μέθοδο επιλογής Block Maxima, ξεκινώντας με πολλά Blocks (τα οποία περιέχουν μικρό αριθμό παρατηρήσεων), σταδιακά μειώνουμε τον αριθμό των Blocks, στα οποία χωρίζουμε το αρχικό δείγμα, και αυξάνουμε τις παρατηρήσεις που περιέχουν έως ότου να μην παρατηρείται καλή προσαρμογή στην GEV. Η διαδικασία αυτή εξασφαλίζει την ισονομία των παρατηρήσεων  $M_n$ , καθώς και την ισορροπία μεταξύ μεροληψίας και διακύμανσης όπως αναφέρθηκε στο Κεφάλαιο 2.1.1.

Έτσι, για την Εκθετική κατανομή με παράμετρο  $\lambda = 1$ , αφού παρήχθη δείγμα μεγέθους  $n = 5.000$ , και μετά από πολλαπλές δοκιμές, ο βέλτιστος αριθμός Block Maxima, που δίνει καλή προσαρμογή στην GEV, είναι 500, με 10 παρατηρήσεις το καθένα μπλοκ. Βάσει αυτών των μέγιστων τιμών από κάθε block και με χρήση της μεθόδου μεγίστης πιθανοφάνειας εκτιμούμε τις παραμέτρους της GEV, την οποία θα πρέπει ακολουθούν προσεγγιστικά. Έτσι, καταλήγουμε στο συμπέρασμα ότι η κατανομή των δεδομένων θα είναι η GEV ( $\mu, \sigma, \xi$ ), με  $(\hat{\mu}, \hat{\sigma}, \hat{\xi}) = (2.35558697, 0.89880337, 0.06392047)$ .

Εν συνεχεία, παρατίθενται τα διαγράμματα καλής προσαρμογής των δεδομένων στην GEV.



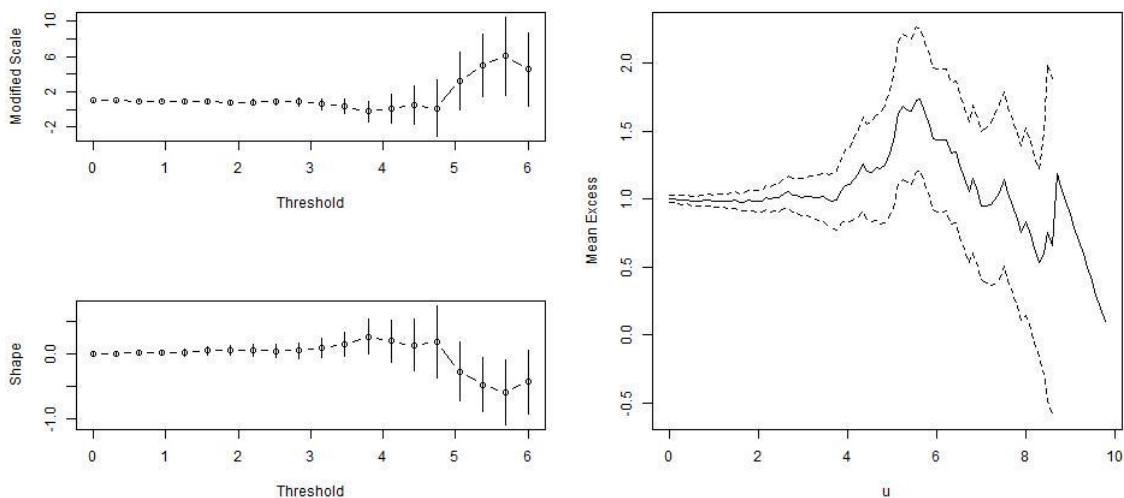
Εικόνα 4.2.1 Διαγράμματα καλής προσαρμογής στην GEV για την Εκθετική ( $\lambda = 1$ )



Από τα PP-plot και QQ-plot διαφαίνεται, ότι τα θεωρητικά και τα παρατηρούμενα σημεία των κατανομών συμπίπτουν, όπως είναι αναμενόμενο εάν η παρατηρούμενη κατανομή ανήκει στην οικογένεια GEV, γεγονός που φαίνεται και από το ιστόγραμμα των Block Maxima (Density Plot). Αντίστοιχα, από το γράφημα στάθμης απόδοσης (Return Level Plot) βλέπουμε ότι η  $z_p(w)$  αυξάνεται γραμμικά ως προς το  $w$ , άρα επιβεβαιώνεται, ότι η παράμετρος  $\xi = 0$ .

Για την εφαρμογή της μεθόδου Peak Over Threshold η επιλογή του βέλτιστου κατωφλιού  $u$  θα γίνει από τα παρακάτω γραφήματα.

Εικόνα 4.2.2 Διαγράμματα για την επιλογή του βέλτιστου  $u$  στην περίπτωση της Εκθετικής ( $\lambda=1$ )

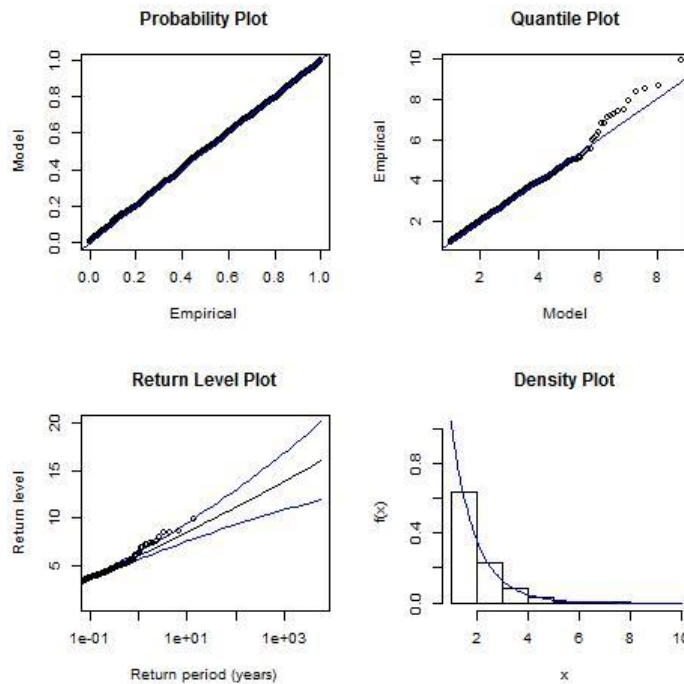


Με χρήση των δύο διαγραμμάτων για το threshold, καταλήγουμε στο συμπέρασμα ότι το μικρότερο  $u$  πέραν του οποίου το  $\xi$  σταθεροποιείται και η εκτίμηση του  $\tilde{\sigma}$

μεταβάλλεται γραμμικά ως προς το  $u$  είναι το 1. Αντίστοιχα συμπεράσματα παίρνουμε και από το Mean Excess Plot, όπου παρατηρούμε ότι η μονάδα είναι η τιμή όπου το διάγραμμα γίνεται οριζόντιο και τα δ.ε δεν ξεφεύγουν. Συνεπώς, θα χρησιμοποιηθούν για τον υπολογισμό των εκτιμητών τα δεδομένα από την αρχική εκθετική κατανομή, τα οποία είναι μεγαλύτερα της μονάδας. Με την εφαρμογή του  $u$  για το πρώτο δείγμα, το ποσοστό των δεδομένων που ακολουθεί GPD με  $(\tilde{\sigma}, \tilde{\xi}) = (0.96482807, 0.02029499)$  είναι 37,10 %. Επομένως, ο MLE που θα υπολογισθεί βάσει της μεθόδου Peak Over Threshold θα χρησιμοποιεί δείγματα με περίπου 1.900 παρατηρήσεις.

Στη συνέχεια, με τον ίδιο γραφικό τρόπο παρατηρούμε ότι τα θεωρητικά και τα παρατηρούμενα σημεία των κατανομών συμπίπτουν, όπως είναι αναμενόμενο όταν η παρατηρούμενη κατανομή είναι η GPD, γεγονός που φαίνεται και από το ιστόγραμμα των Peaks Over Threshold (Density Plot). Αντίστοιχα, από το γράφημα στάθμης απόδοσης (Return Level Plot) βλέπουμε ότι η  $z_p(w)$  αυξάνεται γραμμικά ως προς το  $w$ , άρα  $\xi = 0$ .

Εικόνα 4.2.3 Διαγράμματα καλής προσαρμογής στην GPD για την Εκθετική ( $\lambda = 1$ )



Παρακάτω, φαίνονται συγκεντρωτικά οι εκτιμώμενες (μέσω προσομοίωσης  $10^5$  επαναλήψεων) παράμετροι των κατανομών των διάφορων εκτιμητών του  $\xi$  από κάθε εκτιμητή για  $k = 25, 50, 100, 500$ .

Πίνακας 4.2.1 Περιγραφικά χαρακτηριστικά της κατανομής των προσομοιωμένων εκτιμητών για την Εκθετική ( $\lambda = 1$ )

<i>MLE</i>	<i>Εκθετική (<math>\lambda=1</math>)</i>			
	<i>MEAN</i>	<i>SE</i>	<i>MSE</i>	<i>DURATION</i>
BM	0,0354	0,0321	0,0023	2.497,56
POT	-0,0016	0,0235	0,0006	1.776,11

	25					50			
	MEAN	SE	MSE	DURATION		MEAN	SE	MSE	DURATION
Pickand's	-0,0236	0,3637	0,1329	190,22	Pickand's	-0,0135	0,2563	0,0659	207,01
Hill's	0,1629	0,0291	0,0274	2.081,59	Hill's	0,1836	0,0229	0,0342	4.354,23
Negative Hill's	-0,3543	0,1423	0,1458	3.856,92	Negative Hill's	-0,3029	0,1025	0,1023	7.381,53
Moment	-0,0831	0,2315	0,0605	4.153,72	Moment	-0,0289	0,1524	0,0241	9.042,61
Moment Ratio	0,1407	0,0277	0,0206	4.180,86	Moment Ratio	0,1588	0,0215	0,0257	9.561,99
Peng	-0,1053	0,2487	0,0729	3.981,83	Peng	-0,0536	0,1667	0,0306	9.316,63
W	-0,1108	0,2136	0,0579	3.489,08	W	-0,0561	0,1429	0,0236	9.670,68
PWM	0,0537	0,2192	0,0509	3.487,19	PWM	0,0271	0,1595	0,0262	9.327,65

	100					500			
	MEAN	SE	MSE	DURATION		MEAN	SE	MSE	DURATION
Pickand's	-0,0064	0,1803	0,0326	191,67	Pickand's	-0,0008	0,0857	0,0065	165,85
Hill's	0,2106	0,0185	0,0447	8.653,50	Hill's	0,3240	0,0124	0,1051	35.653,47
Negative Hill's	-0,2587	0,0750	0,0725	18.523,23	Negative Hill's	-0,1860	0,0398	0,0362	72.332,47
Moment	1,37E-05	0,1042	0,0109	19.136,84	Moment	0,0369	0,0479	0,0037	75.112,51
Moment Ratio	0,1808	0,0167	0,0330	16.247,91	Moment Ratio	0,2653	0,0099	0,0705	71.269,61
Peng	-0,0298	0,1156	0,0143	6.309,24	Peng	-0,0218	0,0548	0,0035	70.038,84
W	-0,0281	0,0988	0,0105	16.003,83	W	-0,0058	0,0444	0,0020	73.465,10
PWM	0,0142	0,1138	0,0131	15.815,80	PWM	0,0027	0,0515	0,0027	70.726,28

Παρατηρούμε ότι ο MLE με βάση τη μέθοδο διαλογής δεδομένων POT παρουσιάζει το μικρότερο MSE, το οποίο ισούται με 0,0006 και σχεδόν πάντα την μικρότερη μεροληψία (έκτος της περίπτωσης  $k=500$ ). Σε κάθε περίπτωση, πλην της  $k=500$ , το αμέσως μικρότερο MSE έχει ο MLE υπολογισμένος με την μέθοδο BM. Ωστόσο, μια άμεση σύγκριση των δύο εκτιμητών δεν θα ήταν δόκιμη, καθώς ο  $MLE^{BM}$  χρησιμοποιεί μόνο 500 δεδομένα, ενώ ο  $MLE^{POT}$  1.900, χωρίς, όμως, να είναι δυνατό να εξαλείψουμε αυτή την διαφορά στα μεγέθη των δειγμάτων, καθώς για αυτό το πλήθος δεδομένων, παρατηρείται η καλύτερη προσαρμογή στην αντίστοιχη οικογένεια κατανομών.

Στην περίπτωση της Εκθετικής, αναμέναμε βάσει θεωρίας, καλύτερη απόδοση από τους Pickand's, Moment, Moment Ratio και Peng. Αρχικά, για τον εκτιμητή Pickand's, παρατηρούμε ότι όσο αυξάνεται το  $k$  τόσο μειώνεται η μεροληψία του, καθώς και το MSE του, και μάλιστα σημαντικά. Ίδια συμπεριφορά παρατηρείται και από τον εκτιμητή Peng, χωρίς ωστόσο να αποδίδει εξίσου καλά. Ο εκτιμητής Moment Ratio δεν αποδίδει όπως αναμενόταν, με την μεροληψία του να εντείνεται με την αύξηση του  $k$ . Τέλος, ο εκτιμητής Moment παρουσιάζει την μικρότερη μεροληψία που σημειώθηκε στην προσομοίωση, με τιμή μόλις 0,0000137, και σε όλες τις περιπτώσεις αρκετά καλό MSE.

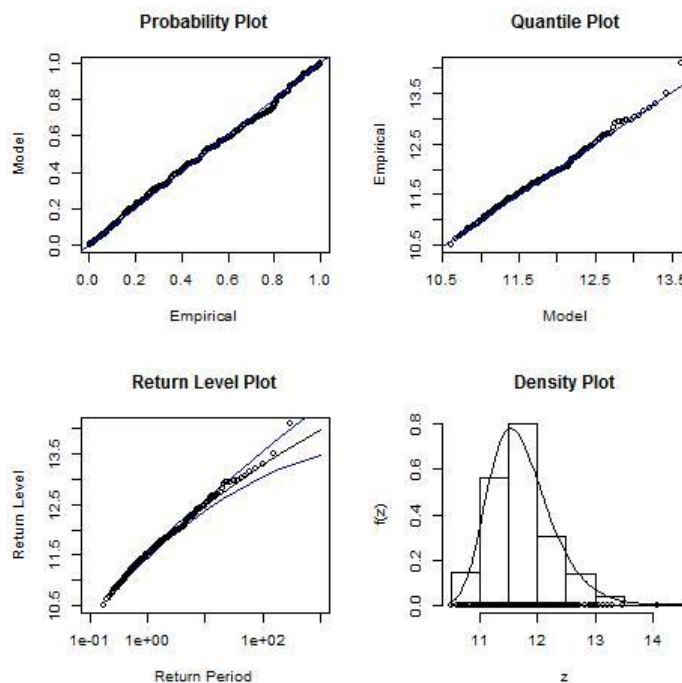
Αξίζει να σημειωθεί, πως αντίθετα με τις προσδοκίες, ο PWM παρουσιάζει πολύ καλές τιμές στα κριτήρια μας, καθώς όσο αυξάνεται το  $k$  μειώνονται τόσο η μεροληψία όσο και το MSE, σε σημείο που φτάνει να γίνεται από τους καλύτερους για  $k=500$ .

Εν κατακλείδι, μια ασυμπτωτική προσέγγιση φαίνεται να δίνει καλύτερα αποτελέσματα στην περίπτωση της Εκθετικής κατανομής με  $\lambda = 1$ . Για τους μη ασυμπτωτικούς εκτιμητές, παρατηρείται πως συμπεριφέρονται καλύτερα για  $k = 500$ , με τον  $W$  να εκτιμά με την μικρότερη μεροληψία από τους υπόλοιπους.

### 4.3 ΚΑΝΟΝΙΚΗ ΚΑΤΑΝΟΜΗ

Στην περίπτωση δεδομένων με κανονική κατανομή εξετάστηκε δείγμα με παραμέτρους  $\mu = 10$  και  $\sigma = 1$ . Θεωρητικά, η κανονική κατανομή λόγω της λεπτής ουράς που έχει αναμένεται να έχει τιμή του δείκτη  $\xi = 0$ . Ωστόσο, παρακάτω, εξαιτίας της μικρής διακύμανσης που έχει επιλεγεί, καθώς και τον περιορισμό σε 5.000 παρατηρήσεις, πιθανώς εμφανίζεται κάποιο μέγιστο στην ουρά της κατανομής κάνοντάς την να έχει ελαφρά αρνητική εκτίμηση του δείκτη. Ο βέλτιστος αριθμός block, στα οποία χρειάστηκε να μοιραστεί το κάθε δείγμα ήταν 300 block με 17 παρατηρήσεις το καθένα. Η προσαρμοσμένη GEV για τα 300 maxima είχε παραμέτρους  $(\hat{\mu}, \hat{\sigma}, \hat{\xi}) = (11.4983474, 0.4728069, -0.0865454)$ . Από τον έλεγχο καλής προσαρμογής της κατανομής με αυτές τις παραμέτρους παίρνουμε την ακόλουθη εικόνα.

Εικόνα 4.3.1 Διαγράμματα καλής προσαρμογής στην GEV για την Κανονική (10, 1)

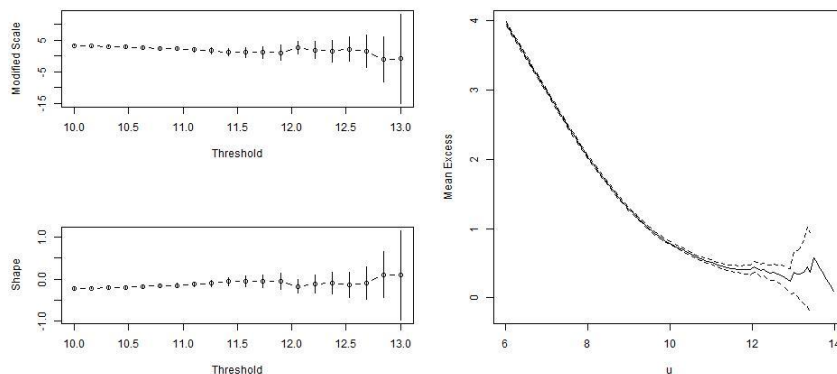


Όπως είναι εμφανές δεν υπάρχει κάποια σημαντική απόκλιση από τις θεωρητικές τιμές του μοντέλου επομένως μπορούμε να υποθέσουμε ικανοποιητική προσαρμογή. Αξιοσημείωτη είναι η ελαφριά τάση της στάθμης απόδοσης να γίνει κοίλη καμπύλη,

πράγμα που δείχνει, ότι η πραγματική τιμή της παραμέτρου  $\xi$  να είναι ελαφριά μικρότερη του μηδενός και η εκτίμηση  $\hat{\xi} = -0.0865454$  που έγινε να μην πρόκειται για μία τυχαία υποεκτίμηση της.

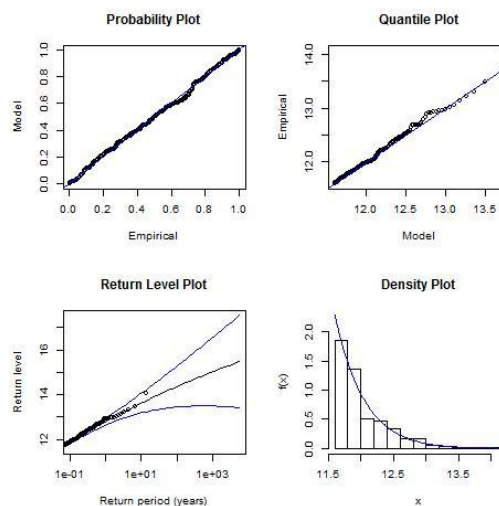
Ακολουθώντας, για την εφαρμογή της μεθόδου POT ελήφθη το σημείο  $u = 11,6$  ως βέλτιστο κατώφλι για τα δεδομένα επιλέγοντας το 5,14% των αρχικών δεδομένων, αφήνοντας περίπου 260 παρατηρήσεις για τον υπολογισμό της παραμέτρου  $\xi$ . Από την προσαρμογή της GPD στα δεδομένα, παίρνουμε εκτίμηση για τις παραμέτρους της κατανομής  $(\hat{\sigma}, \hat{\xi}) = (0.43759874, -0.04787455)$ .

Εικόνα 4.3.2 Διαγράμματα για την επιλογή του βέλτιστου  $u$  στην περίπτωση της Κανονικής (10, 1)



Ο έλεγχος καλής προσαρμογής της κατανομής της GEV με τις παραμέτρους που εκτιμήθηκαν προηγουμένως δίνει την παρακάτω εικόνα.

Εικόνα 4.3.3 Διαγράμματα καλής προσαρμογής στην GPD για την Κανονική (10, 1)



Από την σύγκριση των θεωρητικών και των παρατηρούμενων ποσοστημορίων και εκατοστημορίων παρατηρούνται αποκλίσεις. Παρά τις δοκιμές που έγιναν δεν βρέθηκε με κανέναν τρόπο καλύτερη προσαρμογή. Αυτό μπορεί να οφείλεται είτε στον τρόπο υπολογισμού του πακέτου είτε στο ότι η κανονική κατανομή παρουσιάζει πολύ λεπτή

ουρά, με αποτέλεσμα να δίνεται η «ψευδαίσθηση» ύπαρξης πεπερασμένου δεξιού άκρου του στηρίγματος. Για να είναι δυνατός ο υπολογισμός του MLE με την Block Maxima, υποθέτουμε καλή προσαρμογή της κατανομής. Ταυτόχρονα, η καμπύλη της στάθμης απόδοσης υποδεικνύει, ότι η παράμετρος  $\xi$  πιθανότατα είναι ελαφρά μικρότερη του μηδενός.

Παρακάτω, φαίνονται συγκεντρωτικά οι εκτιμώμενες (μέσω προσομοίωσης  $10^5$  επαναλήψεων) παράμετροι των κατανομών των διάφορων εκτιμητών του  $\xi$  από κάθε εκτιμητή για  $k = 25, 50, 100, 500$ .

Πίνακας 4.3.1 Περιγραφικά χαρακτηριστικά της κατανομής των προσομοιωμένων εκτιμητών για την Κανονική (10, 1)

<i>Κανονική (10, 1)</i>									
<i>MLE</i>	<i>MEAN</i>	<i>SE</i>	<i>MSE</i>	<i>DURATION</i>		<i>MEAN</i>	<i>SE</i>	<i>MSE</i>	<i>DURATION</i>
BM	-0,1492	0,0394	0,0238						2.152,66
POT	-0,1410	0,0594	0,0234						647,45

<i>25</i>					<i>50</i>				
	<i>MEAN</i>	<i>SE</i>	<i>MSE</i>	<i>DURATION</i>		<i>MEAN</i>	<i>SE</i>	<i>MSE</i>	<i>DURATION</i>
Pickand's	-0,1512	0,3575	0,1506	188,16	Pickand's	-0,1644	0,2519	0,0905	186,84
Hill's	0,0247	0,0045	0,0006	1.882,05	Hill's	0,0269	0,0034	0,0007	3.651,07
Negative Hill's	-0,3904	0,1440	0,1732	3.308,32	Negative Hill's	-0,3454	0,1042	0,1301	6.734,42
Moment	-0,1819	0,2281	0,0851	3.486,48	Moment	-0,1404	0,1505	0,0424	6.906,19
Moment Ratio	0,0219	0,0047	0,0005	3.861,51	Moment Ratio	0,0240	0,0036	0,0006	7.517,28
Peng	-0,1847	0,2309	0,0874	4.144,22	Peng	-0,1433	0,1529	0,0439	7.390,10
W	-0,1852	0,2253	0,0851	3.740,36	W	-0,1433	0,1491	0,0428	7.409,22
PWM	-0,0192	0,2267	0,0518	3.662,45	PWM	-0,0635	0,1655	0,0314	7.635,86

<i>100</i>					<i>500</i>				
	<i>MEAN</i>	<i>SE</i>	<i>MSE</i>	<i>DURATION</i>		<i>MEAN</i>	<i>SE</i>	<i>MSE</i>	<i>DURATION</i>
Pickand's	-0,1961	0,1770	0,0698	188,29	Pickand's	-0,4204	0,0776	0,1828	189,05
Hill's	0,0297	0,0027	0,0009	7.243,33	Hill's	0,0405	0,0016	0,0016	36.003,32
Negative Hill's	-0,3079	0,0772	0,1008	13.549,19	Negative Hill's	-0,2567	0,0429	0,0677	68.530,88
Moment	-0,1287	0,1042	0,0274	13.868,09	Moment	-0,1651	0,0478	0,0296	69.415,52
Moment Ratio	0,0264	0,0027	0,0007	15.029,58	Moment Ratio	0,0347	0,0014	0,0012	75.334,68
Peng	-0,1320	0,1060	0,0287	14.484,69	Peng	-0,1710	0,0488	0,0316	74.937,17
W	-0,1311	0,1034	0,0279	14.785,15	W	-0,1675	0,0472	0,0303	76.540,45
PWM	-0,0950	0,1196	0,0233	15.241,08	PWM	-0,1717	0,0563	0,0326	71.668,73

Από τα αποτελέσματα επικυρώνεται η υποψία, ότι η εκτιμώμενη (MLE) τιμή της παραμέτρου  $\xi$  είναι ελαφρά μικρότερη το μηδενός από τις μέσες τιμές των εκτιμήσεων μέσω πιθανοφάνειας, οι οποίες αποδίδουν τιμή του  $\xi$  περίπου -0,14. Ωστόσο, οι εκτιμητές

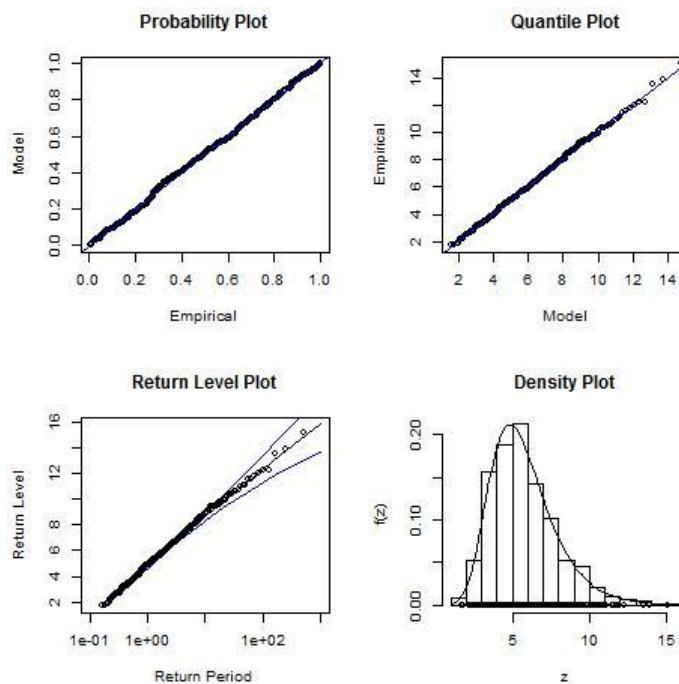
των Hill's και Λόγου Ροπών εμφανίζονται πιο αποδοτικοί στην εκτίμηση του μηδενικού δείκτη  $\xi$  καθώς έχουν σταθερά την μικρότερη μεροληψία στις εκτιμήσεις τους και διατηρούν ελάχιστο μέσο τετραγωνικό σφάλμα, το οποίο αυξάνεται, όμως, καθώς η τιμή του  $k$  αυξάνεται. Η μόνη διαφορά σε αυτό το σημείο είναι το γεγονός, ότι ο εκτιμητής Λόγου Ροπών δεν προβλέπεται από τη θεωρία να είναι ο πιο αποδοτικός, όταν η τιμή του δείκτη  $\xi$  είναι μηδενική. Ικανοποιητικός είναι και ο εκτιμητής Ροπών Σταθμισμένης Πιθανότητας, ο οποίος έχει την ελάχιστη ανάμεσα στους εκτιμητές με αρνητική μεροληψία. Αξιολογώντας το χρόνο για την εκτίμηση του αποτελέσματος, ο εκτιμητής Hill's και ειδικά για μικρότερες τιμές  $k$  εμφανίζεται να είναι η καταλληλότερη επιλογή.

#### 4.4 ΚΑΤΑΝΟΜΗ ΓΑΜΜΑ

Εξετάζοντας την περίπτωση των δεδομένων από κατανομή Γάμμα δοκιμάστηκαν δύο κατανομές με παραμέτρους  $\alpha = 1$  και  $\beta = 0,5$  και  $\beta = 1,5$  αντίστοιχα ανά κατανομή. Η Γάμμα, όπως και οι άλλες κατανομές με συνάρτηση πυκνότητας πιθανότητας ανάλογης ενός εκθετικού μέρους  $e^{\beta x}$ , έχει λεπτή ουρά και αντίστοιχα δείκτη  $\xi = 0$ .

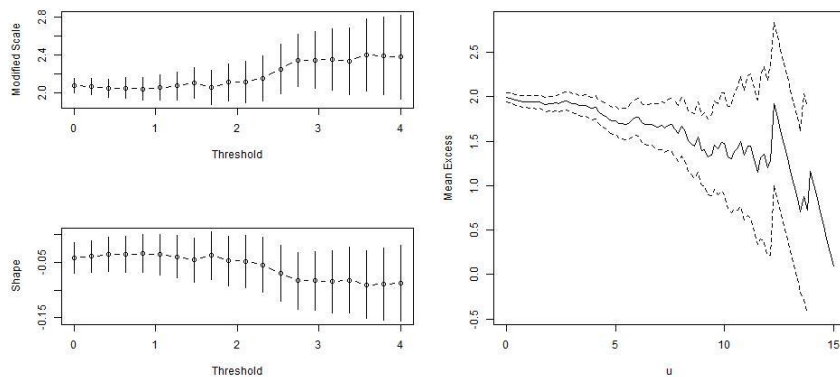
Για την Γάμμα (1, 0.5) ο βέλτιστος αριθμός block ήταν 500 τα οποία αποτελούνται από 10 παρατηρήσεις. Η παράμετροι της GEV, που προσαρμόστηκε ήταν  $(\hat{\mu}, \hat{\sigma}, \hat{\xi}) = (4.80323922, 1.73920507, -0.02553793)$ . Από τις εκτιμήσεις ήδη φαίνεται, ότι η τιμή του δείκτη  $\xi$  είναι πολύ κοντά στο μηδέν και από τα παρακάτω γραφήματα ελέγχου καλής προσαρμογής φαίνεται, ότι μπορούμε να υποθέσουμε την καλή προσαρμογή και από το σχήμα του γραφήματος στάθμης απόδοσης την τιμή του  $\xi$ , που αναμένεται να είναι μηδέν.

Εικόνα 4.4.1 Διαγράμματα καλής προσαρμογής στην GEV για την Γάμμα (1, 0.5)



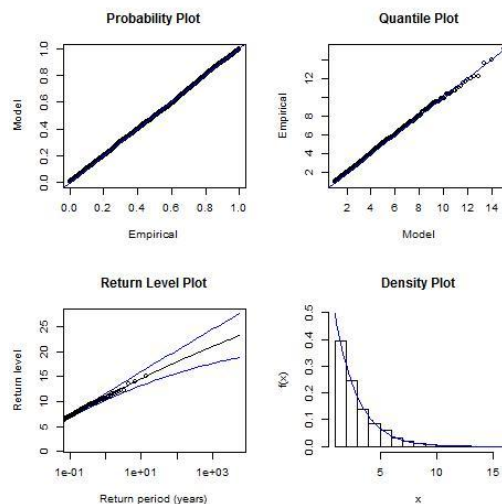
Στην εφαρμογή της μεθόδου POT ακολουθώντας την υπόδειξη των παρακάτω γραφημάτων για την επιλογή του βέλτιστου κατωφλιού  $u$  κρατήθηκε η τιμή  $u = 1$  τιμή, η οποία επιλέγει το 61,6% των παρατηρήσεων. Επομένως τα δείγματα που θα χρησιμοποιούνται για τον υπολογισμό του εκτιμητή μεγίστης πιθανοφάνειας αναμένεται να έχουν 3.000 παρατηρήσεις περίπου.

Εικόνα 4.4.2 Διαγράμματα για την επιλογή του βέλτιστου  $u$  στην περίπτωση της Γάμμα (1, 0.5)



Στο ακόλουθο σχήμα με τα διαγράμματα ελέγχου καλής προσαρμογής διαπιστώνουμε, ότι οι θεωρητικές τιμές του μοντέλου της κατανομής συμπίπτουν εξαιρετικά με τις παρατηρούμενες, ενώ το διάγραμμα στάθμη απόδοσης εμφανίζεται γραμμικό, επομένως μπορούμε να υποθέσουμε την καλή προσαρμογή της κατανομής και ότι η τιμή του δείκτη  $\eta$  είναι μηδέν.

Εικόνα 4.4.3 Διαγράμματα καλής προσαρμογής στην GPD για την Γάμμα (1, 0.5)



Παρακάτω, φαίνονται συγκεντρωτικά οι εκτιμώμενες (μέσω προσομοίωσης  $10^5$  επαναλήψεων) παράμετροι των κατανομών των διάφορων εκτιμητών του  $\xi$  από κάθε εκτιμητή για  $k = 25, 50, 100, 500$ .



Πίνακας 4.4.1 Περιγραφικά χαρακτηριστικά της κατανομής των προσομοιωμένων εκτιμητών για την Γάμμα (1, 0,5)

**Γάμμα (1, 0,5)**

<i>MLE</i>	<i>MEAN</i>	<i>SE</i>	<i>MSE</i>	<i>DURATION</i>
BM	0,0354	0,0322	0,0023	2.354,11
POT	-0,0010	0,0182	0,0003	2.850,81

	<b>25</b>					<b>50</b>			
	<i>MEAN</i>	<i>SE</i>	<i>MSE</i>	<i>DURATION</i>		<i>MEAN</i>	<i>SE</i>	<i>MSE</i>	<i>DURATION</i>
Pickand's	-0,0259	0,3619	0,1317	249,30	Pickand's	-0,0120	0,2562	0,0658	257,91
Hill's	0,1627	0,0292	0,0273	2.152,88	Hill's	0,1834	0,0230	0,0342	4.202,95
Negative Hill's	-0,3543	0,1427	0,1459	3.846,02	Negative Hill's	-0,3030	0,1027	0,1024	7.768,40
Moment	-0,0838	0,2320	0,0609	4.047,82	Moment	-0,0298	0,1522	0,0240	7.852,15
Moment Ratio	0,1405	0,0276	0,0205	4.357,95	Moment Ratio	0,1586	0,0215	0,0256	8.679,99
Peng	-0,1060	0,2491	0,0733	4.648,99	Peng	-0,0546	0,1663	0,0306	8.851,88
W	-0,1113	0,2141	0,0582	4.781,07	W	-0,0568	0,1424	0,0235	8.798,02
PWM	0,0531	0,2194	0,0510	4.273,46	PWM	0,0259	0,1596	0,0261	8.466,28

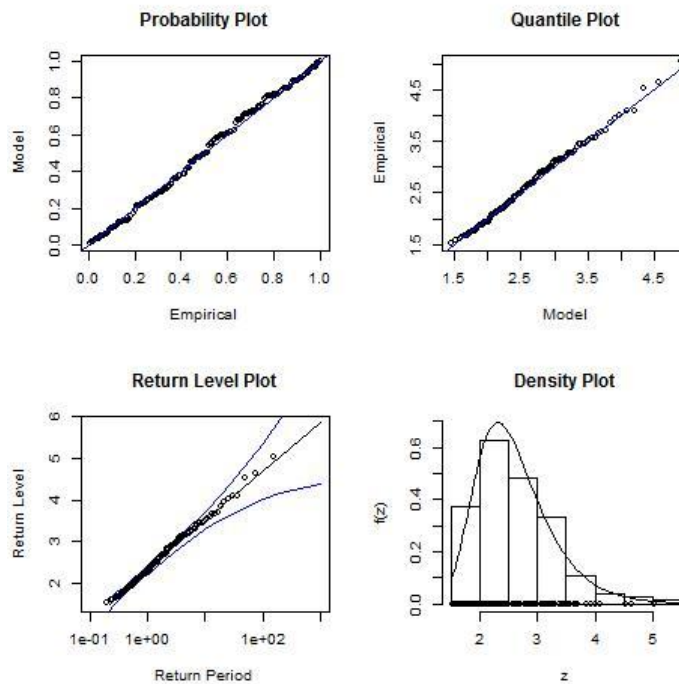
	<b>100</b>					<b>500</b>			
	<i>MEAN</i>	<i>SE</i>	<i>MSE</i>	<i>DURATION</i>		<i>MEAN</i>	<i>SE</i>	<i>MSE</i>	<i>DURATION</i>
Pickand's	-0,0069	0,1807	0,0327	252,82	Pickand's	-0,0012	0,0806	0,0065	254,25
Hill's	0,2106	0,0185	0,0447	8.346,40	Hill's	0,3241	0,0125	0,1052	40.949,39
Negative Hill's	-0,2590	0,0753	0,0728	15.117,16	Negative Hill's	-0,1862	0,0400	0,0363	78.847,98
Moment	-0,0009	0,1044	0,0109	15.682,77	Moment	0,0367	0,0479	0,0036	82.829,73
Moment Ratio	0,1807	0,0168	0,0329	17.487,07	Moment Ratio	0,2653	0,0100	0,0705	87.428,84
Peng	-0,0307	0,1157	0,0143	16.773,03	Peng	-0,0220	0,0548	0,0035	87.224,68
W	-0,0290	0,0989	0,0106	17.164,26	W	-0,0061	0,0445	0,0020	86.549,48
PWM	0,0133	0,1138	0,0131	16.867,59	PWM	0,0026	0,0514	0,0027	84.379,39

Από τα συνολικά αποτελέσματα μπορεί να διαπιστωθεί, ότι η μέθοδος επιλογής POT είχε καλύτερη απόδοση από την αντίστοιχη με Block Maxima καθ' ότι έχει παράξει εκτιμήσεις με την μικρότερη μεροληψία με τη μέση τιμή της κατανομής τους να είναι μόλις -0,001. Η μέθοδος των Block Maxima υστερεί ακόμα ενός εκτιμητή, αυτόν του Pickand's, ο οποίος φαίνεται να παράγει εκτιμήσεις με μικρότερη μεροληψία, αν και υπάρχει αρκετά μεγαλύτερη μεταβλητότητα στην κατανομή των εκτιμήσεών του και παρουσιάζει ένα από τα χειρότερα μέσο τετραγωνικό σφάλμα. Ωστόσο, για μεγαλύτερες τιμές του  $k$  το σφάλμα αυτό μειώνεται αρκετά και ταυτόχρονα εμφανίζεται ο εκτιμητής με τη μέθοδο των ροπών να έχει την ελάχιστη μεροληψία, και η τυπική απόκλιση του παραμένει αρκετά μεγαλύτερη.

Στην περίπτωση της Γάμμα (1, 1.5) αναμένεται, ομοίως η τιμή του δείκτη  $\xi$  να είναι, επίσης, μηδενική. Το βέλτιστο πλήθος Block για αυτήν την κατανομή υπολογίστηκε στα 150 Block περιέχοντας 34 παρατηρήσεις το καθένα.

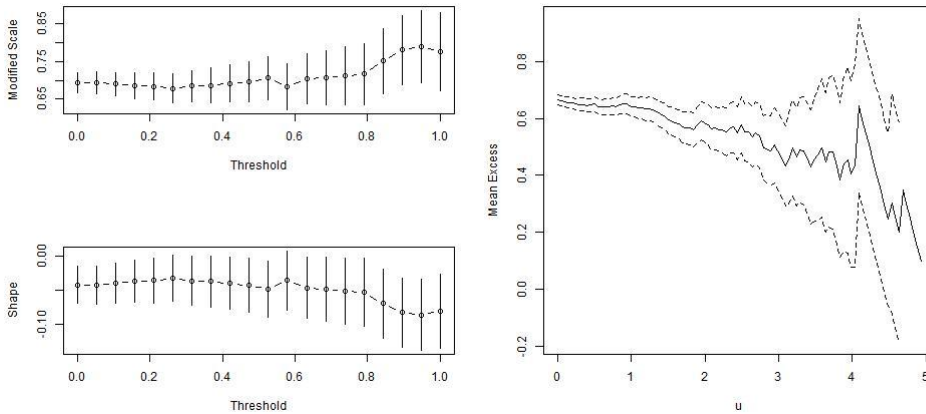
Όπως φαίνεται από τα ακόλουθα διαγράμματα ελέγχου καλής προσαρμογής τα 150 Block Maxima, που έχουν επιλεγεί μπορούν να θεωρηθούν, ότι προέρχονται από μία GEV με παραμέτρους εκτιμημένες ως  $(\hat{\mu}, \hat{\sigma}, \hat{\xi}) = (2.31229259, 0.53056498, -0.009845492)$ . Ταυτόχρονα, στο διάγραμμα στάθμης απόδοσης η  $z_p(w)$  αυξάνεται γραμμικά ως προς το  $w$ , επομένως ο δείκτης  $\xi = 0$ .

Εικόνα 4.4.4 Διαγράμματα καλής προσαρμογής στην GEV για την Γάμμα(1, 1.5)



Εξετάζοντας τα ακόλουθα γραφήματα για την εύρεση του βέλτιστου κατωφλιού  $\nu$  για την εφαρμογή της μεθόδου POT καταλήγουμε, στην χρήση του σημείου  $\nu = 0,5$ .

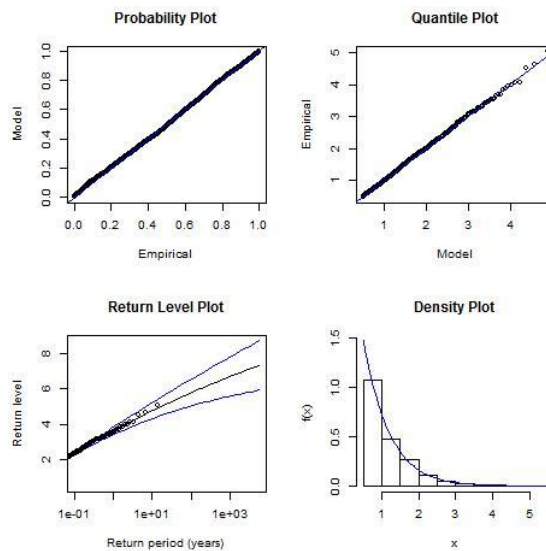
Εικόνα 4.4.5 Διαγράμματα για την επιλογή του βέλτιστου  $u$  στην περίπτωση της Γάμμα (1, 1.5)



Εφαρμόζοντας το βέλτιστο κατώφλι, που βρέθηκε, το ποσοστό των δεδομένων που ακολουθεί GPD με  $(\tilde{\sigma}, \tilde{\xi}) = (0.68208815, -0.04891637)$  είναι 47,5%. Επομένως, ο MLE που θα υπολογισθεί βάσει της μεθόδου Peak Over Threshold θα χρησιμοποιεί δείγματα με περίπου 2.400 παρατηρήσεις.

Από τα αποτελέσματα των γραφικών ελέγχων προσαρμογής του δείγματος στην GPD με τις αντίστοιχες παραμέτρους που εκτιμήθηκαν, βλέπουμε, ότι υπάρχει τέλεια ταύτιση των θεωρητικών και των παρατηρούμενων τιμών, καθώς και το διάγραμμα στάθμης απόδοσης έχει τη συνάρτηση  $z_p(w)$  να αυξάνεται γραμμικά ως προς το  $w$ , επομένως και σε αυτήν την περίπτωση ο δείκτης  $\xi = 0$ .

Εικόνα 4.4.6 Διαγράμματα καλής προσαρμογής στην GPD για την Γάμμα (1, 1.5)



Παρακάτω, φαίνονται συγκεντρωτικά οι εκτιμώμενες (μέσω προσομοίωσης  $10^5$  επαναλήψεων) παράμετροι των κατανομών των διάφορων εκτιμητών του  $\xi$  από κάθε εκτιμητή για  $k = 25, 50, 100, 500$ .

Πίνακας 4.4.2 Περιγραφικά χαρακτηριστικά της κατανομής των προσομοιωμένων εκτιμητών για την Γάμμα (1, 1.5)

<i>Γάμμα (1, 1.5)</i>				
<i>MLE</i>	<i>MEAN</i>	<i>SE</i>	<i>MSE</i>	<i>DURATION</i>
BM	0,0075	0,0608	0,0038	1.239,85
POT	-0,0013	0,0207	0,0004	2.418,56

	<i>25</i>					<i>50</i>			
	<i>MEAN</i>	<i>SE</i>	<i>MSE</i>	<i>DURATION</i>		<i>MEAN</i>	<i>SE</i>	<i>MSE</i>	<i>DURATION</i>
Pickand's	-0,0259	0,3619	0,1317	255,39	Pickand's	-0,0120	0,2562	0,0658	259,26
Hill's	0,1627	0,0292	0,0273	2.167,20	Hill's	0,1834	0,0230	0,0342	4.236,69
Negative Hill's	-0,3543	0,1427	0,1459	3.898,28	Negative Hill's	-0,3030	0,1027	0,1024	7.873,23
Moment	-0,0838	0,2320	0,0609	4.095,40	Moment	-0,0298	0,1522	0,0240	7.913,88
Moment Ratio	0,1405	0,0276	0,0205	4.382,95	Moment Ratio	0,1586	0,0215	0,0256	8.771,00
Peng	-0,1060	0,2491	0,0733	4.727,28	Peng	-0,0546	0,1663	0,0306	8.479,84
W	-0,1113	0,2141	0,0582	4.541,89	W	-0,0568	0,1424	0,0235	8.496,24
PWM	0,0531	0,2194	0,0510	4.224,39	PWM	0,0259	0,1596	0,0261	8.403,90

	<i>100</i>					<i>500</i>			
	<i>MEAN</i>	<i>SE</i>	<i>MSE</i>	<i>DURATION</i>		<i>MEAN</i>	<i>SE</i>	<i>MSE</i>	<i>DURATION</i>
Pickand's	-0,0069	0,1807	0,0327	258,16	Pickand's	-0,0012	0,0806	0,0065	254,83
Hill's	0,2106	0,0185	0,0447	8.447,04	Hill's	0,3241	0,0125	0,1052	41.362,84
Negative Hill's	-0,2590	0,0753	0,0728	15.286,61	Negative Hill's	-0,1862	0,0400	0,0363	79.719,40
Moment	-0,0009	0,1044	0,0109	15.818,80	Moment	0,0367	0,0479	0,0036	83.828,17
Moment Ratio	0,1807	0,0168	0,0329	17.677,00	Moment Ratio	0,2653	0,0100	0,0705	88.073,98
Peng	-0,0307	0,1157	0,0143	16.981,63	Peng	-0,0220	0,0548	0,0035	88.042,53
W	-0,0290	0,0989	0,0106	17.100,57	W	-0,0061	0,0445	0,0020	86.156,03
PWM	0,0133	0,1138	0,0131	16.880,77	PWM	0,0026	0,0514	0,0027	84.732,27

Από τα αποτελέσματα, παρατηρείται η ίδια συμπεριφορά με την προηγούμενη περίπτωση. Οι εκτιμήσεις με τις μεθόδους επιλογής δεδομένων POT φαίνονται να είναι οι πιο αποδοτικές, καθώς έχουν το μικρότερο μέσο τετραγωνικό σφάλμα και ελάχιστη μεροληψία. Η μέθοδος Block Maxima είναι μια καλή εναλλακτική καθώς έχει γενικά αρκετά καλά περιγραφικά στατιστικά, ωστόσο για μεγαλύτερες τιμές του  $k$  οι εκτιμητές των Pickand's, W και Ροτών Σταθμισμένης Πιθανότητας εμφανίζονται να έχουν καλύτερη από αυτόν πρόβλεψη.

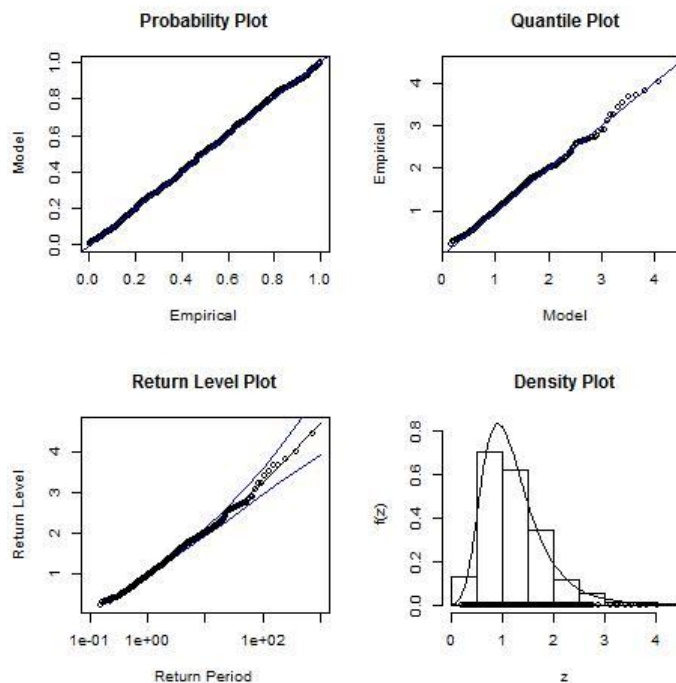
#### 4.5 ΚΑΤΑΝΟΜΗ WEIBULL

Όπως και με την κατανομή Γάμμα δοκιμάστηκαν δύο περιπτώσεις της κατανομής Weibull, η οποία δεν έχει πεπερασμένο μέγιστο στο στήριγμά της και συνάρτηση πυκνότητας πιθανότητας ανάλογη του εκθετικού  $e^{-(x/\lambda)^r}$ . Επομένως βάσει θεωρίας θα

έχει μηδενική τιμή για το δείκτη  $\xi$ . Οι περιπτώσεις της κατανομής, που χρησιμοποιήθηκαν, είχαν παράμετρο λόγου  $\lambda = 1$  και σχήματος  $\tau = 0.5$  και  $\tau = 1.5$  αντίστοιχα.

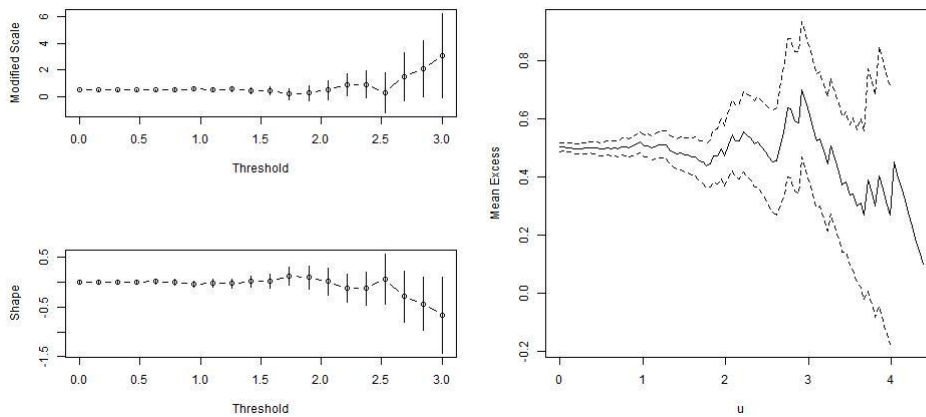
Σχετικά με την κατανομή Weibull (1, 0.5) ο βέλτιστος αριθμός block για τα δεδομένα ήταν 750 με το καθένα να περιέχει 7 παρατηρήσεις. Οι παράμετροι της GEV, που προσαρμόστηκε ήταν  $(\hat{\mu}, \hat{\sigma}, \hat{\xi}) = (0.94575881, 0.44230109, 0.05878767)$ , ενώ η προσαρμογή της κατανομής στα δεδομένα αντίστοιχα, από τα ακόλουθα διαγράμματα ελέγχου, τα οποία μας δείχνουν την καλή προσαρμογή της κατανομής GEV στις 750 παρατηρήσεις που επιλέχθηκαν με τη μέθοδο. Αντίστοιχα, το γράφημα στάθμης απόδοσης απεικονίζει τη  $z_p(w)$  αυξάνεται γραμμικά ως προς το  $w$ , άρα επιβεβαιώνεται, ότι η παράμετρος  $\xi = 0$ .

Εικόνα 4.5.1 Διαγράμματα καλής προσαρμογής στην GEV για την Weibull (1, 0.5)



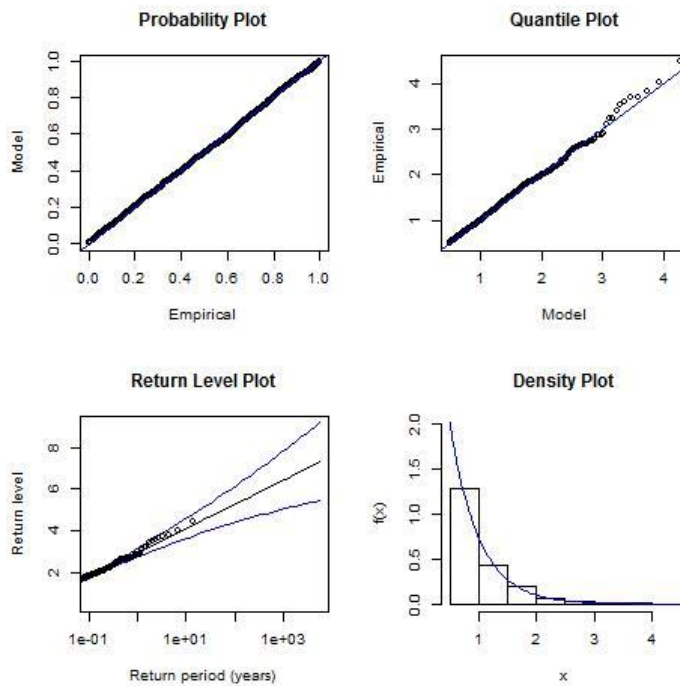
Στην εφαρμογή της μεθόδου POT αρχικά παίρνουμε τα ακόλουθα σχήματα για την επιλογή του βέλτιστου κατωφλιού  $u$  για την επιλογή των δεδομένων.

Εικόνα 4.5.2 Διαγράμματα για την επιλογή του βέλτιστου  $u$  στην περίπτωση της Weibull (1, 0.5)



Από το σχήμα των γραφημάτων είναι εμφανές, ότι το κατώφλι θα πρέπει να είναι μια μικρή τιμή αρκετά κοντά στο μηδέν. Στην εφαρμογή της μεθόδου επιλέχθηκε κατώφλι  $u = 0.5$  και οι αντίστοιχες εκτιμήσεις για τις παραμέτρους της GPD ήταν  $(\hat{\sigma}, \hat{\xi}) = (0.498152083, 0.001795324)$ . Αυτές οι εκτιμήσεις βασίστηκαν στο 37,12% του αρχικού δείγματος και, επομένως, τα ακόλουθα δείγματα στα οποία θα βασιστούν οι υπόλοιπες εκτιμήσεις θα αποτελούνται από 1.800 περίπου παρατηρήσεις. Η προσαρμογή των επιλεγμένων δεδομένων από το πρώτο δείγμα φαίνονται στα ακόλουθα διαγράμματα ελέγχου καλής προσαρμογής.

Εικόνα 4.5.3 Διαγράμματα καλής προσαρμογής στην GPD για την Weibull (1, 0.5)



Τα διαγράμματα PP και QQ δείχνουν τέλεια ταύτιση των θεωρητικών και παρατηρούμενων ποσοστημορίων και εκατοστημορίων αντίστοιχα, ενώ από το

ιστόγραμμα φαίνεται η προσαρμογή της συνάρτησης πυκνότητας πιθανότητας. Από το διάγραμμα στάθμης απόδοσης βλέπουμε, ότι η  $z_p(w)$  αυξάνεται γραμμικά ως προς το  $w$ , επομένως, συμπεραίνουμε, ότι ο δείκτης  $\xi$  είναι μηδενικός.

Στη συνέχεια, παρατίθενται συγκεντρωτικά τα βασικά περιγραφικά στατιστικά της κατανομής των 100.000 αποτελεσμάτων από κάθε εκτιμητή για κάθε  $k = 25, 50, 100, 500$ .

Πίνακας 4.5.1 Περιγραφικά χαρακτηριστικά της κατανομής των προσομοιωμένων εκτιμητών για την Weibull (1, 0.5)

<i>Weibull (1, 0.5)</i>				
<i>MLE</i>	<i>MEAN</i>	<i>SE</i>	<i>MSE</i>	<i>DURATION</i>
BM	0,0614	0,0266	0,0045	5.955,80
POT	-0,0016	0,0236	0,0006	3.153,26

	<i>25</i>					<i>50</i>			
	<i>MEAN</i>	<i>SE</i>	<i>MSE</i>	<i>DURATION</i>		<i>MEAN</i>	<i>SE</i>	<i>MSE</i>	<i>DURATION</i>
Pickand's	-0,0223	0,3625	0,1319	213,29	Pickand's	-0,0131	0,2550	0,0652	226,75
Hill's	0,1628	0,0291	0,0273	2.323,47	Hill's	0,1836	0,0230	0,0342	4.420,95
Negative Hill's	-0,3546	0,1427	0,1461	4.214,96	Negative Hill's	-0,3032	0,1025	0,1024	8.510,70
Moment	-0,0832	0,2312	0,0604	4.652,21	Moment	-0,0295	0,1514	0,0238	8.344,50
Moment Ratio	0,1406	0,0276	0,0205	4.363,48	Moment Ratio	0,1588	0,0215	0,0257	8.608,35
Peng	-0,1054	0,2483	0,0728	4.316,06	Peng	-0,0543	0,1656	0,0304	8.545,40
W	-0,1108	0,2133	0,0578	3.855,11	W	-0,0566	0,1418	0,0233	7.118,99
PWM	0,0536	0,2193	0,0510	3.556,15	PWM	0,0263	0,1588	0,0259	7.067,19

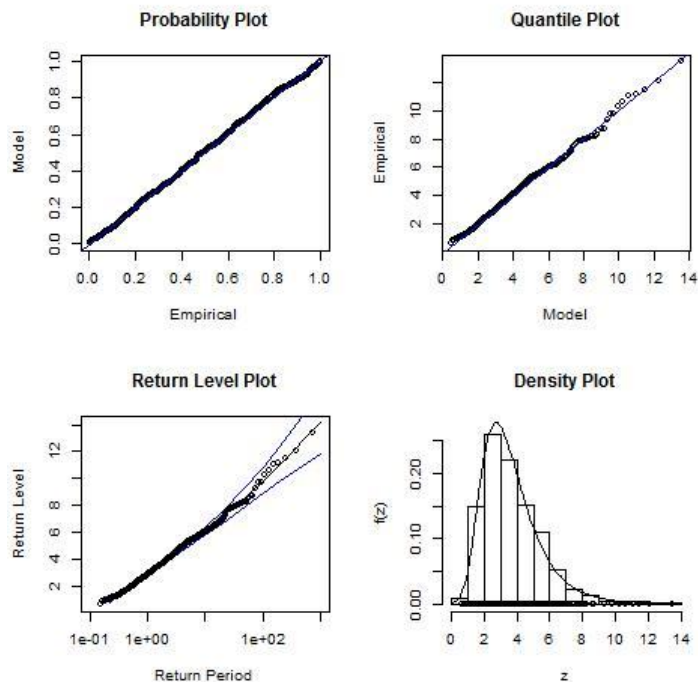
	<i>100</i>					<i>500</i>			
	<i>MEAN</i>	<i>SE</i>	<i>MSE</i>	<i>DURATION</i>		<i>MEAN</i>	<i>SE</i>	<i>MSE</i>	<i>DURATION</i>
Pickand's	-0,0072	0,1806	0,0327	224,39	Pickand's	-0,0007	0,0812	0,0066	225,81
Hill's	0,2106	0,0185	0,0447	8.966,42	Hill's	0,3240	0,0125	0,1052	44.467,27
Negative Hill's	-0,2589	0,0751	0,0727	17.496,28	Negative Hill's	-0,1861	0,0398	0,0362	87.745,77
Moment	-0,0005	0,1047	0,0110	16.908,97	Moment	0,0369	0,0478	0,0036	89.423,34
Moment Ratio	0,1808	0,0168	0,0330	17.104,69	Moment Ratio	0,2653	0,0099	0,0705	87.025,39
Peng	-0,0303	0,1161	0,0144	17.118,69	Peng	-0,0219	0,0546	0,0035	83.715,63
W	-0,0286	0,0992	0,0106	14.054,10	W	-0,0059	0,0443	0,0020	70.072,98
PWM	0,0136	0,1141	0,0132	14.076,36	PWM	0,0027	0,0513	0,0026	72.098,62

Σε αυτήν την περίπτωση, η μέθοδος POT έχει αποδώσει την μικρότερη μεροληψία, ενώ, ταυτόχρονα, έχει και το μικρότερο μέσο τετραγωνικό σφάλμα από όλους τους υπόλοιπους εκτιμητές. Η μέθοδος των Block Maxima εμφάνισε σχετικά μεγαλύτερη μεροληψία, αλλά δεν υπήρχε κάποια άλλη σχετική παρατήρηση, καθώς το μέσο τετραγωνικό σφάλμα υπολογίσθηκε σε σύνηθες ελάχιστο επίπεδο. Αρκετά μικρή

μεροληψία έδειξαν και οι εκτιμητές του Pickand's και με τη μέθοδο των Ροπών. Ο εκτιμητής του Pickand's είχε καλύτερη απόδοση για μεγαλύτερα  $k$ , ενώ αντιθέτως αυτός των Ροπών είχε την καλύτερη απόδοση για περίπου  $k = 100$ .

Στη δεύτερη περίπτωση χρησιμοποιήθηκε η κατανομή Weibull (1, 1.5) ο βέλτιστος αριθμός block για την εφαρμογή της μεθόδου Block Maxima ήταν 750, όπως και στην προηγούμενη περίπτωση. Αντίστοιχα οι παράμετροι της προσαρμοσμένης GEV κατανομής ήταν  $(\hat{\mu}, \hat{\sigma}, \hat{\xi}) = (2.83706260, 1.32680955, 0.05868797)$ . Η GEV είχε ικανοποιητική προσαρμογή στις παρατηρήσεις αυτές, με τα διαγράμματα PP και QQ να μην εμφανίζουν σημαντικές αποκλίσεις από τις θεωρητικές τιμές, ενώ ταυτόχρονα το ιστόγραμμα των παρατηρήσεων δείχνει αυτή τη σχέση ανάμεσα στη θεωρητική συνάρτηση πυκνότητας και τα δεδομένα. Από το διάγραμμα στάθμης απόδοσης βλέπουμε, ότι η  $z_p(w)$  αυξάνεται γραμμικά ως προς το  $w$ , επομένως, συμπεραίνουμε, ότι ο δείκτης  $\xi$  είναι μηδενικός.

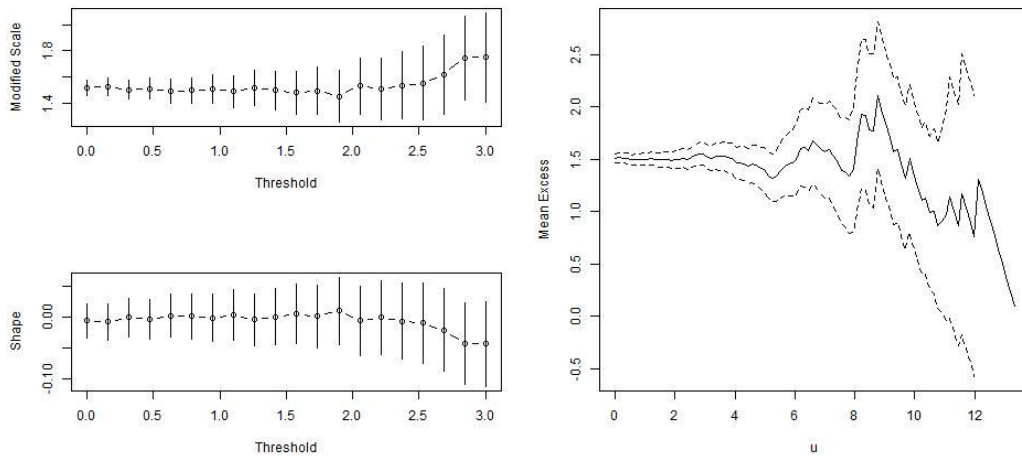
Εικόνα 4.5.4 Διαγράμματα καλής προσαρμογής στην GEV για την Weibull (1, 1.5)



Για την εφαρμογή της μεθόδου POT χρησιμοποιήθηκαν τα ακόλουθα διαγράμματα απ' όπου είναι εμφανές, ότι το βέλτιστο κατώφλι  $u$  βρίσκεται περίπου στην περιοχή τιμών 0,5 και 1. Για την περίπτωση που ελέγχθηκε, επιλέχθηκε η τιμή  $u = 1$  ως βέλτιστο κατώφλι.

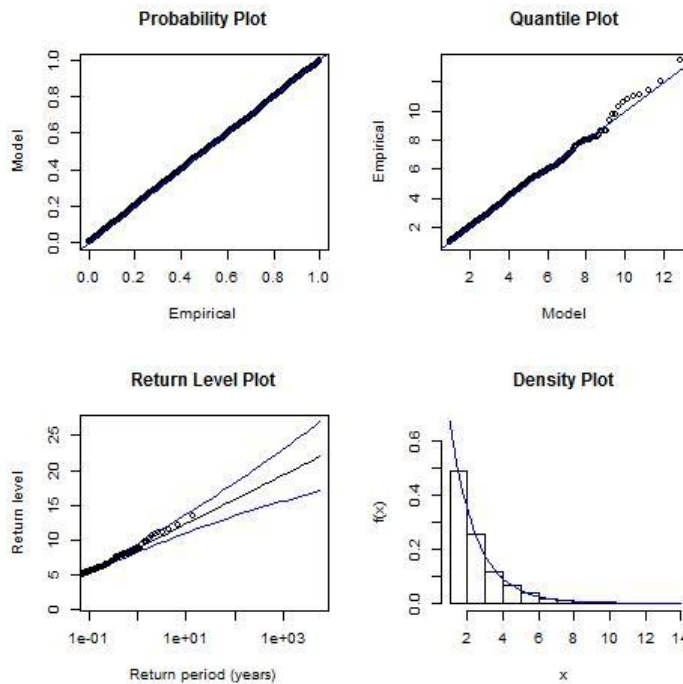


Εικόνα 4.5.5 Διαγράμματα για την επιλογή του βέλτιστου  $u$  στην περίπτωση της Weibull (1, 1.5)



Οι παράμετροι της προσαρμοσμένης κατανομής GPD σε αυτήν την περίπτωση υπολογίστηκαν βάσει του αρχικού δείγματος ως  $(\hat{\sigma}, \hat{\xi}) = (1.493968915, 0.002001766)$ . Οι υπολογισμοί βάσει του κατωφλιού, που χρησιμοποιήθηκε ήταν το 51,82% των δεδομένων και, επομένως, οι εκτιμήσεις του MLE βασίζονται σε περίπου 2.600 παρατηρήσεις κάθε φορά. Η προσαρμογή της κατανομής GPD στα δεδομένα κατά αντιστοιχία ήταν ικανοποιητική όπως φαίνεται από τα παρακάτω διαγράμματα, όπως και στην προηγούμενη περίπτωση.

Εικόνα 4.5.6 Διαγράμματα καλής προσαρμογής στην GPD για την Weibull (1, 1.5)



Παρακάτω, φαίνονται συγκεντρωτικά οι εκτιμώμενες (μέσω προσομοίωσης  $10^5$  επαναλήψεων) παράμετροι των κατανομών των διάφορων εκτιμητών του  $\xi$  από κάθε εκτιμητή για  $k = 25, 50, 100, 500$ .

Πίνακας 4.5.2 Περιγραφικά χαρακτηριστικά της κατανομής των προσομοιωμένων εκτιμητών για την Weibull (1, 1.5)

<i>Weibull (1, 1.5)</i>				
<i>MLE</i>	<i>MEAN</i>	<i>SE</i>	<i>MSE</i>	<i>DURATION</i>
BM	0,0614	0,0266	0,0045	3.116,14
POT	-0,0011	0,0199	0,0004	2.172,24

	<i>25</i>					<i>50</i>			
	<i>MEAN</i>	<i>SE</i>	<i>MSE</i>	<i>DURATION</i>		<i>MEAN</i>	<i>SE</i>	<i>MSE</i>	<i>DURATION</i>
Pickand's	-0,0223	0,3625	0,1319	223,03	Pickand's	-0,0131	0,2550	0,0652	225,31
Hill's	0,1628	0,0291	0,0273	2.308,35	Hill's	0,1836	0,0230	0,0342	4.402,60
Negative Hill's	-0,3546	0,1427	0,1461	4.186,79	Negative Hill's	-0,3032	0,1025	0,1024	8.467,97
Moment	-0,0832	0,2312	0,0604	4.631,17	Moment	-0,0295	0,1514	0,0238	8.404,18
Moment Ratio	0,1406	0,0276	0,0205	4.392,58	Moment Ratio	0,1588	0,0215	0,0257	8.650,43
Peng	-0,1054	0,2483	0,0728	4.296,44	Peng	-0,0543	0,1656	0,0304	8.527,49
W	-0,1108	0,2133	0,0578	3.866,60	W	-0,0566	0,1418	0,0233	7.080,37
PWM	0,0536	0,2193	0,0510	3.546,37	PWM	0,0263	0,1588	0,0259	7.053,66

	<i>100</i>					<i>500</i>			
	<i>MEAN</i>	<i>SE</i>	<i>MSE</i>	<i>DURATION</i>		<i>MEAN</i>	<i>SE</i>	<i>MSE</i>	<i>DURATION</i>
Pickand's	-0,0072	0,1806	0,0327	222,87	Pickand's	-0,0007	0,0812	0,0066	226,29
Hill's	0,2106	0,0185	0,0447	8.891,24	Hill's	0,3240	0,0125	0,1052	44.276,13
Negative Hill's	-0,2589	0,0751	0,0727	17.403,50	Negative Hill's	-0,1861	0,0398	0,0362	87.116,09
Moment	-0,0005	0,1047	0,0110	16.771,15	Moment	0,0369	0,0478	0,0036	88.889,26
Moment Ratio	0,1808	0,0168	0,0330	17.161,22	Moment Ratio	0,2653	0,0099	0,0705	87.066,53
Peng	-0,0303	0,1161	0,0144	17.015,50	Peng	-0,0219	0,0546	0,0035	83.225,12
W	-0,0286	0,0992	0,0106	14.007,80	W	-0,0059	0,0443	0,0020	69.864,59
PWM	0,0136	0,1141	0,0132	14.037,09	PWM	0,0027	0,0513	0,0026	71.928,46

Όπως και στην προηγούμενη περίπτωση και εδώ ο καλύτερος εκτιμητής ήταν ο μεγίστης πιθανοφάνειας μετά από εφαρμογή της μεθόδου POT. Η μόνη περίπτωση, όπου κάποιος άλλος εκτιμητής εμφανίστηκε με μικρότερη μεροληψία ήταν αυτός του Pickand's και μόνο για μεγάλη τιμή του  $k$ , όπως και αυτός των ροπών, αλλά μόνο για  $k = 100$  καθώς περαιτέρω αύξηση προσέθετε μεροληψία. Σχετικά στα ίδια επίπεδα ήταν και οι εκτιμητές W και Ροπών Σταθμισμένης Πιθανότητας σε αντίθεση με τον εκτιμητή Hill's και Negative Hill's που σε κάθε περίπτωση ήταν οι χειρότεροι.

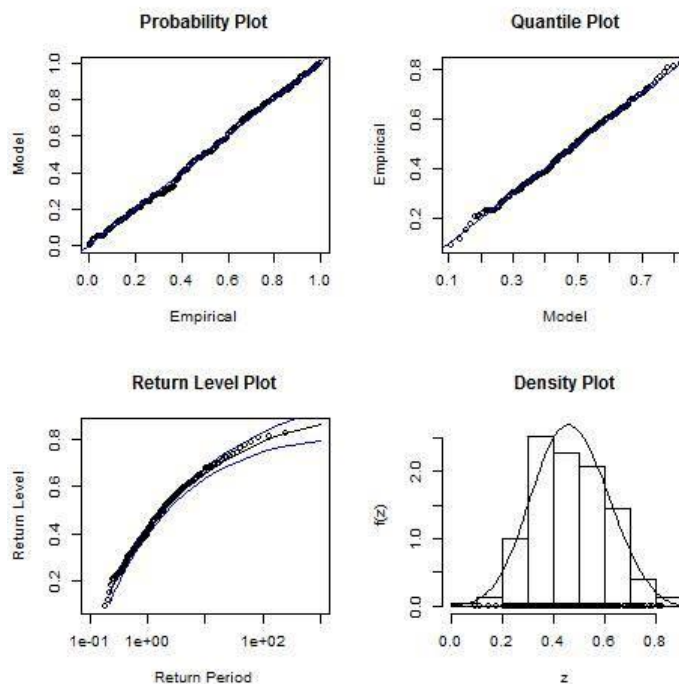
## 4.6 ΚΑΤΑΝΟΜΗ ΒΗΤΑ

Στην κατανομή Beta θα εξετασθούν 2 περιπτώσεις, με παραμέτρους  $(\alpha, \beta) = (0.5, 4), (5, 1)$ .

Αρχικά, για την κατανομή Beta με παραμέτρους  $(\alpha, \beta) = (0.5, 4)$  (που αντιστοιχεί σε  $\xi = -0.25$ ) ελέγχουμε την προσαρμογή των δεδομένων που παρήχθησαν με την μέθοδο επιλογής Block Maxima. Αφού παρήχθη δείγμα μεγέθους  $n = 5.000$ , και με κατάλληλες δοκιμές, καταλήξαμε ότι ο βέλτιστος αριθμός Blocks για αυτήν είναι 250, με 20 παρατηρήσεις το καθένα. Βάσει αυτών των παρατηρήσεων και με χρήση της μεθόδου μεγίστης πιθανοφάνειας εκτιμούμε τις παραμέτρους της GEV, την οποία θα πρέπει ακολουθούν προσεγγιστικά. Έτσι, καταλήγουμε στο συμπέρασμα ότι η κατανομή των δεδομένων θα είναι η GEV με  $(\hat{\mu}, \hat{\sigma}, \hat{\xi}) = (0.4154798, 0.1423271, -0.2694280)$ .

Εν συνεχεία, παρατίθενται τα διαγράμματα καλής προσαρμογής των δεδομένων στην GEV.

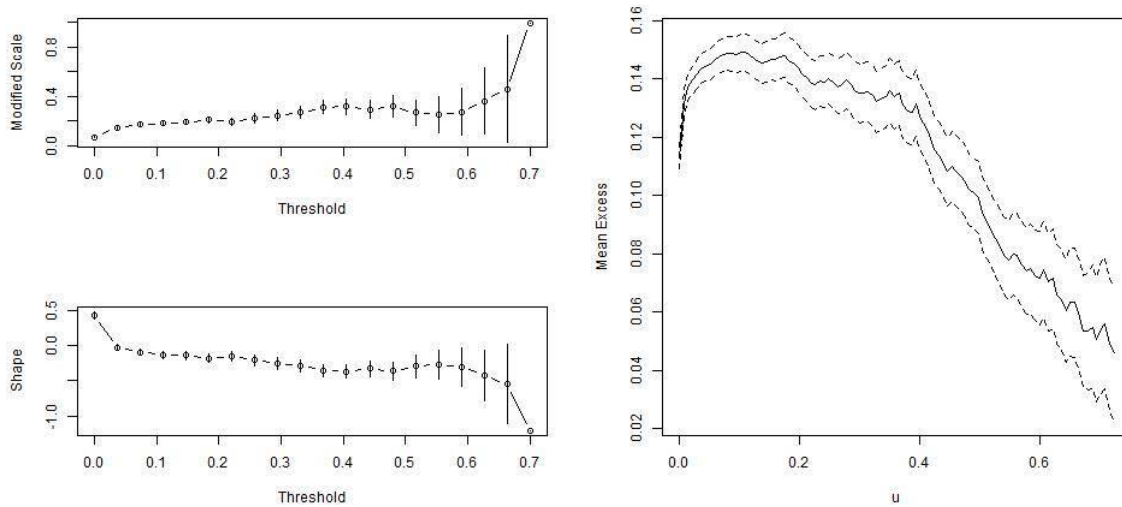
Εικόνα 4.6.1 Διαγράμματα καλής προσαρμογής στην GEV για την Βήτα (0.5, 4)



Από τα PP-plot και QQ-plot διαφαίνεται ότι τα θεωρητικά και τα παρατηρούμενα σημεία των κατανομών συμπίπτουν, όπως είναι αναμενόμενο όταν η παρατηρούμενη κατανομή είναι η GEV, γεγονός που φαίνεται και από το ιστόγραμμα των Block Maxima (Density Plot). Αντίστοιχα, από το γράφημα στάθμης απόδοσης (Return Level Plot) βλέπουμε ότι η  $z_p(w)$  είναι κοίλη συνάρτηση του  $w$ , άρα  $\xi < 0$ .

Για την εφαρμογή της μεθόδου Peak Over Threshold η επιλογή του βέλτιστου κατωφλιού  $u$  θα γίνει από τα παρακάτω γραφήματα.

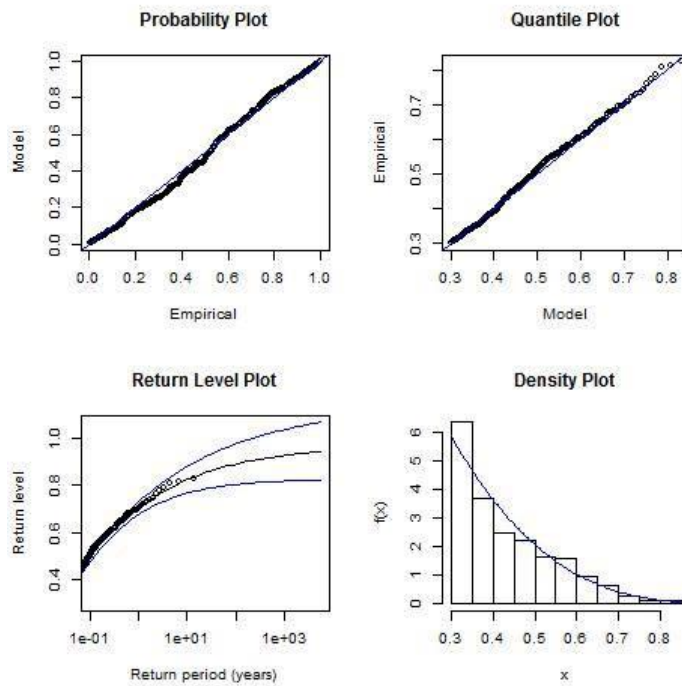
Εικόνα 4.6.2 Διαγράμματα για την επιλογή του βέλτιστου  $u$  στην περίπτωση της Βήτα (0.5, 4)



Με χρήση των δύο διαγραμμάτων για το threshold, καταλήγουμε στο συμπέρασμα ότι το μικρότερο  $u$  πέραν του οποίου το  $\xi$  σταθεροποιείται και η εκτίμηση του  $\hat{\sigma}$  μεταβάλλεται γραμμικά ως προς το  $u$  είναι το 0,3. Αντίστοιχα συμπεράσματα παίρνουμε και από το Mean Excess Plot, όπου παρατηρούμε ότι η μονάδα είναι η τιμή όπου το διάγραμμα γίνεται οριζόντιο και τα δ.ε δεν ξεφεύγουν. Συνεπώς, θα χρησιμοποιηθούν για τον υπολογισμό των εκτιμητών τα δεδομένα από την αρχική κατανομή Βήτα, τα οποία είναι μεγαλύτερα της τιμής αυτής. Με την εφαρμογή του  $u$  για το πρώτο δείγμα, το ποσοστό των δεδομένων που ακολουθεί GPD με  $(\hat{\sigma}, \hat{\xi}) = (0.1695769, -0.2520536)$  είναι 10,24 %. Επομένως, ο MLE που θα υπολογισθεί βάσει της μεθόδου Peak Over Threshold θα χρησιμοποιεί δείγματα με περίπου 1.020 παρατηρήσεις.

Στη συνέχεια, με τον ίδιο γραφικό τρόπο παρατηρούμε ότι τα θεωρητικά και τα παρατηρούμενα σημεία των κατανομών συμπίπτουν, όπως είναι αναμενόμενο όταν η παρατηρούμενη κατανομή είναι η GPD, γεγονός που φαίνεται και από το ιστόγραμμα των Peaks Over Threshold (Density Plot). Αντίστοιχα, από το γράφημα στάθμης απόδοσης (Return Level Plot) βλέπουμε ότι η  $z_p(w)$  είναι κοίλη συνάρτηση του  $w$ , άρα  $\xi < 0$ .

Εικόνα 4.6.3 Διαγράμματα καλής προσαρμογής στην GPD για την Βήτα (0.5 , 4)



Παρακάτω, φαίνονται συγκεντρωτικά οι εκτιμώμενες (μέσω προσομοίωσης  $10^5$  επαναλήψεων) παράμετροι των κατανομών των διάφορων εκτιμητών του  $\xi$  από κάθε εκτιμητή για  $k = 25, 50, 100, 500$ .

Πίνακας 4.6.1 Περιγραφικά χαρακτηριστικά της κατανομής των προσομοιωμένων εκτιμητών για την Βήτα(0.5 , 4)

**Beta (0.5, 4) ( $\xi = -0.25$ )**

<i>MLE</i>	<i>MEAN</i>	<i>SE</i>	<i>MSE</i>	<i>DURATION</i>
BM	-0,1774	0,0403	0,0069	2.359,00
POT	-0,2049	0,0384	0,0035	1.204,64

	<b>25</b>					<b>50</b>			
	<i>MEAN</i>	<i>SE</i>	<i>MSE</i>	<i>DURATION</i>		<i>MEAN</i>	<i>SE</i>	<i>MSE</i>	<i>DURATION</i>
Pickand's	-0,2361	0,3556	0,1267	584,85	Pickand's	-0,2090	0,2509	0,0646	584,20
Hill's	0,0983	0,0167	0,1216	3.707,62	Hill's	0,1239	0,0149	0,1400	7.095,35
Negative Hill's	-0,4648	0,1447	0,0671	7.286,45	Negative Hill's	-0,4221	0,1020	0,0400	14.401,16
Moment	-0,3113	0,2605	0,0716	7.212,77	Moment	-0,2540	0,1694	0,0287	13.986,73
Moment Ratio	0,0778	0,0122	0,1076	4.499,08	Moment Ratio	0,0983	0,0108	0,1214	8.730,53
Peng	-0,3318	0,2692	0,0792	4.516,57	Peng	-0,2796	0,1773	0,0323	8.769,14
W	-0,3213	0,2444	0,0648	5.590,48	W	-0,2684	0,1568	0,0249	11.009,61
PWM	-0,1399	0,2386	0,0691	5.582,62	PWM	-0,1784	0,1735	0,0352	10.997,04

	100					500			
	MEAN	SE	MSE	DURATION		MEAN	SE	MSE	DURATION
Pickand's	-0,1819	0,1779	0,0363	584,41	Pickand's	-0,0070	0,0809	0,0656	584,00
Hill's	0,1594	0,0136	0,1678	14.104,86	Hill's	0,3223	0,0126	0,3277	70.250,18
Negative Hill's	-0,3831	0,0726	0,0230	27.941,29	Negative Hill's	-0,3094	0,0348	0,0047	118.794,16
Moment	-0,2207	0,1169	0,0145	27.859,86	Moment	-0,1458	0,0539	0,0138	139.524,02
Moment Ratio	0,1256	0,0096	0,1412	17.411,59	Moment Ratio	0,2446	0,0082	0,2447	88.151,70
Peng	-0,2545	0,1241	0,0154	17.517,16	Peng	-0,2235	0,0599	0,0043	88.441,65
W	-0,2404	0,1059	0,0113	21.818,27	W	-0,1925	0,0440	0,0052	114.989,99
PWM	-0,1946	0,1246	0,0186	21.800,59	PWM	-0,1778	0,0547	0,0082	114.949,05

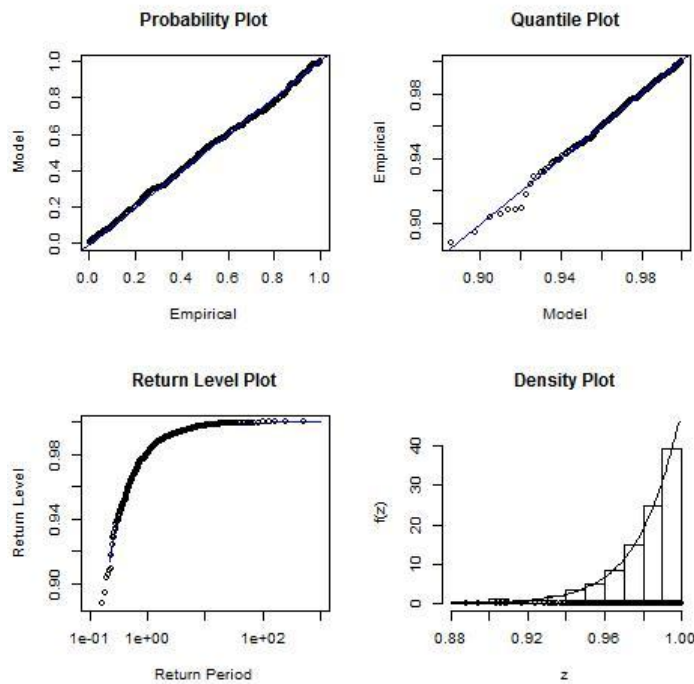
Για την Beta (0.5, 4) και στην περίπτωση, όπου εκτιμούμε τον δείκτη ακραίων τιμών κοντά στο  $-0,25$ , αναμένουμε βασιζόμενοι στην θεωρία να αποδίδουν καλύτερα οι εκτιμητές που προτείνονται για  $\xi \in \mathbb{R}$ , καθώς δεν υπάρχει κάποιος που προβλέπεται να δίνει καλύτερες εκτιμήσεις κοντά στην συγκεκριμένη τιμή. Από τις τιμές που παρουσιάζονται στον παραπάνω πίνακα παρατηρούμε ότι ο Peng έχει πολύ μικρή διάρκεια υπολογισμού και καθώς το  $k$  αυξάνεται τείνει να βελτιώνεται, σε σημείο που φτάνει να δίνει πάρα πολύ καλές εκτιμήσεις, με πολύ μικρό MSE. Πολύ καλές εκτιμήσεις δίνει, επίσης, και ο εκτιμητής Moment, για μικρές τιμές του  $k$ , απαιτώντας όμως περισσότερο χρόνο για τον υπολογισμό του. Επίσης, ο εκτιμητής W για  $k = 50$  &  $100$  δίνει και αυτό μικρό MSE και μεροληψία, ενώ οι εκτιμητές Pickand's και PWM τείνουν να υπερεκτιμούν τον δείκτη.

Στην παραπάνω περίπτωση, παρατηρούμε ότι μία εκτίμηση με χρήση της μεθόδου Μέγιστης Πιθανοφάνειας με δεδομένα τα οποία αποκομίστηκαν με χρήση τόσο της POT όσο και της BM, δίνει εκτιμήσεις με πολύ μικρό MSE, ωστόσο και οι δύο παρουσιάζουν πολύ μεγάλη μεροληψία.

Ομοίως, θα γίνει έλεγχος της προσαρμογής των δεδομένων που παρήχθησαν με την κατανομή Beta με παραμέτρους  $(\alpha, \beta) = (5, 1)$  (που αντιστοιχεί σε  $\xi = -1$ ) για την μέθοδο επιλογής Block Maxima. Αφού παρήχθη δείγμα μεγέθους  $n = 5.000$ , και με κατάλληλες δοκιμές, καταλήξαμε ότι ο βέλτιστος αριθμός Blocks για αυτήν είναι 500, με 10 παρατηρήσεις το καθένα. Βάσει αυτών των παρατηρήσεων και με χρήση της μεθόδου μέγιστης πιθανοφάνειας εκτιμούμε τις παραμέτρους της GEV, την οποία θα πρέπει ακολουθούν προσεγγιστικά. Έτσι, καταλήγουμε στο συμπέρασμα ότι η κατανομή των δεδομένων θα είναι η GEV με  $(\hat{\mu}, \hat{\sigma}, \hat{\xi}) = (0.98054100 \ 0.01887995 \ -0.97040015)$ .

Εν συνεχεία, παρατίθενται τα διαγράμματα καλής προσαρμογής των δεδομένων στην GEV.

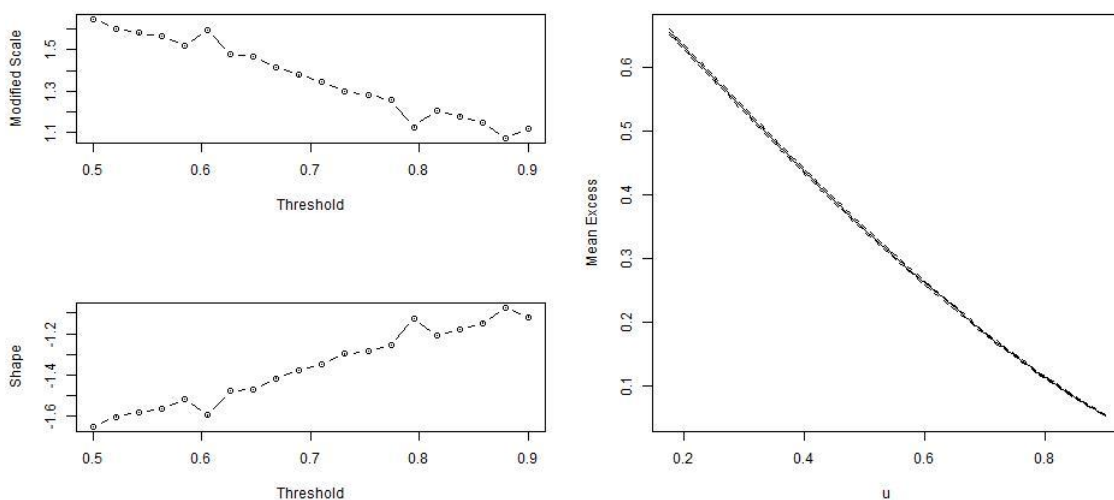
Εικόνα 4.6.4 Διαγράμματα καλής προσαρμογής στην GEV για την Βήτα (5, 1)



Από τα PP-plot και QQ-plot διαφαίνεται, ότι τα θεωρητικά και τα παρατηρούμενα σημεία των κατανομών συμπίπτουν, όπως είναι αναμενόμενο όταν η παρατηρούμενη κατανομή είναι η GEV, γεγονός που φαίνεται και από το ιστόγραμμα των Block Maxima (Density Plot). Αντίστοιχα, από το γράφημα στάθμης απόδοσης (Return Level Plot) βλέπουμε ότι η  $z_p(w)$  είναι κοίλη συνάρτηση του  $w$ , άρα  $\xi < 0$ .

Για την εφαρμογή της μεθόδου Peak Over Threshold η επιλογή του βέλτιστου κατωφλιού  $u$  θα γίνει από τα παρακάτω γραφήματα.

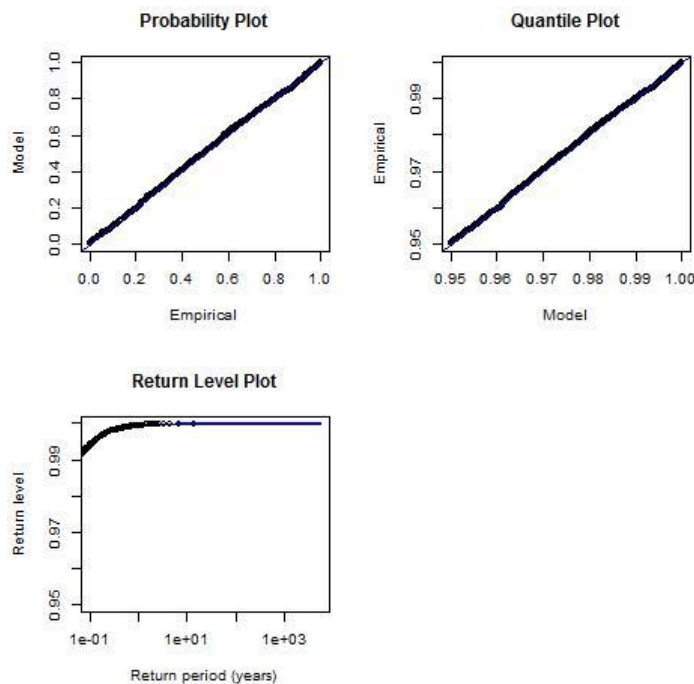
Εικόνα 4.6.5 Διαγράμματα για την επιλογή του βέλτιστου  $u$  στην περίπτωση της Βήτα (5, 1)



Με χρήση των δύο διαγραμμάτων για το threshold, καταλήγουμε στο συμπέρασμα ότι το μικρότερο  $u$  πέραν του οποίου το  $\xi$  σταθεροποιείται και η εκτίμηση του  $\hat{\sigma}$  μεταβάλλεται γραμμικά ως προς το  $u$  είναι το 0,95. Αντίστοιχα συμπεράσματα παίρνουμε και από το Mean Excess Plot, όπου παρατηρούμε ότι η μονάδα είναι η τιμή όπου το διάγραμμα γίνεται οριζόντιο και τα δ.ε δεν ξεφεύγουν. Συνεπώς, θα χρησιμοποιηθούν για τον υπολογισμό των εκτιμητών τα δεδομένα από την αρχική κατανομή Βήτα, τα οποία είναι μεγαλύτερα της τιμής αυτής. Με την εφαρμογή του  $u$  για το πρώτο δείγμα, το ποσοστό των δεδομένων που ακολουθεί GPD με  $(\hat{\sigma}, \hat{\xi}) = (0.05111134, -1.02232122)$  είναι 23,20 %. Επομένως, ο MLE που θα υπολογισθεί βάσει της μεθόδου Peak Over Threshold θα χρησιμοποιεί δείγματα με περίπου 1.160 παρατηρήσεις.

Στη συνέχεια, με τον ίδιο γραφικό τρόπο παρατηρούμε ότι τα θεωρητικά και τα παρατηρούμενα σημεία των κατανομών συμπίπτουν, όπως είναι αναμενόμενο όταν η παρατηρούμενη κατανομή είναι η GPD, γεγονός που φαίνεται και από τα διαγράμματα PP και QQ των Peaks Over Threshold<sup>3</sup>. Το διάγραμμα στάθμης απόδοσης εμφανίζει κοίλη καμπύλη επομένως μπορούμε να διαπιστώσουμε ότι η παράμετρος  $\xi$  είναι αρνητική ( $\xi < 0$ ).

Εικόνα 4.6.6 Διαγράμματα καλής προσαρμογής στην GPD για την Βήτα (5, 1)



<sup>3</sup> Η R είχε αδυναμία να παράξει το Density Plot (Error in plot.window(xlim, ylim, "", ...): need finite 'ylim' values)



Παρακάτω, φαίνονται συγκεντρωτικά οι εκτιμώμενες (μέσω προσομοίωσης  $10^5$  επαναλήψεων) παράμετροι των κατανομών των διάφορων εκτιμητών του  $\xi$  από κάθε εκτιμητή για  $k = 25, 50, 100, 500$ .

Πίνακας 4.6.2 Περιγραφικά χαρακτηριστικά της κατανομής των προσομοιωμένων εκτιμητών για την Βήτα (5, 1)

<i>Bήτα (5, 1) (<math>\xi = -1</math>)</i>									
<i>MLE</i>	<i>MEAN</i>	<i>SE</i>	<i>MSE</i>	<i>DURATION</i>		<i>MEAN</i>	<i>SE</i>	<i>MSE</i>	<i>DURATION</i>
BM	-0,9295	0,3217	0,1085			7.213,96			
POT	-1,06178	0,0360	0,0051			7.576,12			

<i>25</i>					<i>50</i>				
	<i>MEAN</i>	<i>SE</i>	<i>MSE</i>	<i>DURATION</i>		<i>MEAN</i>	<i>SE</i>	<i>MSE</i>	<i>DURATION</i>
Pickand's	-1,0228	0,3571	0,1280	299,00	Pickand's	-1,0246	0,2512	0,0637	300,69
Hill's	0,0005	0,0001	1,0010	2.139,71	Hill's	0,0010	0,0002	1,0021	4.087,71
Negative Hill's	-0,9619	0,1960	0,0399	3.846,37	Negative Hill's	-0,9822	0,1399	0,0199	7.733,45
Moment	-1,1518	0,5334	0,3076	4.320,72	Moment	-1,0717	0,3400	0,1208	8.580,71
Moment Ratio	0,0003	0,0001	1,0007	4.193,37	Moment Ratio	0,0007	0,0001	1,0014	8.591,80
Peng	-1,1520	0,5334	0,3077	4.083,91	Peng	-1,0720	0,3401	0,1209	8.004,35
W	-1,1517	0,5332	0,3073	2.771,21	W	-1,0716	0,3398	0,1206	5.451,50
PWM	-0,8000	0,3858	0,1888	2.885,53	PWM	-0,8980	0,2913	0,0952	5.974,01

<i>100</i>					<i>500</i>				
	<i>MEAN</i>	<i>SE</i>	<i>MSE</i>	<i>DURATION</i>		<i>MEAN</i>	<i>SE</i>	<i>MSE</i>	<i>DURATION</i>
Pickand's	-1,0401	0,1766	0,0328	299,48	Pickand's	-1,2306	0,0791	0,0594	301,15
Hill's	0,0020	0,0002	1,0041	8.166,39	Hill's	0,0107	0,0006	1,0216	40.343,63
Negative Hill's	-0,9938	0,0996	0,0100	15.711,85	Negative Hill's	-1,0188	0,0451	0,0024	78.332,41
Moment	-1,0398	0,2303	0,0546	15.828,27	Moment	-1,0500	0,1032	0,0131	77.444,28
Moment Ratio	0,0014	0,0001	1,0027	16.960,07	Moment Ratio	0,0071	0,0003	1,0142	80.022,28
Peng	-1,0404	0,2303	0,0547	15.922,03	Peng	-1,0536	0,1034	0,0136	79.720,98
W	-1,0397	0,2299	0,0544	10.777,47	W	-1,0497	0,1025	0,0130	56.223,14
PWM	-0,9530	0,2130	0,0476	12.383,85	PWM	-1,0318	0,1000	0,0110	54.926,17

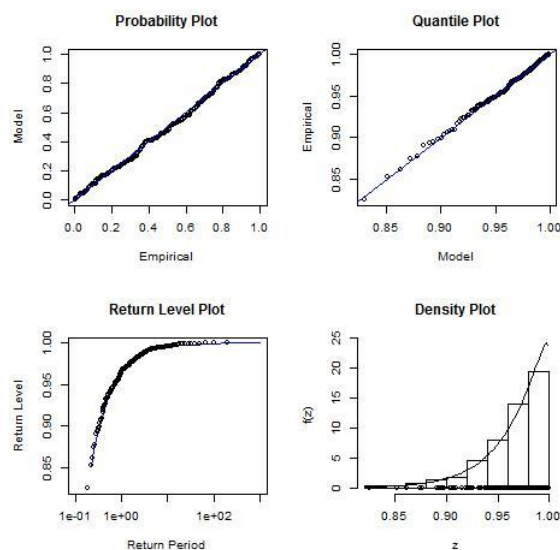
Για την Beta (5, 1) αναμένουμε να αποδίδει καλύτερα ο εκτιμητής Negative Hill's, μιας και προτείνεται για  $\xi < -1/2$ . Πράγματι, σε κάθε  $k$ , ο Negative Hill's παρουσιάζει μία από τις χαμηλότερες μεροληψίες, καθώς και το χαμηλότερο MSE. Μάλιστα για  $k = 500$  γίνεται βέλτιστος με μεροληψία μόλις στο 0,0188 και MSE 0,0024. Αξίζει να σημειωθεί ότι στην προκειμένη περίπτωση, όπου ο δείκτης ακραίων τιμών έχει τιμή -1, κανείς άλλος εκτιμητής δεν πλησιάζει την απόδοση του Negative Hill's. Ένας ακόμα εκτιμητής που βάσει της προσομοίωσης έχει καλά αποτελέσματα, είναι ο Pickand's, ωστόσο οι τιμές του MSE του δεν πλησιάζουν αυτές του Negative Hill's και η μεροληψία του αυξάνεται όσο μεγαλώνει το  $k$ . Σε αντίθεση με τους παραπάνω, ο εκτιμητής Moment

Ratio αδυνατεί να πλησιάσει την τιμή  $-1$ , παρόλο που προτείνεται για  $\xi \in \mathbb{R}$ . Τέλος, οι εκτιμητές Moment, Peng,  $W$  και ο PWM παρουσιάζουν καλύτερες τιμές, καθώς το  $k$  αυξάνεται (για μικρότερα  $k$  οι τρεις πρώτοι τείνουν να υποεκτιμούν, ενώ ο PWM να υπερεκτιμά την ποσότητα) και για  $k = 500$  παρουσιάζουν καλή μεροληψία και μέτριο MSE. Μία τελευταία παρατήρηση, είναι ότι η εκτιμήτρια Μέγιστης Πιθανοφάνειας, με χρήση της POT παρουσιάζει πάρα πολύ καλό MSE.

#### 4.7 ΟΜΟΙΟΜΟΡΦΗ ΚΑΤΑΝΟΜΗ

Η ομοιόμορφη κατανομή που εξετάστηκε ήταν η συνεχής στο διάστημα  $(0, 1)$  και εξ ορισμού με πεπερασμένο άνω άκρο στις τιμές που μπορεί να πάρει μια μεταβλητή. Η θεωρητική τιμή της παραμέτρου  $\xi$  σε αυτήν την περίπτωση είναι  $\xi = -1$ . Ο ελάχιστος διαχωρισμός των αρχικών δεδομένων βάσει του δείγματος 5.000 παρατηρήσεων από την κατανομή, που να έχει καλή προσαρμογή στην GEV ήταν αυτός των 200 block μεγέθους 25 παρατηρήσεων, επομένως καταλήγουμε στην εξέταση για την προσαρμογή 200 παρατηρήσεων. Η εκτίμηση με τη μέθοδο μέγιστης πιθανοφάνειας των παραμέτρων της προσαρμοσμένης κατανομής των 200 maxima που θα χρησιμοποιήσουμε έδωσε τις ακόλουθες τιμές  $(\hat{\mu}, \hat{\sigma}, \hat{\xi}) = (0.96444014, 0.03332199, -0.94318413)$ . Η καλή προσαρμογή των δεδομένων φαίνεται παρακάτω.

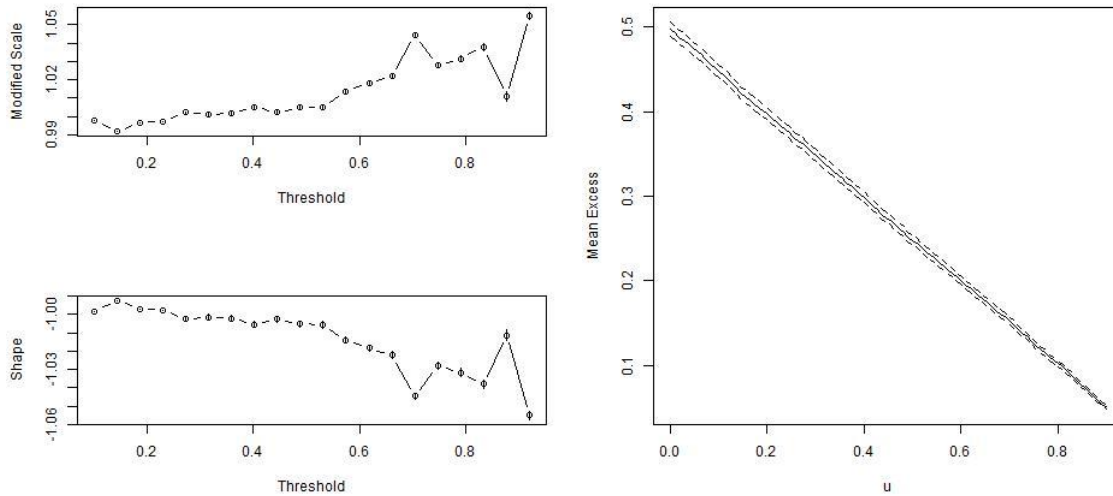
Εικόνα 4.7.1 Διαγράμματα καλής προσαρμογής στην GEV για την Ομοιόμορφη  $(0, 1)$



Από τα γραφήματα PP και QQ φαίνεται, ότι υπάρχει καλή προσαρμογή των δεδομένων στην GEV, ενώ ταυτόχρονα φαίνεται και από το ιστόγραμμα των δεδομένων. Το διάγραμμα στάθμης απόδοσης εμφανίζει κοίλη καμπύλη επομένως μπορούμε να διαπιστώσουμε ότι η παράμετρος  $\xi$  είναι αρνητική ( $\xi < 0$ ).

Ακολούθως, η επιλογή του βέλτιστου κατωφλιού για την επιλογή με την χρήσης της μεθόδου POT

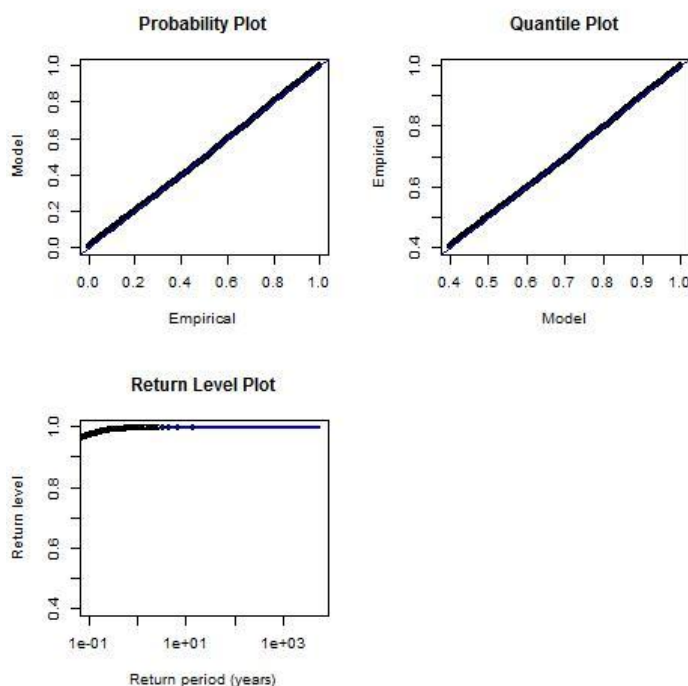
Εικόνα 4.7.2 Διαγράμματα για την επιλογή του βέλτιστου  $u$  στην περίπτωση της Ομοιόμορφης  $(0, 1)$



Από την εικόνα των διαγραμμάτων δεν είναι αρκετά εμφανής η τιμή, για την οποία θα βελτιστοποιείται η προσαρμογή των δεδομένων στην GPD, ωστόσο, λόγω της σχετικής σταθερότητας του πρώτου γραφήματος κοντά στην περιοχή του 0,4 επιλέχθηκε αυτή ως βέλτιστη τιμή για το κατώφλι, πράγμα που σημαίνει, ότι το πλήθος των δεδομένων που θα επιλέγονται για την ανάλυση αναμένεται να είναι το 60%, δηλαδή 3.000 παρατηρήσεις κάθε φορά. Από την προσαρμογή της κατανομής στα δεδομένα στην GPD παίρνουμε τις ακόλουθες εκτιμήσεις  $(\tilde{\sigma}, \tilde{\xi}) = (0.6050933, 1.0089080)$ . Ακολούθως έχουμε την εξέταση της καλής προσαρμογής της κατανομής στα δεδομένα<sup>4</sup>.

<sup>4</sup> Η R είχε αδυναμία να παράξει το Density Plot (Error in plot.window(xlim, ylim, "", ...) : need finite 'ylim' values)

Εικόνα 4.7.3 Διαγράμματα καλής προσαρμογής στην GPD για την Ομοιόμορφη (0, 1)



Παρακάτω, φαίνονται συγκεντρωτικά οι εκτιμώμενες (μέσω προσομοίωσης  $10^5$  επαναλήψεων) παράμετροι των κατανομών των διάφορων εκτιμητών του  $\xi$  από κάθε εκτιμητή για  $k = 25, 50, 100, 500$ .

Πίνακας 4.7.1 Περιγραφικά χαρακτηριστικά της κατανομής των προσομοιωμένων εκτιμητών για την Ομοιόμορφη (0, 1)

**Ομοιόμορφη (0, 1) ( $\xi = -1$ )**

<b>MLE</b>	<b>MEAN</b>	<b>SE</b>	<b>MSE</b>	<b>DURATION</b>
BM	-1,0003	0,1406	0,0198	6.360,42
POT	-1,0041	0,0194	0,0004	7.327,22

	<b>25</b>					<b>50</b>			
	<b>MEAN</b>	<b>SE</b>	<b>MSE</b>	<b>DURATION</b>		<b>MEAN</b>	<b>SE</b>	<b>MSE</b>	<b>DURATION</b>
Pickand's	-1,0152	0,3567	0,1275	279,43	Pickand's	-1,0069	0,2493	0,0622	276,62
Hill's	0,0026	0,0006	1,0052	3.168,49	Hill's	0,0051	0,0008	1,0103	6.335,99
Negative Hill's	-0,9595	0,1960	0,0401	6.070,01	Negative Hill's	-0,9796	0,1400	0,0200	12.369,84
Moment	-1,1443	0,5345	0,3065	6.286,46	Moment	-1,0690	0,3405	0,1207	12.592,62
Moment Ratio	0,0017	0,0004	1,0035	6.386,96	Moment Ratio	0,0034	0,0005	1,0068	12.654,95
Peng	-1,1452	0,5347	0,3070	12.893,51	Peng	-1,0707	0,3408	0,1212	12.893,51
W	-1,1440	0,5334	0,3053	4.177,68	W	-1,0687	0,3393	0,1199	7.741,35
PWM	-0,7940	0,3864	0,1917	3.868,03	PWM	-0,8955	0,2909	0,0956	7.728,06

	100					500			
	MEAN	SE	MSE	DURATION		MEAN	SE	MSE	DURATION
Pickand's	-1,0038	0,1770	0,0313	277,99	Pickand's	-1,0008	0,0789	0,0062	275,56
Hill's	0,0102	0,0012	1,0206	12.645,67	Hill's	0,0537	0,0028	1,1103	62.735,43
Negative Hill's	-0,9897	0,0994	0,0100	24.989,29	Negative Hill's	-0,9978	0,0446	0,0020	124.523,73
Moment	-1,0329	0,2319	0,0549	25.295,75	Moment	-1,0069	0,1021	0,0105	125.295,20
Moment Ratio	0,0068	0,0007	1,0137	25.219,95	Moment Ratio	0,0355	0,0017	1,0722	125.930,90
Peng	-1,0363	0,2323	0,0553	25.179,28	Peng	-1,0251	0,1030	0,0112	121.274,16
W	-1,0326	0,2303	0,0541	15.403,07	W	-1,0062	0,0987	0,0098	76.876,61
PWM	-0,9463	0,2137	0,0486	15.426,90	PWM	-0,9890	0,0972	0,0096	78.448,04

Από τα αποτελέσματα φαίνεται η πολύ ακριβής εκτίμηση, η οποία δίνεται από τον εκτιμητή μεγίστης πιθανοφάνειας μετά από εφαρμογή της μεθόδου POT, με το ελάχιστο μέσο τετραγωνικό σφάλμα. Η μέθοδος των Block Maxima εμφανίζει οριακά μικρότερη μεροληψία στις εκτιμήσεις, ωστόσο παρουσιάζει ταυτόχρονα και μεγαλύτερη μεταβλητότητα από την POT.

Αναφορικά με τους υπόλοιπους εκτιμητές οι καλύτεροι είναι οι Pickand's και Negative Hill's, όπως αναμενόταν από τη θεωρία. Η μεροληψία που φαίνεται να έχει ο Pickand's είναι επίσης οριακά μικρότερη από αυτή του Negative Hill's, όμως ο δεύτερος δίνει συστηματικά καλύτερες εκτιμήσεις με μικρότερη μεταβλητότητα και αρκετά μικρότερο μέσο τετραγωνικό σφάλμα. Οι εκτιμητές Peng & W καθώς και ο εκτιμητής Ροπών Σταθμισμένης Πιθανότητας παρουσιάζουν τόσο μεγαλύτερη μεροληψία όσο και μεταβλητότητα στα αποτελέσματα παρά την αναμενόμενη απόδοση, που θεωρήθηκε στην αρχή. Αντίθετα, οι εκτιμητές των Hill's και Λόγου Ροπών, καθώς προβλεπόταν, είχαν τη μέγιστη μεροληψία στα αποτελέσματά τους και είναι οι πλέον ακατάλληλοι για αυτήν την περίπτωση. Συνολικά, ο πιο εύρηστος εκτιμητής ξεχωρίζει να είναι αυτός του Pickand's, ο οποίος παραμένει – παρά την σχετικά μεγαλύτερη μεταβλητότητα στα αποτελέσματά του – αρκετά γρήγορος στον υπολογισμό ακόμα και για αρκετά μεγάλες τιμές του  $k$ , αν και ο Negative Hill's έχει σταθερά μικρότερο μέσο τετραγωνικό σφάλμα από αυτόν για κάθε τιμή του  $k$ .

#### 4.8 ΚΑΤΑΝΟΜΗ BURR

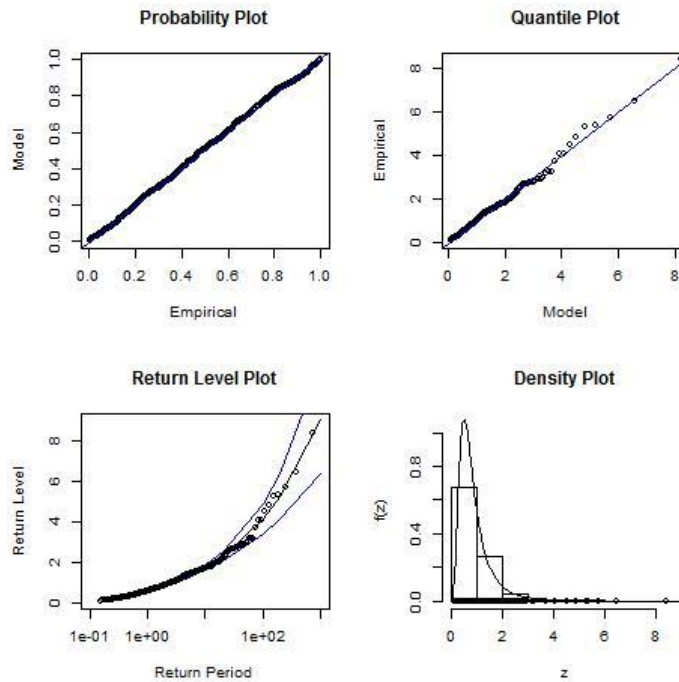
Στην κατανομή Burr θα εξετασθούν 3 περιπτώσεις, με παραμέτρους  $(\alpha, \gamma, \theta) = (4, 1, 1), (1, 2, 1), (1, 1, 1)$ .

Αρχικά, για την κατανομή Burr με παραμέτρους  $(\alpha, \gamma, \theta) = (4, 1, 1)$  (που αντιστοιχεί σε  $\xi = 0.25$ ) ελέγχουμε την προσαρμογή των δεδομένων που παρήχθησαν με την μέθοδο επιλογής Block Maxima. Αφού παρήχθη δείγμα μεγέθους  $n = 5000$ , και με κατάλληλες δοκιμές, καταλήξαμε, ότι ο βέλτιστος αριθμός Blocks για αυτήν είναι 750, με 7 παρατηρήσεις το καθένα. Βάσει αυτών των παρατηρήσεων και με χρήση της μεθόδου μεγίστης πιθανοφάνειας εκτιμούμε τις παραμέτρους της GEV, την οποία θα πρέπει

ακολουθούν προσεγγιστικά. Έτσι, καταλήγουμε στο συμπέρασμα ότι η κατανομή των δεδομένων θα είναι η GEV με  $(\hat{\mu}, \hat{\sigma}, \hat{\xi}) = (0.6010610, 0.3551495, 0.3057976)$ .

Εν συνεχεία, παρατίθενται τα διαγράμματα καλής προσαρμογής των δεδομένων στην GEV.

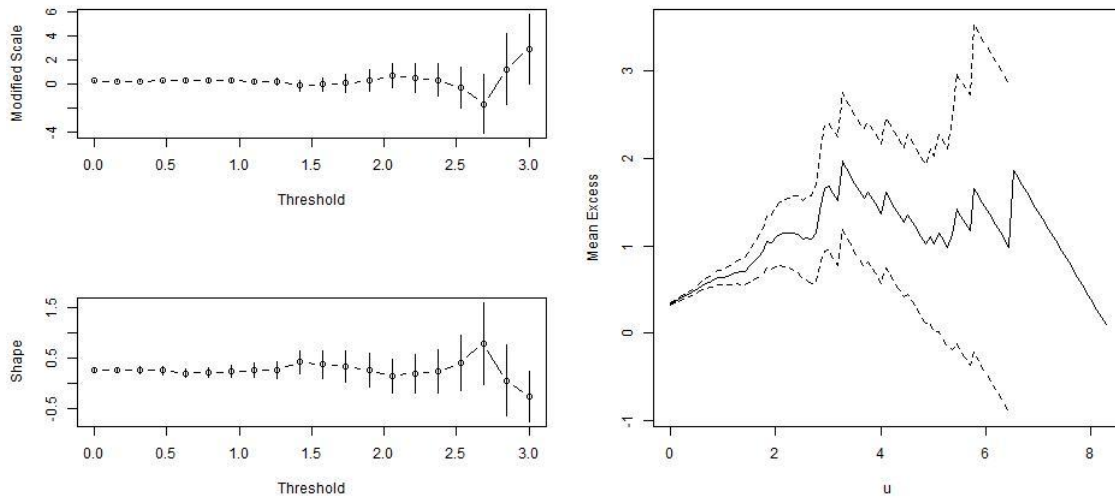
Εικόνα 4.8.1 Διαγράμματα καλής προσαρμογής στην GEV για την Burr (4, 1, 1)



Από τα PP-plot και QQ-plot διαφαίνεται, ότι τα θεωρητικά και τα παρατηρούμενα σημεία των κατανομών συμπίπτουν, όπως είναι αναμενόμενο, όταν η παρατηρούμενη κατανομή είναι η GEV, γεγονός που φαίνεται και από το ιστόγραμμα των Block Maxima (Density Plot). Αντίστοιχα, από το γράφημα στάθμης απόδοσης (Return Level Plot) βλέπουμε, ότι η  $z_p(w)$  σχηματίζει κυρτή καμπύλη, επομένως συμπεραίνουμε, ότι ισχύει  $\xi > 0$ .

Για την εφαρμογή της μεθόδου Peak Over Threshold η επιλογή του βέλτιστου κατωφλιού  $u$  έγινε από τα παρακάτω γραφήματα.

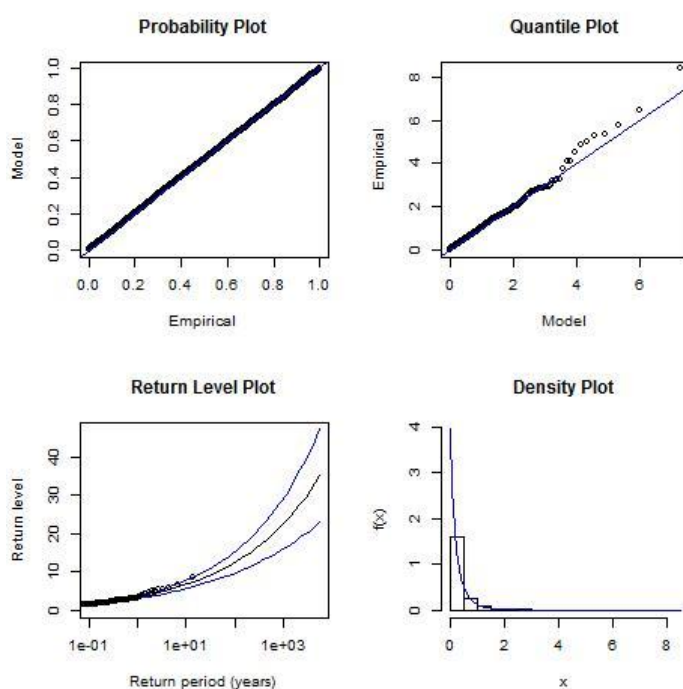
Εικόνα 4.8.2 Διαγράμματα για την επιλογή του βέλτιστου  $u$  στην περίπτωση της  $Burr(4, 1, 1)$



Μετά από την εξέταση των δύο διαγραμμάτων για το threshold, καταλήγουμε στο συμπέρασμα, ότι το μικρότερο  $u$ , πέραν του οποίου το  $\xi$  σταθεροποιείται και η εκτίμηση του  $\hat{\sigma}$  μεταβάλλεται γραμμικά ως προς το  $u$ , είναι το 0. Αντίστοιχα συμπεράσματα παίρνουμε και από το Mean Excess Plot, όπου παρατηρούμε, ότι η μονάδα είναι η τιμή, όπου το διάγραμμα γίνεται οριζόντιο και τα δ.ε δεν ξεφεύγουν. Συνεπώς, θα χρησιμοποιηθούν για τον υπολογισμό των εκτιμητών τα δεδομένα από την αρχική κατανομή Burr, τα οποία είναι μεγαλύτερα της τιμής αυτής. Με την εφαρμογή του  $u$  για το πρώτο δείγμα, το ποσοστό των δεδομένων που ακολουθεί GPD με  $(\hat{\sigma}, \hat{\xi}) = (0.2532676, 0.2447924)$  είναι 100 %. Επομένως, ο MLE που θα υπολογισθεί βάσει της μεθόδου Peak Over Threshold θα χρησιμοποιεί δείγματα με το σύνολο των 5000 αρχικών δεδομένων από την κατανομή  $Burr(4, 1, 1)$ .

Στη συνέχεια, με τον ίδιο γραφικό τρόπο παρατηρούμε, ότι τα θεωρητικά και τα παρατηρούμενα σημεία των κατανομών συμπίπτουν, όπως είναι αναμενόμενο, όταν η παρατηρούμενη κατανομή είναι η GPD, γεγονός που φαίνεται και από το ιστόγραμμα των Peaks Over Threshold (Density Plot). Αντίστοιχα, από το γράφημα στάθμης απόδοσης (Return Level Plot) βλέπουμε ότι η  $z_p(w)$  αυξάνεται γραμμικά ως προς το  $w$ , επομένως συμπεραίνουμε και από εδώ, ότι ισχύει  $\xi > 0$ .

Εικόνα 4.8.3 Διαγράμματα καλής προσαρμογής στην GPD για την Burr (4, 1, 1)



Παρακάτω, φαίνονται συγκεντρωτικά οι εκτιμώμενες (μέσω προσομοίωσης  $10^5$  επαναλήψεων) παράμετροι των κατανομών των διάφορων εκτιμητών του  $\xi$  από κάθε εκτιμητή για  $k = 25, 50, 100, 500$ .

Πίνακας 4.8.1 Περιγραφικά χαρακτηριστικά της κατανομής των προσομοιωμένων εκτιμητών για την Burr (4, 1, 1)

**Burr (4, 1, 1) ( $\xi = 0.25$ )**

	<i>MLE</i>	<i>MEAN</i>	<i>SE</i>	<i>MSE</i>	<i>DURATION</i>
BM		0,3050	0,0315	0,0040	8.705,10
POT		0,2495	0,0177	0,0003	10.632,70

	<b>25</b>					<b>50</b>			
	<i>MEAN</i>	<i>SE</i>	<i>MSE</i>	<i>DURATION</i>		<i>MEAN</i>	<i>SE</i>	<i>MSE</i>	<i>DURATION</i>
Pickand's	0,2251	0,3747	0,1410	582,78	Pickand's	0,2354	0,2637	0,0697	581,78
Hill's	0,3191	0,0607	0,0085	3.858,99	Hill's	0,3369	0,0449	0,0096	7.343,00
Negative Hill's	-0,2622	0,1362	0,2809	7.505,03	Negative Hill's	-0,2066	0,0968	0,2178	14.916,96
Moment	0,1572	0,2271	0,0602	7.340,22	Moment	0,2152	0,1516	0,0242	14.324,16
Moment Ratio	0,2935	0,0706	0,0069	4.662,38	Moment Ratio	0,3115	0,0523	0,0065	9.088,60
Peng	0,1317	0,2675	0,0855	4.644,13	Peng	0,1899	0,1837	0,0374	9.060,24
W	0,0717	0,1933	0,0691	4.633,58	W	0,1383	0,1367	0,0312	9.020,52
PWM	0,2517	0,2153	0,0463	4.647,95	PWM	0,2471	0,1601	0,0256	9.041,78



	100					500			
	MEAN	SE	MSE	DURATION		MEAN	SE	MSE	DURATION
Pickand's	0,2421	0,1868	0,0350	581,29	Pickand's	0,2492	0,0840	0,0071	582,09
Hill's	0,3605	0,0335	0,0133	14.257,59	Hill's	0,4660	0,0186	0,0470	71.832,54
Negative Hill's	-0,1612	0,0697	0,1739	28.775,42	Negative Hill's	-0,0916	0,0345	0,1179	122.350,14
Moment	0,2479	0,1063	0,0113	28.667,50	Moment	0,2943	0,0486	0,0043	143.309,25
Moment Ratio	0,3314	0,0383	0,0081	18.129,12	Moment Ratio	0,4073	0,0185	0,0251	91.387,95
Peng	0,2189	0,1305	0,0180	18.095,88	Peng	0,2356	0,0603	0,0038	91.216,45
W	0,1788	0,1037	0,0158	18.029,34	W	0,2248	0,0603	0,0043	91.066,05
PWM	0,2477	0,1181	0,0140	18.050,36	PWM	0,2492	0,0553	0,0031	91.129,53

Για την Burr (4, 1, 1) και στην περίπτωση, όπου εκτιμούμε τον δείκτη ακραίων τιμών κοντά στο 0,25, αναμένουμε βασιζόμενοι στην θεωρία να αποδίδουν τα βέλτιστα οι εκτιμητές Hill's και PWM, μιας και προτείνονται για  $\xi > 0$ , και ο W, που προτείνεται για  $\xi < 1/2$ . Παρά τις προσδοκίες μόνο ο PWM εμφανίζει αρκετά μικρή μεροληψία, η οποία τείνει να εξαληφθεί για μεγαλύτερες τιμές του  $k$  για τον τύπο του. Ο Hill's, αντίθετα, τείνει να υπερεκτιμά τον δείκτη ακραίων τιμών, με την μεροληψία του να αυξάνεται καθώς το  $k$  μεγαλώνει, σε αντίθεση με τον W, ο οποίος τείνει να υποεκτιμά τον δείκτη, πράγμα που διορθώνεται με την επιλογή αρκετά μεγάλου  $k$  για τον αντίστοιχο τύπο του.

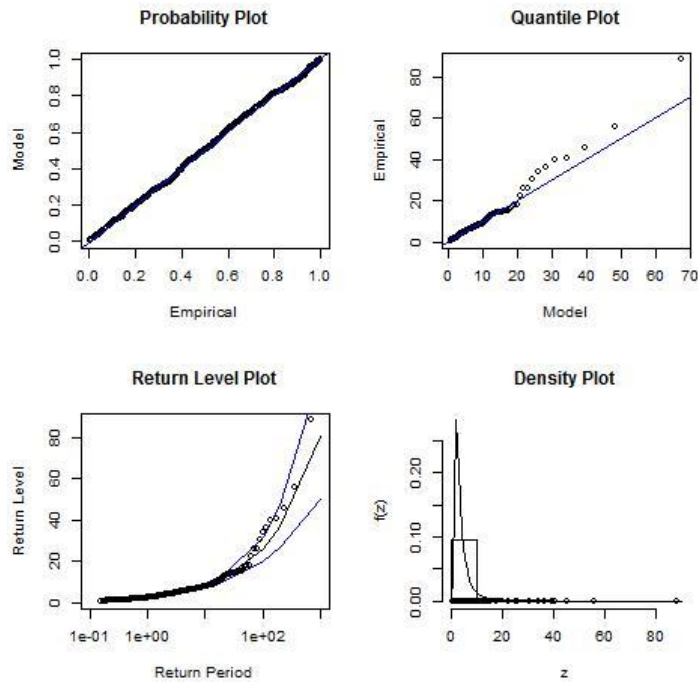
Όσον αφορά τους εκτιμητές που συνιστούνται για όλο το διάστημα  $\mathbb{R}$ , ο Peng παρουσιάζει μεγάλη μεροληψία και υποεκτιμά την παράμετρο  $\xi$ , ενώ ομοίως με προηγούμενος, χρειάζεται αρκετά μεγάλη τιμή  $k$  για να βελτιωθεί η μεροληψία που παρουσιάζει. Οι εκτιμητές Moment και Moment Ratio υποεκτιμούν και υπερεκτιμούν κατά αντιστοιχία τον δείκτη με μόνο τον Moment να εμφανίζει κάποια βελτίωση και μόνο για τιμή  $k = 100$ . Σταθερότερη επίδοση όσον αφορά και μεροληψία και το μέσο τετραγωνικό σφάλμα ήταν ο εκτιμητής του Pickand, ο οποίος είχε ελάχιστη μεροληψία και MSE για  $k = 500$  χωρίς να έχει κάποιο κόστος στο χρόνο υπολογισμού του.

Συνολικά, παρατηρούμε, ότι ο MLE με βάση την POT παρουσιάζει την μικρότερη μεροληψία, με μέση τιμή  $\mu = 0,2495$  και το ελάχιστο μέσο τετραγωνικό σφάλμα από όλους τους εκτιμητές.

Ομοίως, θα γίνει έλεγχος της προσαρμογής των δεδομένων που παρήχθησαν με την Burr (1, 2, 1) (με αντίστοιχο  $\xi = 0,5$ ) για την μέθοδο επιλογής Block Maxima. Έτσι, αφού παρήχθη δείγμα μεγέθους  $n = 5000$ , και με κατάλληλες δοκιμές, καταλήξαμε ότι ο βέλτιστος αριθμός Blocks για αυτήν είναι 700, με 7 παρατηρήσεις το καθένα. Βάσει αυτών των παρατηρήσεων και με χρήση της μεθόδου μεγίστης πιθανοφάνειας εκτιμούμε τις παραμέτρους της GEV, την οποία θα πρέπει ακολουθούν προσεγγιστικά. Έτσι, καταλήγουμε στο συμπέρασμα ότι η κατανομή των δεδομένων θα είναι η GEV με  $(\hat{\mu}, \hat{\sigma}, \hat{\xi}) = (2.5410154, 1.4010163, 0.4807127)$ .

Εν συνεχεία, παρατίθενται τα διαγράμματα καλής προσαρμογής των δεδομένων στην GEV.

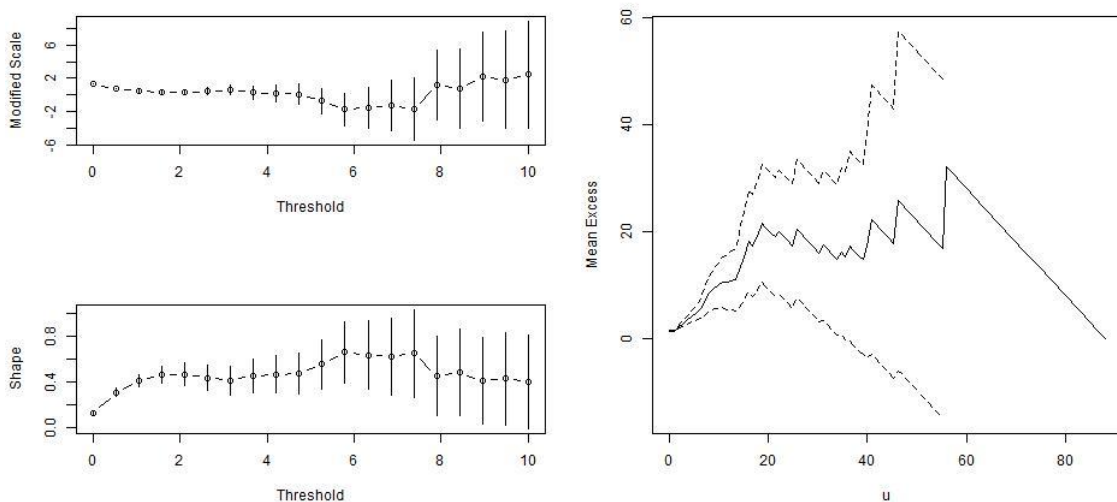
Εικόνα 4.8.4 Διαγράμματα καλής προσαρμογής στην GEV για την Burr (1, 2, 1)



Από τα PP-plot και QQ-plot διαφαίνεται ότι τα θεωρητικά και τα παρατηρούμενα σημεία των κατανομών συμπίπτουν, όπως είναι αναμενόμενο όταν η παρατηρούμενη κατανομή είναι η GEV, γεγονός που φαίνεται και από το ιστόγραμμα των Block Maxima (Density Plot). Αντίστοιχα, από το γράφημα στάθμης απόδοσης (Return Level Plot) βλέπουμε ότι η  $z_p(w)$  αυξάνεται γραμμικά ως προς το  $w$ , άρα  $\xi > 0$ .

Για την εφαρμογή της μεθόδου Peak Over Threshold η επιλογή του βέλτιστου κατωφλιού  $u$  θα γίνει από τα παρακάτω γραφήματα.

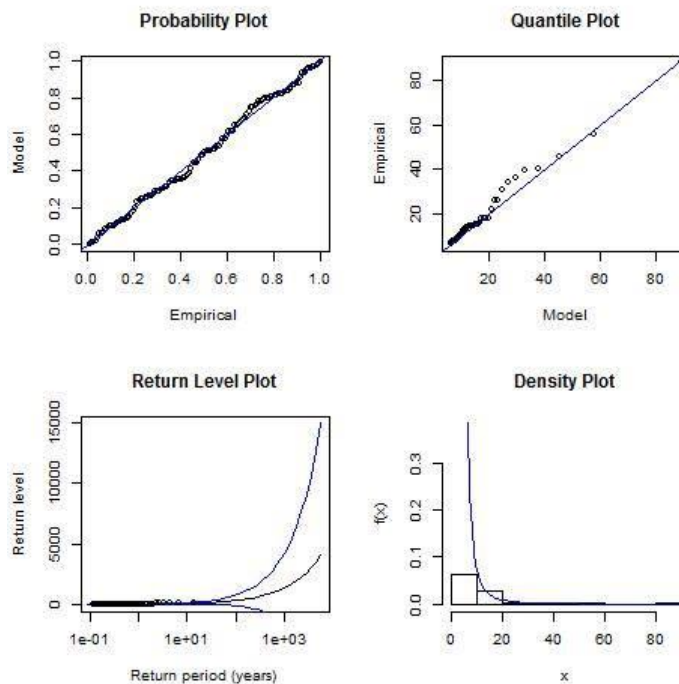
Εικόνα 4.8.5 Διαγράμματα για την επιλογή του βέλτιστου  $u$  στην περίπτωση της Burr (1, 2, 1)



Με χρήση των δύο διαγραμμάτων για το threshold, καταλήγουμε στο συμπέρασμα ότι το μικρότερο  $u$  πέραν του οποίου το  $\hat{\xi}$  σταθεροποιείται και η εκτίμηση του  $\hat{\sigma}$  μεταβάλλεται γραμμικά ως προς το  $u$  είναι το 6,5. Αντίστοιχα συμπεράσματα παίρνουμε και από το Mean Excess Plot, όπου παρατηρούμε ότι η μονάδα είναι η τιμή όπου το διάγραμμα γίνεται οριζόντιο και τα δ.ε δεν ξεφεύγουν. Συνεπώς, θα χρησιμοποιηθούν για τον υπολογισμό των εκτιμητών τα δεδομένα από την αρχική κατανομή Burr, τα οποία είναι μεγαλύτερα της τιμής αυτής. Με την εφαρμογή του  $u$  για το πρώτο δείγμα, το ποσοστό των δεδομένων που ακολουθεί GPD με  $(\hat{\sigma}, \hat{\xi}) = (0.4542395, 0.1595531)$  είναι 2,34 %. Επομένως, ο MLE που θα υπολογισθεί βάσει της μεθόδου Peak Over Threshold θα χρησιμοποιεί δείγματα με περίπου 120 παρατηρήσεις.

Στη συνέχεια, με τον ίδιο γραφικό τρόπο παρατηρούμε ότι τα θεωρητικά και τα παρατηρούμενα σημεία των κατανομών συμπίπτουν, όπως είναι αναμενόμενο όταν η παρατηρούμενη κατανομή είναι η GPD, γεγονός που φαίνεται και από το ιστόγραμμα των Peaks Over Threshold (Density Plot). Αντίστοιχα, από το γράφημα στάθμης απόδοσης (Return Level Plot) βλέπουμε ότι η  $z_p(w)$  αυξάνεται γραμμικά ως προς το  $w$ , άρα  $\xi > 0$ .

Εικόνα 4.8.6 Διαγράμματα καλής προσαρμογής στην GPD για την Burr (1, 2, 1)



Παρακάτω, φαίνονται συγκεντρωτικά οι εκτιμώμενες (μέσω προσομοίωσης  $10^5$  επαναλήψεων) παράμετροι των κατανομών των διάφορων εκτιμητών του  $\xi$  από κάθε εκτιμητή για  $k = 25, 50, 100, 500$ .

Πίνακας 4.8.2 Περιγραφικά χαρακτηριστικά της κατανομής των προσομοιωμένων εκτιμητών για την Burr (1, 2, 1)

<b>Burr (1, 2, 1) (<math>\xi = 0.5</math>)</b>				
<b>MLE</b>	<b>MEAN</b>	<b>SE</b>	<b>MSE</b>	<b>DURATION</b>

BM	0,4729	0,0368	0,0021	4.173,67
POT	0,4782	0,1438	0,0212	594,69

	25					50			
	MEAN	SE	MSE	DURATION		MEAN	SE	MSE	DURATION
Pickand's	0,4674	0,3905	0,1536	208,01	Pickand's	0,4735	0,2747	0,0761	196,25
Hill's	0,5014	0,1001	0,0100	1.818,53	Hill's	0,5030	0,0710	0,0051	3.498,17
Negative Hill's	-0,1947	0,1257	0,4984	3.316,80	Negative Hill's	-0,1409	0,0860	0,4181	6.721,30
Moment	0,3892	0,2358	0,0679	3.528,95	Moment	0,4437	0,1589	0,0284	7.101,78
Moment Ratio	0,4816	0,1297	0,0172	4.074,51	Moment Ratio	0,4917	0,0955	0,0092	7.905,81
Peng	0,3694	0,3073	0,1115	3.468,91	Peng	0,4325	0,2158	0,0511	6.859,14
W	0,2011	0,1696	0,1181	3.478,78	W	0,2725	0,1200	0,0662	6.876,43
PWM	0,4259	0,2167	0,0524	3,362.53	PWM	0,4450	0,1692	0,0317	6,779.89

	100					500			
	MEAN	SE	MSE	DURATION		MEAN	SE	MSE	DURATION
Pickand's	0,4700	0,1943	0,0386	194,91	Pickand's	0,3663	0,0859	0,0253	195,91
Hill's	0,5052	0,0502	0,0025	6.895,75	Hill's	0,5268	0,0230	0,0012	34.476,17
Negative Hill's	-0,0999	0,0587	0,3634	14.783,41	Negative Hill's	-0,0441	0,0246	0,2966	72.335,71
Moment	0,4714	0,1117	0,0133	15.031,69	Moment	0,4939	0,0510	0,0026	68.715,11
Moment Ratio	0,4978	0,0689	0,0048	14.488,68	Moment Ratio	0,5124	0,0309	0,0011	73.591,94
Peng	0,4640	0,1545	0,0252	13.590,65	Peng	0,4794	0,0708	0,0054	72.424,89
W	0,3189	0,0911	0,0411	15.224,28	W	0,3789	0,0554	0,0177	69.946,26
PWM	0,4615	0,1330	0,0192	13,910.77	PWM	0,4768	0,0754	0,0062	71,038.91

Για την Burr (1, 2, 1) αναμένουμε, όπως και προηγουμένως, οι Hill's και PWM να έχουν καλή απόδοση. Παρόλα αυτά, επειδή η σύγκλιση του PWM στο  $\xi$  είναι πλέον πιο αργή στο διάστημα  $1/2 < \xi < 1$ , η απόδοση αυτή δεν θα είναι εξίσου καλή. Αυτό επιβεβαιώνεται από τα αποτελέσματα της προσομοίωσης που φαίνονται στον παραπάνω πίνακα. Μάλιστα, φανερώνουν, μία τάση υποεκτίμησης του δείκτη ακραίων τιμών, η οποία ωστόσο βελτιώνεται καθώς το  $k$  μεγαλώνει. Αντίθετα ο εκτιμητής Hill's παρουσιάζει πολύ καλή αμεροληψία, ειδικά για μικρές τιμές του  $k$  και μερικά από τα καλύτερα MSE, με μικρότερο όλων της περίπτωσης  $k = 500$  όπου παίρνει την τιμή 0,0012.

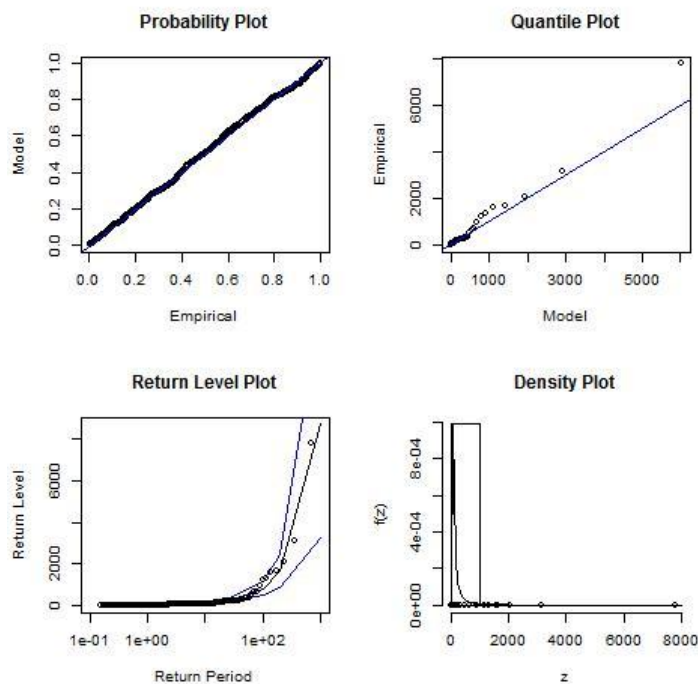
Παρατηρούμε ότι οι εκτιμητές που αποδίδουν καλά σε όλο το  $\mathbb{R}$ , Pickand's και W, παρουσιάζουν μεγάλη μεροληψία και τείνουν να υποεκτιμούν την παράμετρο  $\xi$ , ενώ οι Peng και Moment δίνουν μικρό MSE καθώς και μικρή μεροληψία όσο αυξάνεται το  $k$ . Ίδια συμπεριφορά παρουσιάζει και ο Moment Ratio, ο οποίος για  $k = 500$  δίνει το καλύτερο MSE με τιμή 0,0011.

Όσον αφορά τις ασυμπτωτικές εκτιμήσεις, ο MLE με βάση την BM παρουσιάζει ένα από τα μικρότερα MSE, με τιμή 0,0021 υποεκτιμώντας τον δείκτη ακραίων τιμών κατά 0,0271.

Ομοίως, θα γίνει έλεγχος της προσαρμογής των δεδομένων που παρήχθησαν με την κατανομή Burr με παραμέτρους  $(\alpha, \gamma, \theta) = (1, 1, 1)$  (με αντίστοιχο  $\xi = 1$ ) για την μέθοδο επιλογής Block Maxima. Αφού παρήχθη δείγμα μεγέθους  $n = 5000$ , και με κατάλληλες δοκιμές, καταλήξαμε ότι ο βέλτιστος αριθμός Blocks για αυτήν είναι 700, με 7 παρατηρήσεις το καθένα. Βάσει αυτών των παρατηρήσεων και με χρήση της μεθόδου μέγιστης πιθανοφάνειας εκτιμούμε τις παραμέτρους της GEV, την οποία θα πρέπει ακολουθούν προσεγγιστικά. Έτσι, καταλήγουμε στο συμπέρασμα ότι η κατανομή των δεδομένων θα είναι η GEV με  $(\hat{\mu}, \hat{\sigma}, \hat{\xi}) = (6.402193, 6.972023, 1.037909)$ .

Εν συνεχεία, παρατίθενται τα διαγράμματα καλής προσαρμογής των δεδομένων στην GEV.

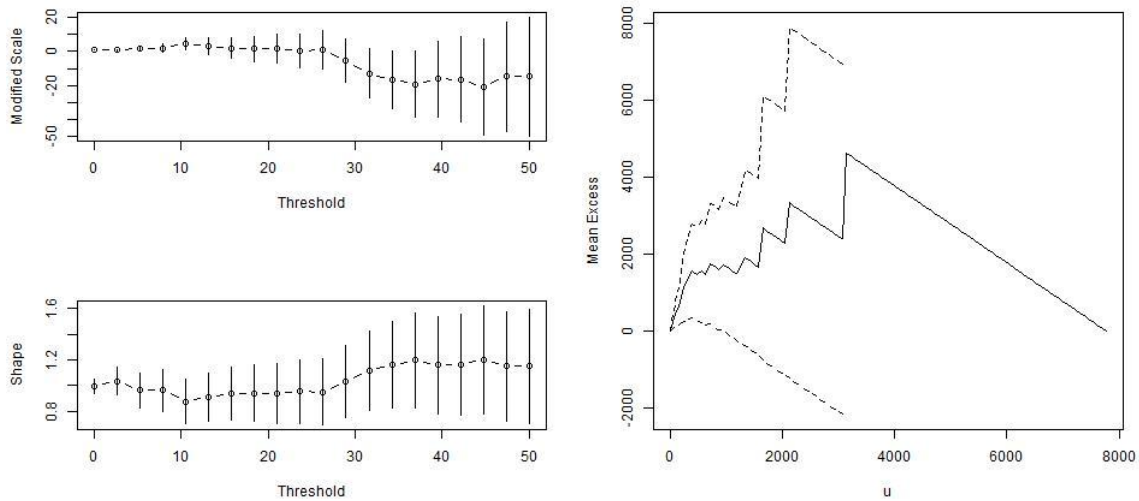
Εικόνα 4.8.7 Διαγράμματα καλής προσαρμογής στην GEV για την Burr (1, 1, 1)



Από τα PP-plot και QQ-plot διαφαίνεται ότι τα θεωρητικά και τα παρατηρούμενα σημεία των κατανομών συμπίπτουν, όπως είναι αναμενόμενο όταν η παρατηρούμενη κατανομή είναι η GEV, γεγονός που φαίνεται και από το ιστόγραμμα των Block Maxima (Density Plot). Αντίστοιχα, από το γράφημα στάθμης απόδοσης (Return Level Plot) βλέπουμε ότι η  $z_p(w)$  αυξάνεται γραμμικά ως προς το  $w$ , άρα  $\xi > 0$ .

Για την εφαρμογή της μεθόδου Peak Over Threshold η επιλογή του βέλτιστου κατωφλιού  $u$  θα γίνει από τα παρακάτω γραφήματα.

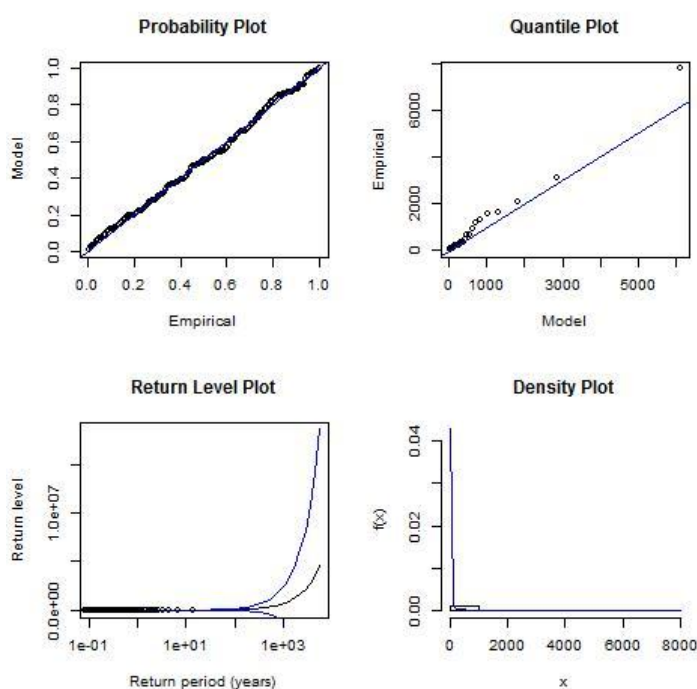
Εικόνα 4.8.8 Διαγράμματα για την επιλογή του βέλτιστου  $u$  στην περίπτωση της Burr (1, 1, 1)



Με χρήση των δύο διαγραμμάτων για το threshold, καταλήγουμε στο συμπέρασμα ότι το μικρότερο  $u$  πέραν του οποίου το  $\xi$  σταθεροποιείται και η εκτίμηση του  $\hat{\sigma}$  μεταβάλλεται γραμμικά ως προς το  $u$  είναι το 32. Αντίστοιχα συμπεράσματα παίρνουμε και από το Mean Excess Plot, όπου παρατηρούμε ότι η μονάδα είναι η τιμή όπου το διάγραμμα γίνεται οριζόντιο και τα δ.ε δεν ξεφεύγουν. Συνεπώς, θα χρησιμοποιηθούν για τον υπολογισμό των εκτιμητών τα δεδομένα από την αρχική κατανομή Burr, τα οποία είναι μεγαλύτερα της τιμής αυτής. Με την εφαρμογή του  $u$  για το πρώτο δείγμα, το ποσοστό των δεδομένων που ακολουθεί GPD με  $(\hat{\sigma}, \hat{\xi}) = (23.34427, 1.10598)$  είναι 3,34 %. Επομένως, ο MLE που θα υπολογισθεί βάσει της μεθόδου Peak Over Threshold θα χρησιμοποιεί δείγματα με περίπου 170 παρατηρήσεις.

Στη συνέχεια, με τον ίδιο γραφικό τρόπο παρατηρούμε ότι τα θεωρητικά και τα παρατηρούμενα σημεία των κατανομών συμπίπτουν, όπως είναι αναμενόμενο όταν η παρατηρούμενη κατανομή είναι η GPD, γεγονός που φαίνεται και από το ιστόγραμμα των Peaks Over Threshold (Density Plot). Αντίστοιχα, από το γράφημα στάθμης απόδοσης (Return Level Plot) βλέπουμε ότι η  $z_p(w)$  αυξάνεται σχηματίζοντας κυρτή καμπύλη ως προς το  $w$ , επομένως η τιμή του δείκτη  $\xi > 0$ .

Εικόνα 4.8.9 Διαγράμματα καλής προσαρμογής στην GPD για την Burr (1, 1, 1)



Παρακάτω, φαίνονται συγκεντρωτικά οι εκτιμώμενες (μέσω προσομοίωσης  $10^5$  επαναλήψεων) παράμετροι των κατανομών των διάφορων εκτιμητών του  $\xi$  από κάθε εκτιμητή για  $k = 25, 50, 100, 500$ .

Πίνακας 4.8.3 Περιγραφικά χαρακτηριστικά της κατανομής των προσομοιωμένων εκτιμητών για την Burr (1, 1, 1)

		<i>Burr (1, 1, 1) (<math>\xi = 1</math>)</i>										
		<i>MLE</i>	<i>MEAN</i>	<i>SE</i>	<i>MSE</i>	<i>DURATION</i>						
BM			1,0118	0,0617	0,0039	5.713,57						
POT			0,9870	0,1646	0,0273	543,30						
		<b>25</b>				<b>50</b>						
		<i>MEAN</i>	<i>SE</i>	<i>MSE</i>	<i>DURATION</i>			<i>MEAN</i>	<i>SE</i>	<i>MSE</i>	<i>DURATION</i>	
Pickand's		0,9675	0,4335	0,1890	205,53	Pickand's		0,9810	0,3054	0,0936	185,83	
Hill's		1,0028	0,2001	0,0401	1.653,95	Hill's		1,0059	0,1420	0,0202	3.170,36	
Negative Hill's		-0,1128	0,1005	1,2483	2.994,38	Negative Hill's		-0,0700	0,0617	1,1488	6.044,51	
Moment		0,8906	0,2925	0,0975	3.185,51	Moment		0,9467	0,2008	0,0432	6.306,57	
Moment		0,9632	0,2594	0,0686	3.352,19	Moment		0,9834	0,1911	0,0368	6.861,65	
Ratio		0,8510	0,4206	0,1991	3.748,93	Ratio		0,9242	0,3018	0,0969	7.976,71	
Peng		0,3444	0,1176	0,4436	3.284,71	Peng		0,4044	0,0734	0,3601	7.348,90	
W		0,6780	0,1938	0,1412	3.413,30	W		0,7282	0,1514	0,0968	6.064,81	
PWM						PWM						

	100					500			
	MEAN	SE	MSE	DURATION		MEAN	SE	MSE	DURATION
Pickand's	0,9899	0,2164	0,0469	185,85	Pickand's	0,9988	0,0975	0,0095	185,78
Hill's	1,0103	0,1004	0,0102	6.247,12	Hill's	1,0537	0,0459	0,0050	31.072,30
Negative Hill's	-0,0419	0,0369	1,0869	12.863,73	Negative Hill's	-0,0116	0,0103	1,0234	66.187,60
Moment	0,9766	0,1415	0,0206	12.540,44	Moment	1,0207	0,0645	0,0046	61.277,39
Moment Ratio	0,9956	0,1379	0,0190	13.514,86	Moment Ratio	1,0248	0,0619	0,0044	64.737,95
Peng	0,9618	0,2175	0,0488	13.855,43	Peng	0,9918	0,0991	0,0099	63.509,04
W	0,4406	0,0473	0,3152	13.084,97	W	0,4804	0,0171	0,2702	65.608,53
PWM	0,7694	0,1203	0,0677	12.032,13	PWM	0,8344	0,0748	0,0330	61.860,10

Βάσει θεωρίας αναμένουμε καλύτερη απόδοση να έχει ο εκτιμητής Hill's. Τα αποτελέσματα που παρουσιάζονται στον παραπάνω πίνακα το επιβεβαιώνουν, μιας και σε όλες σχεδόν τις περιπτώσεις έχει την μικρότερη μεροληψία (πλην της περίπτωσης  $k = 500$ ) και από τα καλύτερα MSE, το οποίο μειώνεται καθώς το  $k$  αυξάνεται. Ο PWM παρουσιάζει πολύ μεγάλη μεροληψία, που βελτιώνεται με την αύξηση του  $k$ . Αντίστοιχη συμπεριφορά παρουσιάζουν οι Peng και W. Οι τρεις παραπάνω εκτιμητές έχουν την τάση να υποεκτιμούν τον δείκτη ακραίων τιμών, τάση που εμφανίζεται και στον εκτιμητή Moment. Με την αύξηση του  $k$ , όχι μόνο διορθώνεται η συμπεριφορά τους, αλλά ο Moment παρουσιάζει, πλέον, ένα από τα μικρότερα MSE. Ο Pickand's, όπως και οι προηγούμενοι εκτιμητές, βελτιώνεται με την αύξηση του  $k$ , με την μεροληψία και το MSE να μειώνονται. Τέλος, το δεύτερο καλύτερο MSE παρουσιάζει ο εκτιμητής Moment Ratio με τιμή 0,0044 για  $k = 500$ .

Παρατηρούμε ότι ο MLE με βάση την BM επιδεικνύει την μικρότερη μεροληψία, με τιμή 0,0118 και το μικρότερο MSE με τιμή 0,0039. Ο MLE με την POT παρουσιάζει μικρή μεροληψία και αρκετά καλό MSE. Παρόλο που στην περίπτωση  $k = 500$  παρουσιάζονται εκτιμητές με πολύ μικρότερο MSE, δεν μπορούμε να αντιπαρέλθουμε της χρήσης του λόγω της ταχύτητας υπολογισμού του.

#### 4.9 ΚΑΤΑΝΟΜΗ FRÉCHET

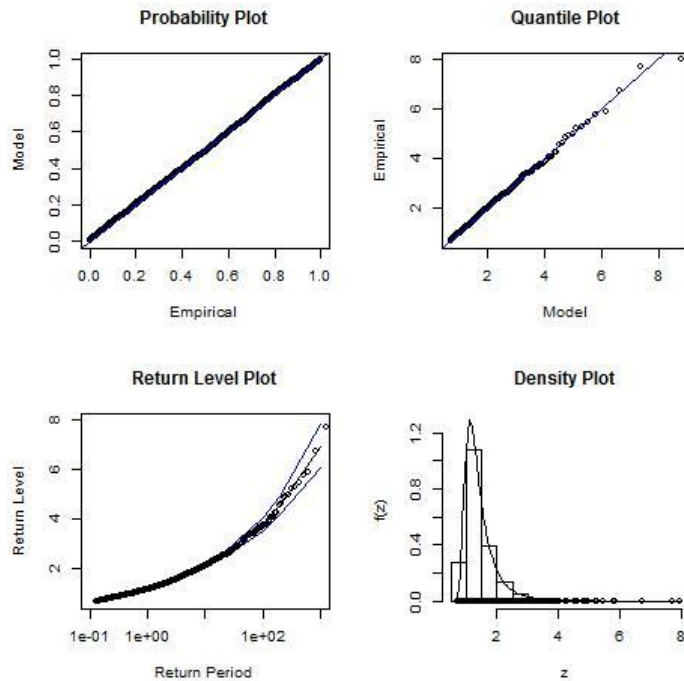
Στην κατανομή Fréchet θα εξετασθούν 3 περιπτώσεις, με  $(\mu, \sigma, \alpha) = (0, 1, 4), (0, 1, 2), (0, 2, 1)$ . Επειδή η Fréchet είναι υποπερίπτωση της GEV, εξακολουθεί να ακολουθεί την οριακή, ακόμα και για πολύ μικρό μέγεθος των blocks. Ως εκ τούτου και στις τρεις περιπτώσεις κρίθηκε ως βέλτιστη επιλογή η χρήση 2.500 blocks, με 2 παρατηρήσεις το καθένα.

Για την κατανομή Fréchet με παραμέτρους  $(\mu, \sigma, \alpha) = (0, 1, 4)$  (που αντιστοιχεί σε  $\xi = 0,25$ ) η προσαρμοσμένη GEV για τα 2.500 maxima είχε παραμέτρους  $(\hat{\mu}, \hat{\sigma}, \hat{\xi}) = (1.1840483, 0.2941643, 0.2618778)$ . Από τα PP-plot και QQ-plot διαφαίνεται ότι τα θεωρητικά και τα παρατηρούμενα σημεία των κατανομών συμπίπτουν, όπως είναι



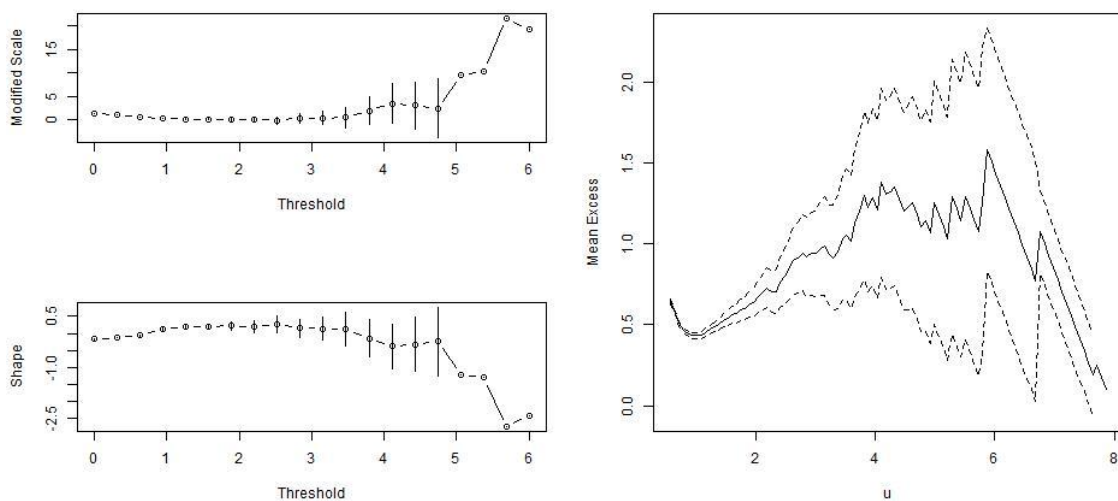
αναμενόμενο όταν η παρατηρούμενη κατανομή είναι η GEV, γεγονός που φαίνεται και από το ιστόγραμμα των Block Maxima (Density Plot). Αντίστοιχα, από το γράφημα στάθμης απόδοσης (Return Level Plot) βλέπουμε ότι η  $z_p(w)$  είναι αύξουσα κυρτή συνάρτηση του  $w$ , άρα  $\xi > 0$

Εικόνα 4.9. Διαγράμματα καλής προσαρμογής στην GEV για την Fréchet (0, 1, 4)



Για την εφαρμογή της μεθόδου Peak Over Threshold η επιλογή του βέλτιστου κατωφλιού  $u$  θα γίνει από τα παρακάτω γραφήματα.

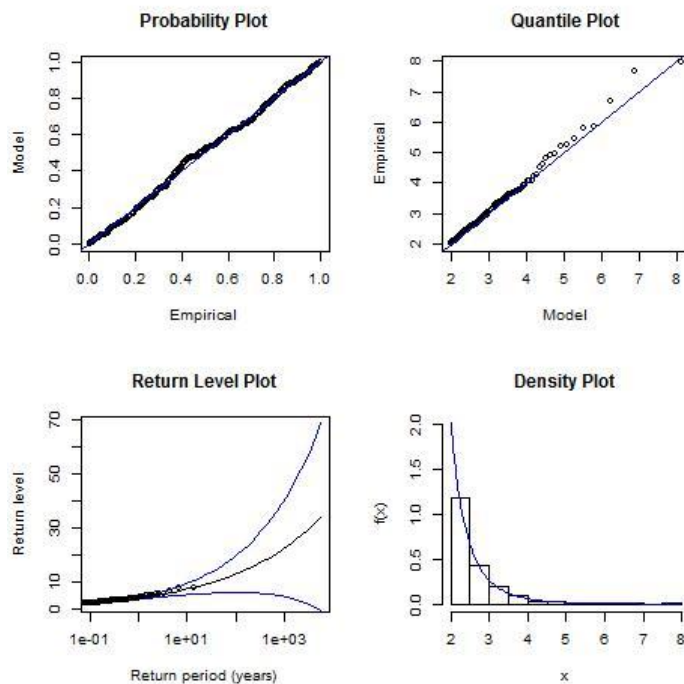
Εικόνα 4.9.2 Διαγράμματα για την επιλογή του βέλτιστου  $u$  στην περίπτωση της Fréchet (0, 1, 4)



Με χρήση των δύο διαγραμμάτων για το threshold, καταλήγουμε στο συμπέρασμα ότι το μικρότερο  $u$  πέραν του οποίου το  $\hat{\xi}$  σταθεροποιείται και η εκτίμηση του  $\hat{\sigma}$  μεταβάλλεται γραμμικά ως προς το  $u$  είναι το 2. Αντίστοιχα συμπεράσματα παίρνουμε και από το Mean Excess Plot, όπου παρατηρούμε ότι η μονάδα είναι η τιμή όπου το διάγραμμα γίνεται οριζόντιο και τα δ.ε δεν ξεφεύγουν. Συνεπώς, θα χρησιμοποιηθούν για τον υπολογισμό των εκτιμητών τα δεδομένα από την αρχική κατανομή Fréchet, τα οποία είναι μεγαλύτερα της τιμής αυτής. Με την εφαρμογή του  $u$  για το πρώτο δείγμα, το ποσοστό των δεδομένων που ακολουθεί GPD με  $(\hat{\sigma}, \hat{\xi}) = (0.4985311, 0.2383379)$  είναι 6,18 %. Επομένως, ο MLE που θα υπολογισθεί βάσει της μεθόδου Peak Over Threshold θα χρησιμοποιεί δείγματα με περίπου 300 παρατηρήσεις.

Στη συνέχεια, με τον ίδιο γραφικό τρόπο παρατηρούμε ότι τα θεωρητικά και τα παρατηρούμενα σημεία των κατανομών συμπίπτουν, όπως είναι αναμενόμενο όταν η παρατηρούμενη κατανομή είναι η GPD, γεγονός που φαίνεται και από το ιστόγραμμα των Peaks Over Threshold (Density Plot). Αντίστοιχα, από το γράφημα στάθμης απόδοσης (Return Level Plot) βλέπουμε ότι η  $z_p(w)$  είναι αύξουσα κυρτή συνάρτηση του  $w$ , άρα  $\xi > 0$ .

Εικόνα 4.9.3 Διαγράμματα καλής προσαρμογής στην GPD για την Fréchet (0, 1, 4)



Παρακάτω, φαίνονται συγκεντρωτικά οι εκτιμώμενες (μέσω προσομοίωσης  $10^5$  επαναλήψεων) παράμετροι των κατανομών των διάφορων εκτιμητών του  $\xi$  από κάθε εκτιμητή για  $k = 25, 50, 100, 500$ .

Πίνακας 4.9.1 Περιγραφικά χαρακτηριστικά της κατανομής των προσομοιωμένων εκτιμητών για την Fréchet (0, 1, 4)

**Fréchet(0, 1, 4) ( $\xi = 0.25$ )**

<i>MLE</i>	<i>MEAN</i>	<i>SE</i>	<i>MSE</i>	<i>DURATION</i>
BM	0,2500	0,0166	0,0003	22.005,19
POT	0,2351	0,0732	0,0056	940,59

	<i>25</i>					<i>50</i>			
	<i>MEAN</i>	<i>SE</i>	<i>MSE</i>	<i>DURATION</i>		<i>MEAN</i>	<i>SE</i>	<i>MSE</i>	<i>DURATION</i>
Pickand's	0,2169	0,3755	0,1421	390,21	Pickand's	0,2281	0,2630	0,0696	386,49
Hill's	0,2502	0,0502	0,0025	3.294,51	Hill's	0,2505	0,0355	0,0013	6.461,11
Negative Hill's	-0,2621	0,1353	0,2806	6.084,73	Negative Hill's	-0,2067	0,0966	0,2180	12.299,37
Moment	0,1387	0,2189	0,0603	6.477,71	Moment	0,1925	0,1473	0,0250	12.767,65
Moment Ratio	0,2404	0,0647	0,0043	6.527,10	Moment Ratio	0,2452	0,0478	0,0023	12.648,50
Peng	0,1289	0,2559	0,0801	6.431,06	Peng	0,1872	0,1771	0,0353	12.676,20
W	0,0712	0,1929	0,0692	6.342,20	W	0,1373	0,1376	0,0316	12.561,82
PWM	0,2512	0,2151	0,0463	6.396,89	PWM	0,2457	0,1611	0,0260	12.671,99

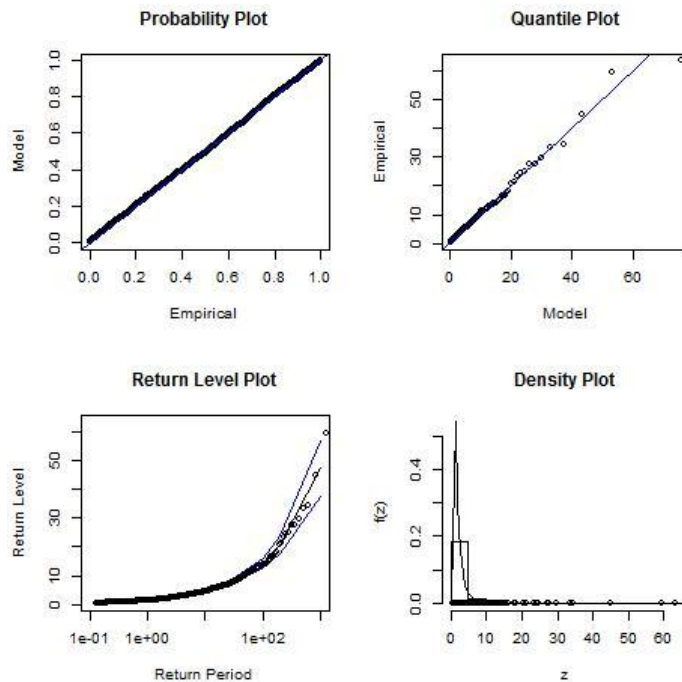
	<i>100</i>					<i>500</i>			
	<i>MEAN</i>	<i>SE</i>	<i>MSE</i>	<i>DURATION</i>		<i>MEAN</i>	<i>SE</i>	<i>MSE</i>	<i>DURATION</i>
Pickand's	0,2262	0,1857	0,0351	385,44	Pickand's	0,1403	0,0823	0,0188	384,57
Hill's	0,2512	0,0251	0,0006	12.773,71	Hill's	0,2566	0,0113	0,0002	62.375,45
Negative Hill's	-0,1617	0,0698	0,1744	24.698,06	Negative Hill's	-0,0930	0,0349	0,1189	124.304,19
Moment	0,2195	0,1029	0,0115	25.548,81	Moment	0,2371	0,0466	0,0023	125.293,84
Moment Ratio	0,2480	0,0345	0,0012	25.220,62	Moment Ratio	0,2527	0,0156	0,0003	125.533,89
Peng	0,2164	0,1255	0,0169	25.164,87	Peng	0,2332	0,0572	0,0036	125.633,72
W	0,1770	0,1042	0,0162	24.998,43	W	0,2184	0,0611	0,0047	125.330,58
PWM	0,2449	0,1187	0,0141	25.093,88	PWM	0,2381	0,0560	0,0033	88.554,49

Από τους εκτιμητές που αναμέναμε να αποδίδουν καλά, ο Hill's φαίνεται να ξεχωρίζει καθολικά με μεροληψία και MSE να πλησιάζουν το μηδέν. Ο εκτιμητής W παρότι δεν είναι ανταγωνιστικός για μικρές τιμές του  $k$ , όσο αυτό αυξάνεται τόσο βελτιώνονται και το MSE και η μεροληψία του, ενώ, ο PWM, σε αντίθεση με τις προσδοκίες, παρότι παρουσιάζει καλή αμεροληψία, έχει αρκετά μεγάλο, μειούμενο όμως, MSE. Εξίσου καλή απόδοση με τον Hill's φαίνεται να έχουν ο εκτιμητής MLE για την μέθοδο BM, καθώς και ο Moment Ratio, με μεροληψία και MSE που όσο αυξάνεται το  $k$  πλησιάζουν το μηδέν. Παρόλα αυτά, ο Moment Ratio είναι πολύ χρονοβόρος και ο MLE ενέχει κινδύνους στην εφαρμογή του, με αποτέλεσμα ο Hill's να είναι προτιμότερος. Οι υπόλοιποι εκτιμητές στο σύνολο τους φαίνεται να υποεκτιμούν τον δείκτη ακραίων τιμών, με καλύτερο όλων τον MLE για την POT με τιμή για το MSE 0,0056 και μεροληψία 0,0149.

Ομοίως, έγινε έλεγχος της προσαρμογής των δεδομένων που παρήχθησαν με την κατανομή Fréchet με παραμέτρους  $(\mu, \sigma, \alpha) = (0, 1, 2)$  (με αντίστοιχο  $\xi = 0,5$ ). Όπως

φαίνεται από τα ακόλουθα διαγράμματα ελέγχου καλής προσαρμογής τα 2.500 Block Maxima, που έχουν επιλεγεί μπορούν να θεωρηθούν, ότι προέρχονται από μία GEV με παραμέτρους εκτιμημένες ως  $(\hat{\mu}, \hat{\sigma}, \hat{\xi}) = (1.4011329, 0.6973275, 0.5135438)$ . Ταυτόχρονα, στο διάγραμμα στάθμης απόδοσης η  $z_p(w)$  είναι αύξουσα κυρτή ως προς το  $w$ , επομένως ο δείκτης  $\xi > 0$ .

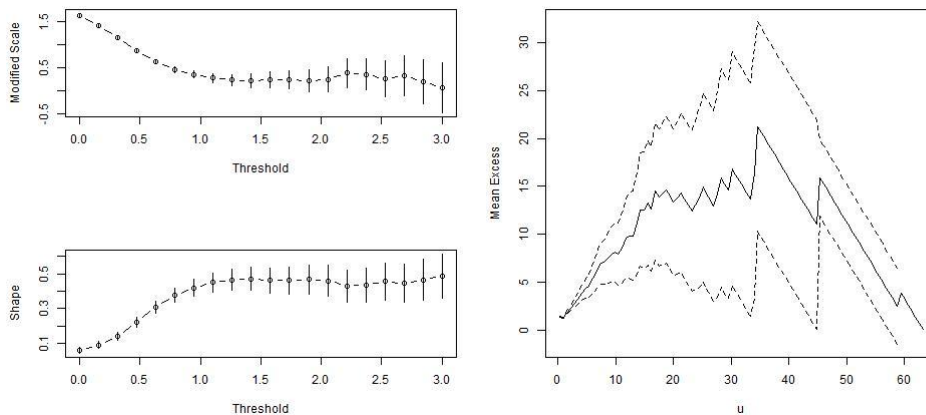
Εικόνα 4.9.4 Διαγράμματα καλής προσαρμογής στην GEV για την Fréchet (0, 1, 2)



Από τα PP-plot και QQ-plot διαφαίνεται ότι τα θεωρητικά και τα παρατηρούμενα σημεία των κατανομών συμπίπτουν, όπως είναι αναμενόμενο όταν η παρατηρούμενη κατανομή είναι η GEV, γεγονός που φαίνεται και από το ιστόγραμμα των Block Maxima (Density Plot). Αντίστοιχα, από το γράφημα στάθμης απόδοσης (Return Level Plot) βλέπουμε ότι η  $z_p(w)$  είναι αύξουσα κυρτή συνάρτηση του  $w$ , άρα  $\xi > 0$ .

Για την εφαρμογή της μεθόδου Peak Over Threshold η επιλογή του βέλτιστου κατωφλιού  $u$  θα γίνει από τα παρακάτω γραφήματα.

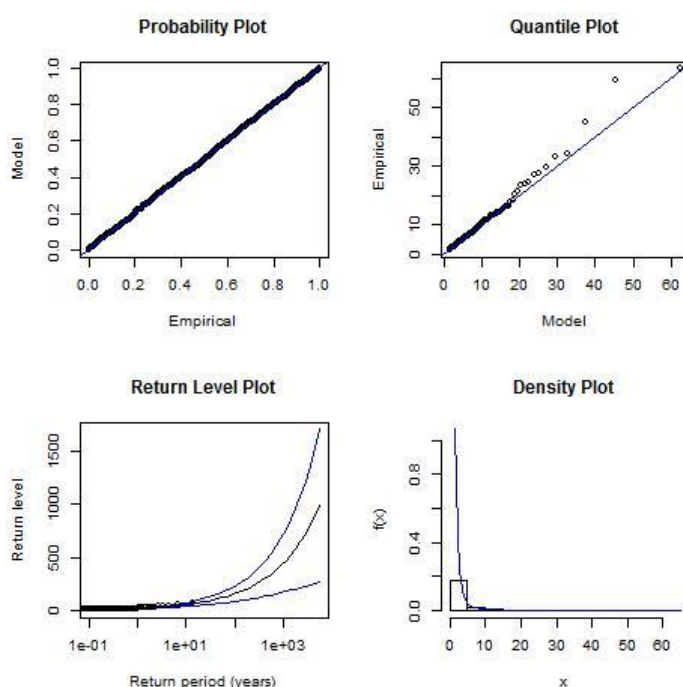
Εικόνα 4.9.5 Διαγράμματα για την επιλογή του βέλτιστου  $u$  στην περίπτωση της Fréchet (0, 1, 2)



Με χρήση των δύο διαγραμμάτων για το threshold, καταλήγουμε στο συμπέρασμα ότι το μικρότερο  $u$  πέραν του οποίου το  $\xi$  σταθεροποιείται και η εκτίμηση του  $\hat{\sigma}$  μεταβάλλεται γραμμικά ως προς το  $u$  είναι το 1,5. Αντίστοιχα συμπεράσματα παίρνουμε και από το Mean Excess Plot, όπου παρατηρούμε ότι η μονάδα είναι η τιμή όπου το διάγραμμα γίνεται οριζόντιο και τα δ.ε δεν ξεφεύγουν. Συνεπώς, θα χρησιμοποιηθούν για τον υπολογισμό των εκτιμητών τα δεδομένα από την αρχική κατανομή Fréchet, τα οποία είναι μεγαλύτερα της τιμής αυτής. Με την εφαρμογή του  $u$  για το πρώτο δείγμα, το ποσοστό των δεδομένων που ακολουθεί GPD με  $(\hat{\sigma}, \hat{\xi}) = (0.9377859, 0.4593706)$  είναι 35,08 %. Επομένως, ο MLE που θα υπολογισθεί βάσει της μεθόδου Peak Over Threshold θα χρησιμοποιεί δείγματα με περίπου 1.750 παρατηρήσεις.

Στη συνέχεια, με τον ίδιο γραφικό τρόπο παρατηρούμε ότι τα θεωρητικά και τα παρατηρούμενα σημεία των κατανομών συμπίπτουν, όπως είναι αναμενόμενο όταν η παρατηρούμενη κατανομή είναι η GPD, γεγονός που φαίνεται και από το ιστόγραμμα των Peaks Over Threshold (Density Plot). Αντίστοιχα, από το γράφημα στάθμης απόδοσης (Return Level Plot) βλέπουμε ότι η  $z_p(w)$  είναι αύξουσα κυρτή συνάρτηση του  $w$ , άρα  $\xi > 0$ .

Εικόνα 4.9.6 Διαγράμματα καλής προσαρμογής στην GPD για την Fréchet (0, 1, 2)



Παρακάτω, φαίνονται συγκεντρωτικά οι εκτιμώμενες (μέσω προσομοίωσης  $10^5$  επαναλήψεων) παράμετροι των κατανομών των διάφορων εκτιμητών του  $\xi$  από κάθε εκτιμητή για  $k = 25, 50, 100, 500$ .

Πίνακας 4.9.2 Περιγραφικά χαρακτηριστικά της κατανομής των προσομοιωμένων εκτιμητών για την Fréchet (0, 1, 2)

**Fréchet (0, 1, 2) ( $\xi = 0.5$ )**

<i>MLE</i>	<i>MEAN</i>	<i>SE</i>	<i>MSE</i>	<i>DURATION</i>
BM	0,5003	0,0194	0,0004	26.451,29
POT	0,4630	0,0348	0,0026	3.813,36

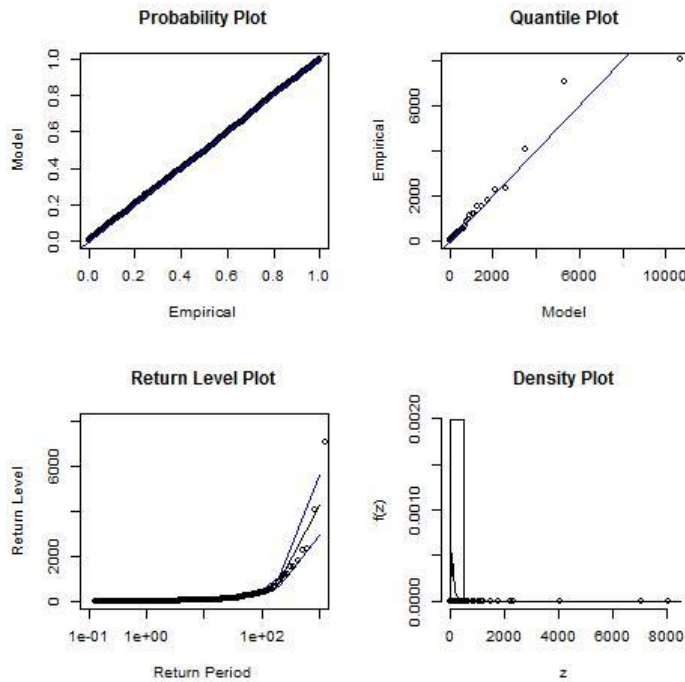
	<b>25</b>					<b>50</b>			
	<i>MEAN</i>	<i>SE</i>	<i>MSE</i>	<i>DURATION</i>		<i>MEAN</i>	<i>SE</i>	<i>MSE</i>	<i>DURATION</i>
Pickand's	0,4655	0,3915	0,1545	250,84	Pickand's	0,4795	0,2746	0,0758	251,25
Hill's	0,5003	0,1005	0,0101	2.475,38	Hill's	0,5010	0,0709	0,0050	4.235,28
Negative Hill's	-0,1945	0,1249	0,4979	3.544,45	Negative Hill's	-0,1408	0,0857	0,4179	9.385,77
Moment	0,3889	0,2354	0,0678	3.686,58	Moment	0,4430	0,1596	0,0287	7.253,58
Moment Ratio	0,4807	0,1294	0,0171	3.701,58	Moment Ratio	0,4904	0,0956	0,0092	9.525,31
Peng	0,3693	0,3064	0,1110	7.066,72	Peng	0,4325	0,2164	0,0514	14.733,82
W	0,2012	0,1691	0,1179	4.527,92	W	0,2723	0,1205	0,0664	7.196,95
PWM	0,4259	0,2163	0,0523	3.653,45	PWM	0,4449	0,1698	0,0319	7.198,04
	<b>100</b>					<b>500</b>			
	<i>MEAN</i>	<i>SE</i>	<i>MSE</i>	<i>DURATION</i>		<i>MEAN</i>	<i>SE</i>	<i>MSE</i>	<i>DURATION</i>

Pickand's	0,4810	0,1939	0,0380	250,53	Pickand's	0,4235	0,0862	0,0133	259,96
Hill's	0,5024	0,0501	0,0025	9.675,07	Hill's	0,5132	0,0226	0,0007	35.821,73
Negative Hill's	-0,0999	0,0587	0,3633	15.315,11	Negative Hill's	-0,0438	0,0245	0,2964	76.621,90
Moment	0,4707	0,1116	0,0133	14.416,33	Moment	0,4937	0,0505	0,0026	76.987,55
Moment Ratio	0,4961	0,0690	0,0048	15.352,25	Moment Ratio	0,5055	0,0313	0,0010	82.157,83
Peng	0,4644	0,1547	0,0252	15.923,65	Peng	0,4860	0,0706	0,0052	71.967,99
W	0,3189	0,0912	0,0411	14.248,90	W	0,3802	0,0548	0,0174	71.576,39
PWM	0,4618	0,1328	0,0191	14.285,71	PWM	0,4815	0,0748	0,0059	70.970,99

Η προσομοίωση δεδομένων από την Fréchet (0, 1, 2) για την αξιολόγηση των εκτιμητών του δείκτη ακραίων τιμών υποδεικνύει τους Hill's και Moment Ratio ως βέλτιστους μεταξύ των εκτιμητών κλειστού τύπου, με MSE που προσεγγίζει το μηδέν, καθώς αυξάνεται το  $k$ , και μεροληψίες πολύ μικρές. Ο εκτιμητής Moment τείνει να υποεκτιμά τον δείκτη, παρόλα αυτά αυξάνοντας το  $k$  η τάση αυτή περιορίζεται και ο MSE μειώνεται σε σημείο που φτάνει για  $k = 500$  να είναι ένας από τους καλύτερους. Αντίστοιχη συμπεριφορά έχουν και οι εκτιμητές Pickand, Peng και W χωρίς, όμως να τον πλησιάζουν σε απόδοση. Παρατηρούμε πως, ο MLE για την POT είναι καθολικά καλύτερος με MSE 0,0004 και μεροληψία 0,0003, απαιτώντας όμως πολύ χρόνο για τον υπολογισμό του. Σε κάθε περίπτωση, πλην της  $k = 500$ , το αμέσως μικρότερο MSE έχει ο MLE υπολογισμένος με την μέθοδο BM.

Στην περίπτωση της Fréchet με παραμέτρους  $(\mu, \sigma, \alpha) = (0, 2, 1)$  (με αντίστοιχο  $\xi = 1$ ) Η εκτίμηση με τη μέθοδο μεγίστης πιθανοφάνειας των παραμέτρων της προσαρμοσμένης κατανομής των 200 maxima που θα χρησιμοποιήσουμε έδωσε τις ακόλουθες τιμές  $(\hat{\mu}, \hat{\sigma}, \hat{\xi}) = (3.931769, 3.926562, 1.012446)$ . Η καλή προσαρμογή των δεδομένων φαίνεται παρακάτω.

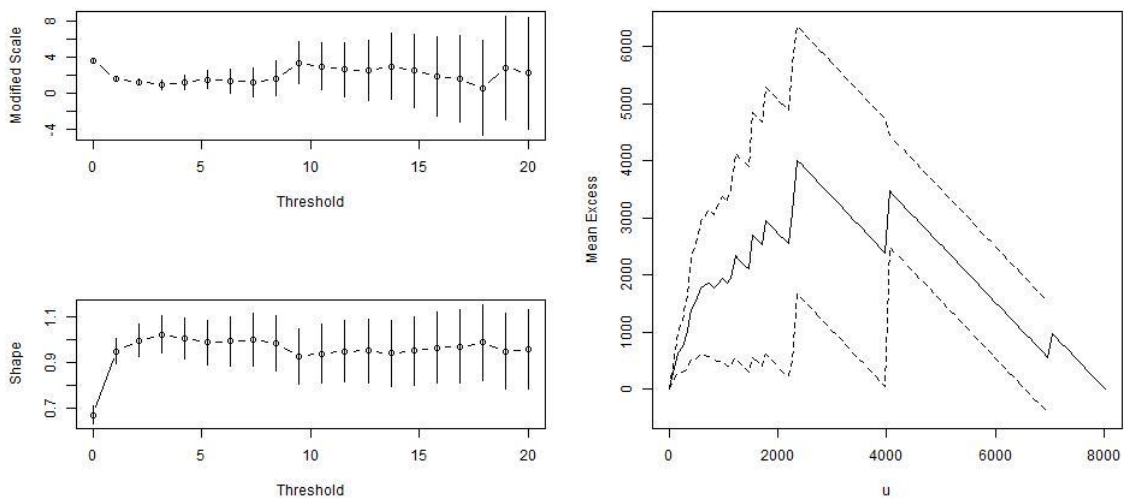
Εικόνα 4.9.7 Διαγράμματα καλής προσαρμογής στην GEV για την Fréchet (0, 2, 1)



Από τα γραφήματα PP και QQ φαίνεται, ότι υπάρχει καλή προσαρμογή των δεδομένων στην GEV, ενώ ταυτόχρονα φαίνεται και από το ιστόγραμμα των δεδομένων. Το διάγραμμα στάθμης απόδοσης εμφανίζει κυρτή καμπύλη επομένως μπορούμε να διαπιστώσουμε, ότι η παράμετρος  $\xi$  είναι θετική ( $\xi > 0$ ).

Για την εφαρμογή της μεθόδου Peak Over Threshold η επιλογή του βέλτιστου κατωφλιού  $u$  θα γίνει από τα παρακάτω γραφήματα.

Εικόνα 4.9.8 Διαγράμματα για την επιλογή του βέλτιστου  $u$  στην περίπτωση της Fréchet (0, 2, 1)

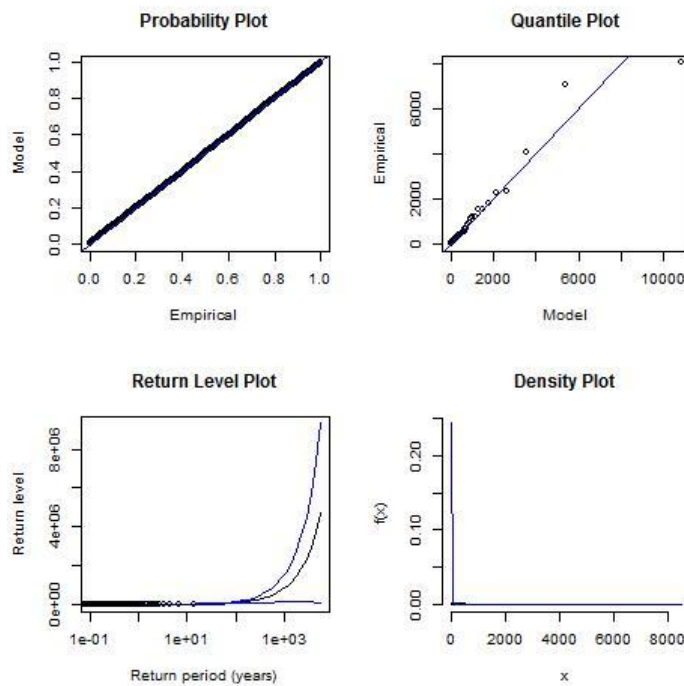




Με χρήση των δύο διαγραμμάτων για το threshold, καταλήγουμε στο συμπέρασμα ότι το μικρότερο  $u$  πέραν του οποίου το  $\xi$  σταθεροποιείται και η εκτίμηση του  $\hat{\sigma}$  μεταβάλλεται γραμμικά ως προς το  $u$  είναι το 3. Αντίστοιχα συμπεράσματα παίρνουμε και από το Mean Excess Plot, όπου παρατηρούμε ότι η μονάδα είναι η τιμή όπου το διάγραμμα γίνεται οριζόντιο και τα δ.ε δεν ξεφεύγουν. Συνεπώς, θα χρησιμοποιηθούν για τον υπολογισμό των εκτιμητών τα δεδομένα από την αρχική κατανομή Fréchet, τα οποία είναι μεγαλύτερα της τιμής αυτής. Με την εφαρμογή του  $u$  για το πρώτο δείγμα, το ποσοστό των δεδομένων που ακολουθεί GPD με  $(\hat{\sigma}, \xi) = (4.100844, 1.013505)$  είναι 48,18 %. Επομένως, ο MLE που θα υπολογισθεί βάσει της μεθόδου Peak Over Threshold θα χρησιμοποιεί δείγματα με περίπου 2.400 παρατηρήσεις.

Στη συνέχεια, με τον ίδιο γραφικό τρόπο παρατηρούμε ότι τα θεωρητικά και τα παρατηρούμενα σημεία των κατανομών συμπίπτουν, όπως είναι αναμενόμενο όταν η παρατηρούμενη κατανομή είναι η GPD, γεγονός που φαίνεται και από το ιστόγραμμα των Peaks Over Threshold (Density Plot). Αντίστοιχα, από το γράφημα στάθμης απόδοσης (Return Level Plot) βλέπουμε ότι η  $z_p(w)$  είναι αύξουσα κυρτή συνάρτηση του  $w$ , άρα  $\xi > 0$ .

Εικόνα 4.9.9 Διαγράμματα καλής προσαρμογής στην GPD για την Fréchet (0, 2, 1)



Παρακάτω, φαίνονται συγκεντρωτικά οι εκτιμώμενες (μέσω προσομοίωσης  $10^5$  επαναλήψεων) παράμετροι των κατανομών των διάφορων εκτιμητών του  $\xi$  από κάθε εκτιμητή για  $k = 25, 50, 100, 500$ .

Πίνακας 4.9.3 Περιγραφικά χαρακτηριστικά της κατανομής των προσομοιωμένων εκτιμητών για την Fréchet (0, 2, 1)

<i>Fréchet (0, 2, 1) (<math>\xi = 1</math>)</i>									
<i>MLE</i>	<i>MEAN</i>	<i>SE</i>	<i>MSE</i>	<i>DURATION</i>		<i>MEAN</i>	<i>SE</i>	<i>MSE</i>	<i>DURATION</i>
BM	1,0003	0,0643	0,0041			37.961,88			
POT	0,9877	0,0403	0,0018			3.343,92			

<i>25</i>					<i>50</i>				
	<i>MEAN</i>	<i>SE</i>	<i>MSE</i>	<i>DURATION</i>		<i>MEAN</i>	<i>SE</i>	<i>MSE</i>	<i>DURATION</i>
Pickand's	0,9626	0,4341	0,1898	336,99	Pickand's	0,9820	0,3053	0,0935	336,63
Hill's	1,0006	0,2010	0,0404	3.241,91	Hill's	1,0020	0,1418	0,0201	6.416,55
Negative Hill's	-0,1126	0,1000	1,2478	6.016,71	Negative Hill's	-0,0700	0,0616	1,1487	12.235,30
Moment	0,8892	0,2925	0,0978	6.424,84	Moment	0,9440	0,2014	0,0437	12.715,36
Moment Ratio	0,9614	0,2588	0,0685	6.529,66	Moment Ratio	0,9809	0,1911	0,0369	12.566,91
Peng	0,8500	0,4193	0,1983	6.330,75	Peng	0,9229	0,3025	0,0974	12.577,18
W	0,3444	0,1174	0,4436	6.277,65	W	0,4042	0,0739	0,3604	12.471,46
PWM	0,6778	0,1938	0,1414	6.383,60	PWM	0,7276	0,1522	0,0974	12.598,36

<i>100</i>					<i>500</i>				
	<i>MEAN</i>	<i>SE</i>	<i>MSE</i>	<i>DURATION</i>		<i>MEAN</i>	<i>SE</i>	<i>MSE</i>	<i>DURATION</i>
Pickand's	0,9902	0,2156	0,0466	333,45	Pickand's	0,9871	0,0968	0,0095	331,45
Hill's	1,0047	0,1002	0,0101	12.716,41	Hill's	1,0264	0,0453	0,0027	62.279,83
Negative Hill's	-0,0419	0,0369	1,0869	24.613,69	Negative Hill's	-0,0116	0,0103	1,0234	124.178,37
Moment	0,9731	0,1413	0,0207	25.470,71	Moment	1,0069	0,0639	0,0041	124.947,02
Moment Ratio	0,9921	0,1381	0,0191	25.068,05	Moment Ratio	1,0109	0,0625	0,0040	125.172,52
Peng	0,9605	0,2177	0,0490	25.035,98	Peng	0,9914	0,0994	0,0100	125.091,55
W	0,4404	0,0476	0,3154	24.889,01	W	0,4803	0,0172	0,2703	124.898,88
PWM	0,7688	0,1207	0,0680	24.950,87	PWM	0,8339	0,0749	0,0332	85.594,47

Για την Fréchet (0, 2, 1) και στην περίπτωση όπου εκτιμούμε τον δείκτη ακραίων τιμών κοντά στην μονάδα, αναμένουμε καλύτερη απόδοση από τον εκτιμητή Hill's. Πράγματι, ο εκτιμητής Hill's παρουσιάζει πολύ μικρή μεροληψία και στις τέσσερις περιπτώσεις που δοκιμάσαμε, καθώς και πολύ μικρή τιμή για το MSE, η οποία μάλιστα είναι η μικρότερη που παρουσιάζουν οι εκτιμητές κλειστού τύπου που συγκρίνονται όταν το  $k = 500$ . Οι εκτιμητές Pickand's, Peng και PWM, υποεκτιμούν τον δείκτη ακραίων τιμών ενώ παράλληλα δεν εμφανίζουν καλό MSE, παρόλο που βελτιώνονται καθώς το  $k$  αυξάνεται. Τάσεις υποεκτίμησης εμφανίζουν και οι εκτιμητές Moment και Moment Ratio, ωστόσο, όταν το  $k$  πάρει την τιμή 500 παρουσιάζουν πάρα πολύ μικρή μεροληψία και MSE, με τον Moment Ratio να είναι οριακά καλύτερος. Τέλος ο εκτιμητής W δεν πλησιάζει τη μονάδα και παρουσιάζει πολύ μεγάλη μεροληψία για όλες τις τιμές του  $k$

Με την χρήση των εκτιμητών Μεγίστης Πιθανοφάνειας, παρατηρούνται άριστα αποτελέσματα, καθώς με την μέθοδο POT εμφανίζεται η μικρότερη τιμή του MSE, ενώ με τη χρήση της BM, ο αντίστοιχος MLE παρουσιάζει την μικρότερη μεροληψία.



## 5 ΣΥΜΠΕΡΑΣΜΑΤΑ

---

Γενικά, στο σύνολο τους, οι εκτιμητές που μελετήθηκαν, μοιράζονται κάποια θετικά χαρακτηριστικά, όπως η ασθενής συνέπεια και η ασυμπτωτική κανονικότητα υπό προϋποθέσεις ανά περίπτωση. Παρόλα αυτά, στην πράξη, με την εφαρμογή σε πραγματικά ή προσομοιωμένα δεδομένα κάποιοι ξεχώρισαν σε συγκεκριμένες περιπτώσεις.

Συγκεκριμένα, όταν ο δείκτης ακραίων τιμών παίρνει τιμές κοντά στο μηδέν, παρατηρήθηκε ότι στις περισσότερες περιπτώσεις οι εκτιμητές Moment και Pickand's απέδιδαν καλύτερα, με τον εκτιμητή Hill's να δίνει ελαφρά καλύτερα αποτελέσματα έναντι των άλλων δύο στην περίπτωση της Κανονικής κατανομής. Ακόμα, με χρήση των εκτιμητών Μέγιστης Πιθανοφάνειας παρατηρήθηκε ότι καλύτερη εκτίμηση δίνουν δεδομένα τα οποία συλλέχθηκαν με την μέθοδο Peak over Threshold.

Για την περίπτωση όπου ο δείκτης ακραίων τιμών είναι ίσος με 0,25, παρατηρείται ότι η καλύτερη ασυμπτωτική μέθοδος παρουσιάζεται με δεδομένα τα οποία συλλέχθηκαν με την μέθοδο Block Maxima. Καλύτερος εκτιμητής είναι ο PWM, ενώ ακολουθούν οι εκτιμητές Hill's και Moment. Όσο για την περίπτωση όπου προβλέπεται ο δείκτης ακραίων τιμών να πάρει την τιμή 0,50, ως καλύτερη ασυμπτωτική μέθοδος παρουσιάζεται και πάλι η Block Maxima, ενώ βέλτιστοι εκτιμητές με κλειστό τύπο είναι οι Hill's και Moment Ratio. Στην τελευταία περίπτωση για θετικό δείκτη ακραίων τιμών, όπου  $\xi = 1$ , πάλι ως αποδοτικότερη ασυμπτωτική μέθοδος κρίνεται η Block Maxima, ενώ καλύτεροι εκτιμητές παρουσιάστηκαν οι Hill's και Moment Ratio.

Τέλος, για αρνητικό δείκτη ακραίων τιμών, παρουσιάζεται ως βέλτιστη ασυμπτωτική μέθοδος εκτίμησης η εκτιμητήρια Μέγιστης Πιθανοφάνειας με χρήση δεδομένων τα οποία συλλέχθηκαν με την μέθοδο Peak over Threshold. Στην περίπτωση όπου  $\xi = -1$ , καλύτεροι εκτιμητές κλειστού τύπου είναι οι Negative Hill's και PWM, ενώ όταν  $\xi = -0.25$ , καλύτερος παρουσιάζεται ο W, με τον PWM να ακολουθεί.

Στη συνέχεια, θα παρουσιάσουμε κάποια γενικά συμπεράσματα για κάθε έναν εκτιμητή. Ξεκινώντας με τον Pickand's, παρατηρούμε ότι στις περισσότερες περιπτώσεις έχει πολύ μικρή μεροληψία, ωστόσο παρουσιάζει συνήθως μεγάλο MSE. Συγκριτικά με τους υπόλοιπους, έδωσε καλύτερα αποτελέσματα για  $\xi = 0$  και το πιο σημαντικό χαρακτηριστικό του είναι πως ανεξαρτήτως του  $k$  έχει πολύ μικρό χρόνο υπολογισμού, μιας και χρησιμοποιεί μόνο τρεις παρατηρήσεις. Στον συγκεκριμένο εκτιμητή, δεν παρατηρείται κάποια τάση με την αλλαγή του  $k$ .

Ο εκτιμητής Hill's, παρόλο που προτείνεται για  $\xi > 0$ , απέδωσε πάρα πολύ καλά και στην περίπτωση του  $\xi = 0$ . Από τους εκτιμητές με κλειστό τύπο έχει τον δεύτερο μικρότερο χρόνο υπολογισμού, ενώ η αύξηση του  $k$ , με εξαίρεση την ακραία περίπτωση του  $k = 500$ , φαίνεται να τον βελτιώνει και ως προς την μεροληψία και ως προς το MSE.

Ο Negative Hill's, απέδωσε τα αναμενόμενα, καθώς έχει πάρα πολύ μικρή μεροληψία και MSE στην περίπτωση  $\hat{\xi} = -1$  και καθώς αυξάνεται το  $k$  τόσο καλύτερα αποτελέσματα δίνει, ακόμα και στην περίπτωση  $k = 500$ . Αρνητικό χαρακτηριστικό του είναι ο αρκετά μεγάλος χρόνος υπολογισμού του.

Ο εκτιμητής των Ροπών, δίνει καλύτερα αποτελέσματα όταν  $\hat{\xi} \geq 0$ , ενώ στις αρνητικές περιπτώσεις παρουσιάζει αρκετά μεγάλη μεροληψία. Ακόμα, αυξάνοντας το  $k$  βελτιώνεται η απόδοση του, ενώ στα αρνητικά του συγκαταλέγεται ο μεγάλος χρόνος υπολογισμού.

Η βελτίωση του παραπάνω εκτιμητή, ο Moment Ratio, δίνει καλύτερα αποτελέσματα για  $\hat{\xi} > 0$  με αρκετά μικρό MSE και μεροληψία, ενώ ο χρόνος υπολογισμού του είναι μεγάλος.

Ο τελευταίος εκτιμητής που χρησιμοποιεί ροπές για τον υπολογισμό του, ο PWM, δίνει πολύ καλά αποτελέσματα για  $\hat{\xi} < 1/2$ , γεγονός που επαληθεύει την θεωρία μιας και η σύγκλιση του για κάθε άλλο  $\xi$  είναι αρκετά αργή, ενώ η αύξηση του  $k$  με εξαίρεση την ακραία περίπτωση του  $k = 500$ , φαίνεται να τον βελτιώνει και ως προς την μεροληψία και ως προς το MSE. Και σε αυτόν παρατηρείται μεγάλος χρόνος υπολογισμού.

Ο εκτιμητής Peng, φαίνεται να δίνει καλές προσεγγίσεις, όχι όμως τις καλύτερες, σε τιμές του δείκτη ακραίων τιμών κοντά στο μηδέν, όχι όμως στο μηδέν. Ο χρόνος υπολογισμού του είναι, επίσης, πολύ μεγάλος και καθώς το  $k$  αυξάνεται τείνει να δίνει πολύ καλύτερα αποτελέσματα, ενώ συνήθως υποεκτιμά τον δείκτη.

Τέλος, ο εκτιμητής W, δίνει αρκετά καλές εκτιμήσεις για  $\hat{\xi} < 1/2$ , ωστόσο δεν δίνει τις καλύτερες και συνήθως υποεκτιμά ελαφρά τον δείκτη. Με χρήση  $k = 500$  βελτιώνεται αισθητά η απόδοση του και παρουσιάζει πολύ καλό MSE. Συγκριτικά με τους προηγούμενους εκτιμητές παρουσιάζει αισθητά μικρότερο χρόνο υπολογισμού.

Συνεχίζοντας στις ασυμπτωτικές μεθόδους εκτίμησης του δείκτη, δεν θα έπρεπε να γίνεται άμεση σύγκριση μεταξύ τους στις παραπάνω περιπτώσεις μιας και η κάθε μία χρησιμοποιεί διαφορετικό πλήθος δεδομένων. Αυτό συμβαίνει διότι σε κάθε περίπτωση επιλέχθηκε ο βέλτιστος αριθμός BM και POT, ώστε να επιτυγχάνεται η καλύτερη προσαρμογή στην εκάστοτε οικογένεια GEV και GPD αντίστοιχα. Η διαφορετικότητα τους διαφαίνεται κυρίως στην κατανομή Fréchet, η οποία συγκαταλέγεται στην οικογένεια GEV και δίνει τη δυνατότητα να γίνει χρήση μεγάλου πλήθους δεδομένων για τον υπολογισμό του MLE, σε αντίθεση με την περίπτωση της POT. Ωστόσο, σε γενικές γραμμές θα μπορούσε να σχολιαστεί ότι για  $\hat{\xi} \leq 0$ , βέλτιστα αποτελέσματα δίνει ο  $MLE^{POT}$ , ενώ για  $\hat{\xi} > 0$  ο  $MLE^{BM}$ .

Μια ερώτηση που εγείρεται από την παρούσα διπλωματική είναι κατά πόσο θα ήταν δόκιμη μια ασυμπτωτική προσέγγιση για την εκτίμηση (MLE) ή θα ήταν προτιμότερη η χρήση εκτιμητών κλειστού τύπου. Η ασυμπτωτική μέθοδος, όπως είδαμε στο προηγούμενο κεφάλαιο, έδινε σταθερά πολύ καλές εκτιμήσεις. Η μεροληψία των MLE ήταν πάντοτε πολύ μικρή, το MSE πολλές φορές προσέγγιζε το μηδέν, ενώ η διάρκεια υπολογισμού

ήταν πολύ μικρή. Παρόλα αυτά, η προσέγγιση αυτή, περιλαμβάνει την εφαρμογή της εκάστοτε μεθόδου επιλογής δεδομένων, η οποία και απαιτεί χρόνο και αυξάνει τις πιθανότητες λάθους.

Η ποικιλία των εκτιμητών, καθώς και η ευρεία χρήση τους ακόμα και σήμερα, δείχνουν, ότι δεν υπάρχει καθολικά βέλτιστος εκτιμητής, παρότι ξεχωρίζουν ανάμεσα στις διάφορες προσεγγίσεις οι εκτιμητές των Pickand, Hill και Ροπών ως οι αρχαιότεροι. Οι περισσότεροι από τους εκτιμητές που ακολούθησαν, είναι κατά κύριο λόγο τροποποιήσεις και βελτιώσεις επί των αρχικών εκτιμητών, οι οποίες σε πολλές περιπτώσεις φαίνεται να αποδίδουν καλύτερα από τους αρχικούς. Η επιλογή κάποιου εκτιμητή για μια ανάλυση επαφίεται στην εξοικείωση του ερευνητή στο αντικείμενο, στον χρόνο και την υπολογιστική ισχύ που διαθέτει και στην ακρίβεια των αποτελεσμάτων που επιθυμεί.

Συμπληρωματικά με αυτήν την διπλωματική θα ήταν ενδιαφέρον να μελετηθούν η εύρεση του βέλτιστου  $k$  ανά περίπτωση, καθώς και η αλλαγή του πλήθους των αρχικών παρατηρήσεων που παίρνουμε από κάθε κατανομή.

## 6 ΒΙΒΛΙΟΓΡΑΦΙΑ

---

- Aban, I. & Meerschaert, M., 2001. *Shifted Hill's Estimator for Heavy Tails*. 4 επιμ. s.l.:Communication in Statistics Part B: Simulation and Computation.
- Balakrishnan, N. & Clifford Cohen, A., 1991. *Order Statistic and Inference Estimation Methods*. San Diego, CA: Academic Press Inc..
- Balakrishnan, N. & Clifford Cohen, A., 1991. *Order Statistics & Inference: Estimation Methods*. San Diego, CA: Academic Press, INC.
- Beirlant, J. και συν., 2004. *Statistics of Extremes Theory and Applications*. Chichester(West Sussex): Jon Wiley & Sons, Ltd.
- Boutsikas, M., 2008-2010. *Extreme Value Theory*. Piraeus: s.n.
- Danielsson, J., Jansen, D. & deVries, C., 1996. *The Method of Moment Ratio Estimator for the Tail Shape Distribution*. s.l.:Communication of Statistics (A) - Theory and Methods.
- Deheuvels, P., de Haan, L., Peng, L. & Pereira, T., 1997. *Comparison of Extreme Value Index Estimators*. s.l.:NEPTUNE T400:EUR-09.
- Dekkers, A. & De Haan, L., 1989. *On The Estimation Of Extreme Value Index And Large Quantile Estimation*. s.l.:Ann. Statist..
- Dekkers, A., Einmal, J. & de Haan, L., 1989. *A Moment Estimator of the Index of an Extreme-Value Distribution*. s.l.:The Annals of Statistics.
- Embrechts, P., Klüppelberg, C. & Mikosch, T., 1997. *Modelling Extremal Events for Insurance and Finance*. s.l.:Springer.
- Falk, M., 1995. *On Testing The Extreme Value Index Via the POT-method*. s.l.:Ann. Statist..
- Fisher, R. & Tippett, L., 1928. *Limiting Forms of the Frequency Distribution of the Largest or Smallest Member of a Sample*. s.l.:Proc. Cambridge Philos. Soc..
- Gnedenko, B., 1943. *Sur la distribution limite du terme maximum d'une série aleatoire '*. s.l.:Ann. Math.
- Greenwood, J., Landwehr, J., Matalas, N. & Wallis, J., 1979. *Probability Weighted Moments: Definition and Relation to Parameters of Several Distributions Expressible in Inverse Form*. s.l.:Water Resources Research.
- Gruber, S. και συν., 2016. *Guide to easy SHARE release 5.0.0*, s.l.: s.n.
- Hosking, J., 1985. *Maximum Likelihood Estimation of the Parameter of the Generalized Extreme Value Distribution*. s.l.:Applied Statistics.
- Hosking, J. & Wallis, J., 1987. *Parameter and Quantile Estimation for the Generalized Pareto Distribution*. s.l.:Technometrics.



Kotz, S. & Nadarajah, S., 2000. *Extreme Values Distributions: Theory and Applications*. s.l.:Imperial College Press.

Landwehr, J., Matalas, N. & Wallis, J., 1979. *Probability Weighted Moments Compared with Some Traditional Techniques in Estimating Gumbel Parameters and Quantiles*. s.l.:Water Resources Research.

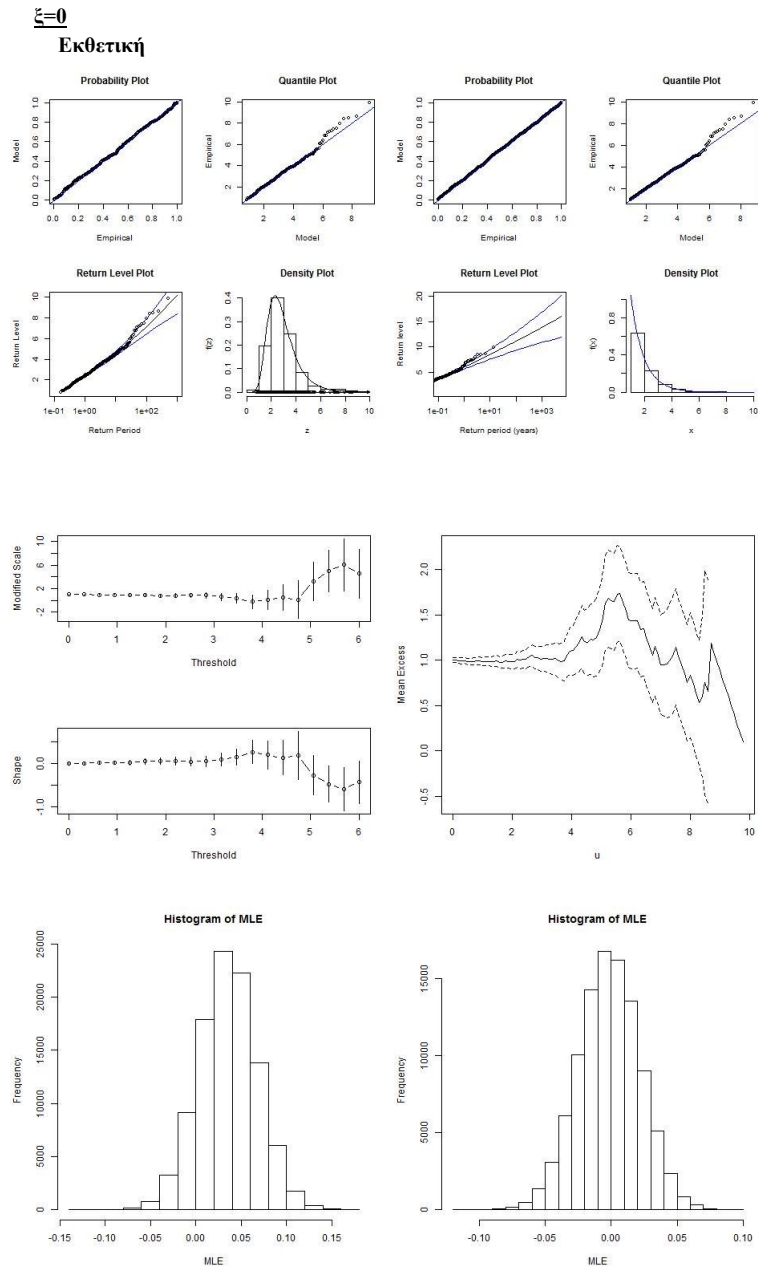
Pichands, J., 1975. *Statistical Inference Using Extreme Order Statistics*. s.l.:The Annals of Statistics.

Smith, R., 1985. *Maximum Likelihood Estimation in a Class of Non-Regular Cases*. s.l.:Biometrika.

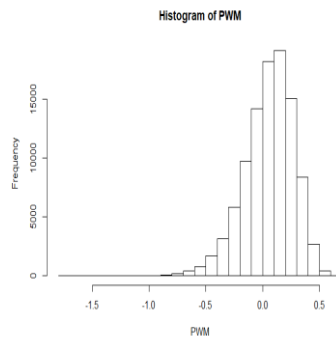
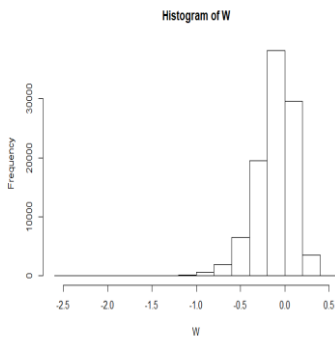
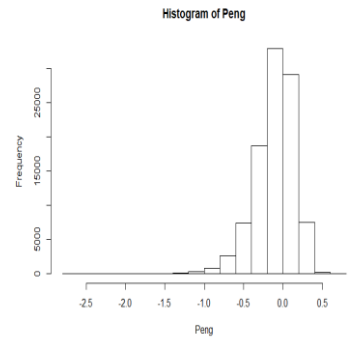
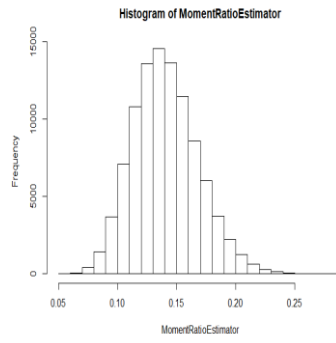
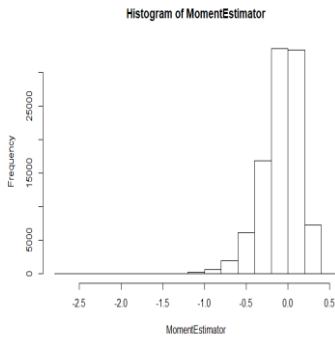
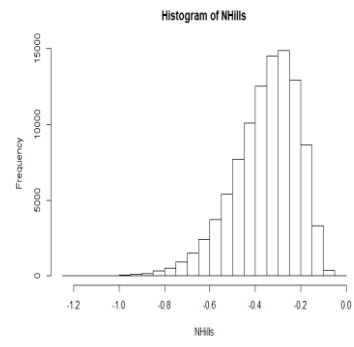
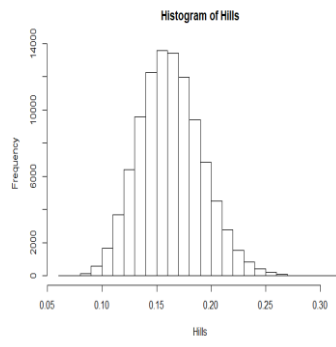
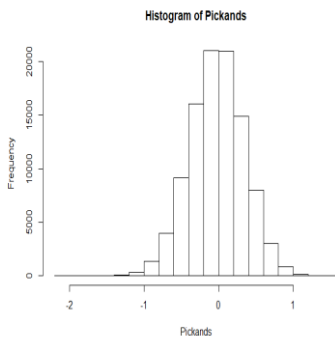


# 7 ΠΑΡΑΡΤΗΜΑ Ι

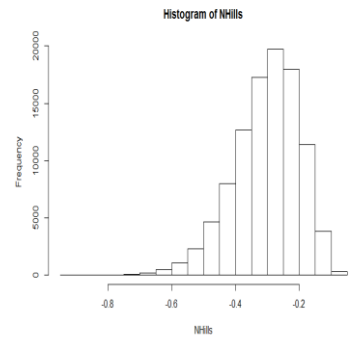
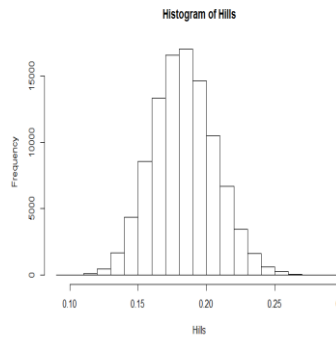
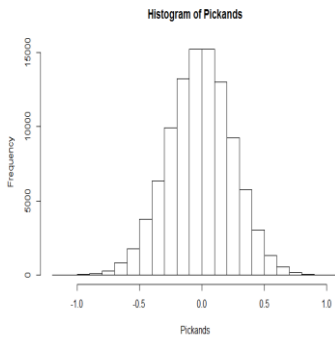
Στην συνέχεια παρατίθενται τα διαγράμματα από τα αποτελέσματα των προσομοιώσεων, συμπεριλαμβάνοντας τα ιστογράμματα των κατανομών ανά εκτιμητή.

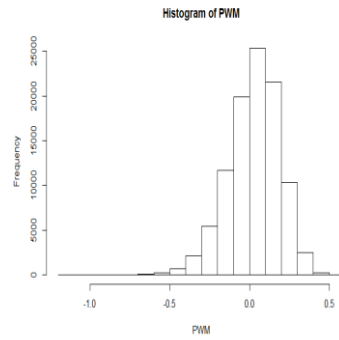
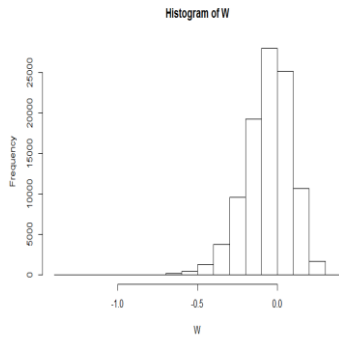
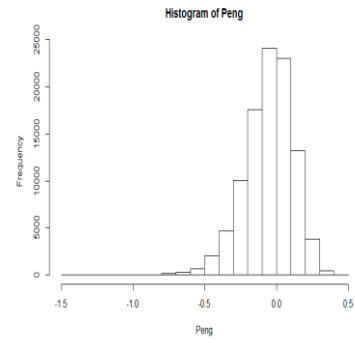
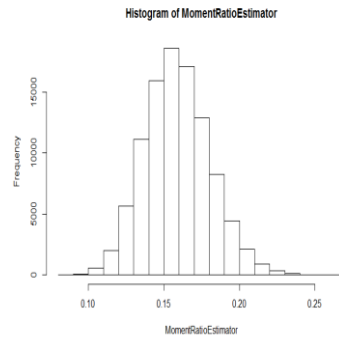
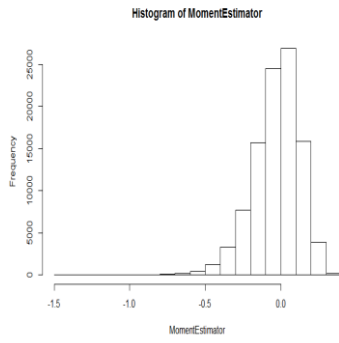


k=25

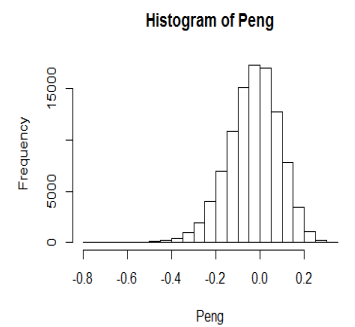
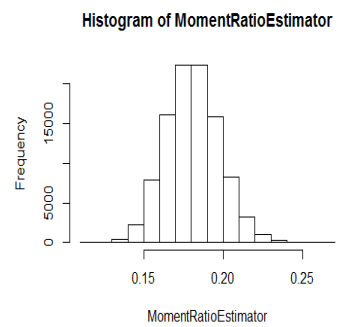
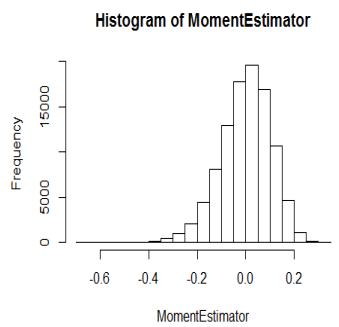
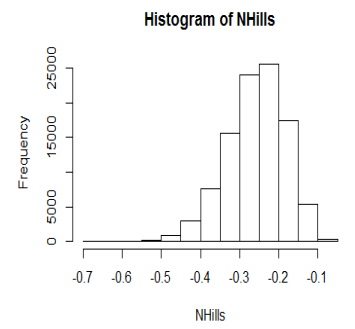
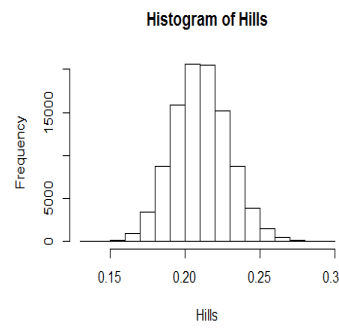
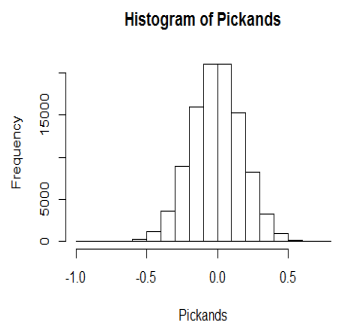


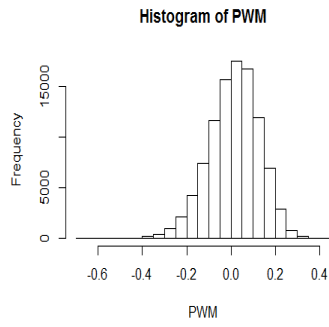
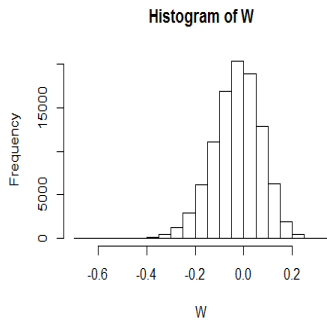
k=50



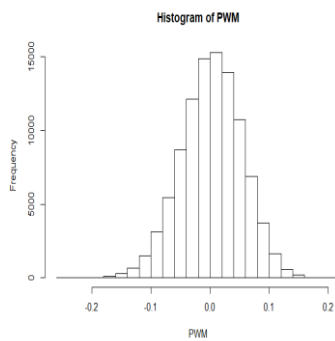
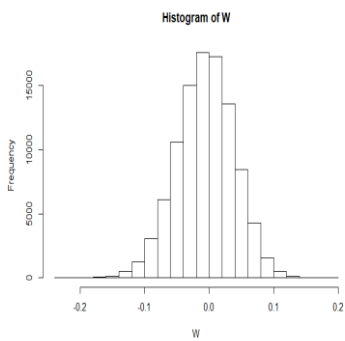
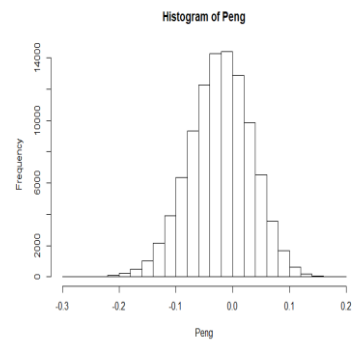
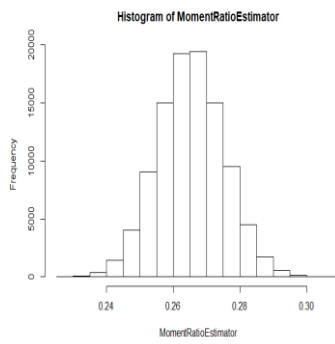
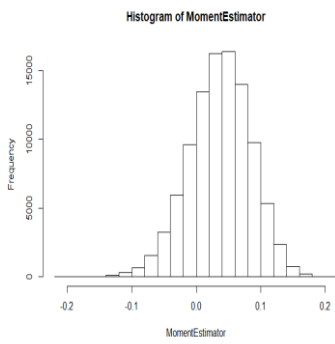
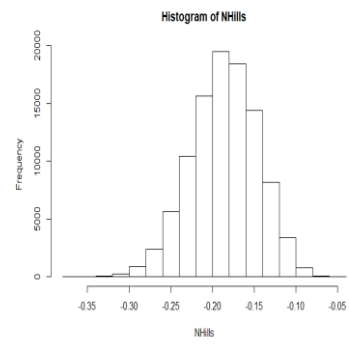
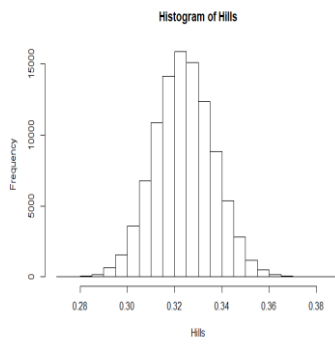
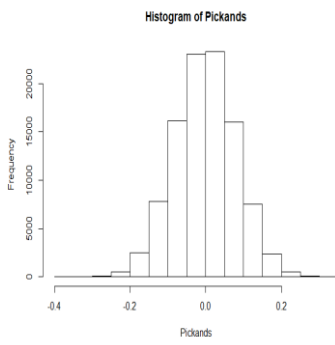


k=100

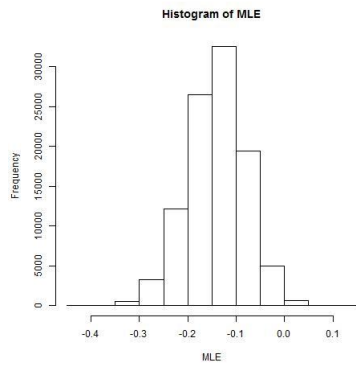
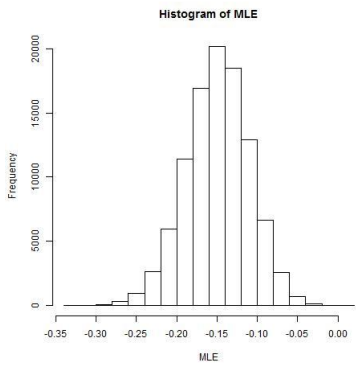
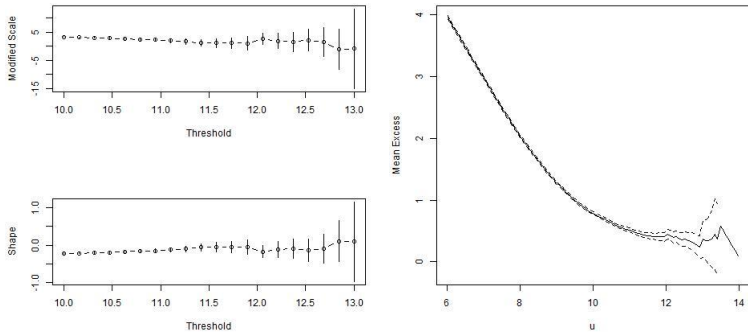
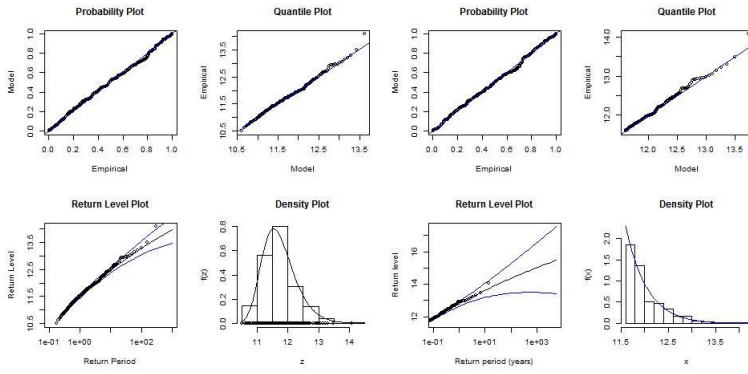




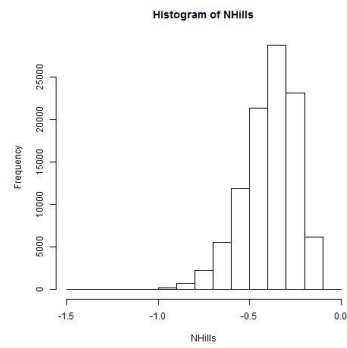
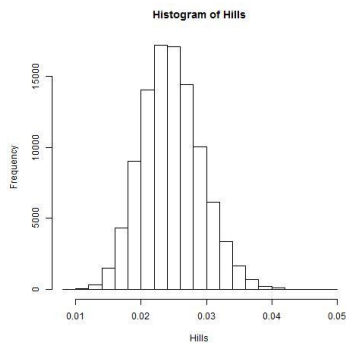
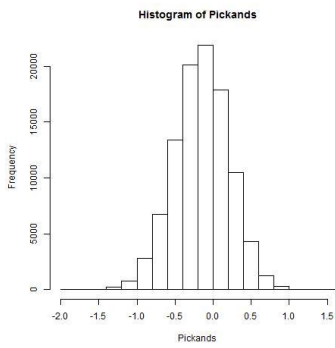
k=500

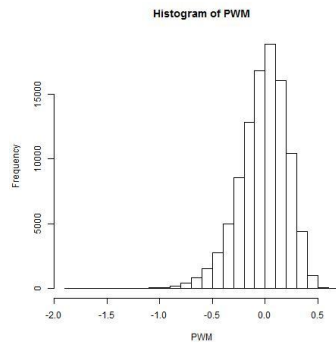
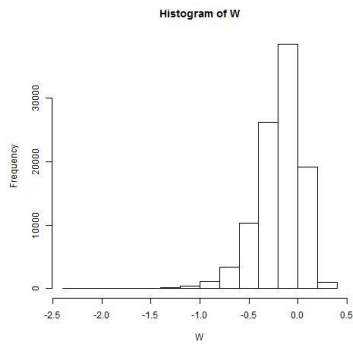
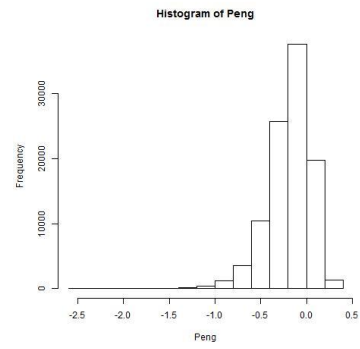
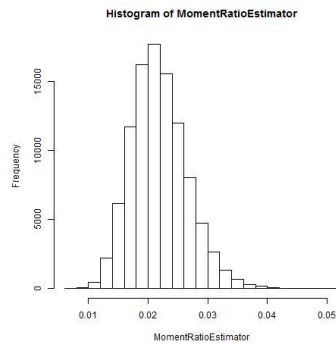
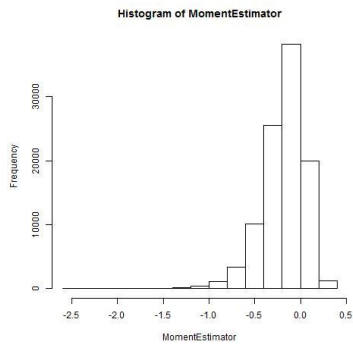


# Κανονική

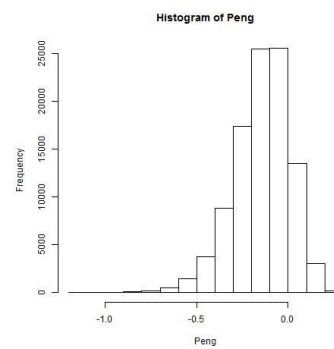
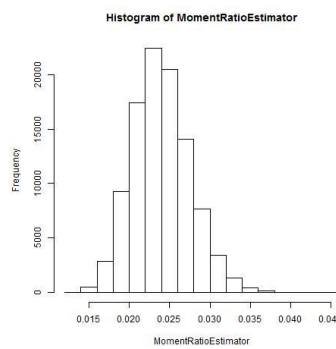
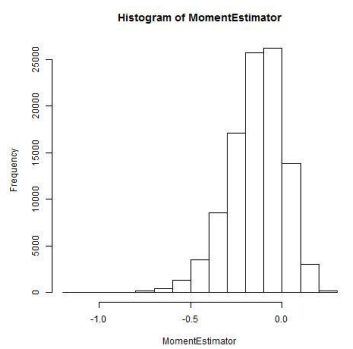
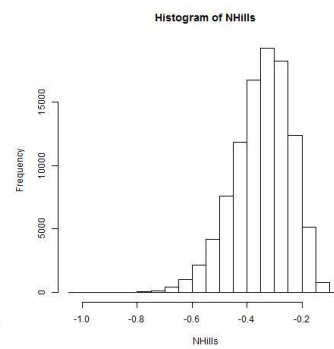
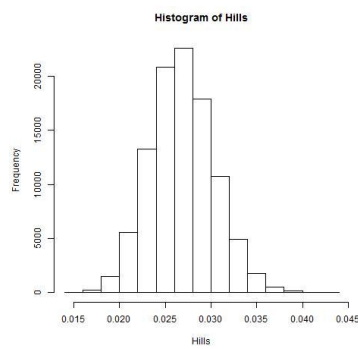
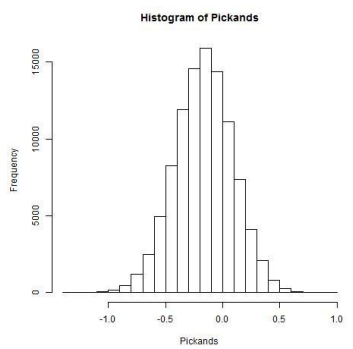


k=25

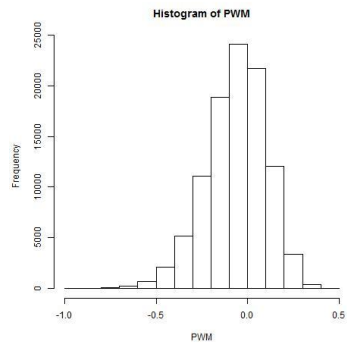
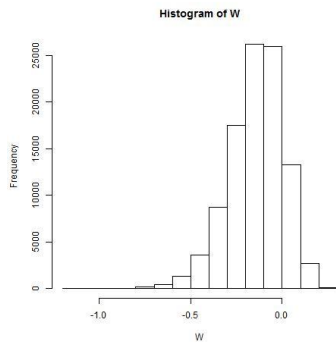




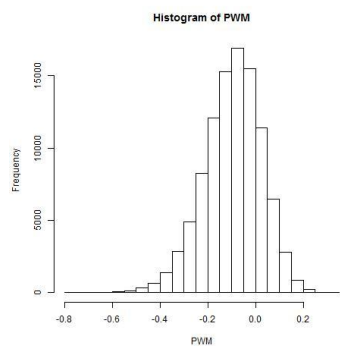
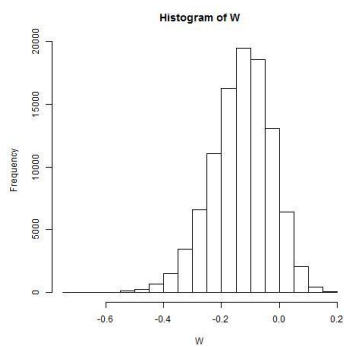
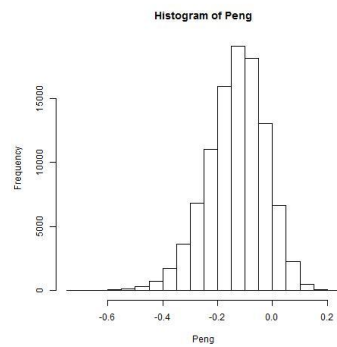
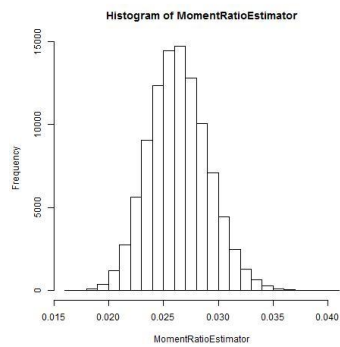
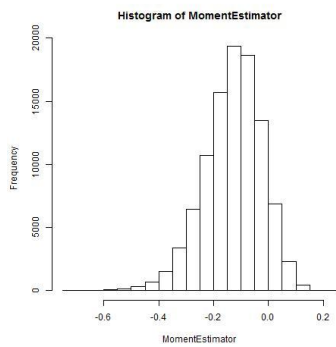
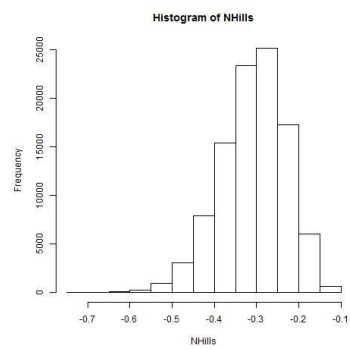
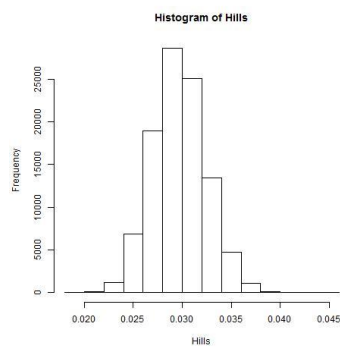
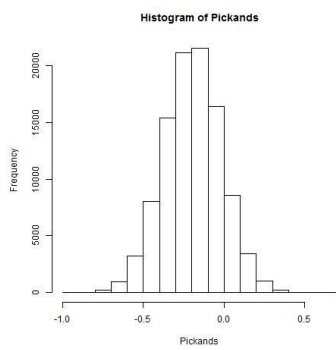
$k=50$



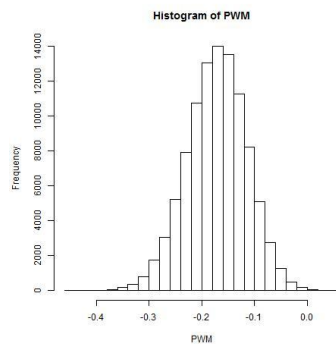
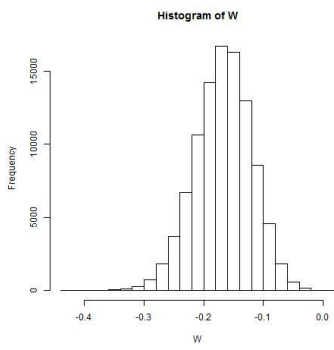
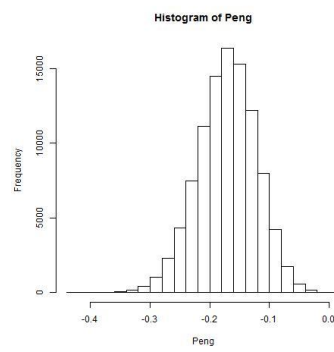
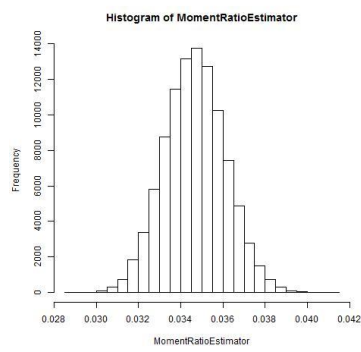
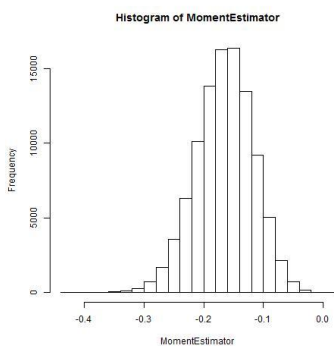
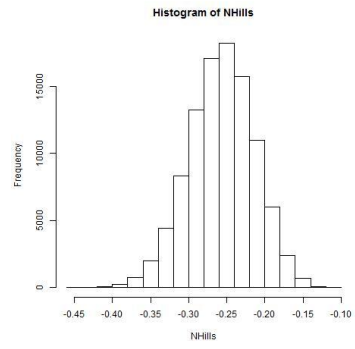
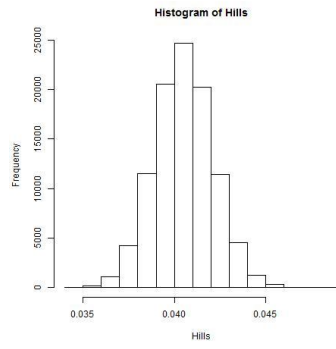
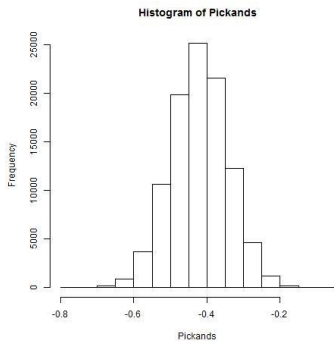




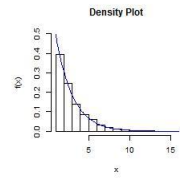
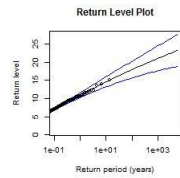
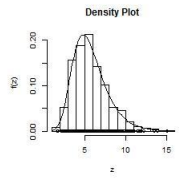
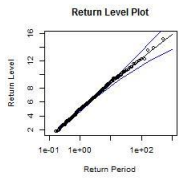
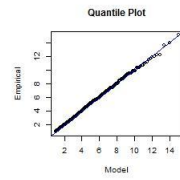
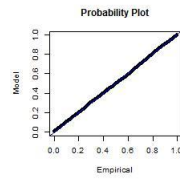
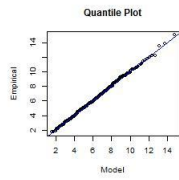
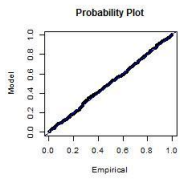
k=100

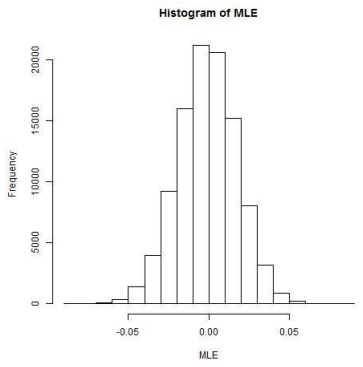
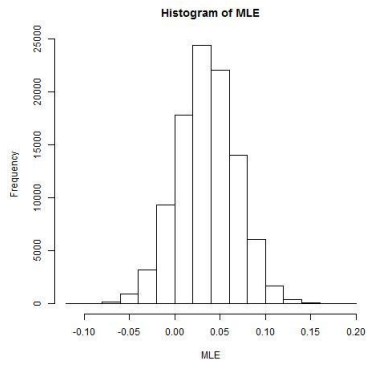
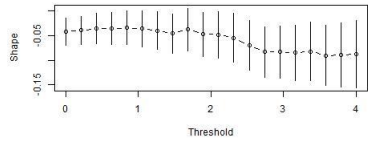
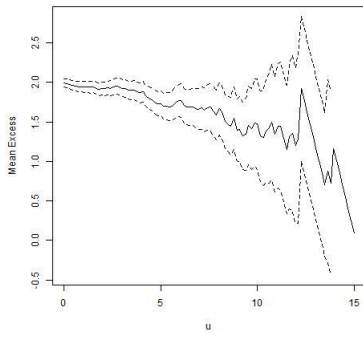
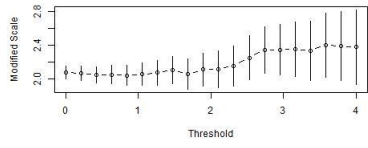


k=500

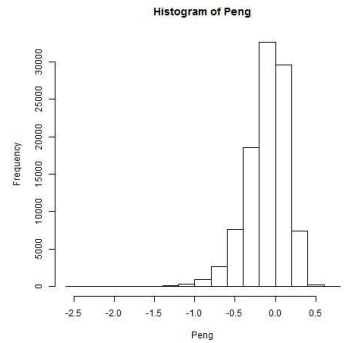
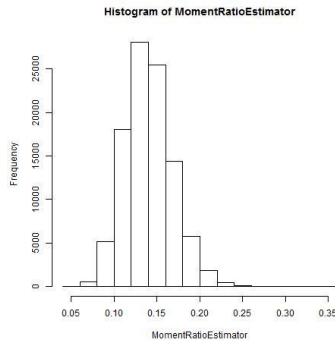
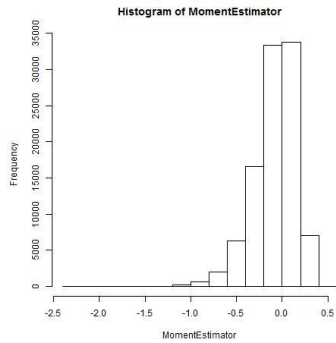
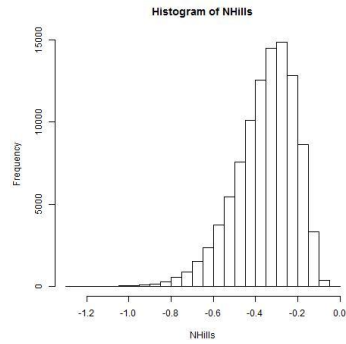
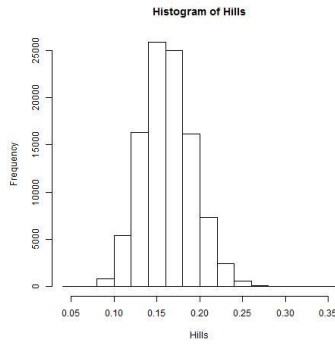
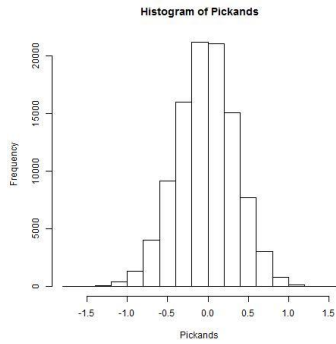


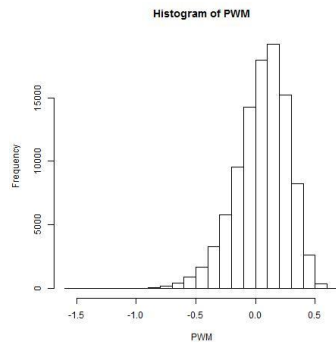
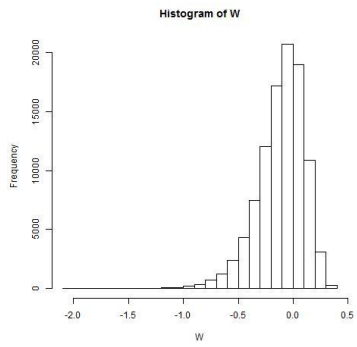
## Γάμμα Γάμμα (1, 0.5)



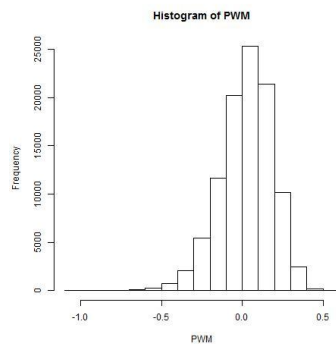
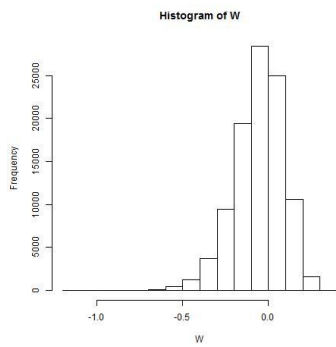
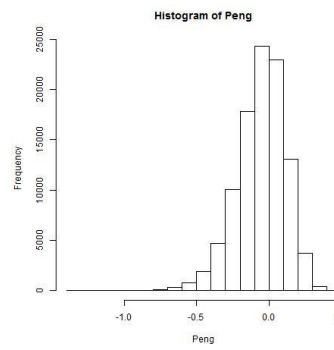
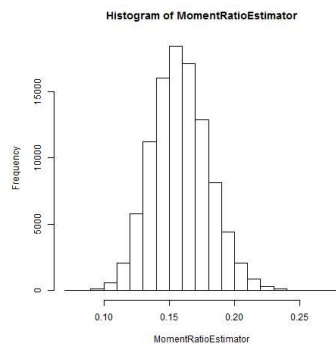
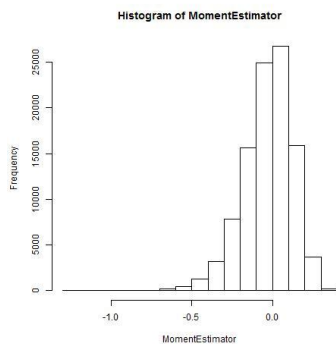
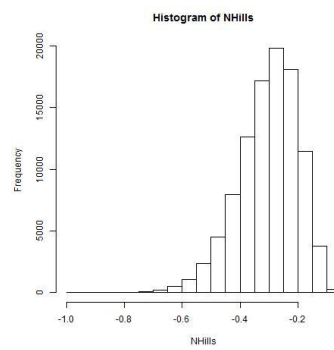
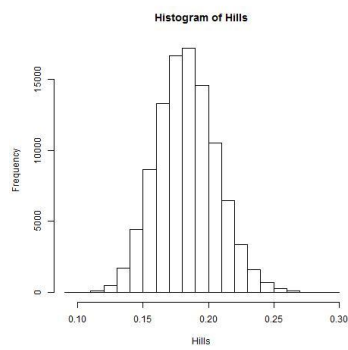
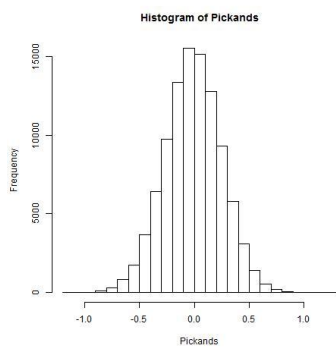


$k=25$

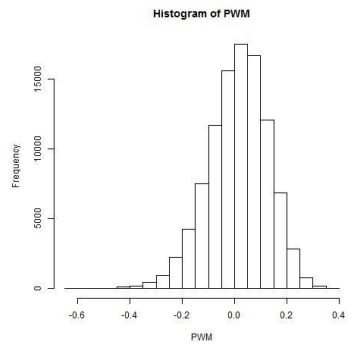
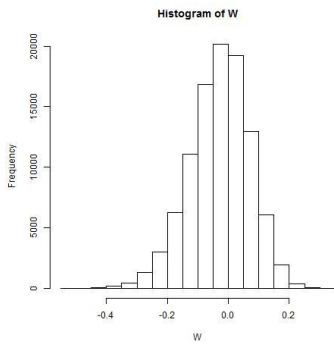
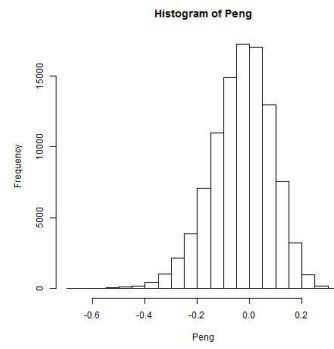
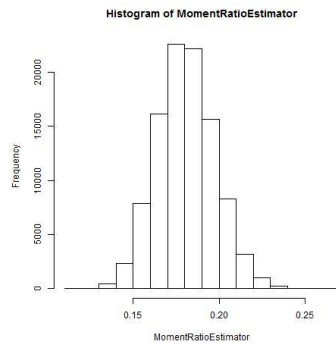
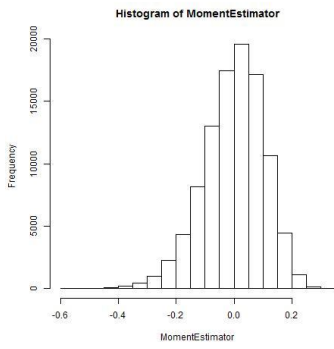
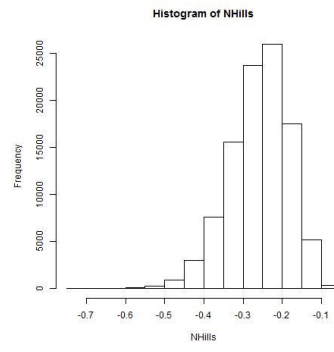
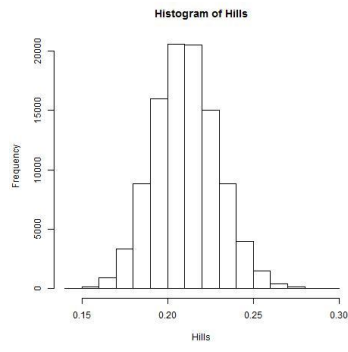
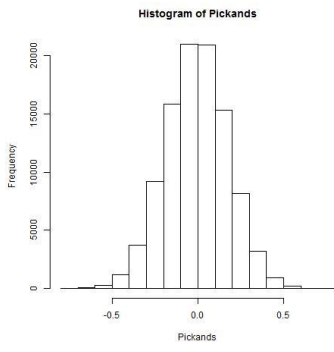




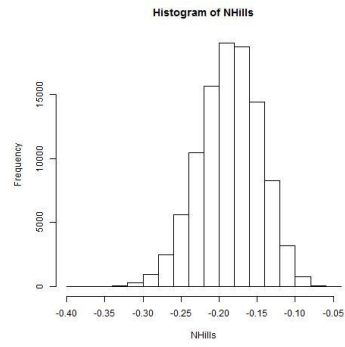
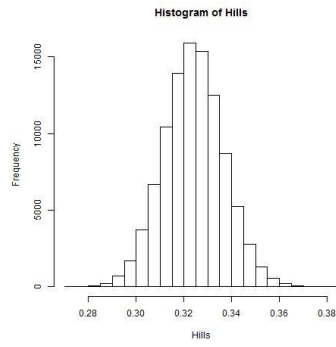
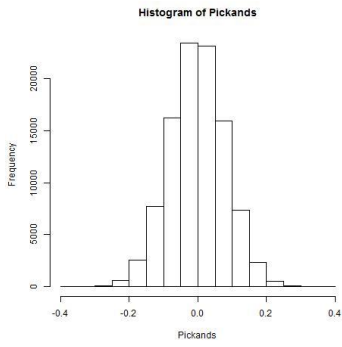
k=50

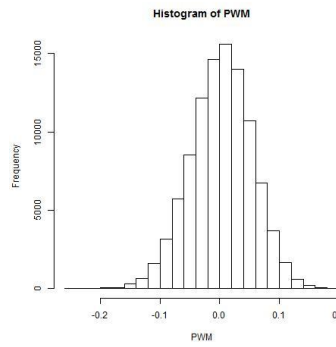
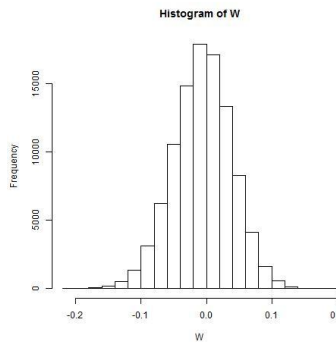
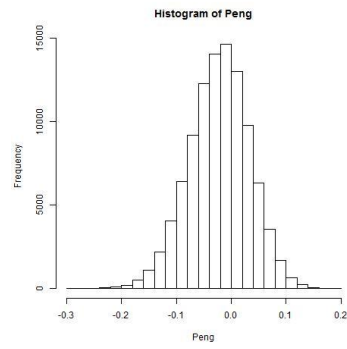
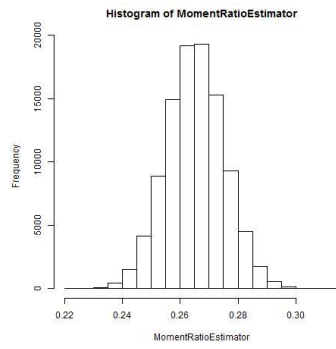
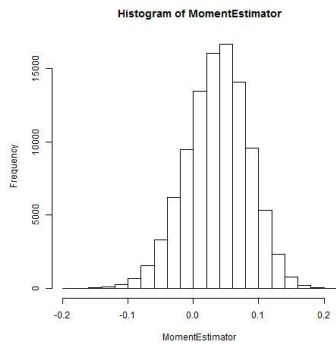


k=100

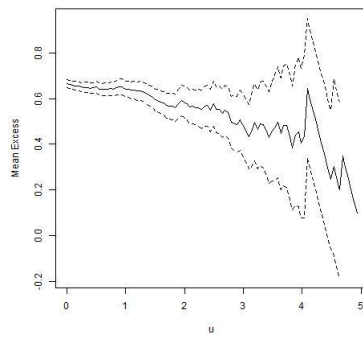
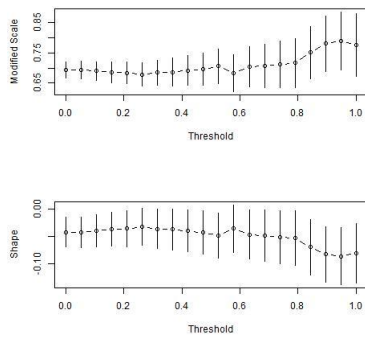
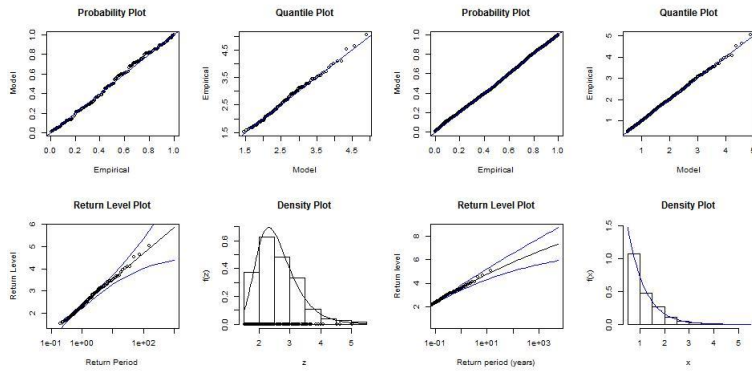


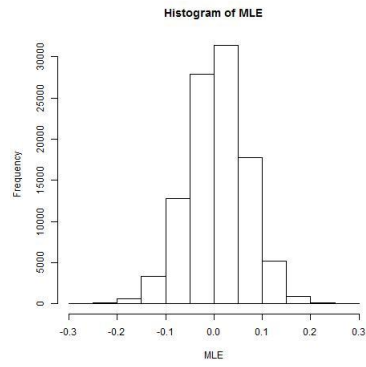
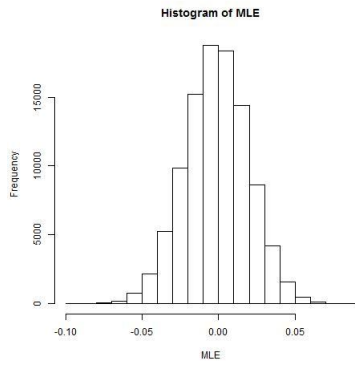
k=500



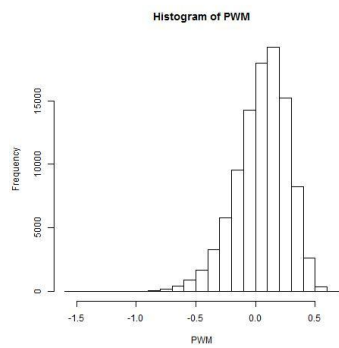
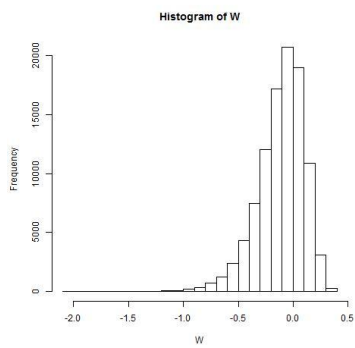
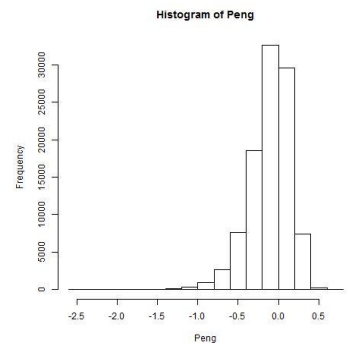
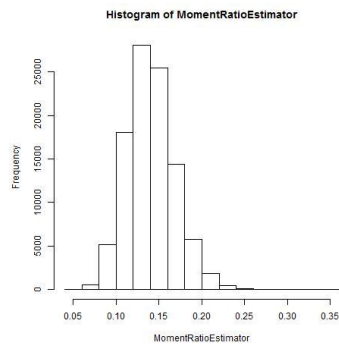
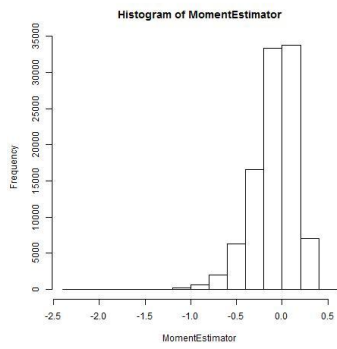
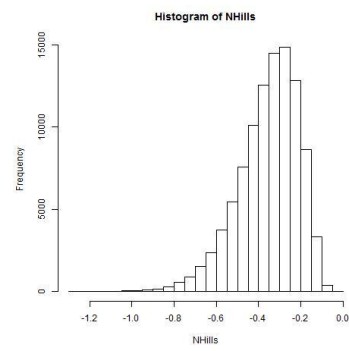
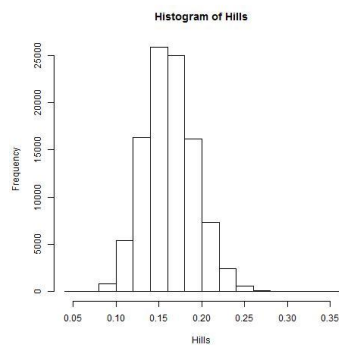
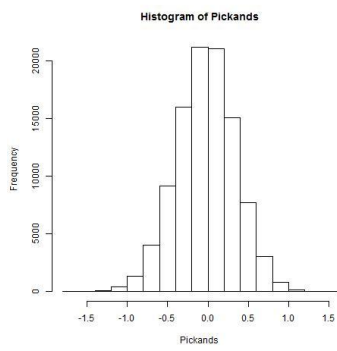


### Γάμμα (1, 1.5)

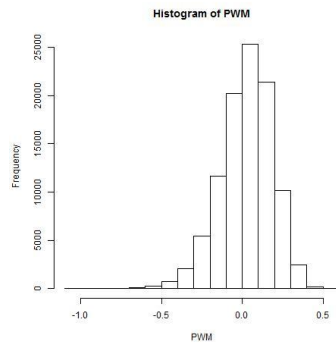
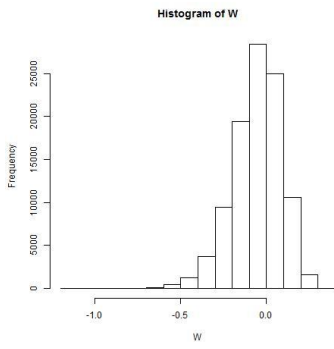
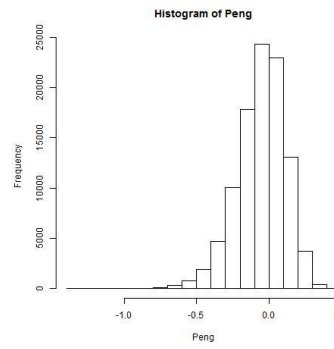
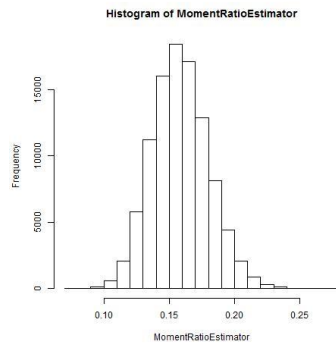
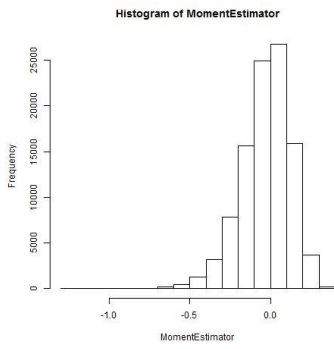
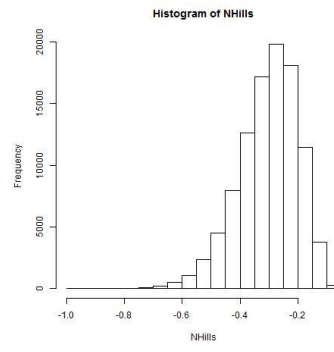
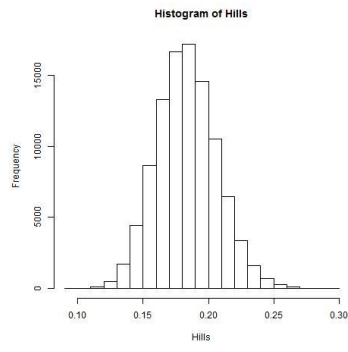
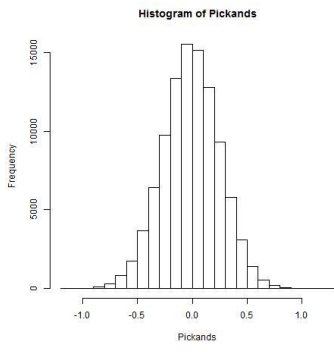




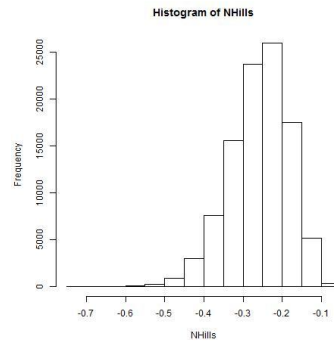
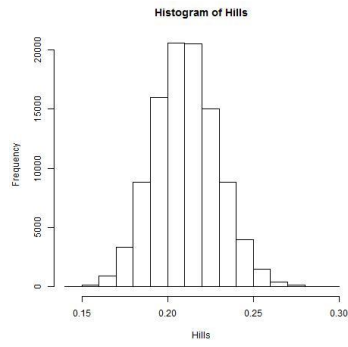
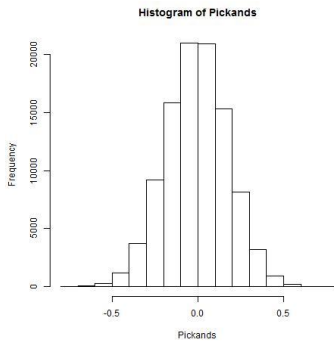
k=25



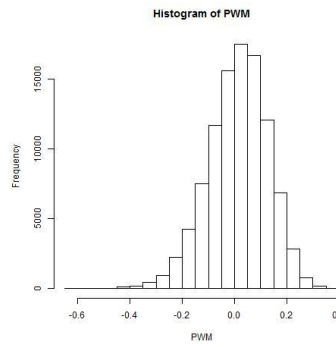
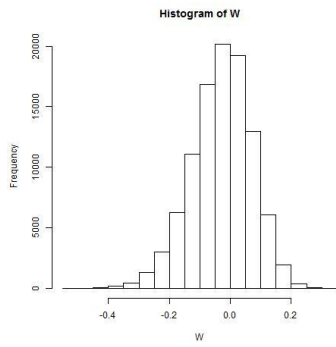
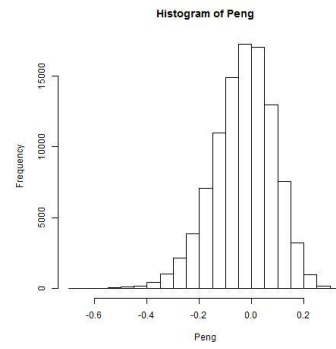
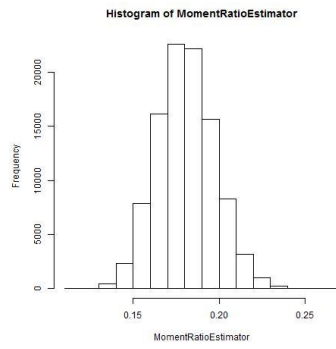
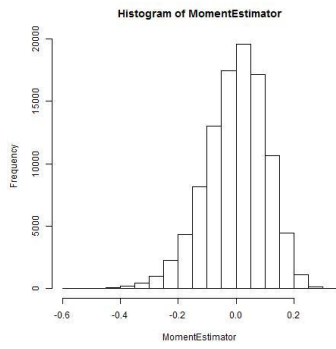
k=50



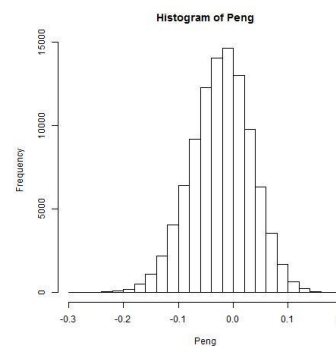
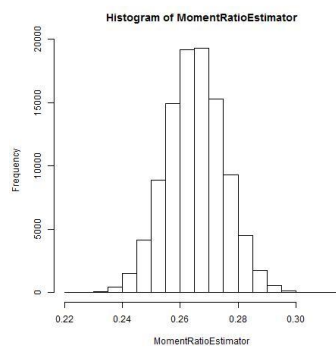
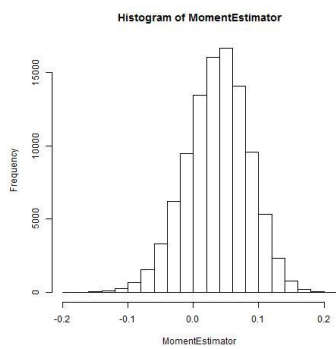
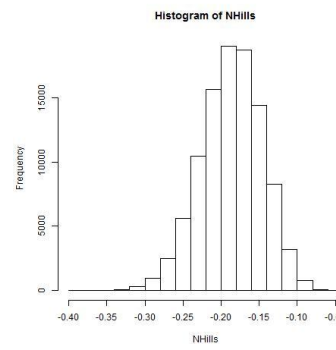
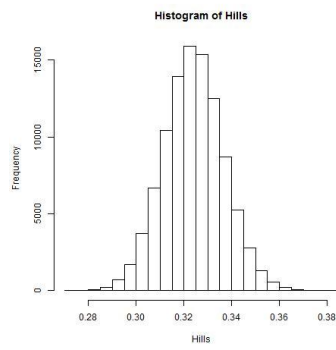
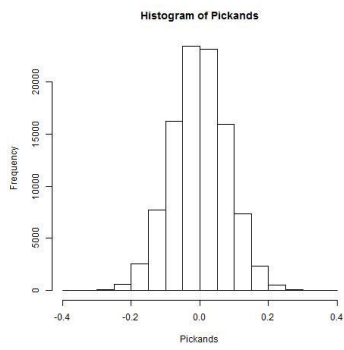
k=100

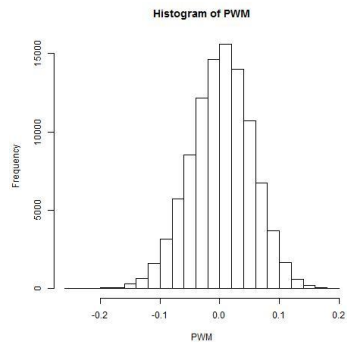
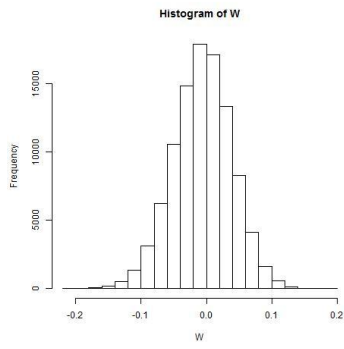




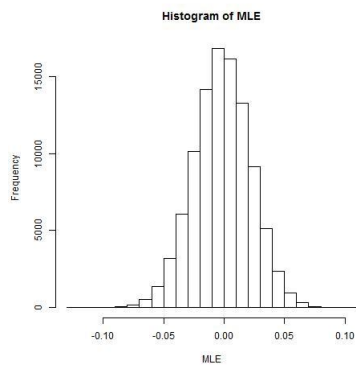
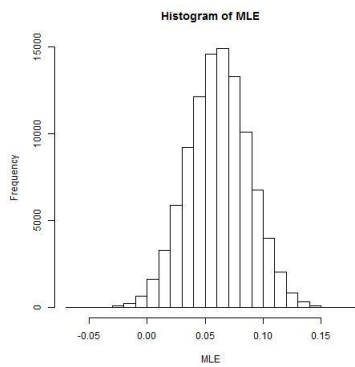
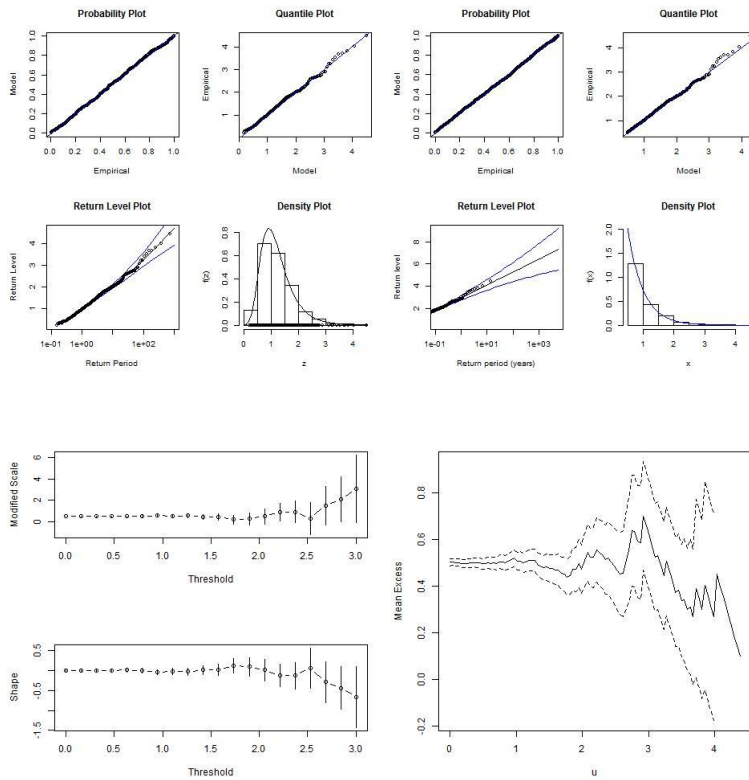


k=500

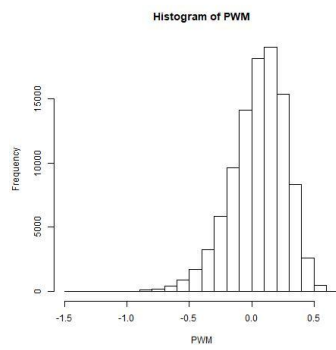
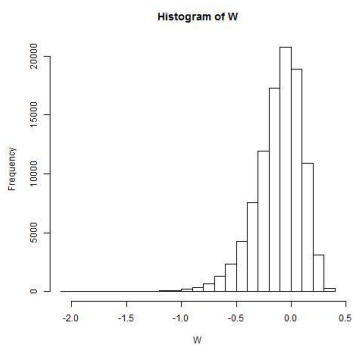
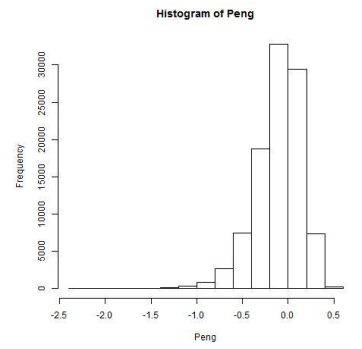
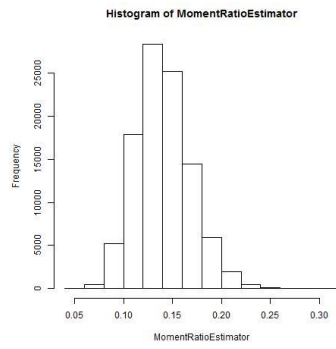
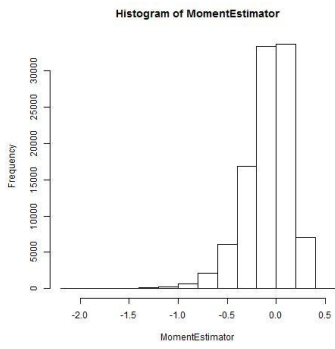
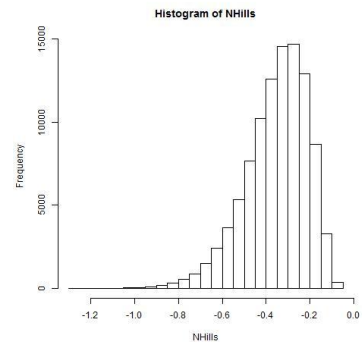
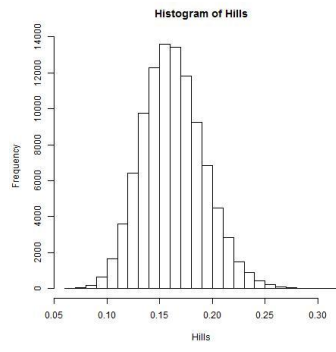
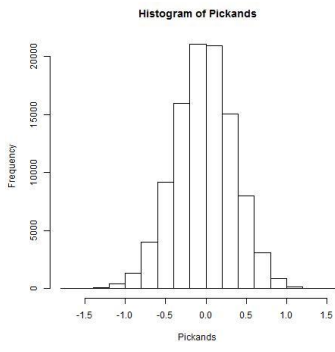




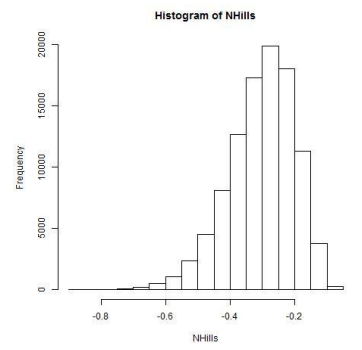
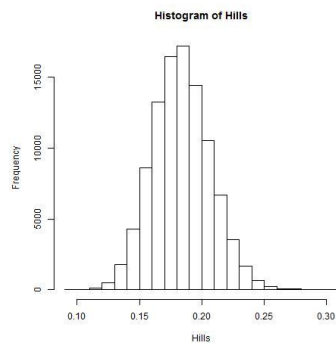
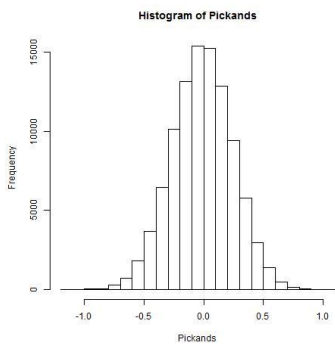
**Weibull  
Weibull (1, 0.5)**

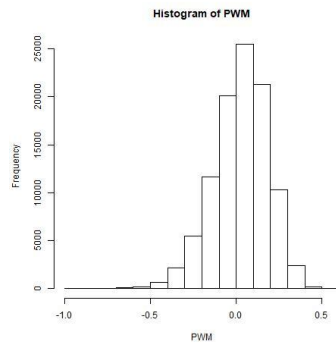
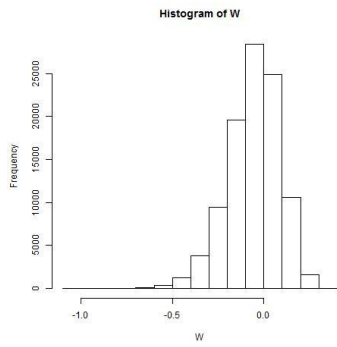
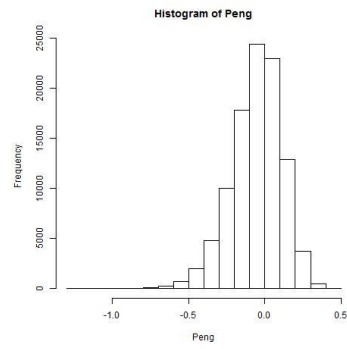
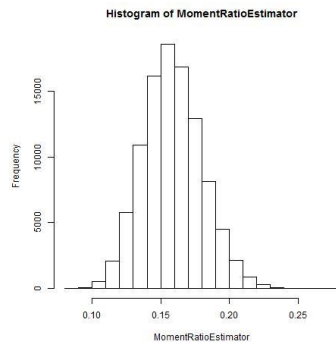
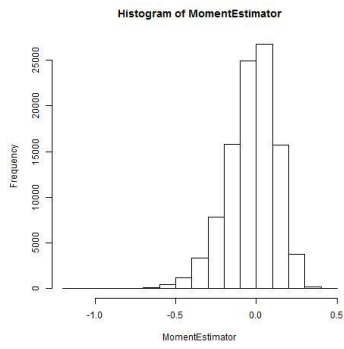


k=25

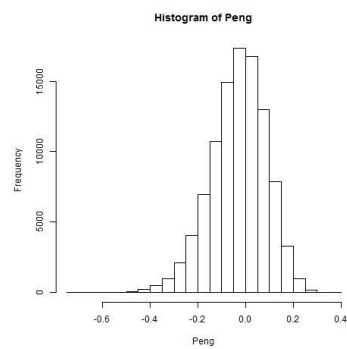
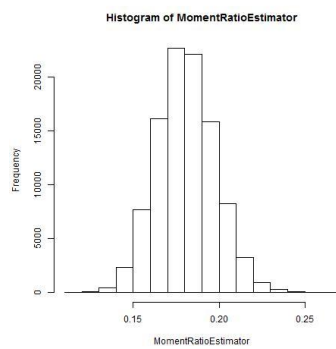
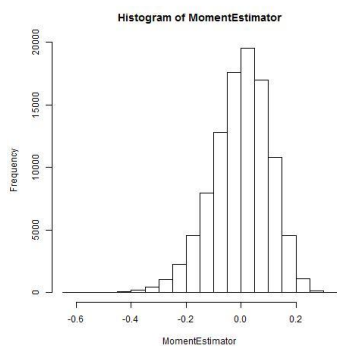
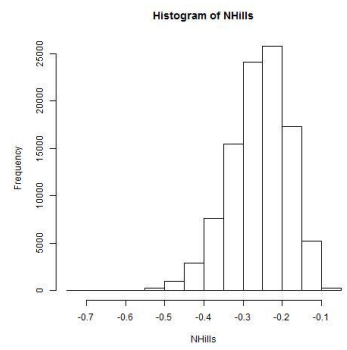
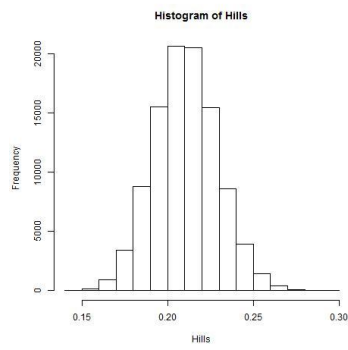
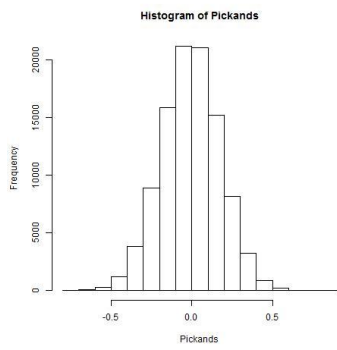


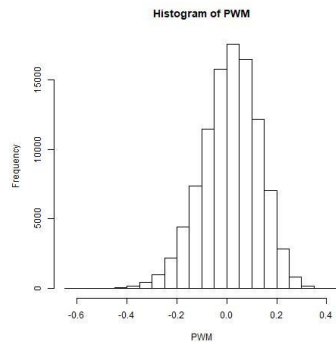
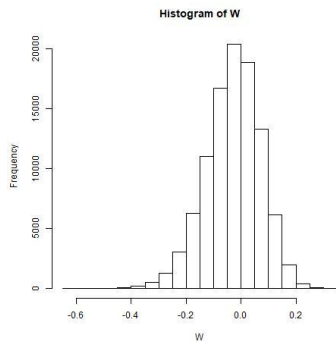
k=50



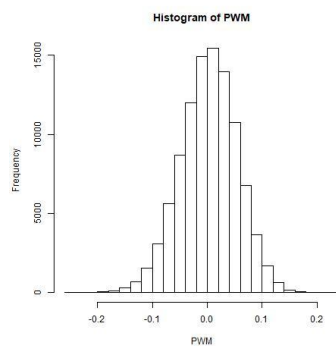
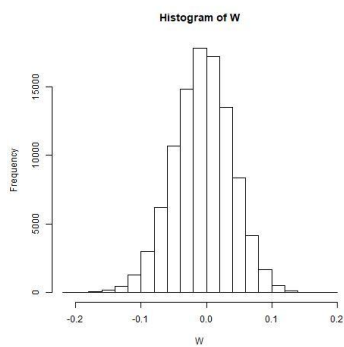
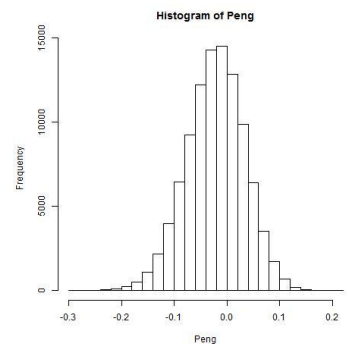
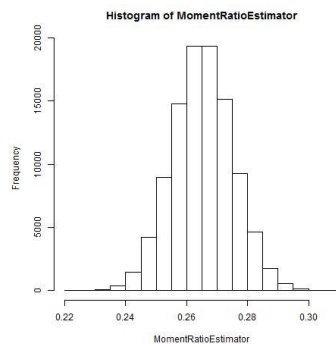
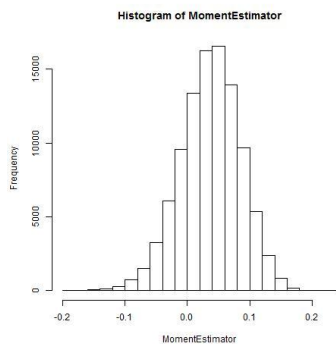
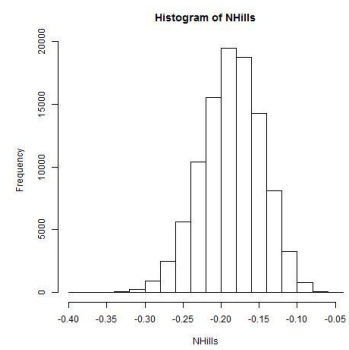
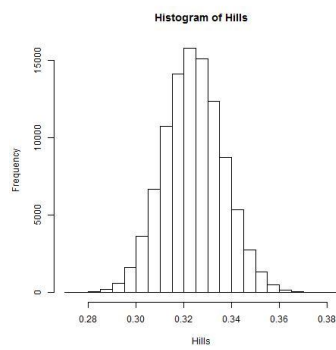
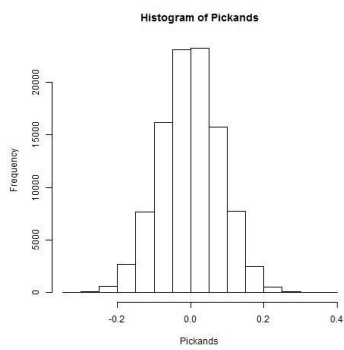


k=100

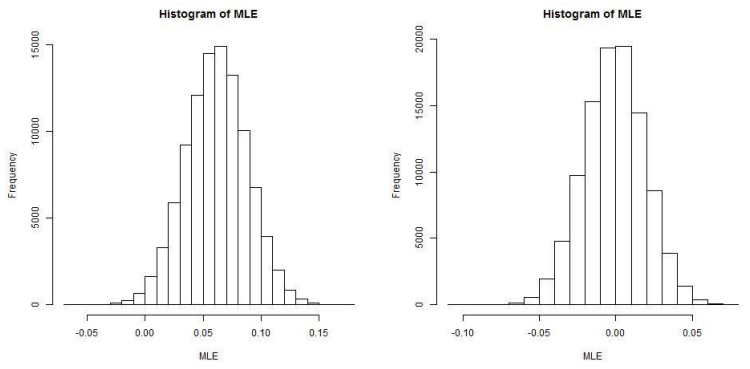
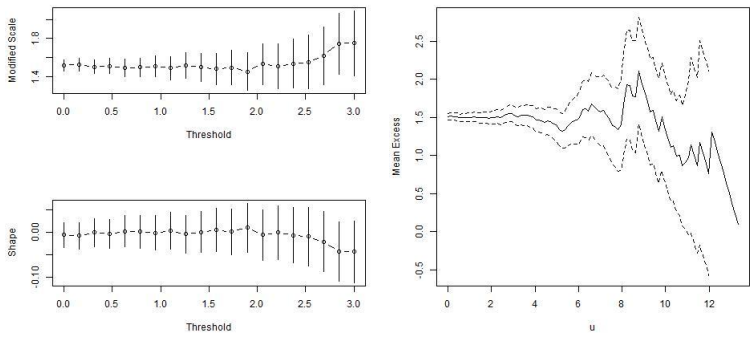
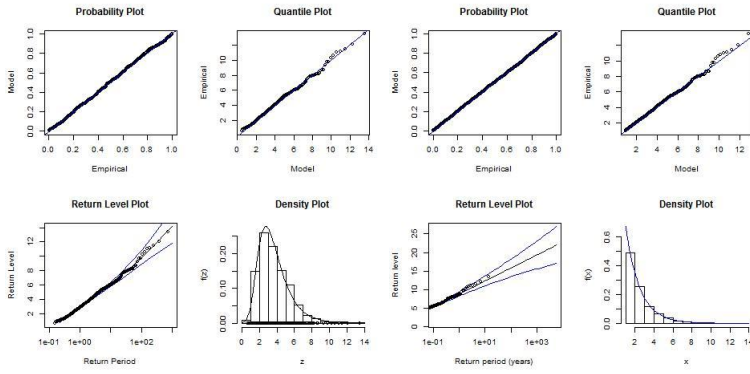




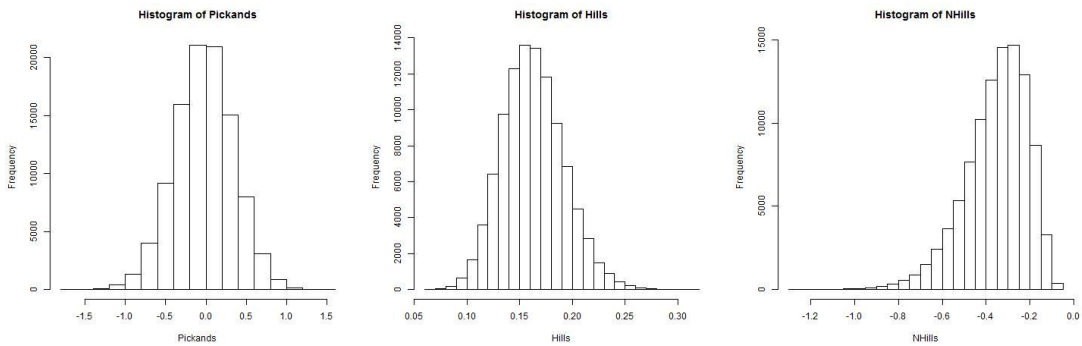
k=500

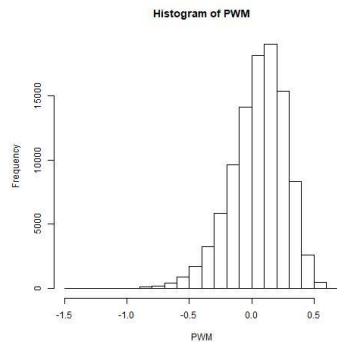
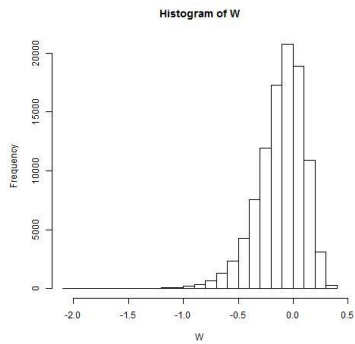
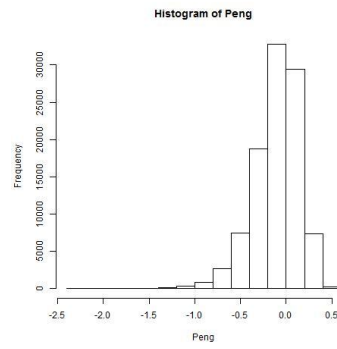
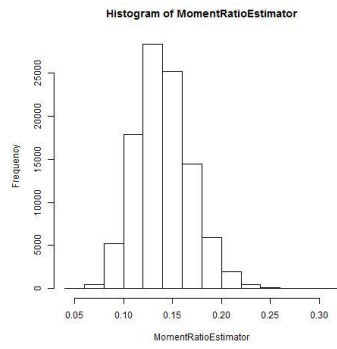
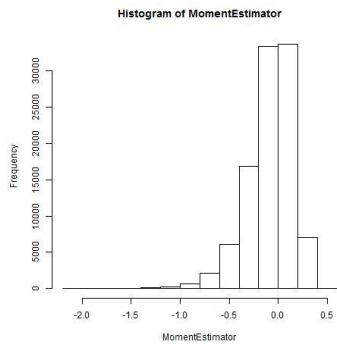


### Weibull (1, 1.5)

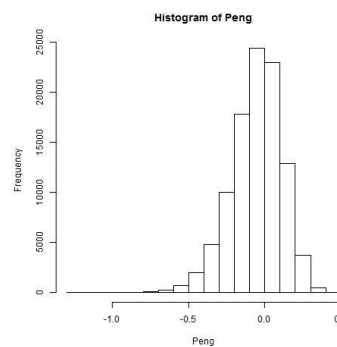
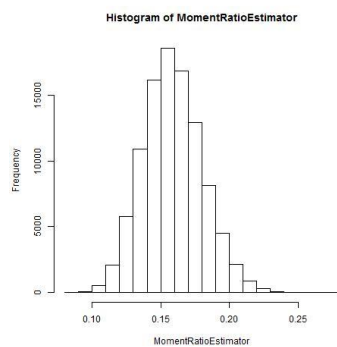
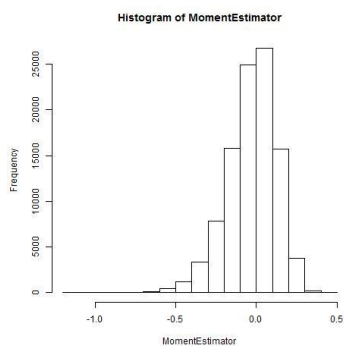
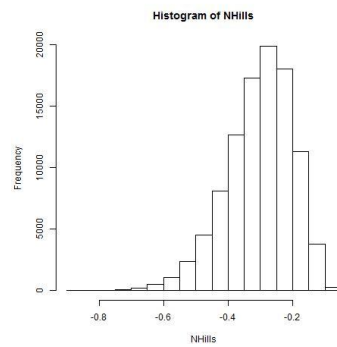
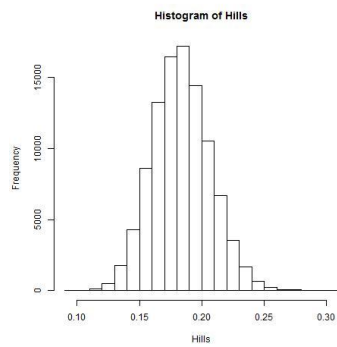
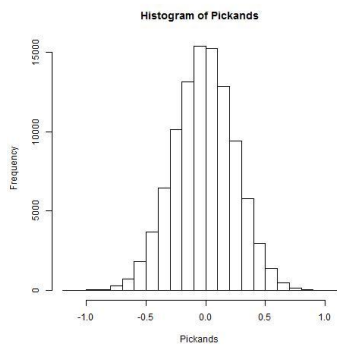


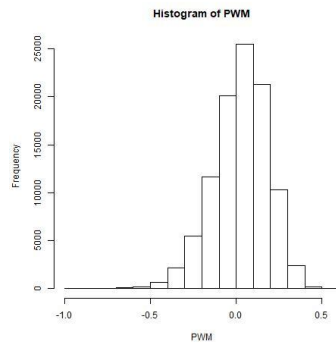
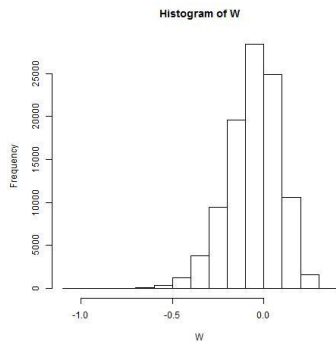
k=25



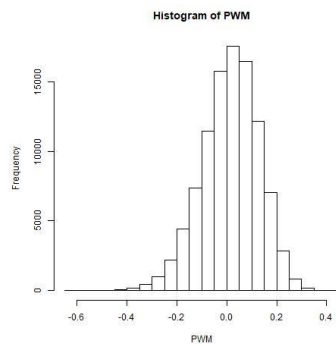
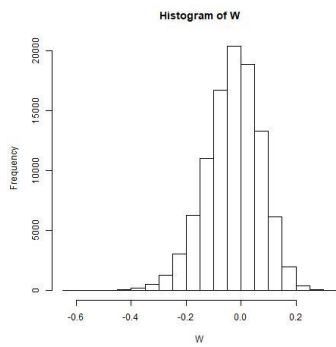
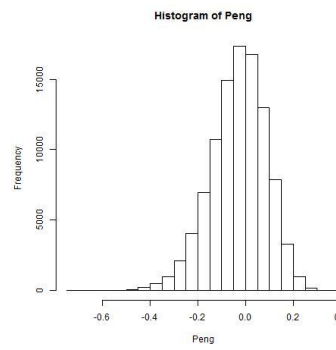
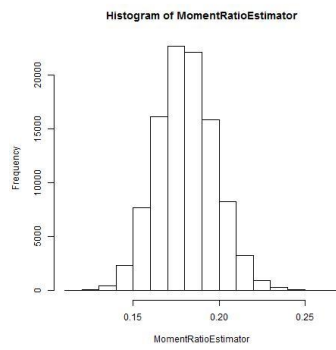
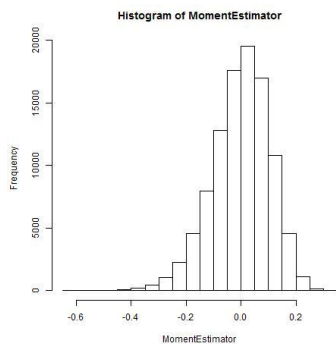
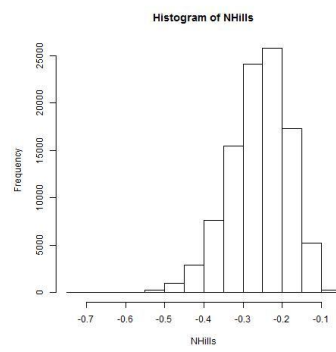
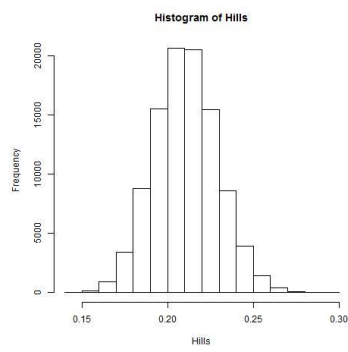
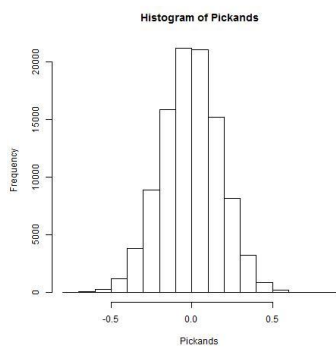


$k=50$



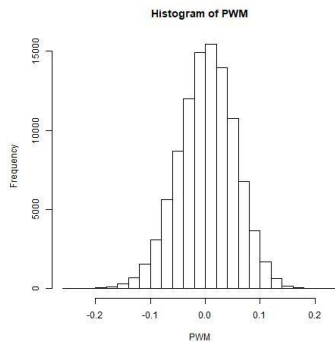
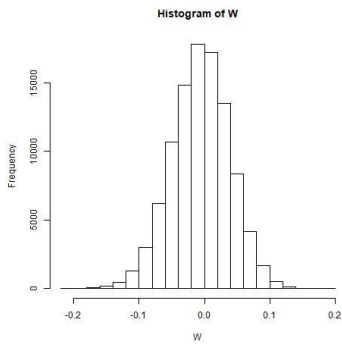
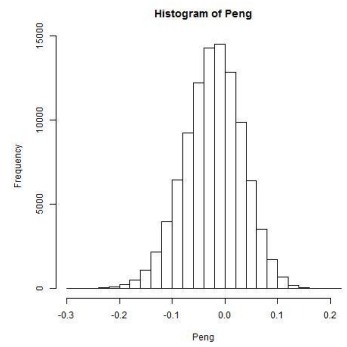
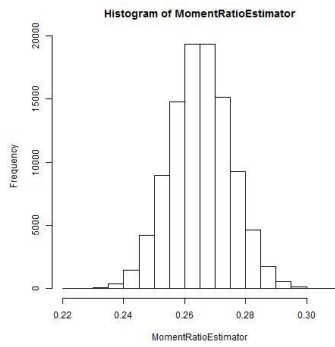
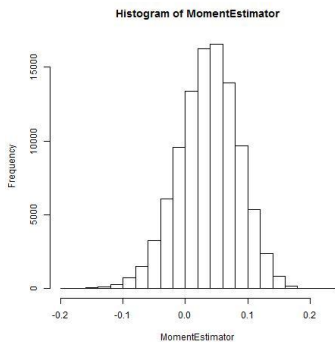
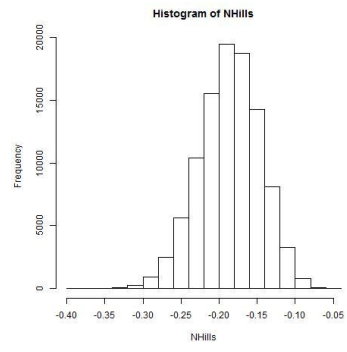
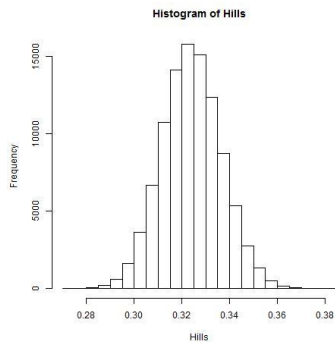
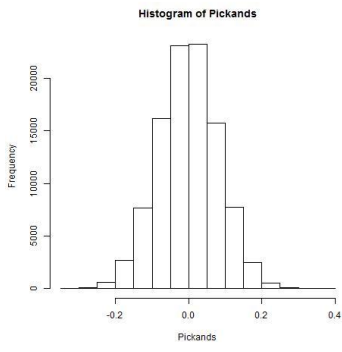


k=100

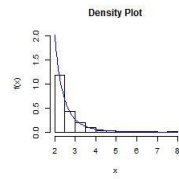
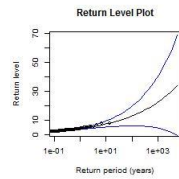
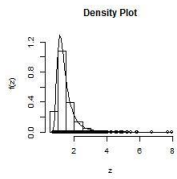
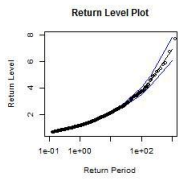
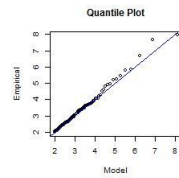
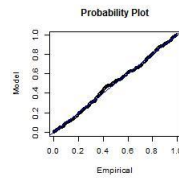
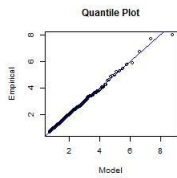
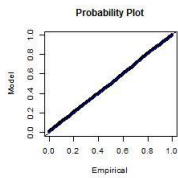


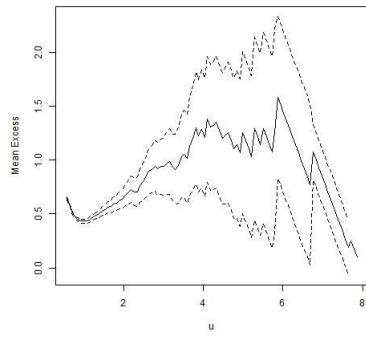
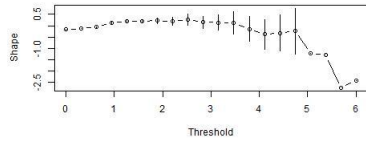
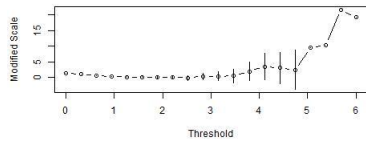


k=500

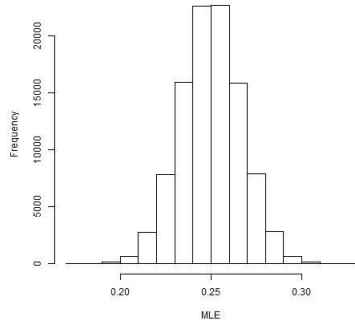


**Fréchet**  
**Fréchet (0, 1, 4)**

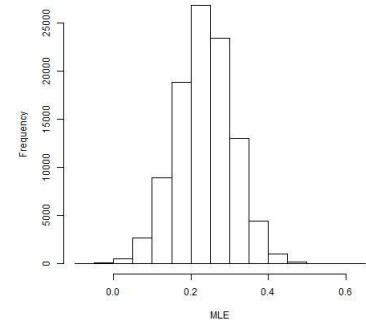




**Histogram of MLE**

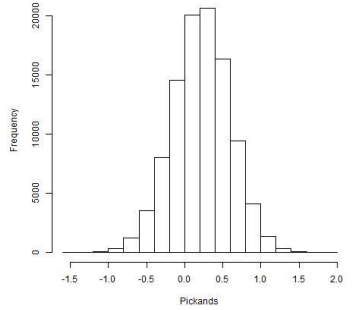


**Histogram of MLE**

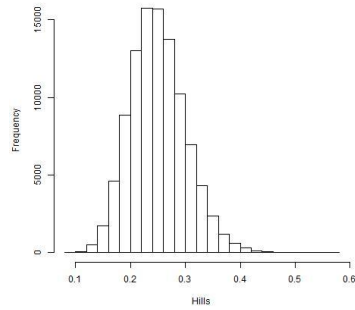


$k=25$

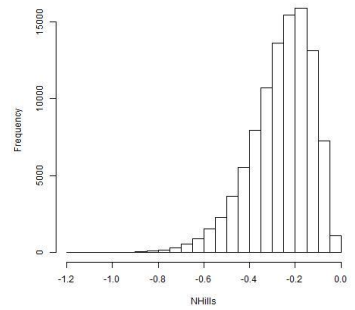
**Histogram of Pickands**



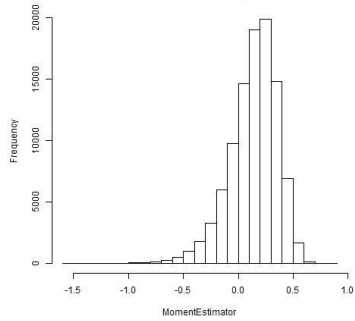
**Histogram of Hills**



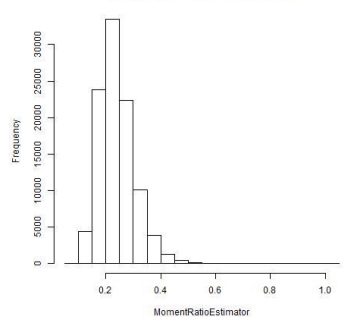
**Histogram of NHills**



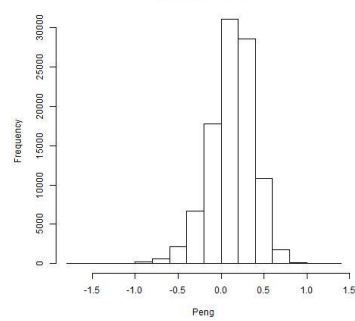
**Histogram of MomentEstimator**

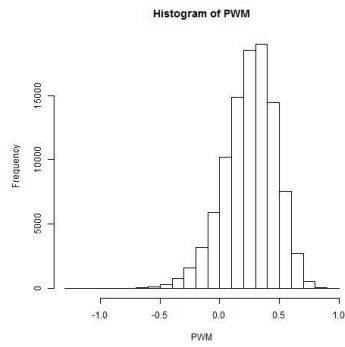
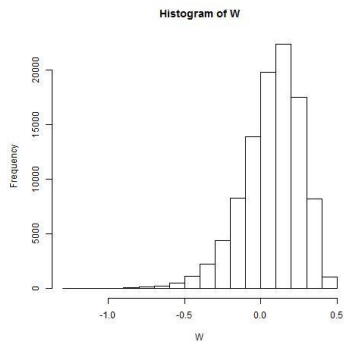


**Histogram of MomentRatioEstimator**

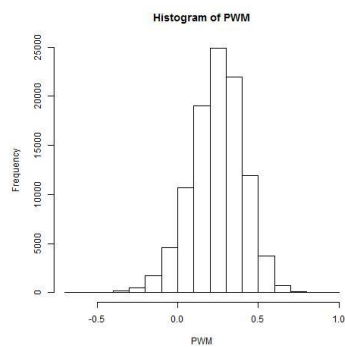
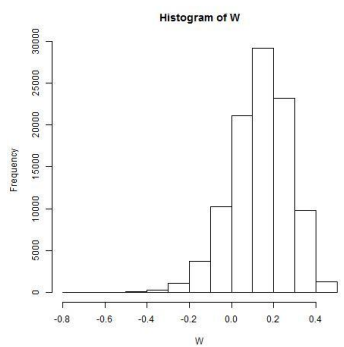
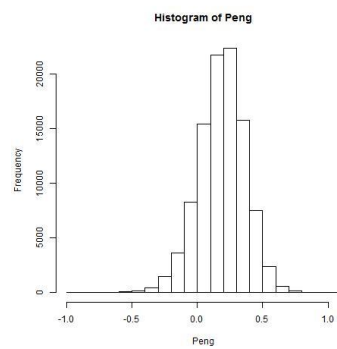
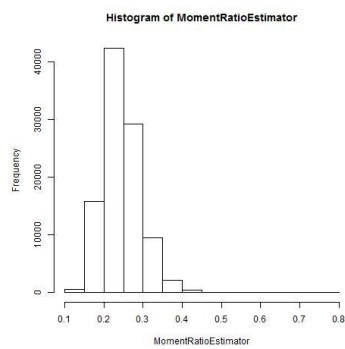
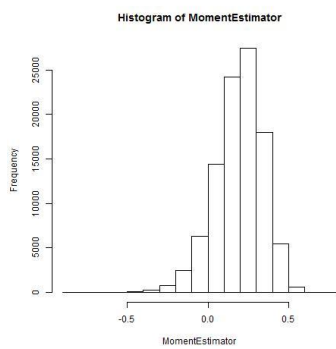
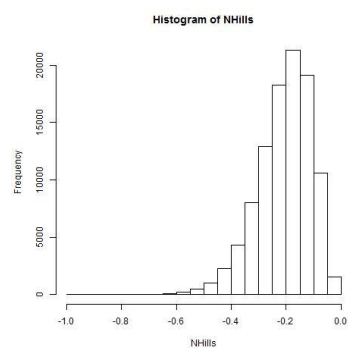
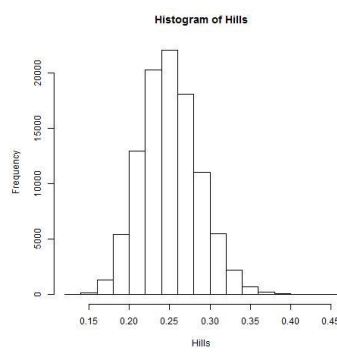
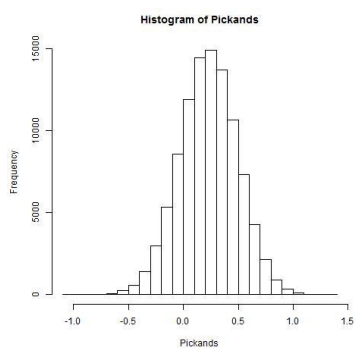


**Histogram of Peng**

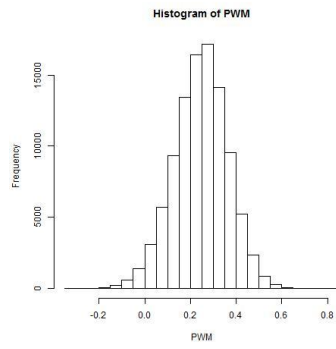
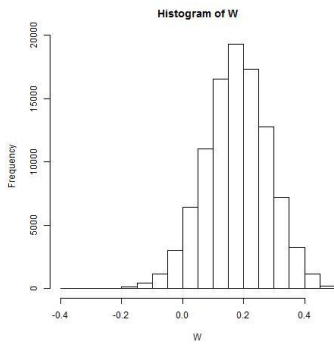
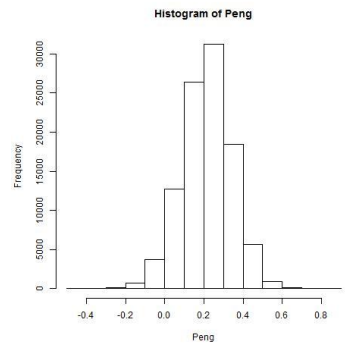
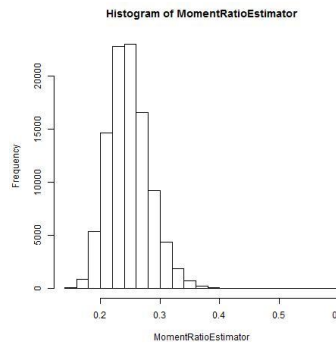
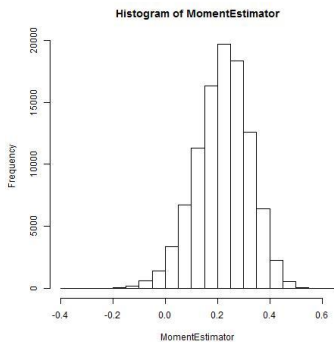
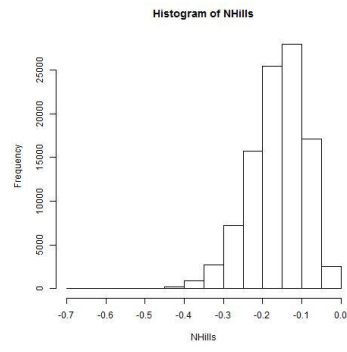
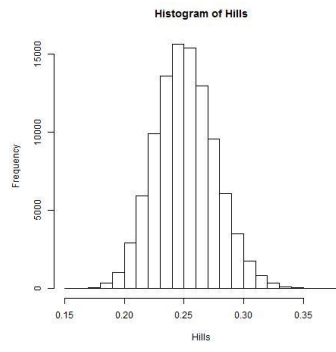
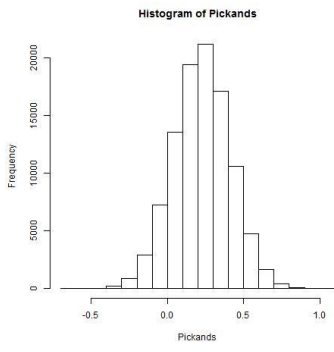




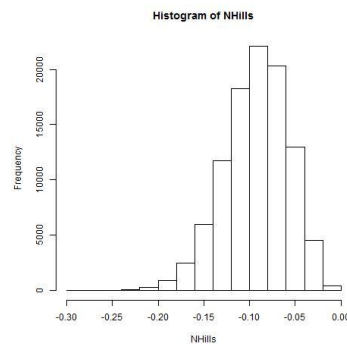
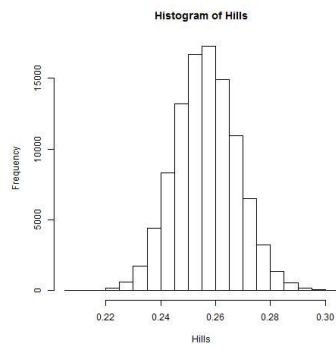
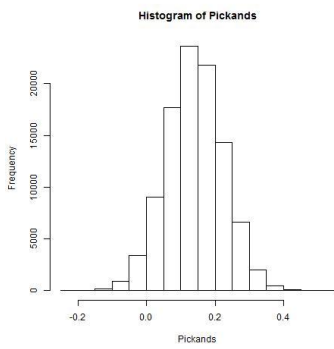
k=50

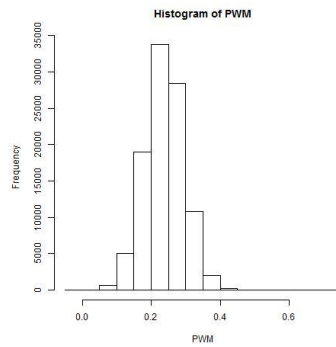
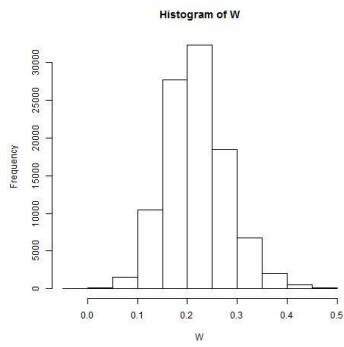
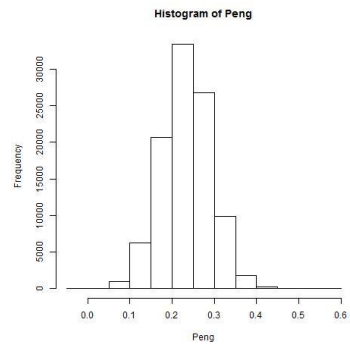
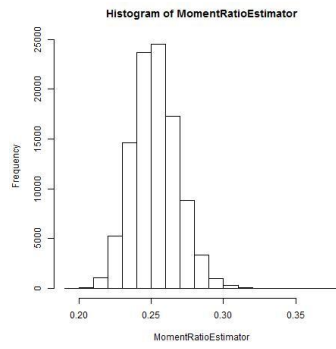
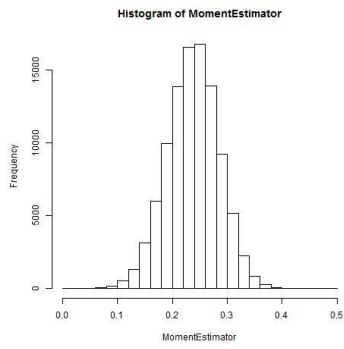


k=100

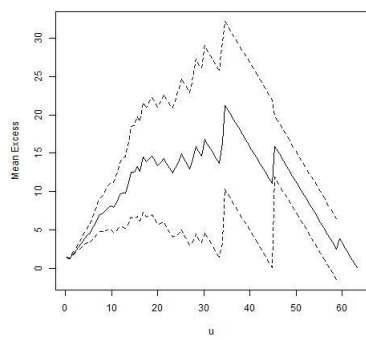
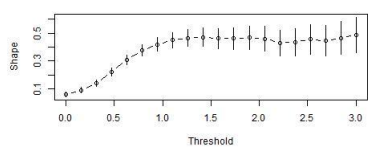
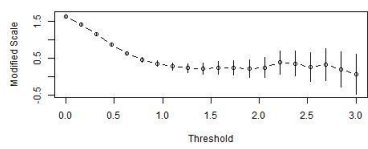
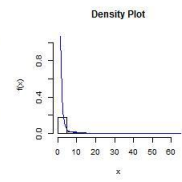
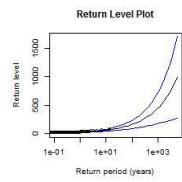
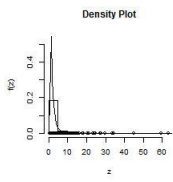
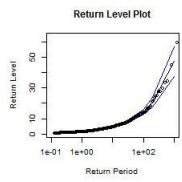
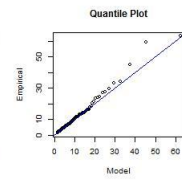
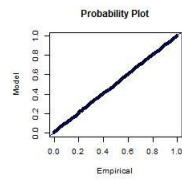
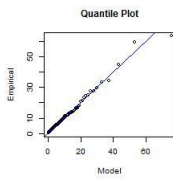
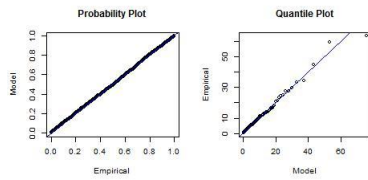


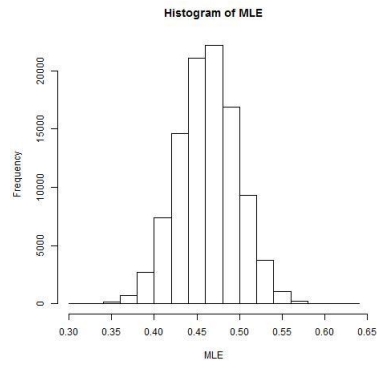
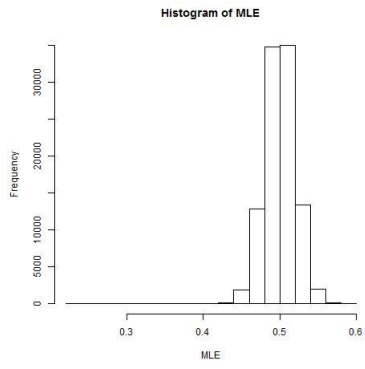
k=500



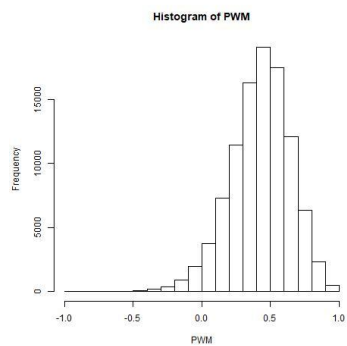
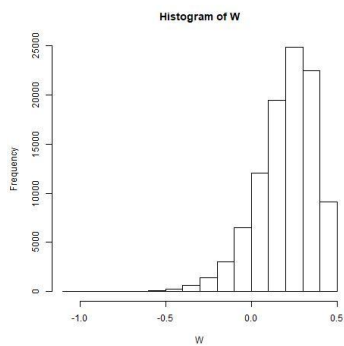
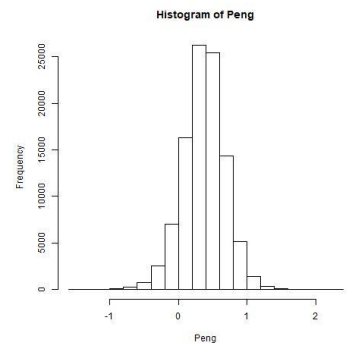
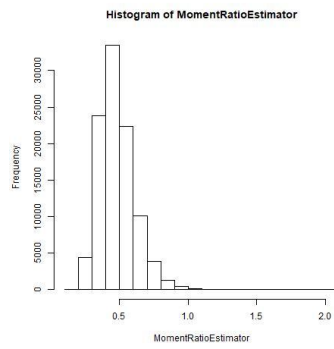
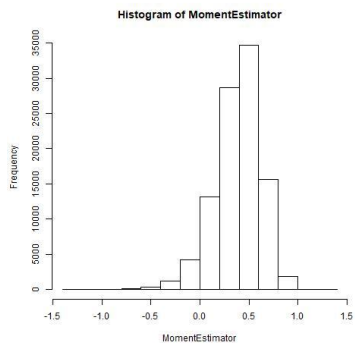
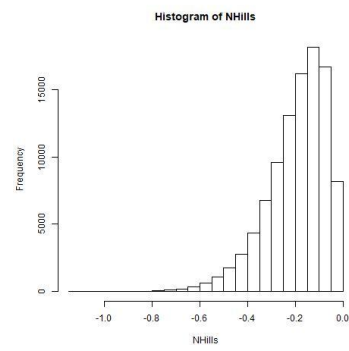
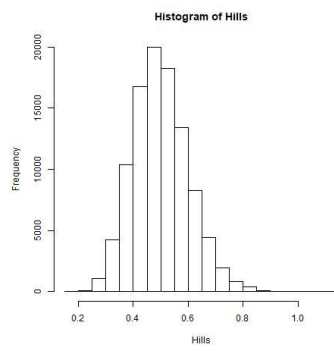
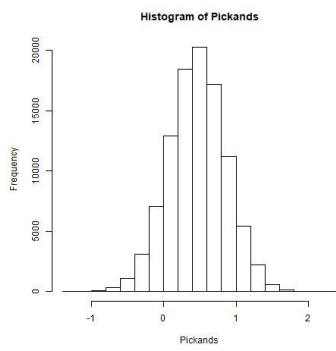


### Fréchet (0, 1, 2)

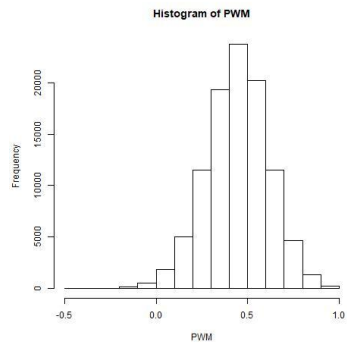
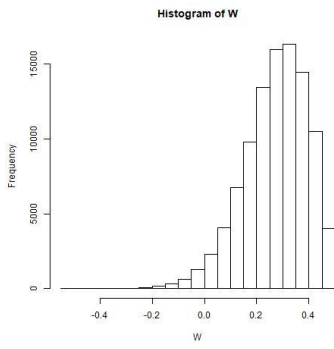
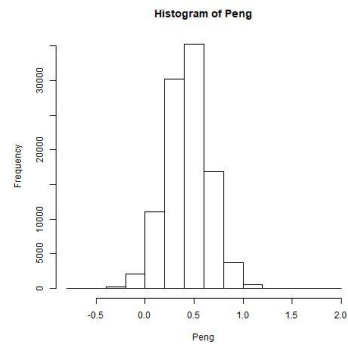
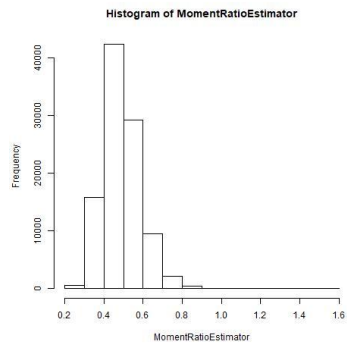
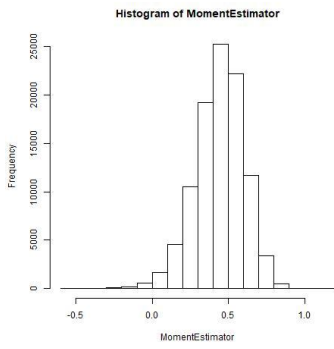
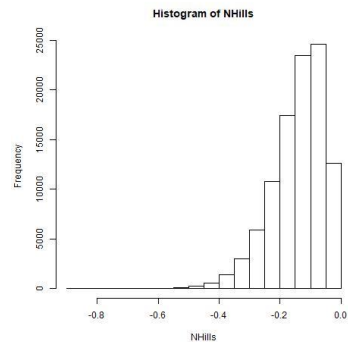
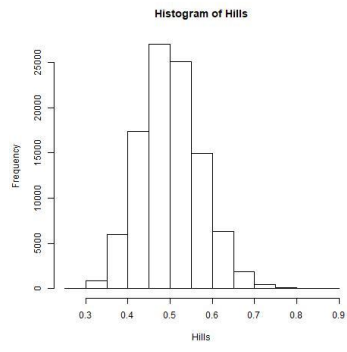
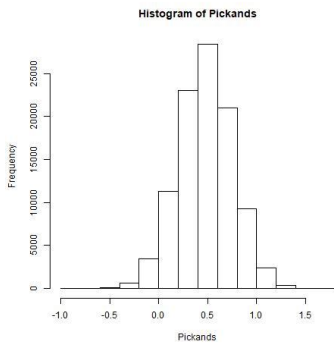




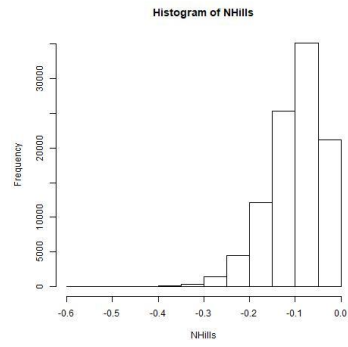
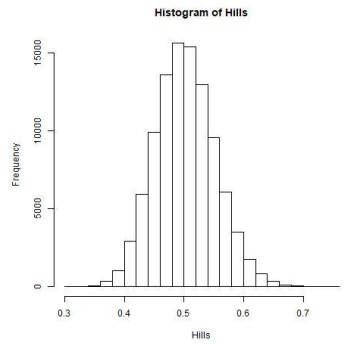
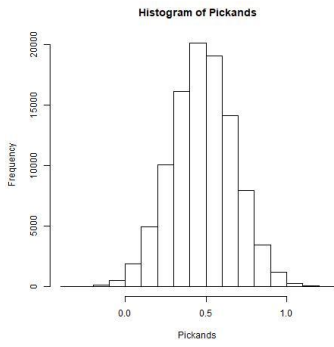
$k=25$

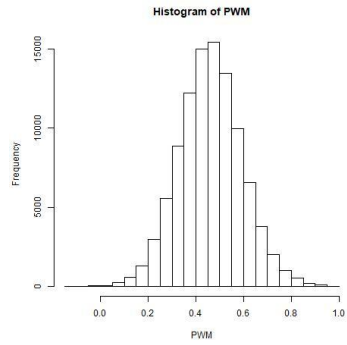
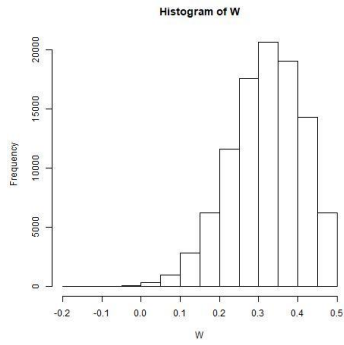
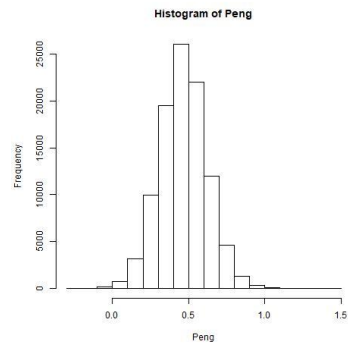
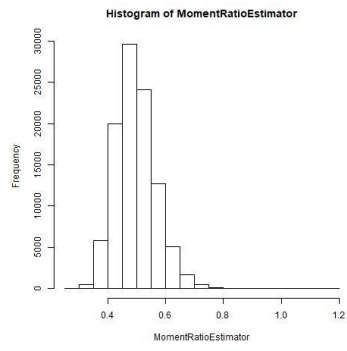
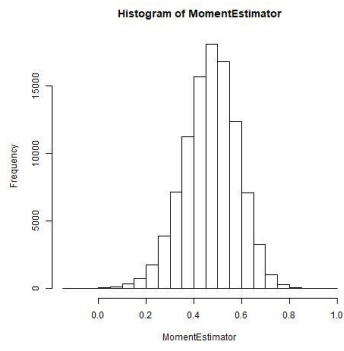


k=50

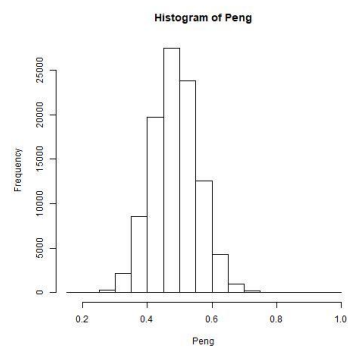
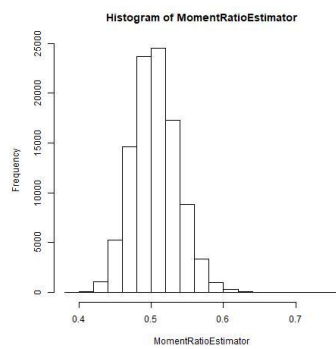
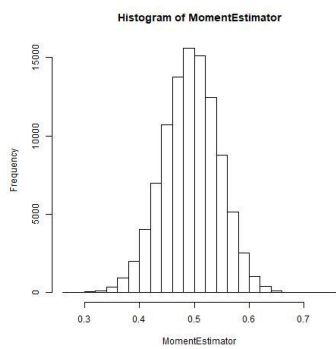
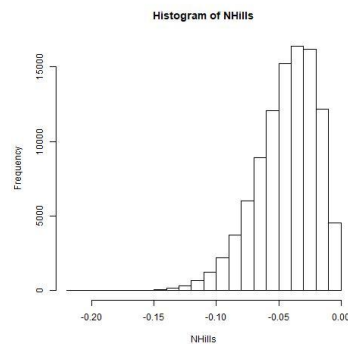
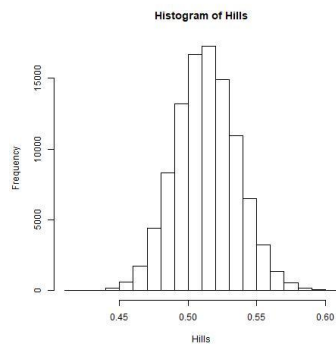
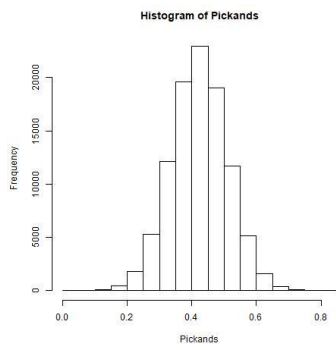


k=100

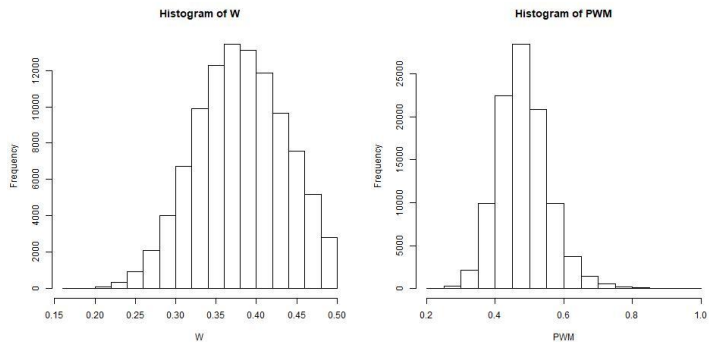




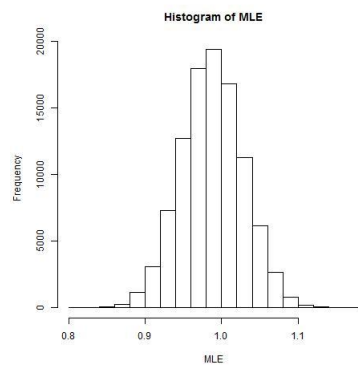
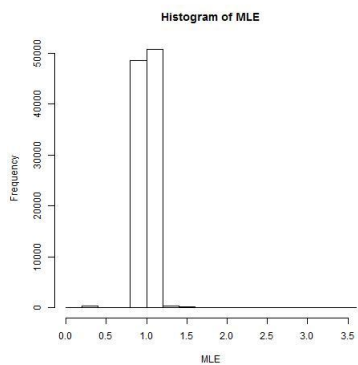
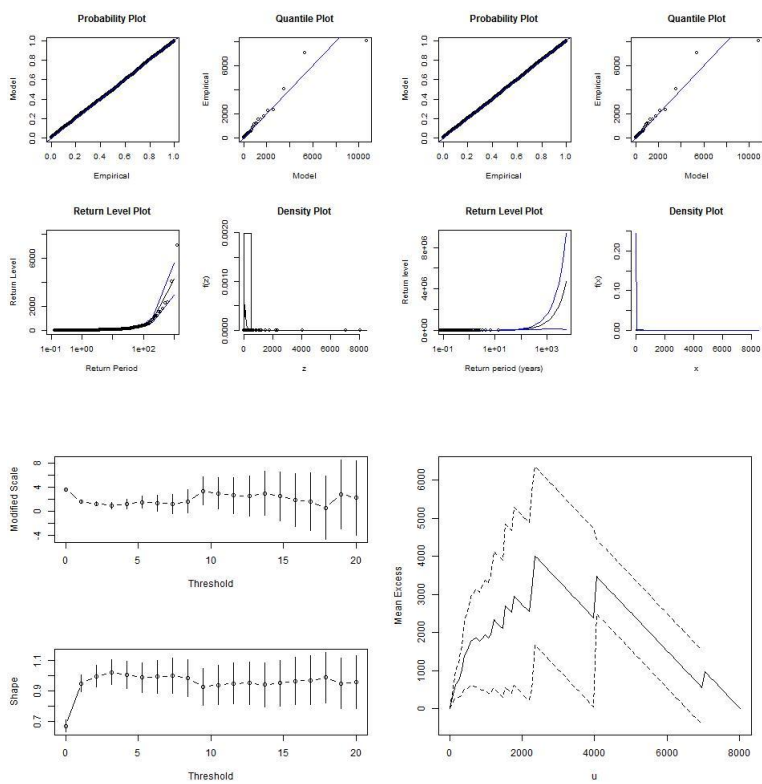
k=500



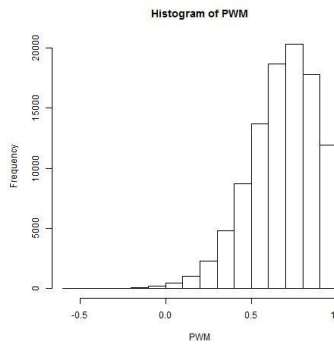
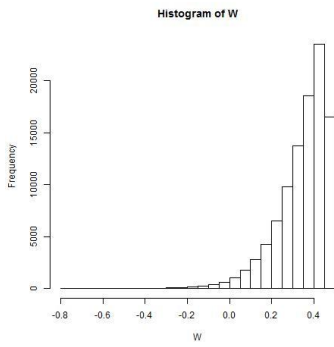
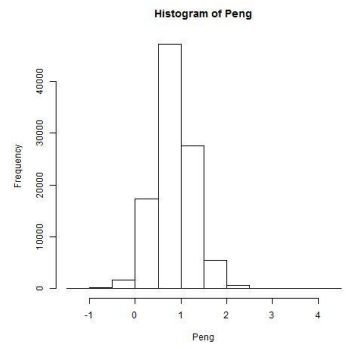
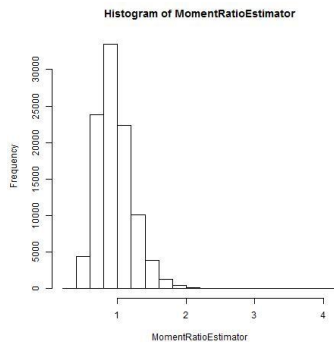
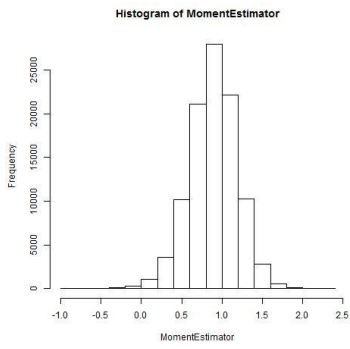
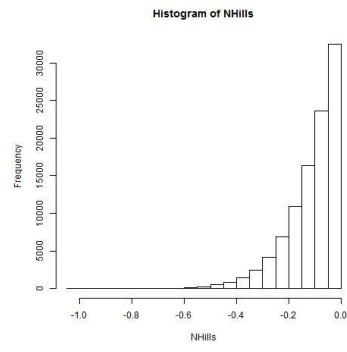
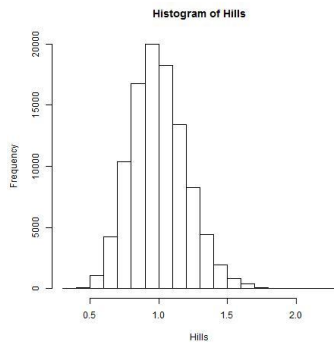
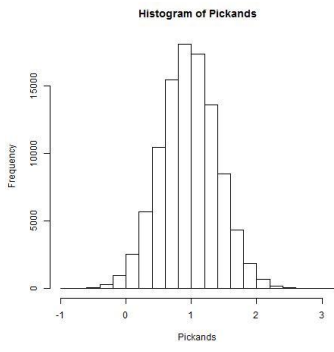




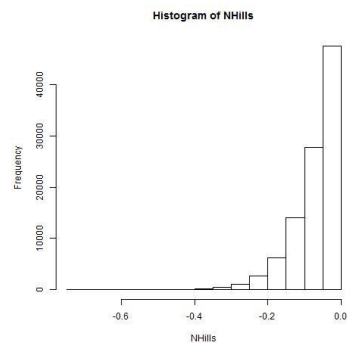
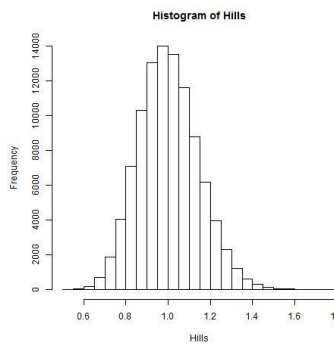
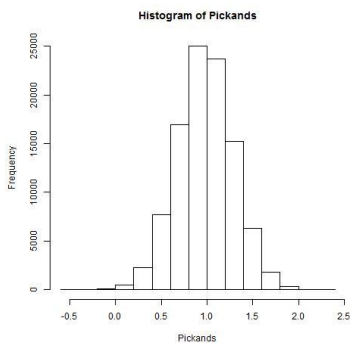
**Fréchet (0, 2, 1)**

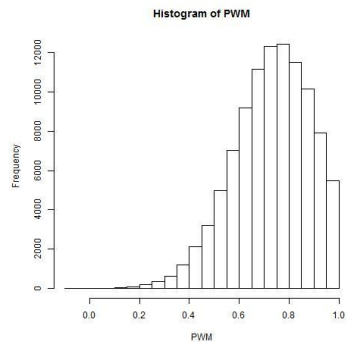
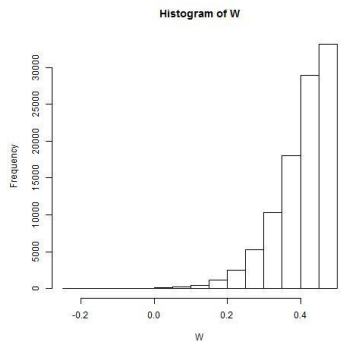
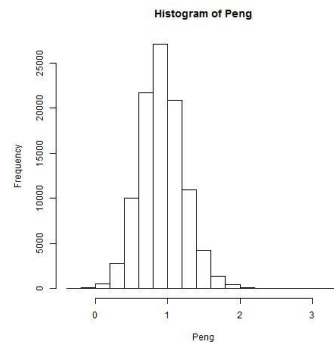
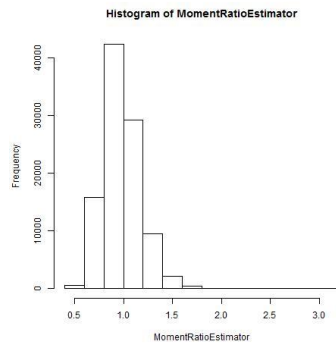
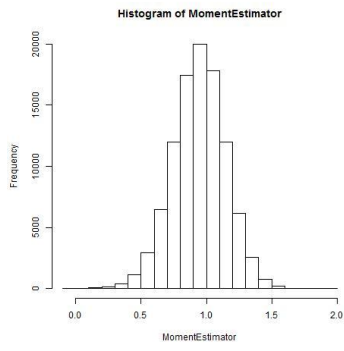


k=25

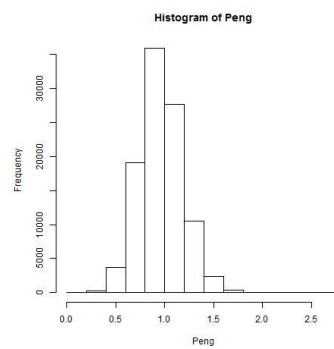
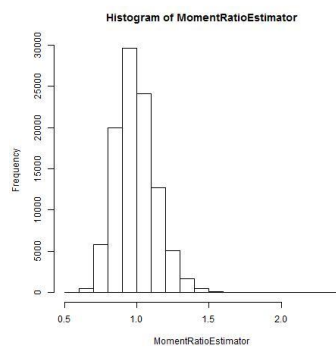
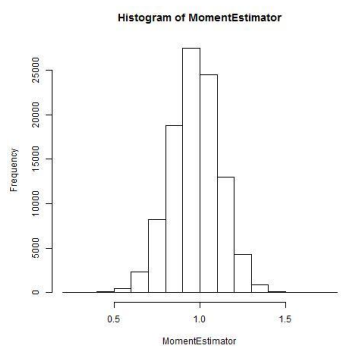
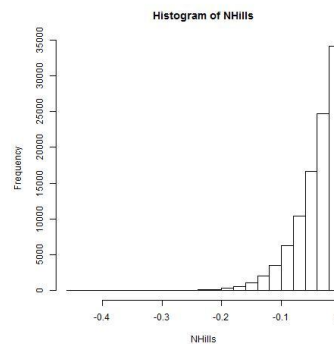
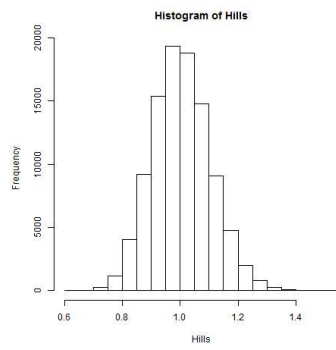
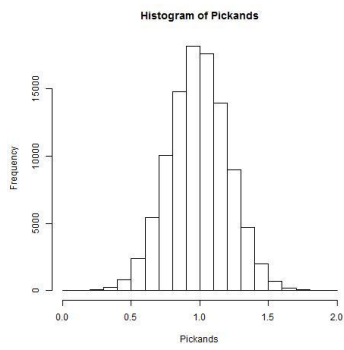


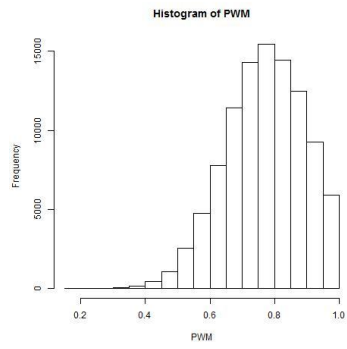
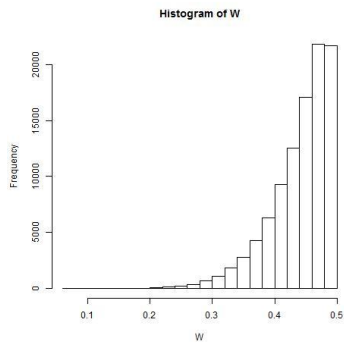
k=50



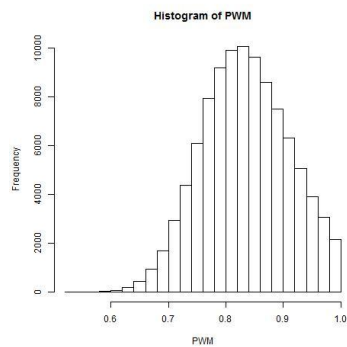
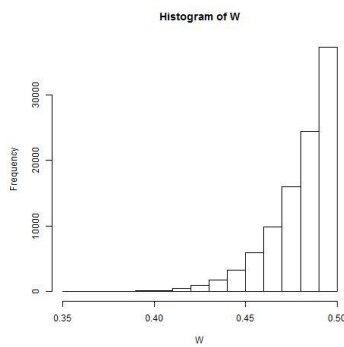
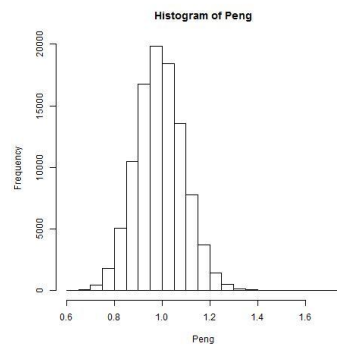
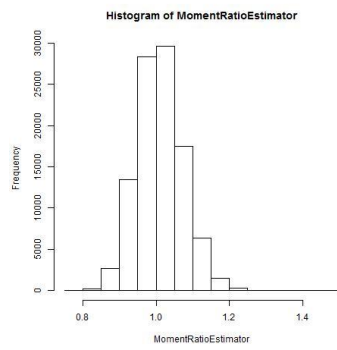
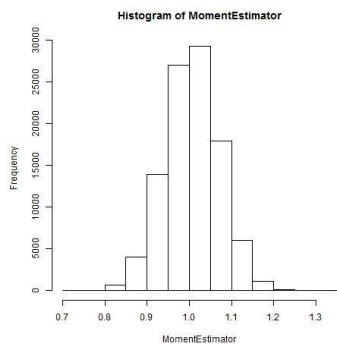
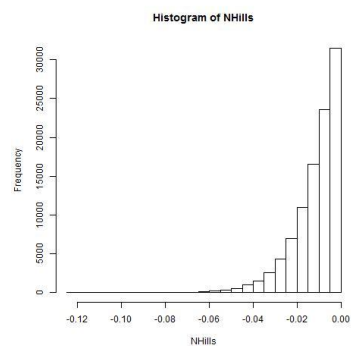
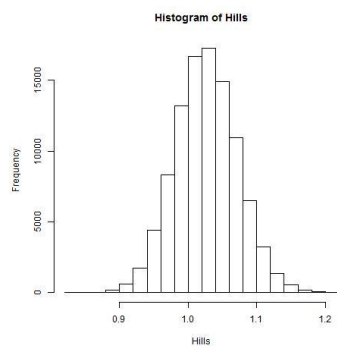
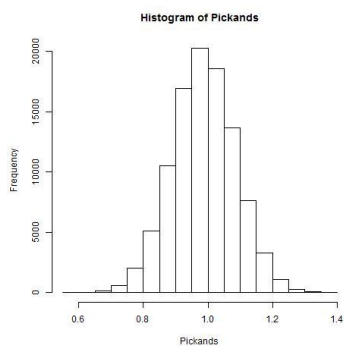


k=100

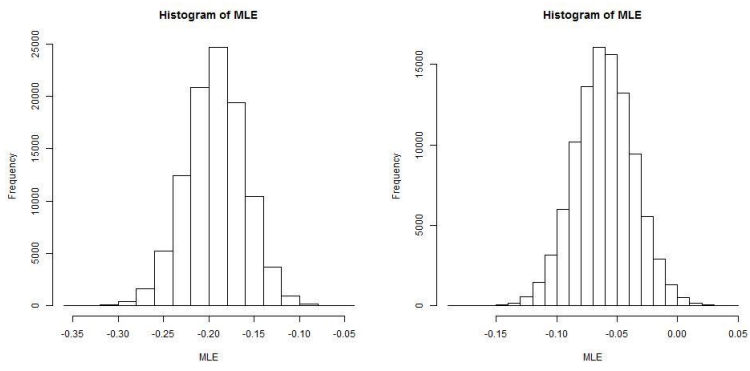
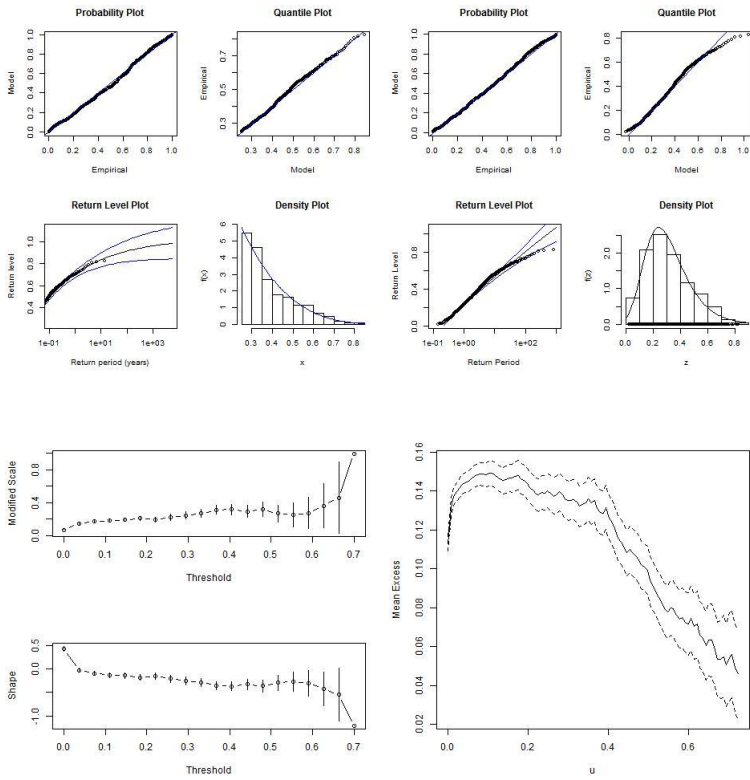




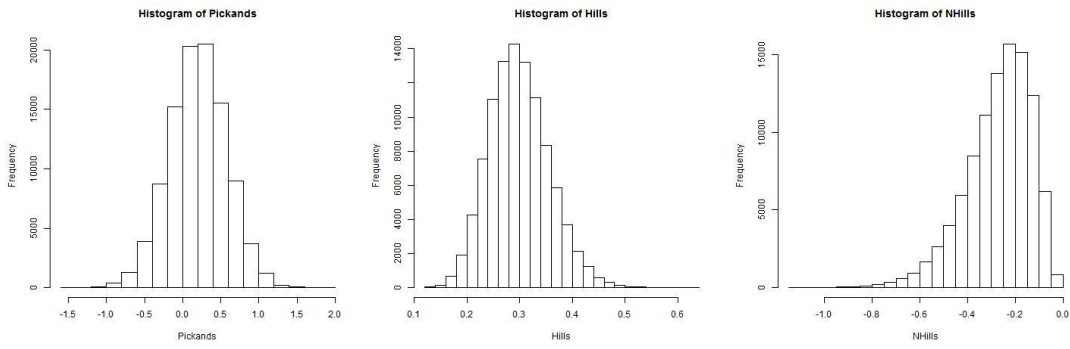
k=500

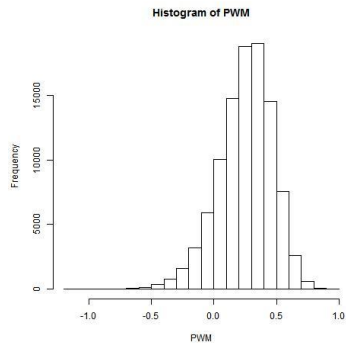
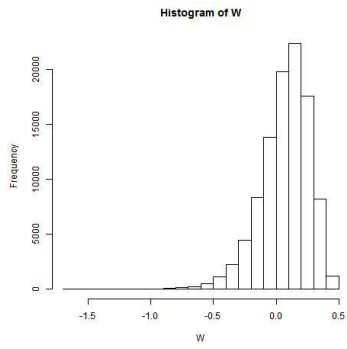
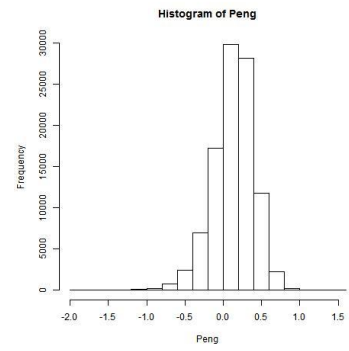
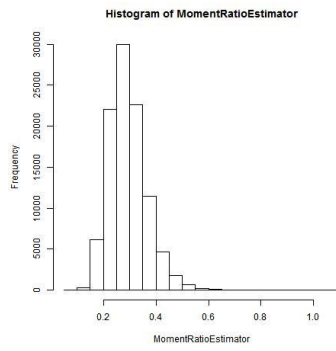
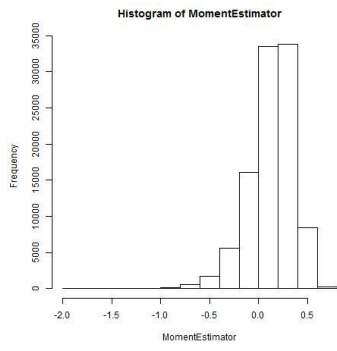


**Burr**  
**Burr (4, 1, 1)**

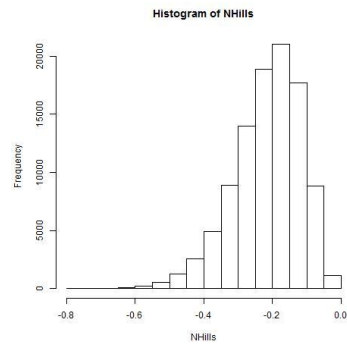
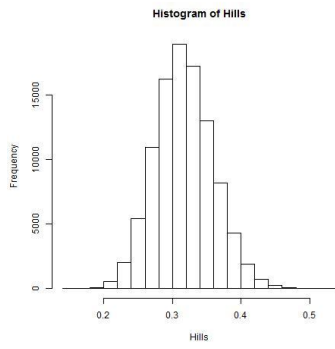
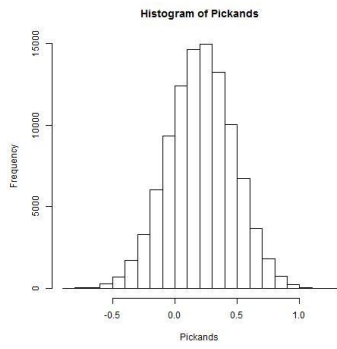


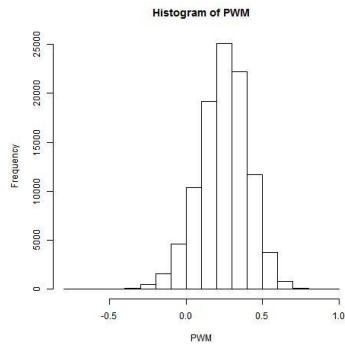
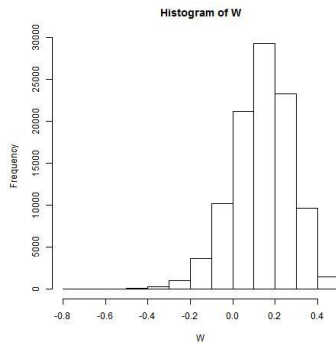
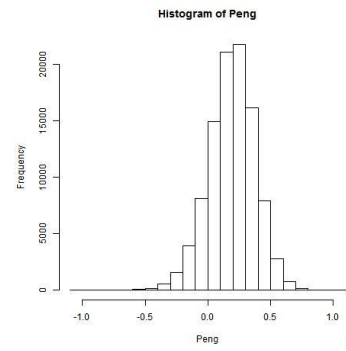
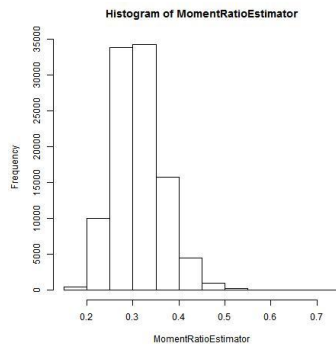
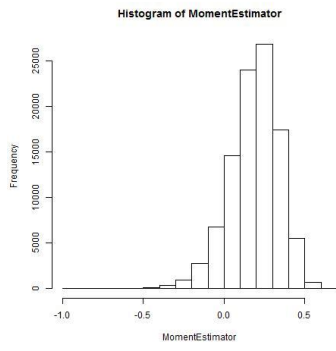
$k=25$



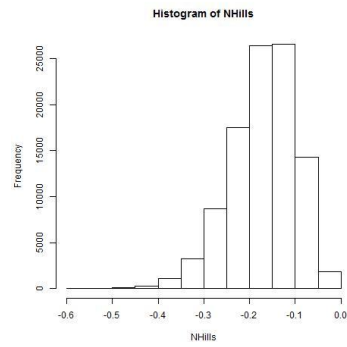
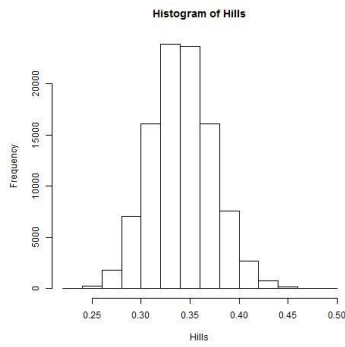
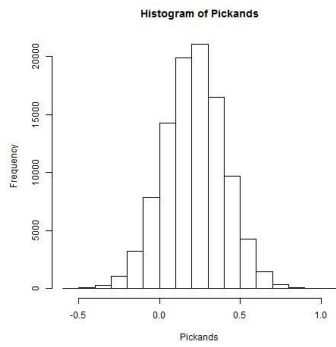


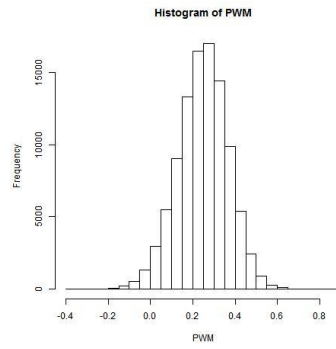
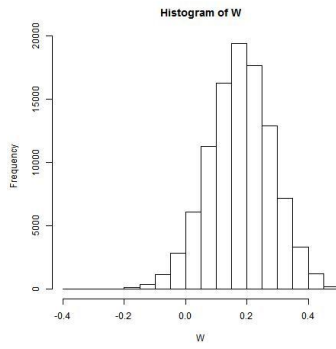
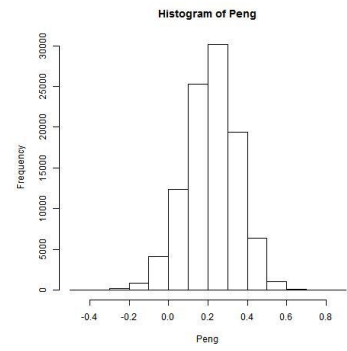
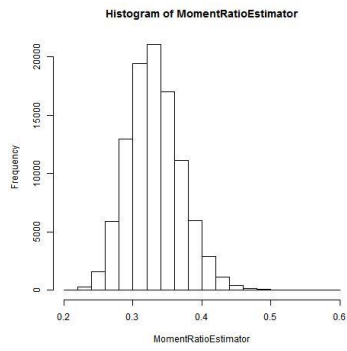
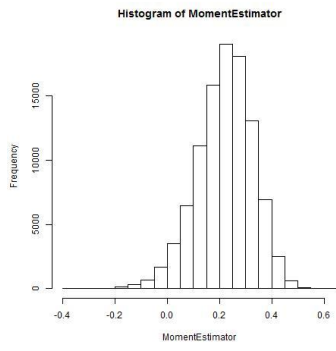
k=50



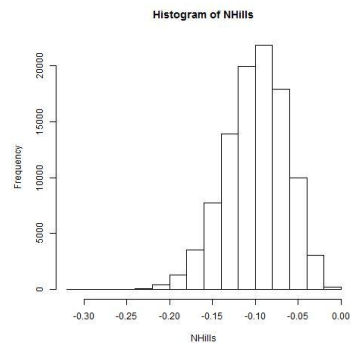
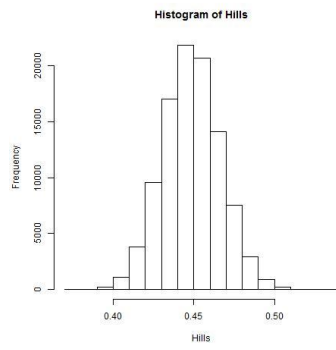
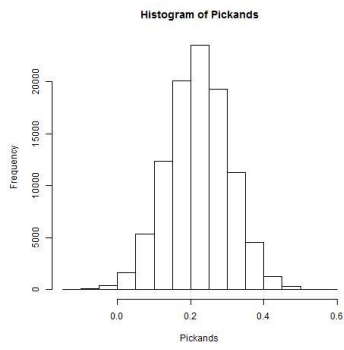


$k=100$

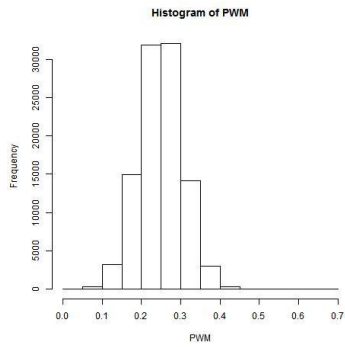
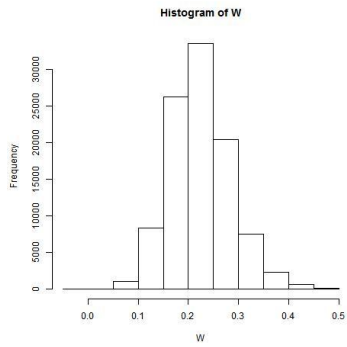
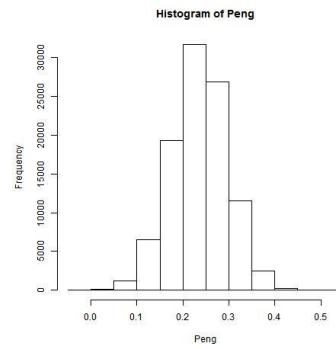
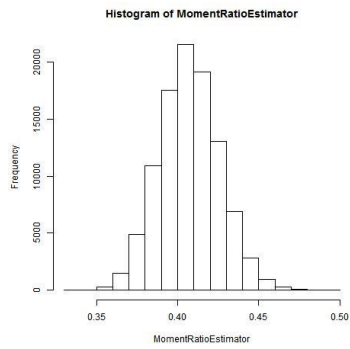
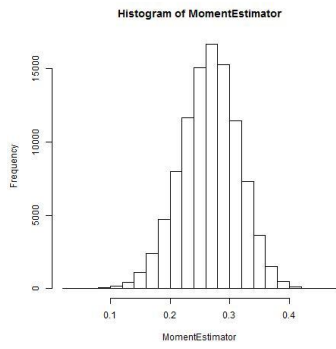




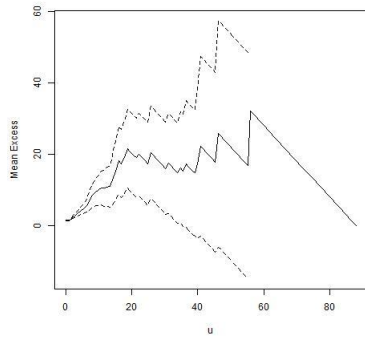
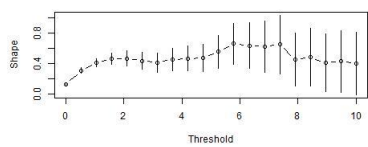
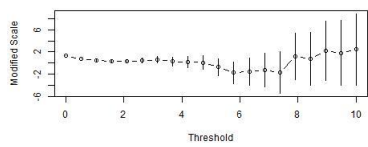
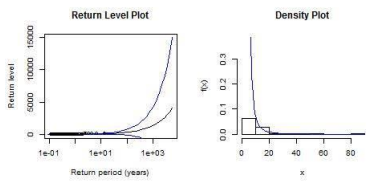
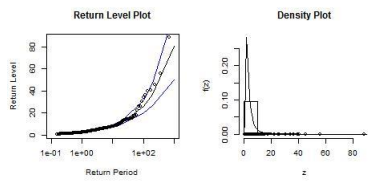
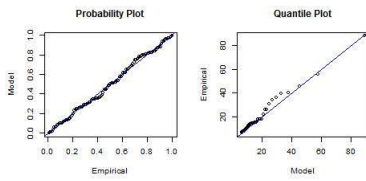
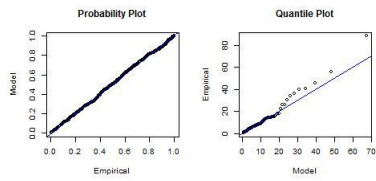
k=500

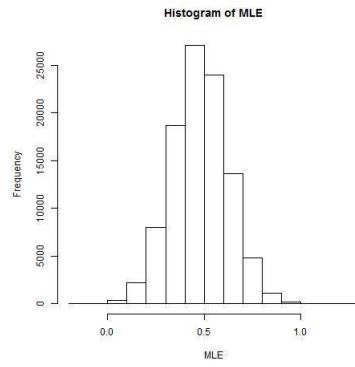
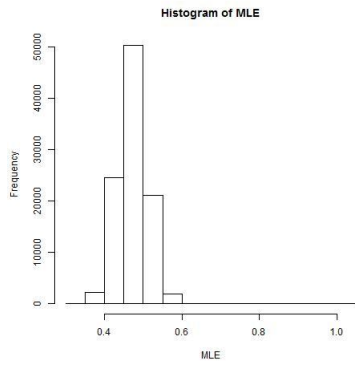




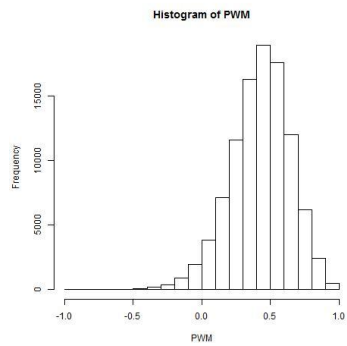
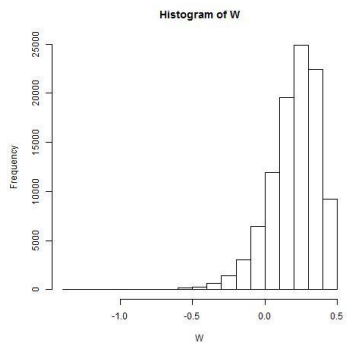
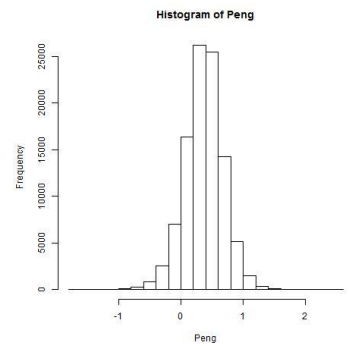
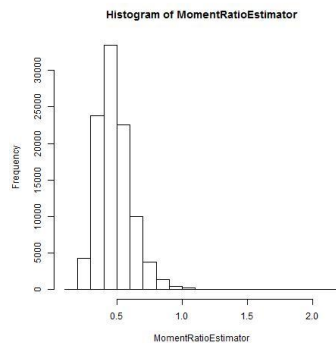
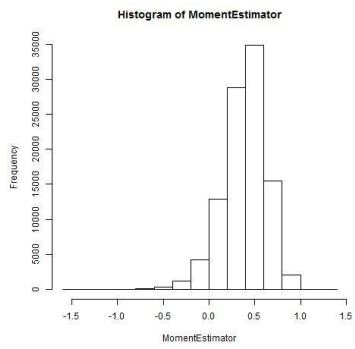
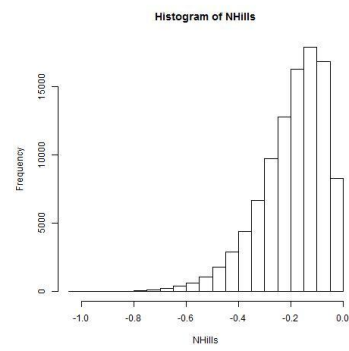
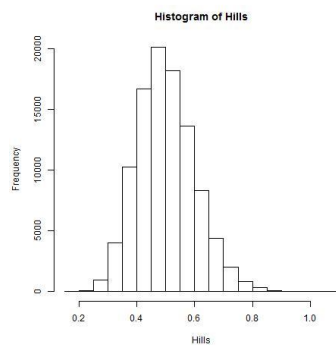
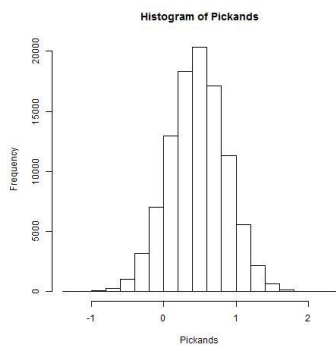


**Burr (1, 2, 1)**

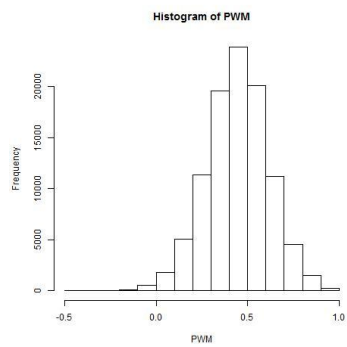
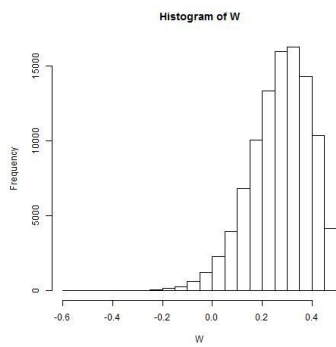
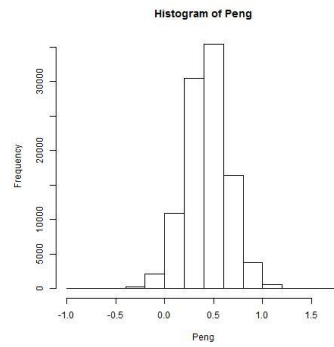
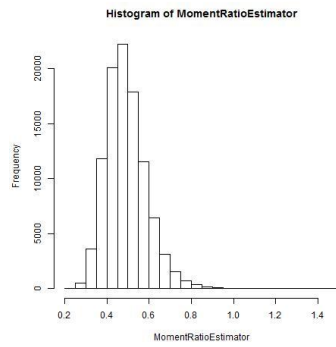
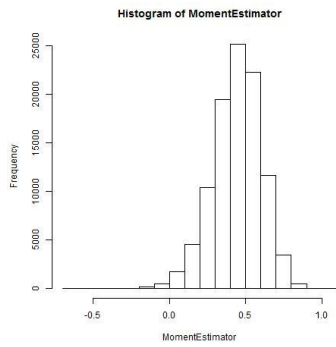
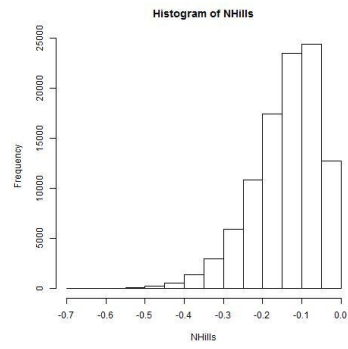
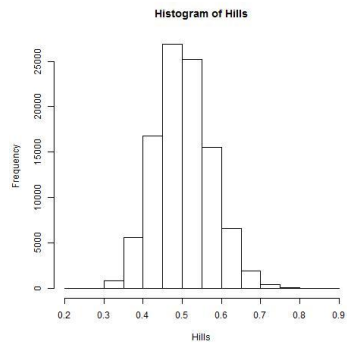
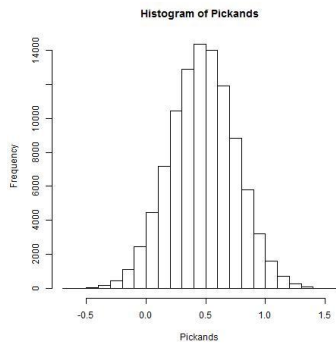




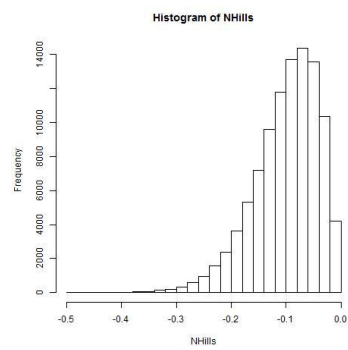
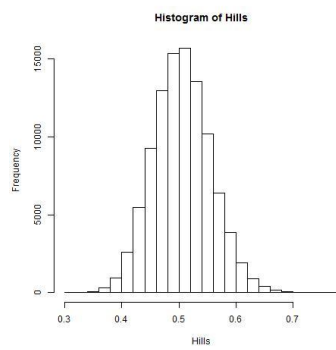
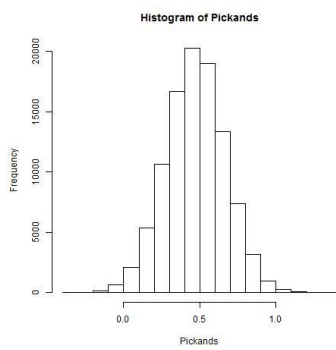
k=25

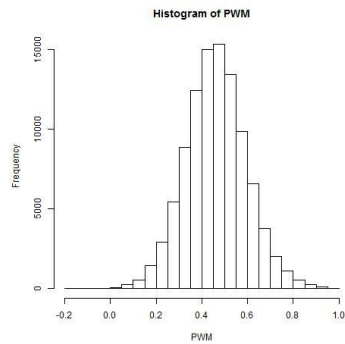
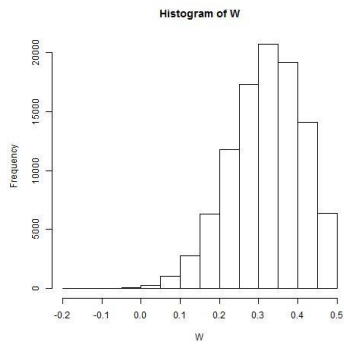
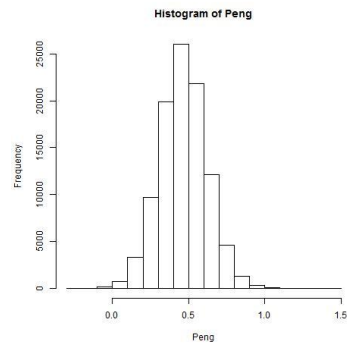
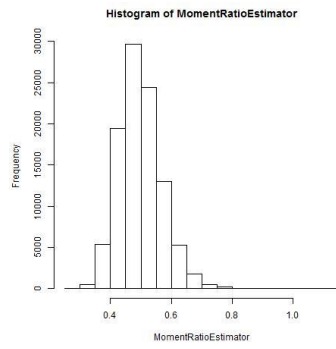
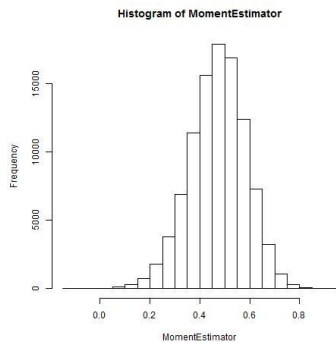


k=50

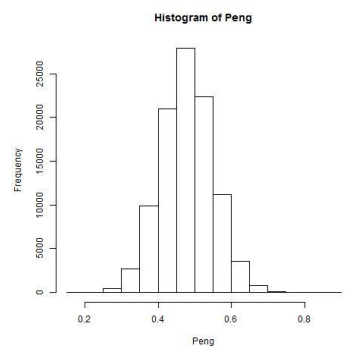
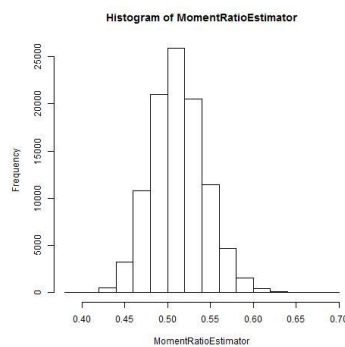
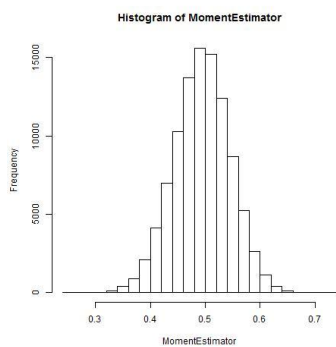
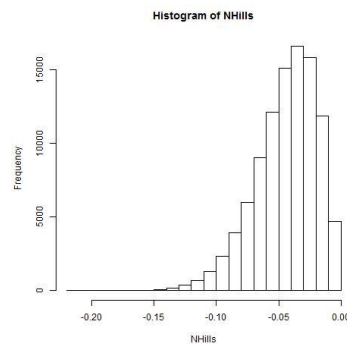
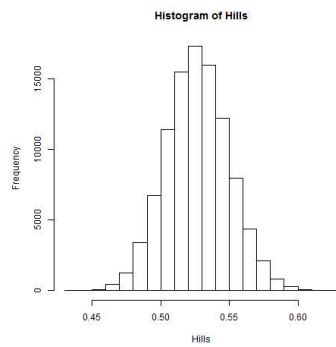
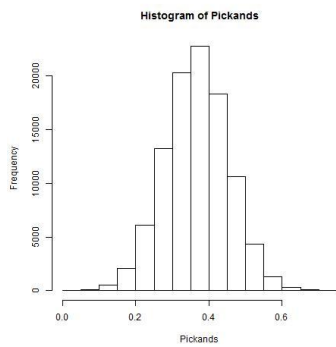


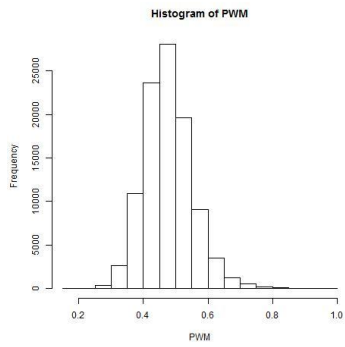
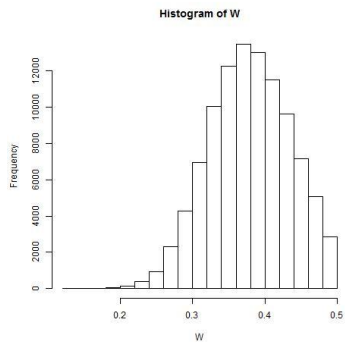
k=100



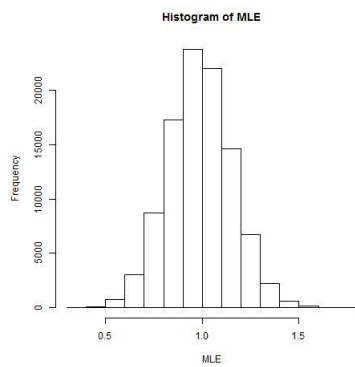
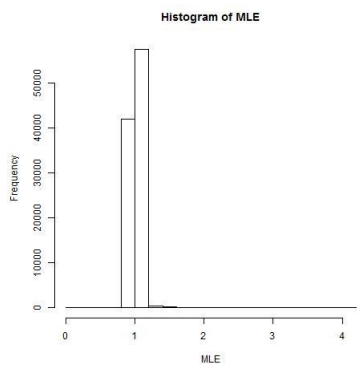
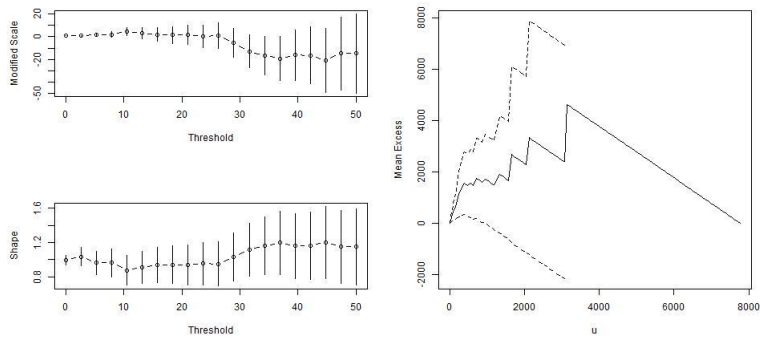
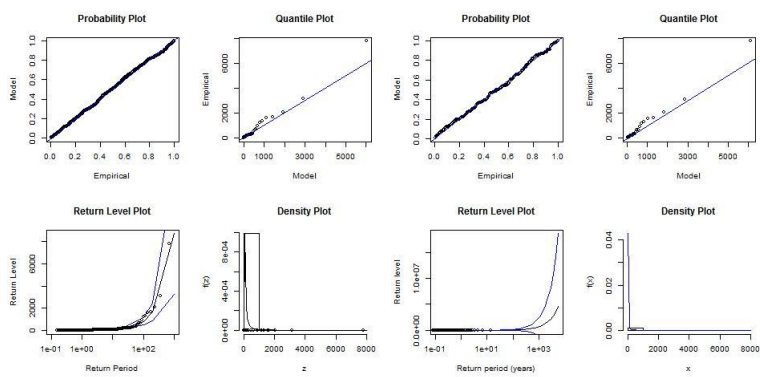


k=500

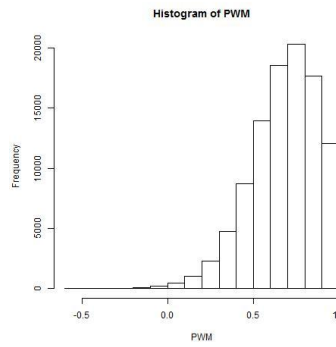
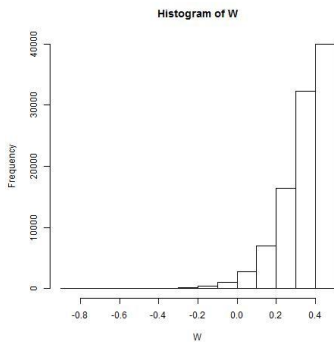
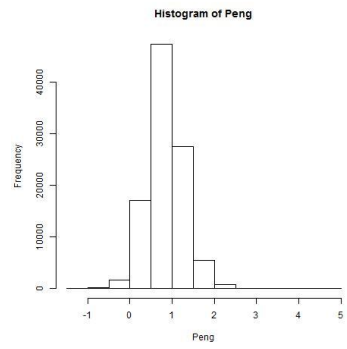
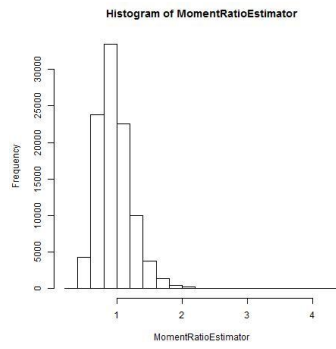
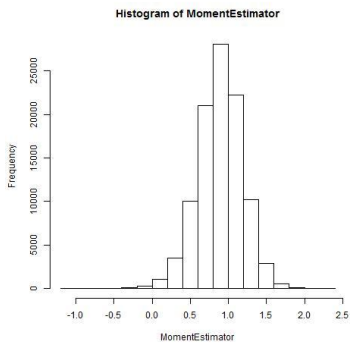
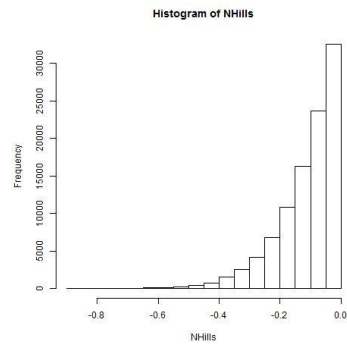
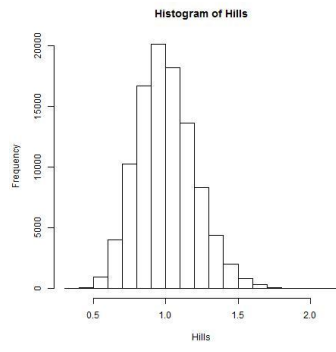
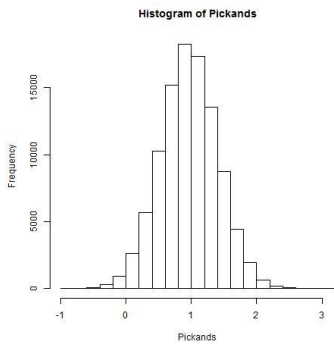




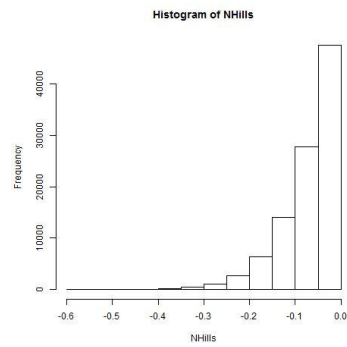
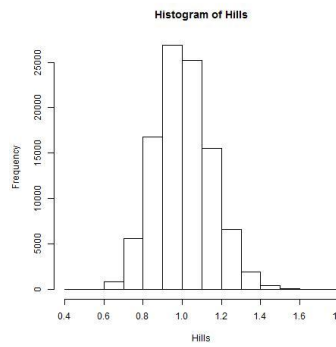
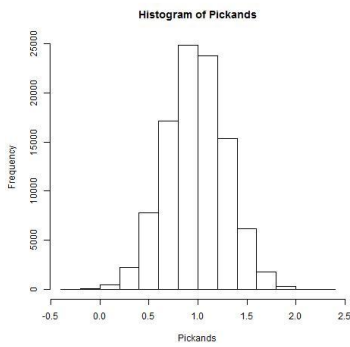
## Burr (1, 1, 1)

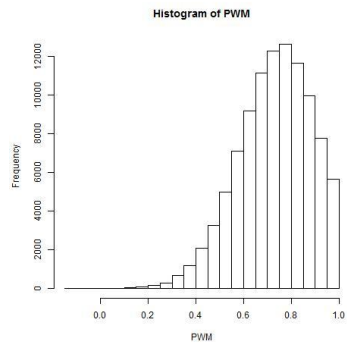
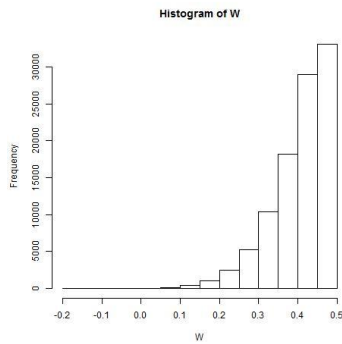
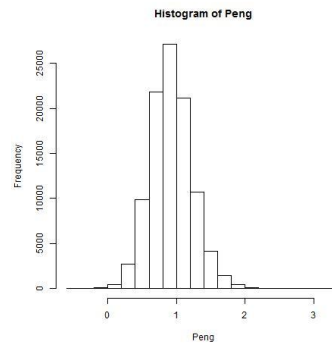
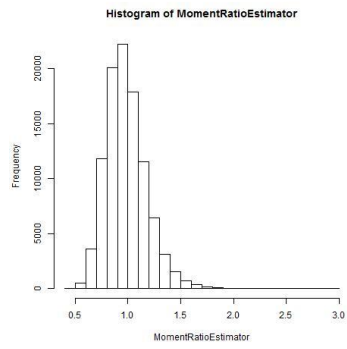
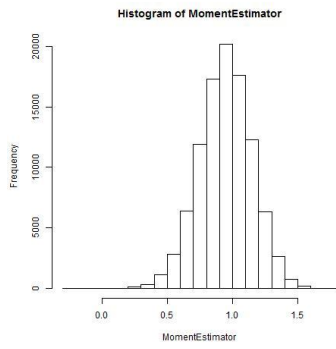


k=25

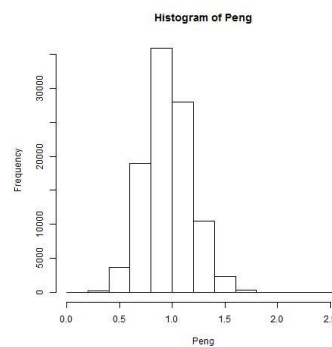
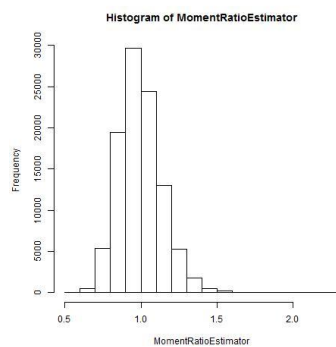
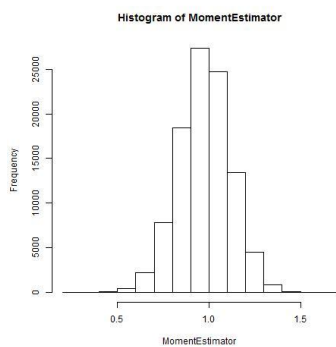
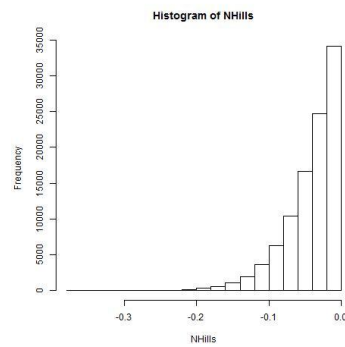
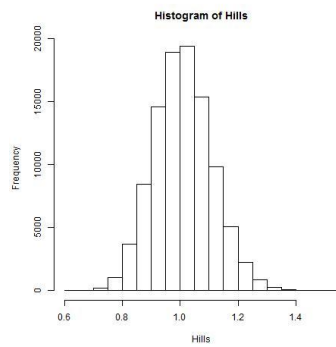
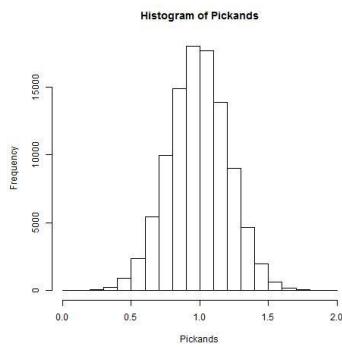


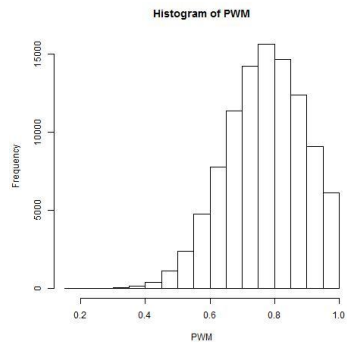
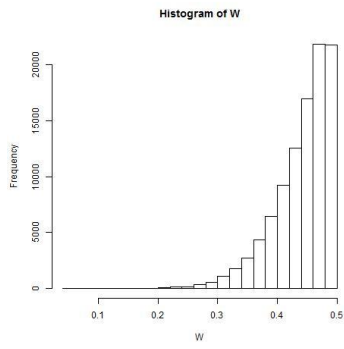
k=50



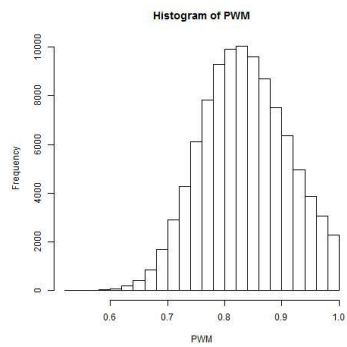
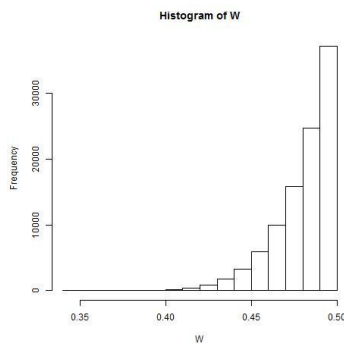
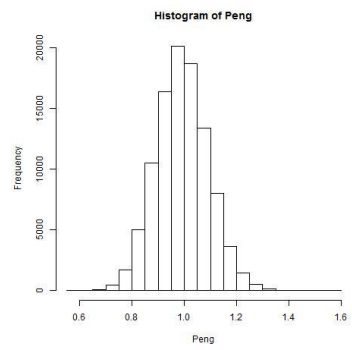
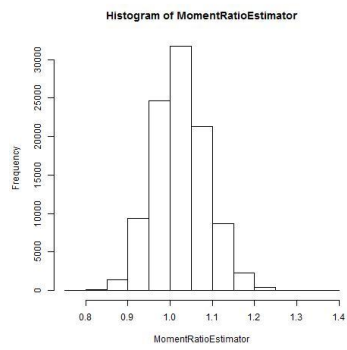
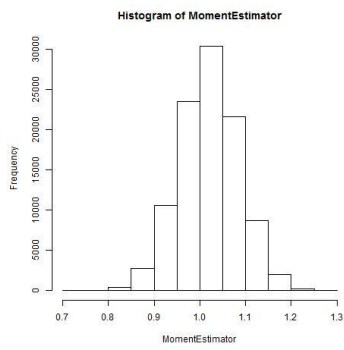
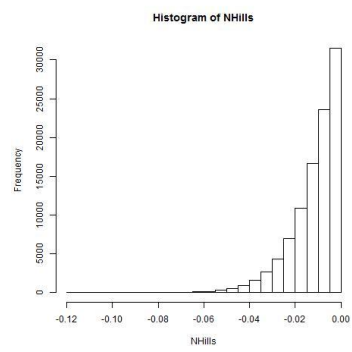
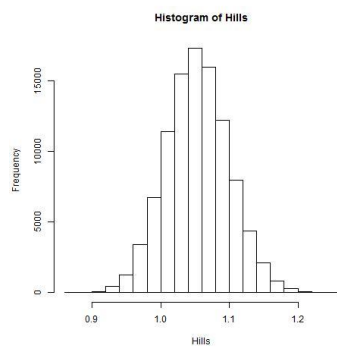
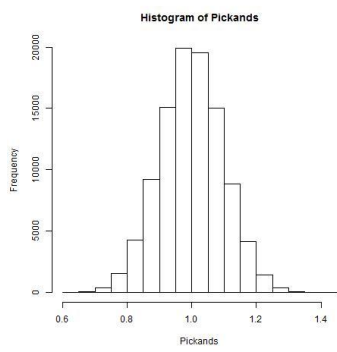


k=100





k=500

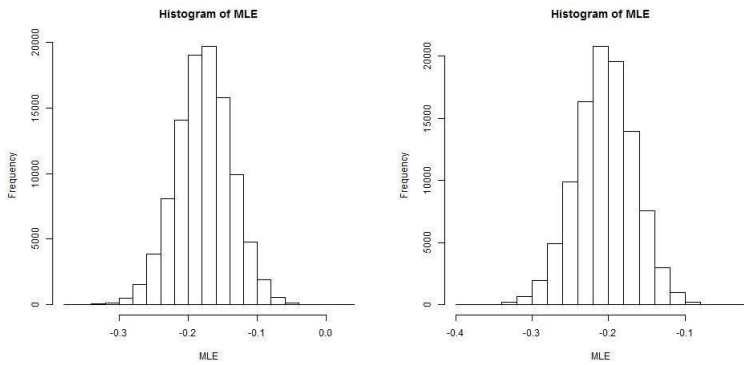
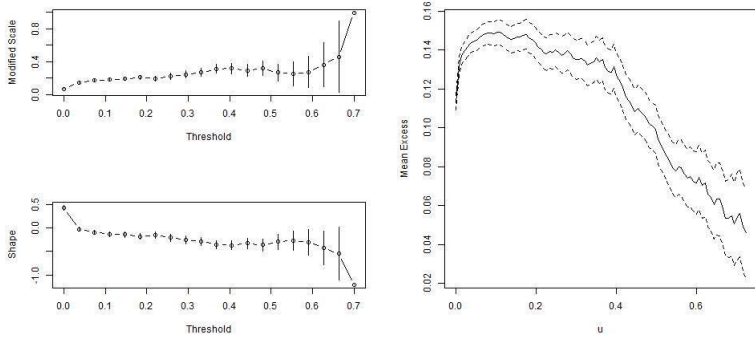
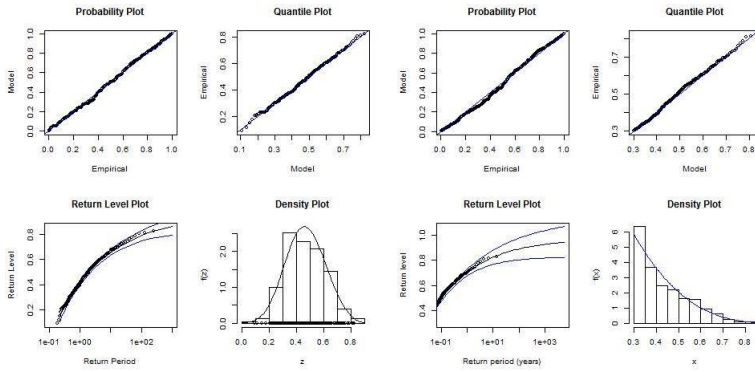




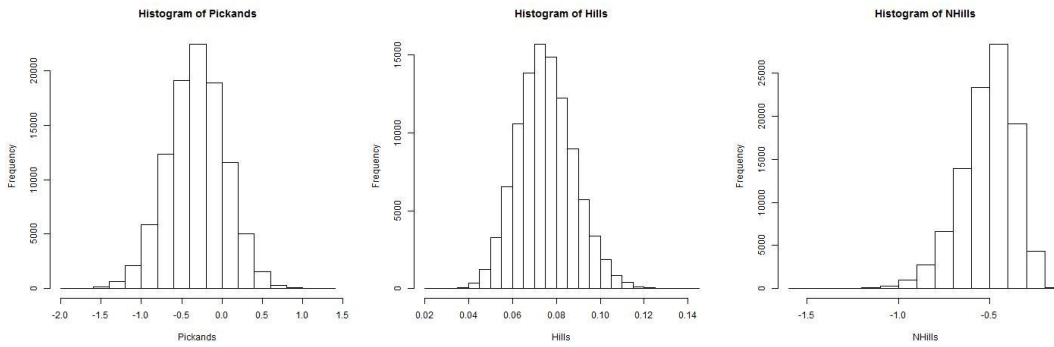
$\beta < 0$

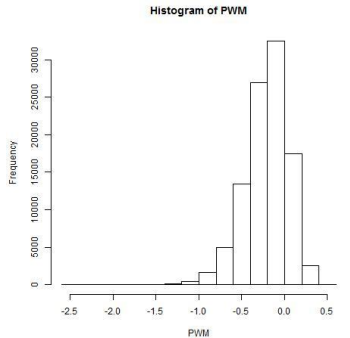
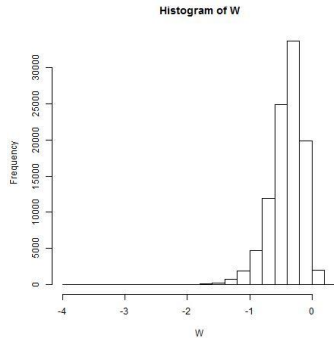
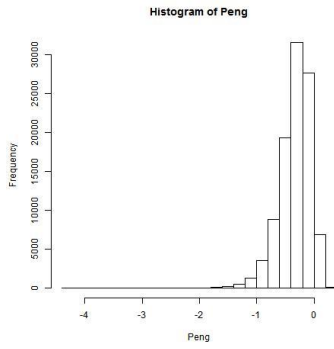
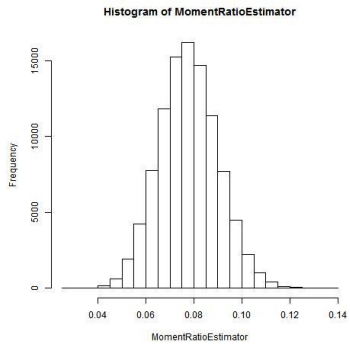
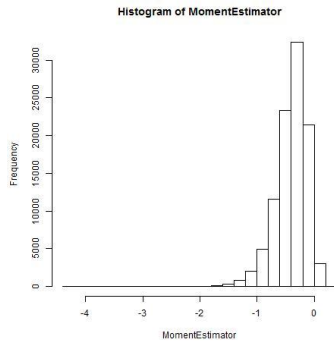
**Beta**

**Beta (0.5, 4)**

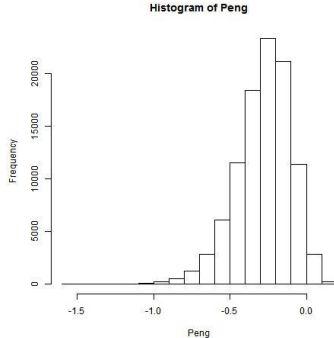
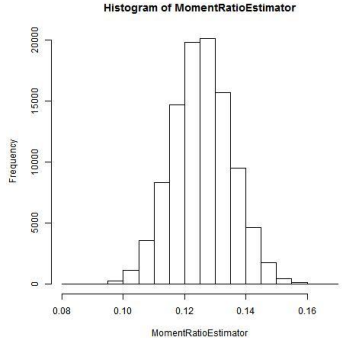
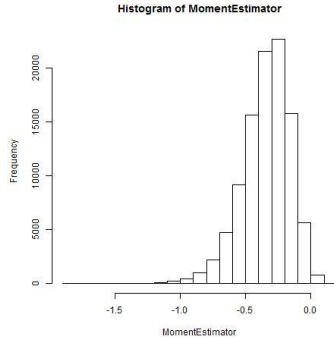
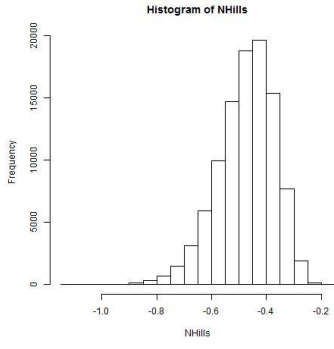
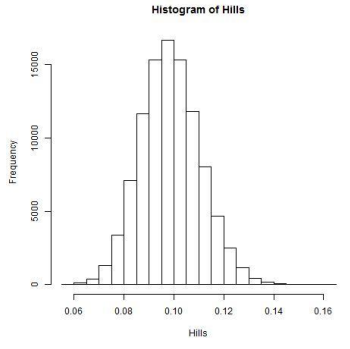
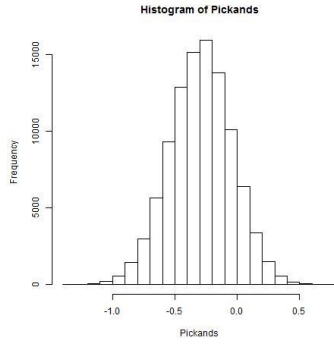


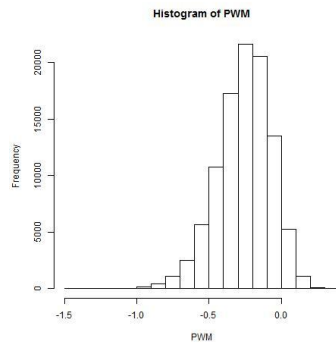
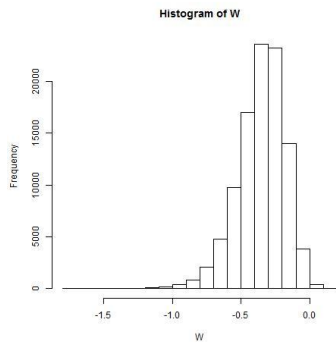
**k=25**



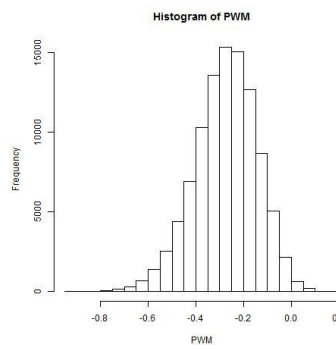
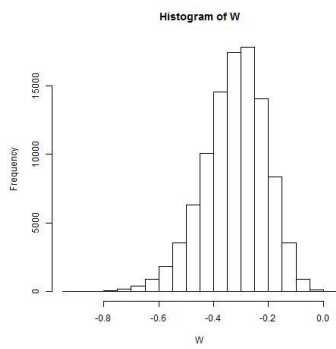
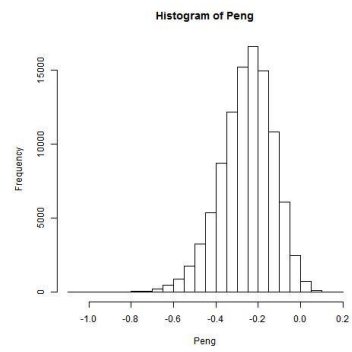
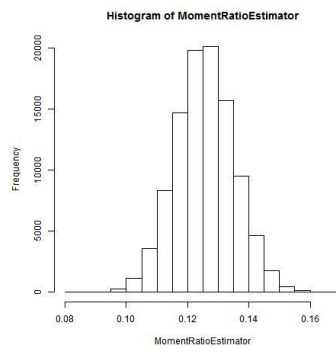
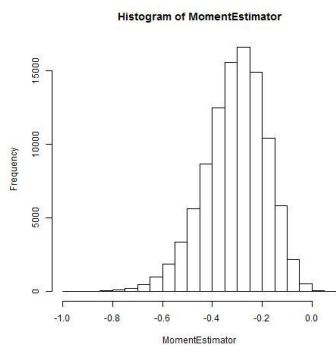
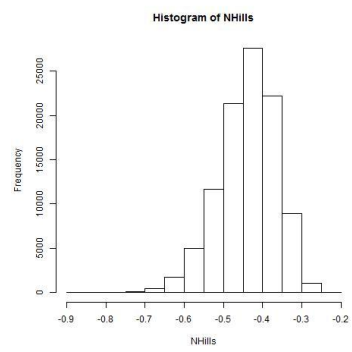
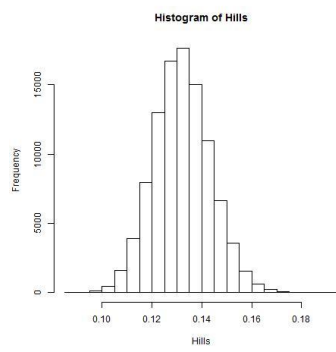
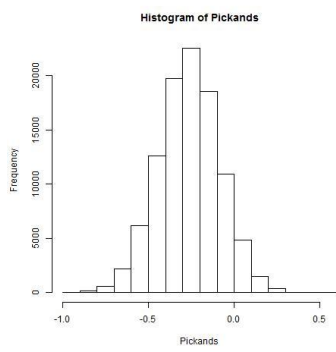


k=50

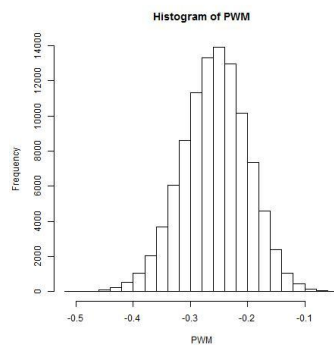
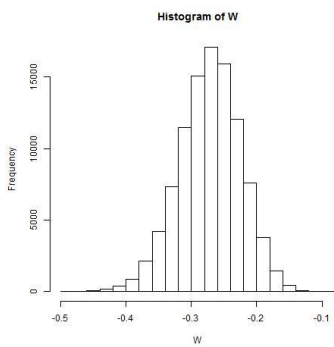
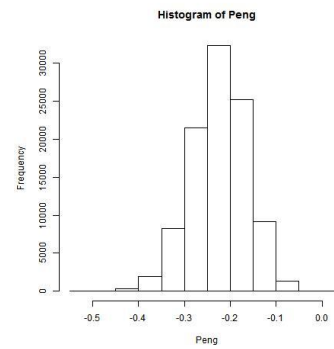
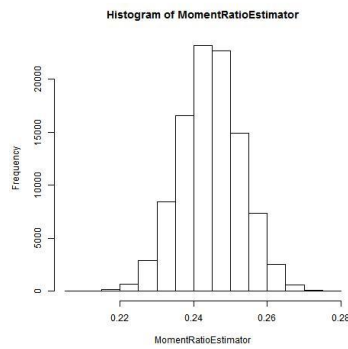
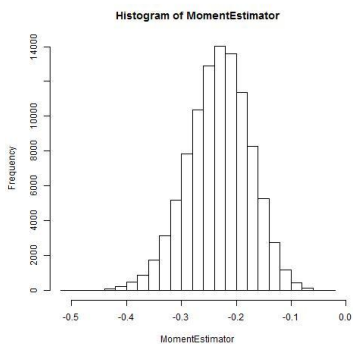
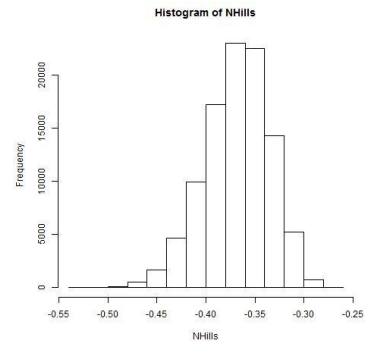
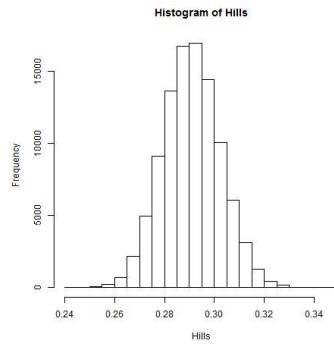
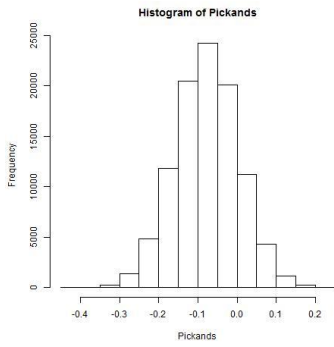




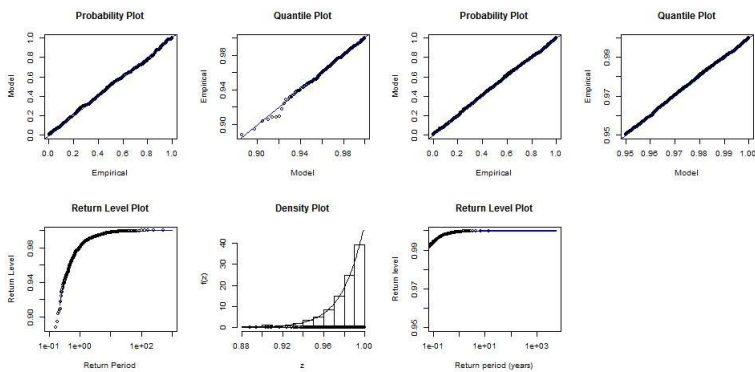
k=100

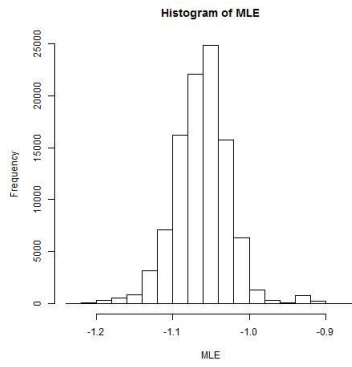
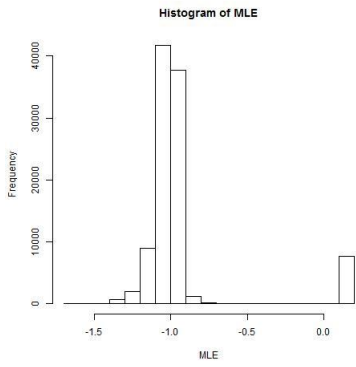
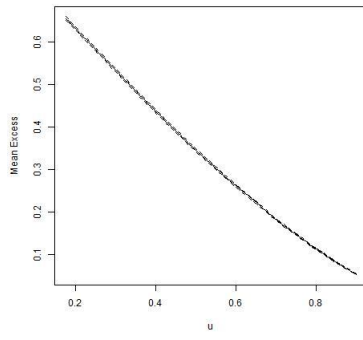
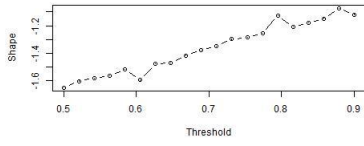
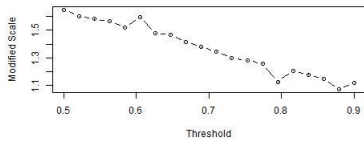


k=500

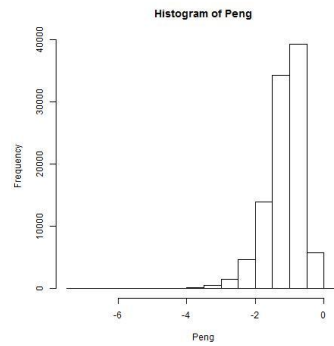
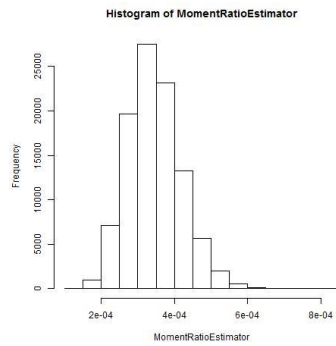
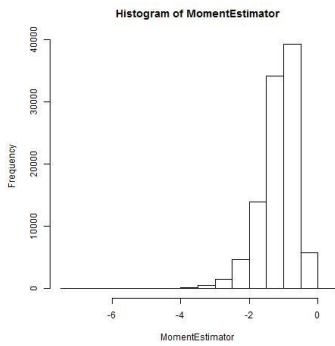
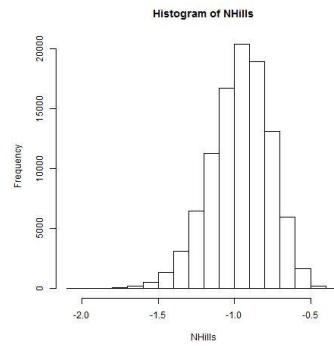
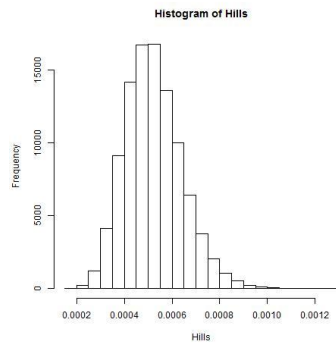
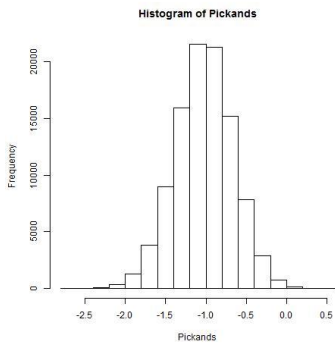


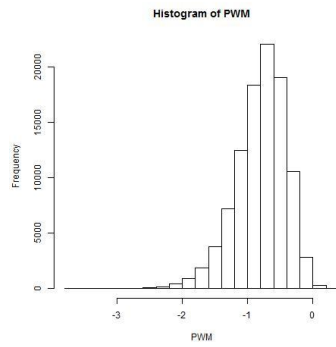
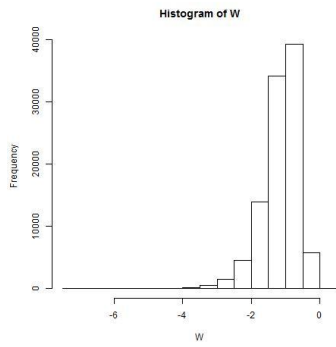
**Beta (5, 1)**



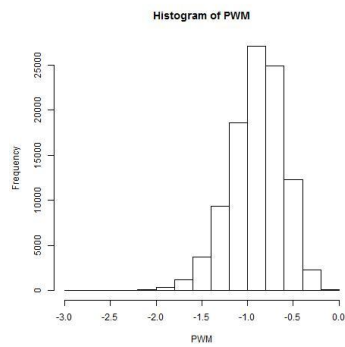
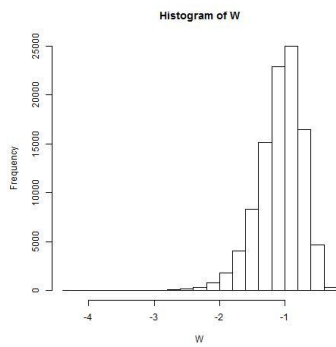
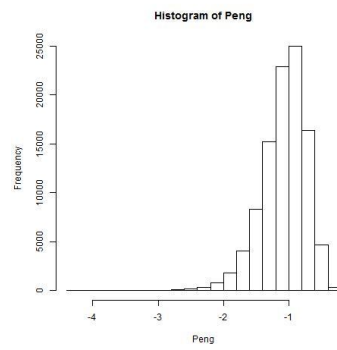
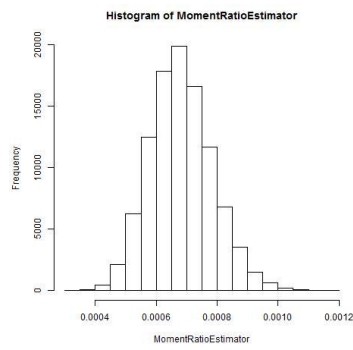
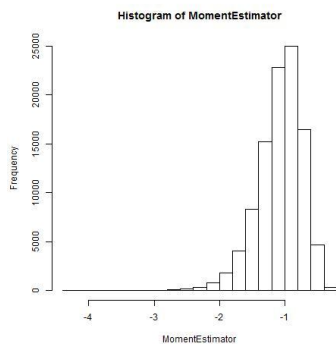
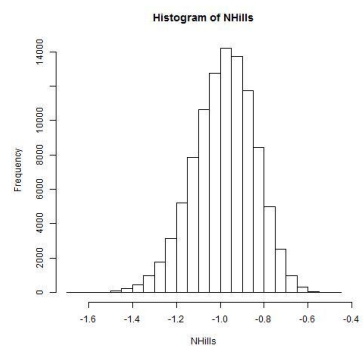
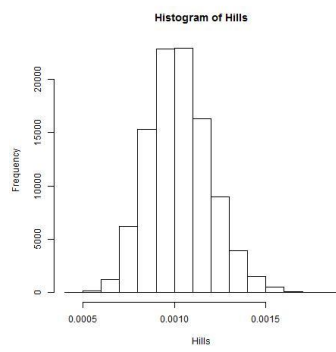
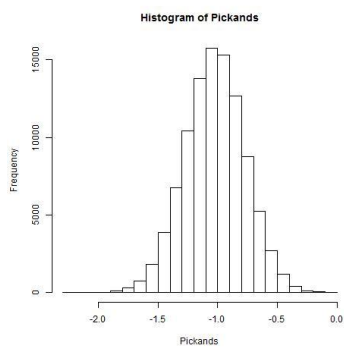


k=25

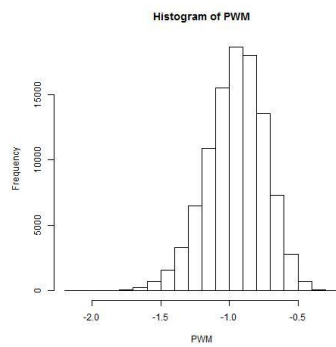
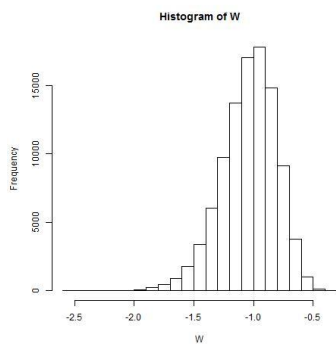
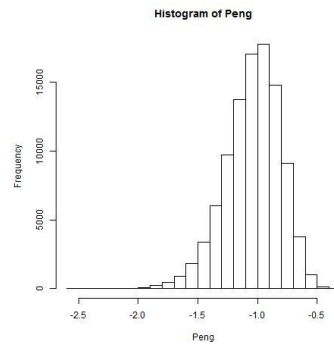
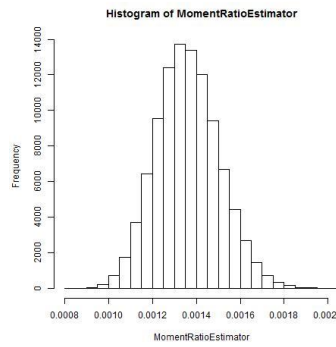
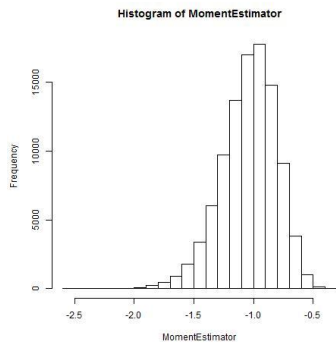
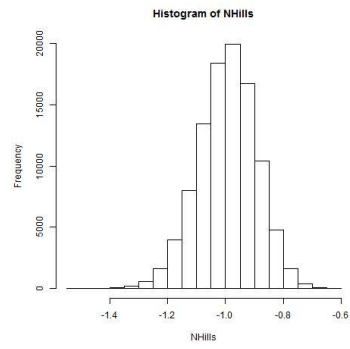
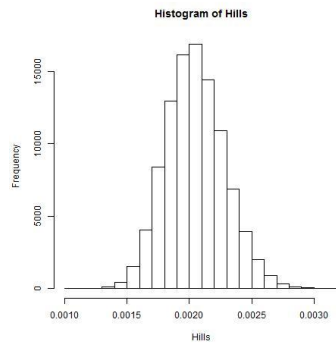
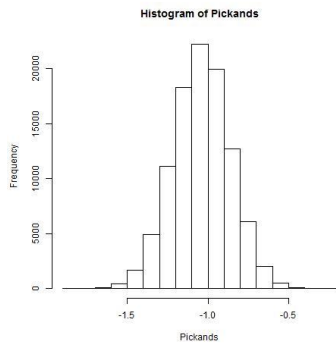




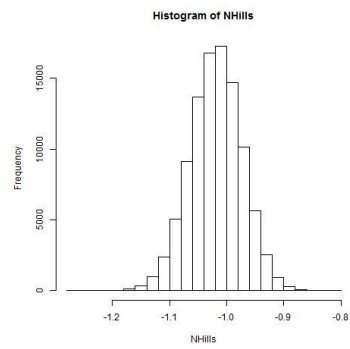
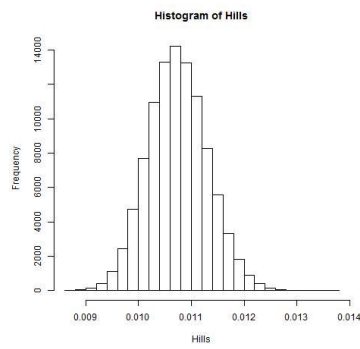
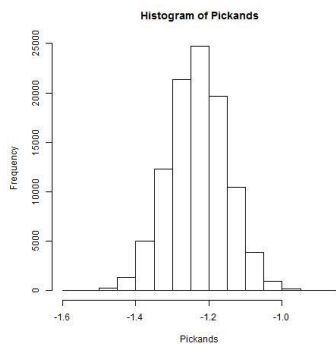
k=50

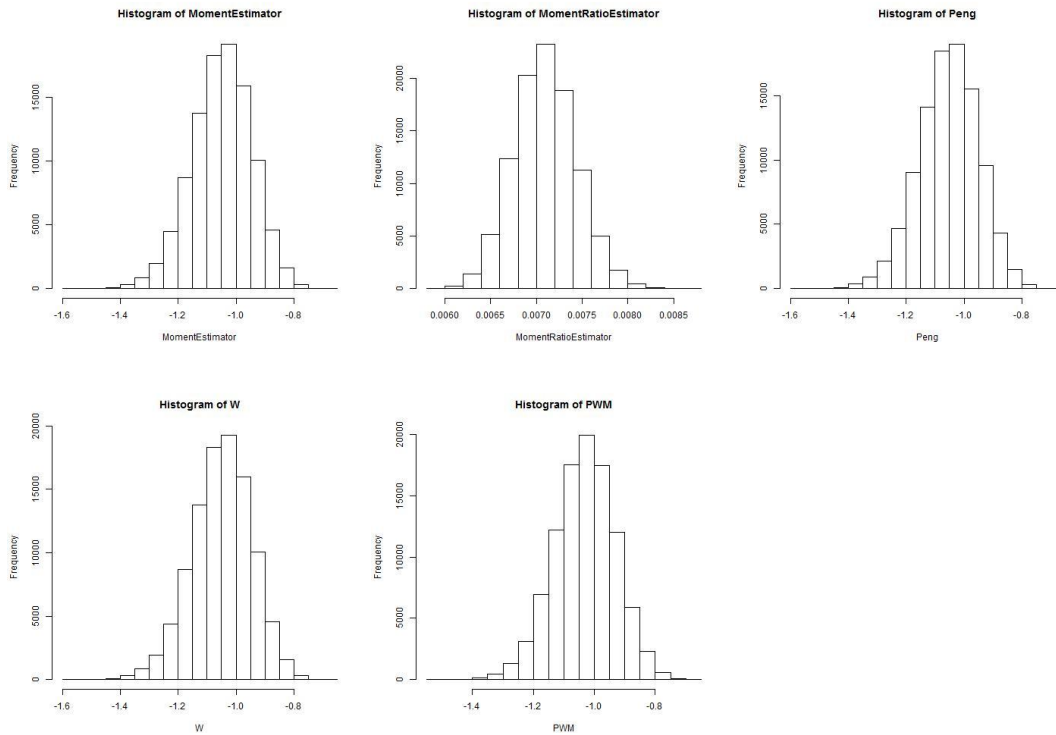


k=100

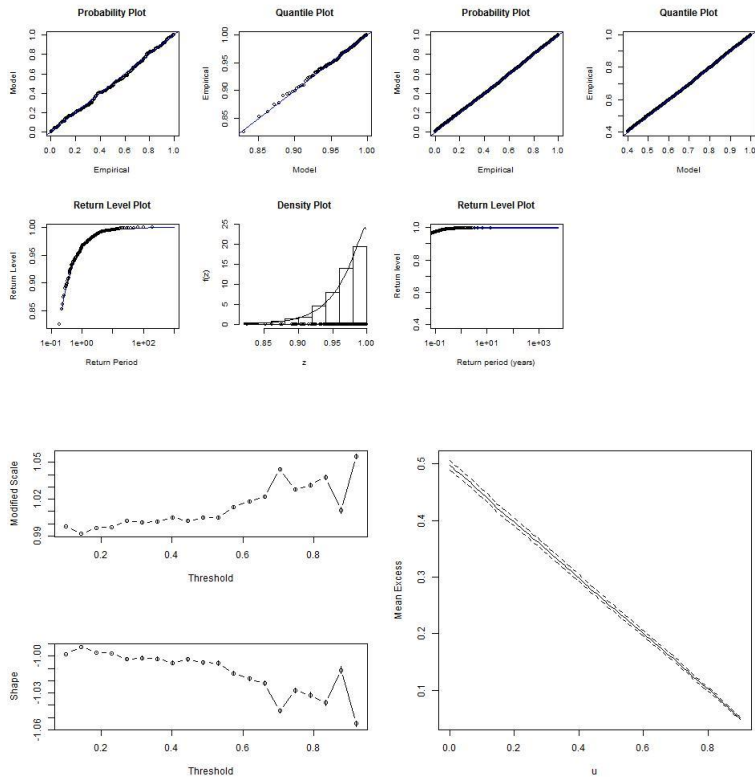


k=500

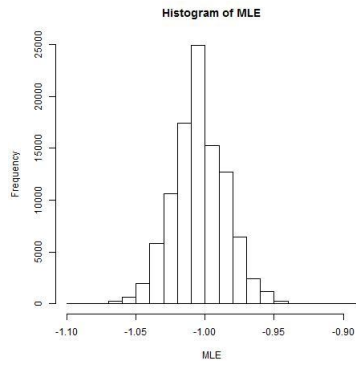
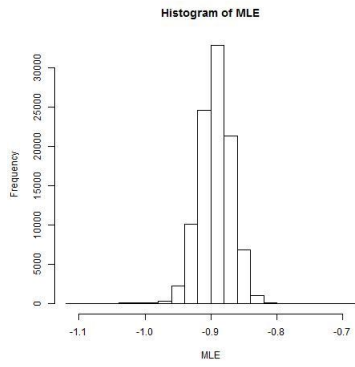




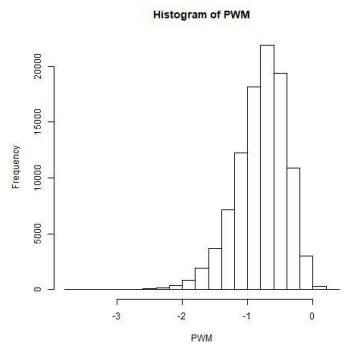
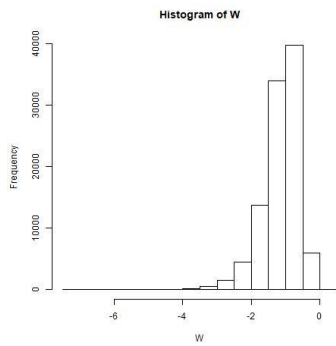
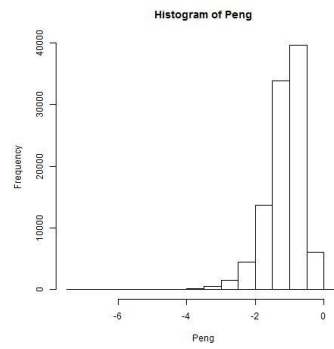
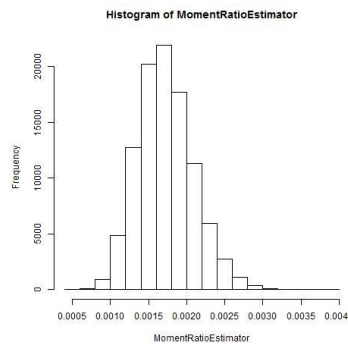
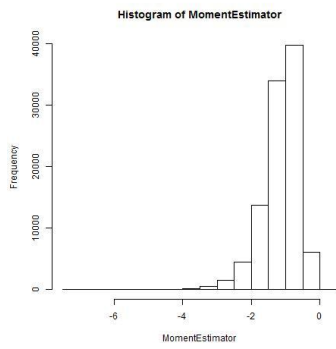
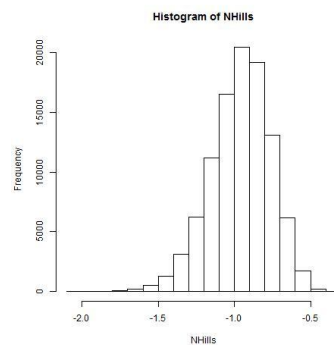
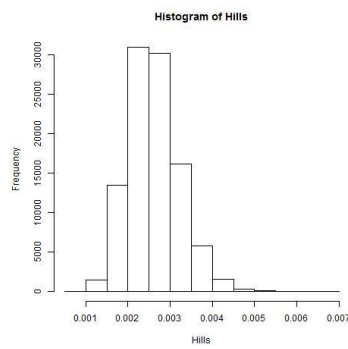
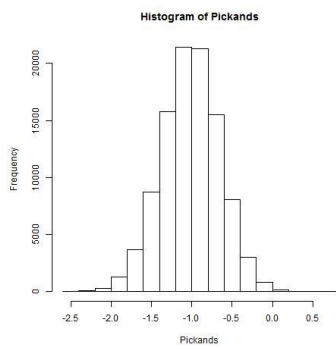
### Uniform



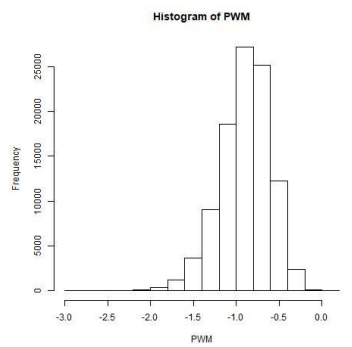
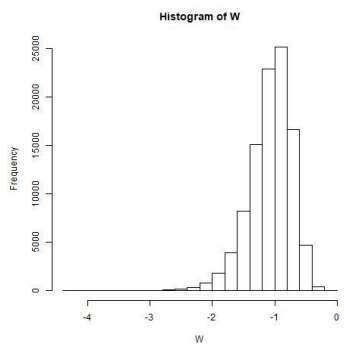
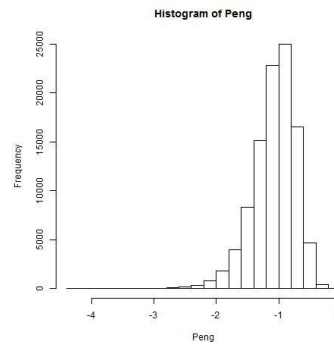
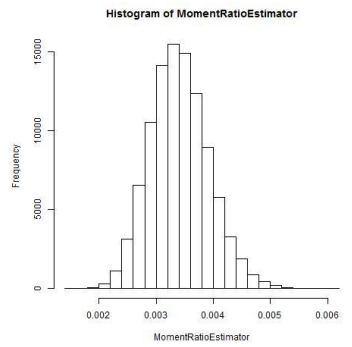
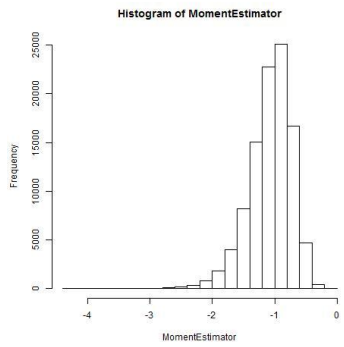
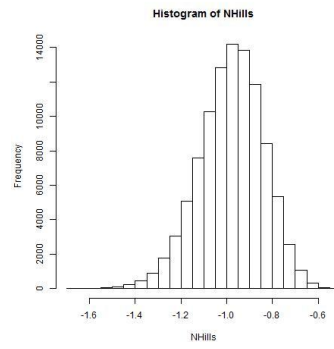
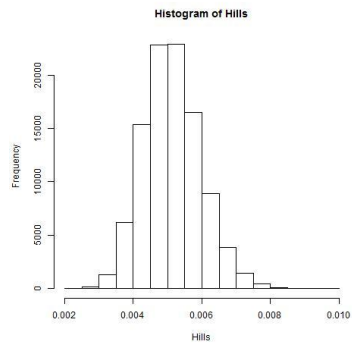
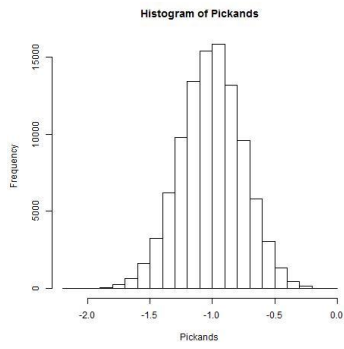




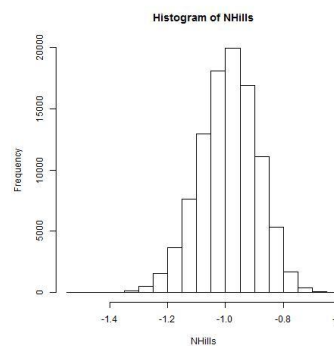
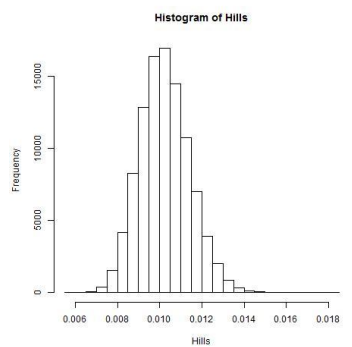
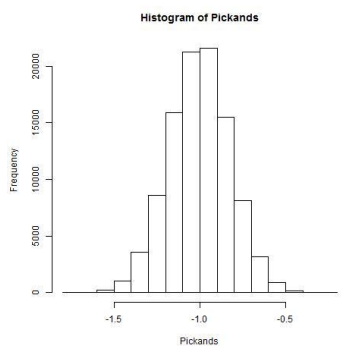
k=25

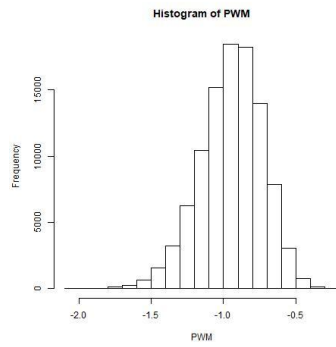
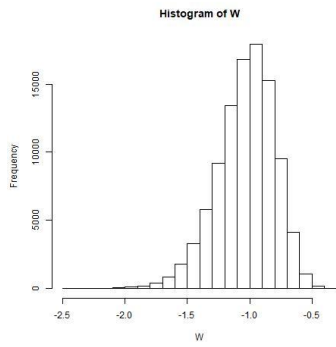
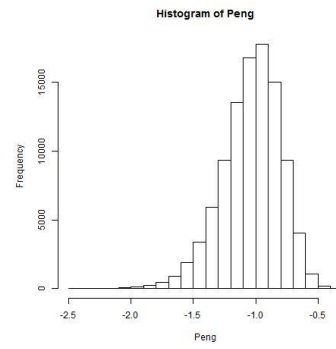
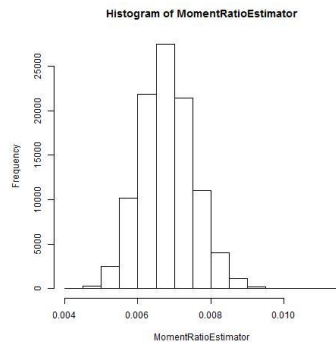
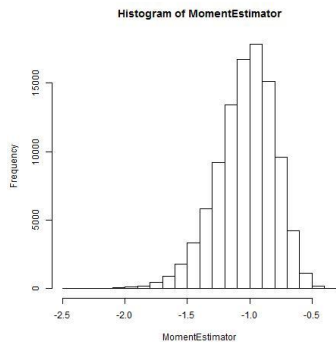


k=50

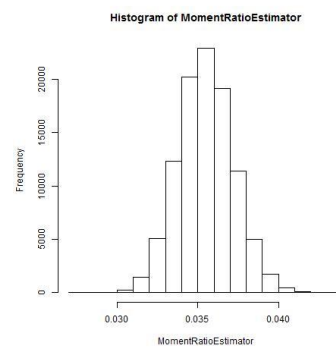
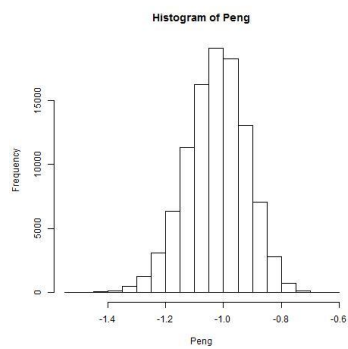
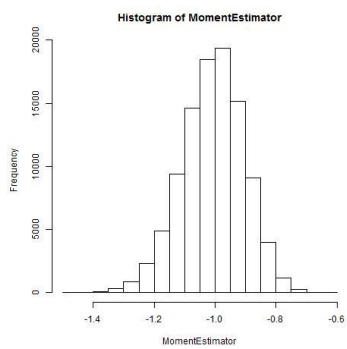
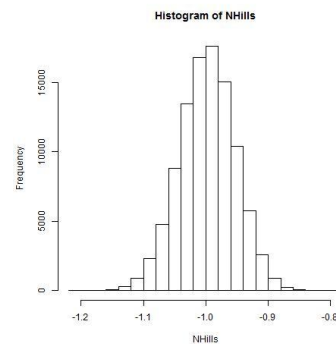
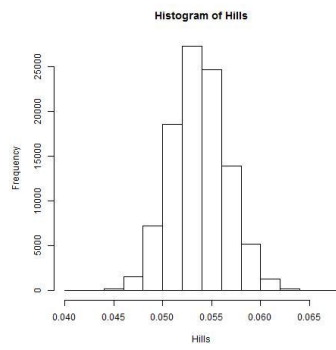
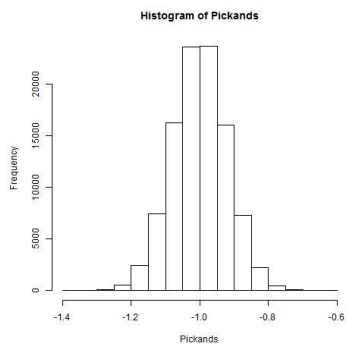


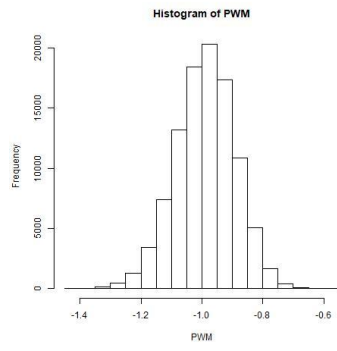
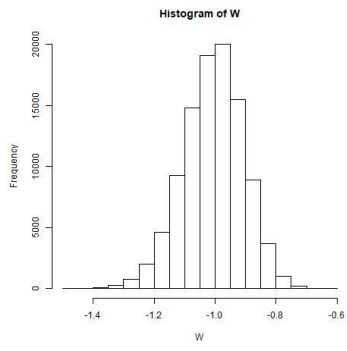
k=100





k=500





## 8 ΠΑΡΑΡΤΗΜΑ II

---

Σε κάθε περίπτωση η παραγωγή των δεδομένων έγινε με τον παρακάτω αλγόριθμο εναλλάσσοντας κάθε φορά σημεία, όπως χρειαζόταν για τις διαφορετικές περιπτώσεις στην θεωρητική τιμή του δείκτη ακραίων τιμών.

Στο σύνολο τα πρόσθετα πακέτα, που χρειάστηκαν για να χρησιμοποιήσουμε συγκεκριμένες συναρτήσεις και κατανομές, φορτώθηκαν στην μνήμη της R με τις παρακάτω εντολές:

```
require(fExtremes); require(extRemes); require(ismev);  
require(actuar); require(VGAM);
```

Η ακόλουθη σειρά εντολών προσδιόριζε το πλήθος  $n$  των δεδομένων που παράγονταν στην εσωτερική εντολή για κάθε περίπτωση κατανομής και ήταν 5000, όπως σημειώθηκε και στην μεθοδολογία. Αντίστοιχα, φαίνεται και ο ορισμός του πλήθους  $l$  των φορών, όπου επαναλαμβανόταν ο βρόχος για την παραγωγή μίας παρατήρησης από έναν εκτιμητή και στο σύνολό τους ήταν 100.000. Τέλος, το αντικείμενο *ksi* κρατήθηκε, ούτως ώστε να χρησιμοποιείται κάθε φορά για τον υπολογισμό του μέσου τετραγωνικού σφάλματος (M.S.E.)

```
n<-5000 ; l = 100000; ksi <- -1;# n = number of initial data;  
m = number of estimator observations; ksi = {-1, -0.25, 0,  
0.25, 0.5, 1}
```

Στη συνέχεια ακολουθούν τα σημεία του κώδικα, από τα οποία παράχθηκαν τα δείγματα για κάθε εκτιμητή. Στα πρώτα κομμάτια φαίνεται η προσαρμογή των GEV και GPD μέσα από τις μεθόδους Block Maxima και Peak Over Threshold αντίστοιχα, ενώ ταυτόχρονα γίνεται και η παραγωγή των  $l = 100000$  παρατηρήσεων από τον εκτιμητή μεγίστης πιθανοφάνειας. Το όρισμα  $k$  κάθε φορά στην μέθοδο Block Maxima ορίζει το πλήθος των Maxima, στα οποία θα καταλήξουμε, και αντίστοιχα το όρισμα  $m = n/k$ .

Τέλος, εφαρμόζουμε την εντολή `proc.time()`, η οποία εμφανίζει το χρόνο από το ρολόι του υπολογιστή. Η εντολή αυτή χρησιμοποιείται στην αρχή και στο τέλος του κώδικα και με την αφαίρεση της τελικής από την αρχική τιμή της, υπολογίζεται το Duration. Σημειώνεται, ότι από το output της συγκεκριμένης εντολής, κρατάμε μόνο την πρώτη τιμή, η οποία παρουσιάζει τα δευτερόλεπτα που απασχολήθηκε ο επεξεργαστής.

```
##### BLOCK MAXIMA #####
set.seed(2017)
data<-rbeta(n, 5, 1)
k<-500
m<-n/k
bm<-rep(0,k)
for(i in 1:k){bm[i]<-max(data[((i-1)*m+1):(i*m)])};
paste("gev.fit(bm)"); # Keeping score with what is the output
jpeg("../Desktop//gev.diag_beta5n1.jpeg");
gev.diag(gev.fit(bm)); dev.off();

ptm <- proc.time() # Saving Starting time
set.seed(2017)
MLE <- rep(0,1)
for(lambda in 1:1){
  data<-rbeta(n, 5, 1)
  k<-500
  m<-n/k
  bm<-rep(0,k)
  for(i in 1:k){bm[i]<-max(data[((i-1)*m+1):(i*m)])}
  MLE[lambda] <- gev.fit(bm, show = F)$mle[3]}
duration <- (proc.time() - ptm)[1];
paste("MLE for BM: ", "Mean =", round(mean(MLE), digits = 4),
      " S.E. =", round(sqrt(var(MLE)), 4), " M.S.E =",
      round(sum((MLE-ksi)^2)/1,4), " Duration =",
      round(duration,4), "seconds")
jpeg("../Desktop//MLE_BM_beta5n1.jpeg"); hist(MLE);
dev.off()
```

Όμοίως και στην περίπτωση της μεθόδου Peak Over Threshold, αναπαράγεται η προηγούμενη διαδικασία. Σε αυτήν την περίπτωση το μόνο επιπλέον όρισμα είναι αυτό

του βέλτιστου κατωφλιού  $u$ , το οποίο καθορίζεται κάθε φορά από τα αποτελέσματα του γραφήματος της εντολής `gpd.fitrange(data)` των δεδομένων.

```
##### POT #####
set.seed(2017)

data<-rbeta(n, 5, 1)

jpeg("../Desktop//gpd.diag_beta5n1.jpeg");gpd.fitrange(data,
0.5,0.96,nint=20); dev.off() # Range caution!!! Depends on
the distribution

u <- 0.95

paste("gpd.fit(data, u), where u=",u); # Keeping score with
what is the output

jpeg("../Desktop//gpd.diag_beta5n1.jpeg");
gpd.diag(gpd.fit(data, u)); dev.off()

jpeg("../Desktop//mrl.plot_beta5n1.jpeg"); mrl.plot(data);
dev.off()

ptm <- proc.time()
set.seed(2017)
MLE <- rep(0,1)
for(lambda in 1:1){
  data<-rbeta(n, 5, 1)
  MLE[lambda] <- gpd.fit(data, u, show = FALSE)$mle[2]}
duration <- (proc.time() - ptm)[1];
paste("MLE for POT: ", "Mean =", round(mean(MLE), digits = 4),
" S.E. =", round(sqrt(var(MLE)), 4), " M.S.E =",
round(sum((MLE-ksi)^2)/1,4), " Duration =",
round(duration,4), "seconds")

jpeg("../Desktop//MLE_POT_beta5n1.jpeg"); hist(MLE);
dev.off()
```

Τέλος, εμφανίζεται ο τρόπος παραγωγής των δεδομένων από τους εκτιμητές, που επιλέχθηκαν για την ανάλυση. Στις παρακάτω περιπτώσεις το όρισμα  $k = \{25, 50, 100, 500\}$  αντιστοιχεί στην παράμετρο  $k$  από τον εκάστοτε τύπο του εκτιμητή.

```
##### PICKAND'S ESTIMATOR #####
ptm <- proc.time()
set.seed(2017)
Pickands <- rep(0,1);k <- 25; #k setting for Pickands
(25,50,100,500)
for(lambda in 1:1){
  data<-rbeta(n, 5, 1)
  Pickands[lambda] <- log((sort(data)[n-k]-sort(data)[n-
2*k])/(sort(data)[n-2*k]-sort(data)[n-4*k]))/log(2)}
duration <- (proc.time() - ptm)[1];
paste("Pickands_25: ", "Mean =", round(mean(Pickands), digits
= 4), " S.E. =", round(sqrt(var(Pickands)), 4), " M.S.E =",
round(sum((Pickands-ksi)^2)/1,4), " Duration =",
round(duration,4), "seconds")
jpeg("../Desktop//Pickands_25_beta5n1.jpeg");
hist(Pickands); dev.off()
ptm <- proc.time()

##### HILL'S ESTIMATOR #####
ptm <- proc.time()
set.seed(2017)
Hills <- rep(0,1);k <- 25; #k setting for Hills
(25,50,100,500)
for(lambda in 1:1){
  data<-rbeta(n, 5, 1)
  h <- 0; for(j in 0:(k-1)){h <- h + (log(sort(data)[n-j])-
log(sort(data)[n-k]))/k;}
  Hills[lambda] <- h}
```



```

duration <- (proc.time() - ptm)[1];

paste("Hills_25: ", "Mean =", round(mean(Hills), digits = 4),
      " S.E. =", round(sqrt(var(Hills)), 4), " M.S.E =",
      round(sum((Hills-ksi)^2)/1,4), " Duration =",
      round(duration,4), "seconds")

jpeg("../Desktop//Hills_25_beta5n1.jpeg"); hist(Hills);
dev.off()

```

##### NEGATIVE HILL'S ESTIMATOR #####

```

ptm <- proc.time()

set.seed(2017)

NHills <- rep(0,1);k <- 25; #k setting for NHills
(25,50,100,500)

for(lambda in 1:1){
  data<-rbeta(n, 5, 1)

  h <- 0; for(j in 1:(k-1)){h <- h + (log(sort(data)[n]-
sort(data)[n-j])-log(sort(data)[n]-sort(data)[n-k]))/k;}

  NHills[lambda] <- h}

duration <- (proc.time() - ptm)[1];

paste("NHills_25: ", "Mean =", round(mean(NHills), digits =
4), " S.E. =", round(sqrt(var(NHills)), 4), " M.S.E =",
round(sum((NHills-ksi)^2)/1,4), " Duration =",
round(duration,4), "seconds")

jpeg("../Desktop//NHills_25_beta5n1.jpeg"); hist(NHills);
dev.off()

```

##### MOMENT ESTIMATOR #####

```

ptm <- proc.time()

set.seed(2017)

MomentEstimator<-rep(0,1);k <- 25; #k setting for
MomentEstimator (25,50,100,500)

for(lambda in 1:1){
  data<-rbeta(n, 5, 1)

```

```

h1<-0; for(j in 0:(k-1)){h1 <- h1 + (log(sort(data)[n-j])
-log(sort(data)[n-k]));}; h1<-h1/k;

h2<-0; for(j in 0:(k-1)){h2 <- h2 + (log(sort(data)[n-j])
-log(sort(data)[n-k]))^2;}; h2<-h2/k;

MomentEstimator[lambda]<-h1+1-(1/2)*(1-h1^2/h2)^(-1)}

duration <- (proc.time() - ptm)[1];

paste("MomentEstimator_25:           ", "Mean           =",
round(mean(MomentEstimator), digits = 4), " S.E. =",
round(sqrt(var(MomentEstimator)), 4), " M.S.E =",
round(sum((MomentEstimator-ksi)^2)/1,4), " Duration =",
round(duration,4), "seconds")

jpeg("../Desktop//MomentEstimator_25_beta5n1.jpeg");
hist(MomentEstimator); dev.off()

```

##### MOMENT RATIO ESTIMATOR #####

```

ptm <- proc.time()

set.seed(2017)

MomentRatioEstimator<-rep(0,1);k <- 25; #k setting for
MomentRatioEstimator (25,50,100,500)

for(lambda in 1:1){

  data<-rbeta(n, 5, 1)

  h1<-0; for(j in 0:(k-1)){h1 <- h1 + (log(sort(data)[n-j])
-log(sort(data)[n-k]));}; h1<-h1/k;

  h2<-0; for(j in 0:(k-1)){h2 <- h2 + (log(sort(data)[n-j])
-log(sort(data)[n-k]))^2;}; h2<-h2/k;

  MomentRatioEstimator[lambda]<-(1/2)*(h2/h1)}

duration <- (proc.time() - ptm)[1];

paste("MomentRatioEstimator_25:           ", "Mean           =",
round(mean(MomentRatioEstimator), digits = 4), " S.E. =",
round(sqrt(var(MomentRatioEstimator)), 4), " M.S.E =",
round(sum((MomentRatioEstimator-ksi)^2)/1,4), " Duration =",
round(duration,4), "seconds")

jpeg("../Desktop//MomentRatioEstimator_25_beta5n1.jpeg");
hist(MomentRatioEstimator); dev.off()

```

```
##### PENG ESTIMATOR #####
ptm <- proc.time()
set.seed(2017)
Peng<-rep(0,1);k <- 25; #k setting for Peng (25,50,100,500)
for(lambda in 1:1){
  data<-rbeta(n, 5, 1)
  h1<-0; for(j in 0:(k-1)){h1 <- h1 + (log(sort(data)[n-j]) -
log(sort(data)[n-k]));}; h1<-h1/k;
  h2<-0; for(j in 0:(k-1)){h2 <- h2 + (log(sort(data)[n-j]) -
log(sort(data)[n-k]))^2;}; h2<-h2/k;
  Peng[lambda]<-(h2/(2*h1))+1-1/2*(1-h1^2/h2)^(-1)}
duration <- (proc.time() - ptm)[1];
paste("Peng_25: ", "Mean =", round(mean(Peng), digits = 4), "
S.E.      =", round(sqrt(var(Peng)), 4), " M.S.E      =",
round(sum((Peng-ksi)^2)/1,4), " Duration      =",
round(duration,4), "seconds")
jpeg("../Desktop//Peng_25_beta5n1.jpeg"); hist(Peng);
dev.off()
```

```
##### W ESTIMATOR #####
ptm <- proc.time()
set.seed(2017)
W<-rep(0,1);k <- 25; #k setting for W (25,50,100,500)
for(lambda in 1:1){
  data<-rbeta(n, 5, 1)
  l1<-0; for(j in 0:(k-1)){l1 <- l1 + (sort(data)[n-j]-
sort(data)[n-k]);}; l1<-l1/k;
  l2<-0; for(j in 0:(k-1)){l2 <- l2 + (sort(data)[n-j]-
sort(data)[n-k])^2;}; l2<-l2/k;
  W[lambda]<-1-(1/2)*(1-l1^2/l2)^(-1)}
duration <- (proc.time() - ptm)[1];
```

```

paste("W_25: ", "Mean =", round(mean(W), digits = 4), " S.E.
=", round(sqrt(var(W)), 4), " M.S.E =", round(sum((W-
ksi)^2)/1,4), " Duration =", round(duration,4), "seconds")

jpeg("../Desktop//W_25_beta5n1.jpeg"); hist(W); dev.off()

##### PWM ESTIMATOR #####

ptm <- proc.time()
set.seed(2017)
PWM<-rep(0,1);k <- 25; #k setting for W (25,50,100,500)
for(lambda in 1:1){
  data<-rbeta(n, 5, 1)
  P<-0; for(j in 0:(k-1)){P <- P + (sort(data)[n-j]-
sort(data)[n-k]);}; P<-P/k;
  Q<-0; for(j in 0:(k-1)){Q <- Q + (sort(data)[n-j]-
sort(data)[n-k])*(j/k);}; Q<-Q/k;
  PWM[lambda] <- 1-(P/(2*Q)-1)^(-1)}
duration <- (proc.time() - ptm)[1];
paste("PWM_25: ", "Mean =", round(mean(PWM), digits = 4), "
S.E. =", round(sqrt(var(PWM)), 4), " M.S.E =",
round(sum((PWM-ksi)^2)/1,4), " Duration =",
round(duration,4), "seconds")
jpeg("../Desktop//PWM_25_beta5n1.jpeg"); hist(PWM);
dev.off()

```

