

ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΕΙΡΑΙΩΣ



ΤΜΗΜΑ ΣΤΑΤΙΣΤΙΚΗΣ & ΑΣΦΑΛΙΣΤΙΚΗΣ ΕΠΙΣΤΗΜΗΣ

ΜΕΤΑΠΤΥΧΙΑΚΟ ΠΡΟΓΡΑΜΜΑ ΣΠΟΥΔΩΝ
ΣΤΗΝ
ΕΦΑΡΜΟΣΜΕΝΗ ΣΤΑΤΙΣΤΙΚΗ

Η ΘΕΩΡΙΑ ΤΩΝ ΑΚΡΑΙΩΝ ΤΙΜΩΝ ΣΤΗΝ
ΑΝΑΛΟΓΙΣΤΙΚΗ ΕΠΙΣΤΗΜΗ ΚΑΙ Η ΕΦΑΡΜΟΓΗ ΤΗΣ
ΣΕ ΑΚΡΑΙΕΣ ΖΗΜΙΕΣ ΑΣΦΑΛΙΣΤΙΚΩΝ
ΧΑΡΤΟΦΥΛΑΚΙΩΝ ΛΟΓΩ ΦΥΣΙΚΩΝ ΚΑΤΑΣΤΡΟΦΩΝ
ΣΤΗΝ ΠΕΡΙΟΧΗ ΤΗΣ ΑΤΤΙΚΗΣ

Μαρινάκης Ιωσήφ

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
Μάιος 2007

ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΕΙΡΑΙΩΣ



ΤΜΗΜΑ ΣΤΑΤΙΣΤΙΚΗΣ & ΑΣΦΑΛΙΣΤΙΚΗΣ ΕΠΙΣΤΗΜΗΣ

ΜΕΤΑΠΤΥΧΙΑΚΟ ΠΡΟΓΡΑΜΜΑ ΣΠΟΥΔΩΝ
ΣΤΗΝ
ΕΦΑΡΜΟΣΜΕΝΗ ΣΤΑΤΙΣΤΙΚΗ

Η ΘΕΩΡΙΑ ΤΩΝ ΑΚΡΑΙΩΝ ΤΙΜΩΝ ΣΤΗΝ
ΑΝΑΛΟΓΙΣΤΙΚΗ ΕΠΙΣΤΗΜΗ ΚΑΙ Η ΕΦΑΡΜΟΓΗ ΤΗΣ
ΣΕ ΑΚΡΑΙΕΣ ΖΗΜΙΕΣ ΑΣΦΑΛΙΣΤΙΚΩΝ
ΧΑΡΤΟΦΥΛΑΚΙΩΝ ΛΟΓΩ ΦΥΣΙΚΩΝ ΚΑΤΑΣΤΡΟΦΩΝ
ΣΤΗΝ ΠΕΡΙΟΧΗ ΤΗΣ ΑΤΤΙΚΗΣ

Μαρινάκης Ιωσήφ

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
Μάιος 2007

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

Τα μέλη της Επιτροπής ήταν:

- Χατζηκωνσταντινίδης Ευστάθιος (Επιβλέπων)
- Πολίτης Κωνσταντίνος
- Πιτσέλης Γεώργιος

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**

**EXTREME VALUE THEORY IN ACTUARIAL
SCIENCE AND ITS APPLICATION IN EXTREME
NATURAL LOSSES THAT HAVE TAKEN PLACE
IN THE AREA OF ATTICA**

By
Joseph Marinakis

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

Piraeus, Greece
May 2007

Ευχαριστίες

Θα ήθελα να ευχαριστήσω ιδιαίτερος τον καθηγητή μου Χατσηκωνσταντινίδη Ε. για την άψογη και άκρως επικοδομητική συνεργασία που είχαμε όλα αυτά τα χρόνια, τόσο στο επίπεδο των προπτυχιακών όσο και στο επίπεδο των μεταπτυχιακών μου σπουδών στο Πανεπιστήμιο του Πειραιά.

Θα ήθελα ακόμη να ευχαριστήσω και τους καθηγητές μου, Πιτσέλη Γ. και Πολίτη Κ., για τις χρήσιμες συμβουλές τους στη συγγραφή της παρούσας εργασίας.

Περίληψη

Είναι κοινή η αίσθηση ότι στις μέρες μας ακραία φυσικά φαινόμενα εμφανίζονται πολύ πιο συχνά από ότι στο παρελθόν και μάλιστα με πολύ μεγαλύτερη ένταση. Σύμφωνα με επίσημα στοιχεία οι οικονομικές απώλειες από ακραία φαινόμενα έχουν ενιαπλασιαστεί τα τελευταία 40 χρόνια παγκοσμίως.

Ο τομέας της ασφάλισης έχει σίγουρα έναν πολύ σημαντικό ρόλο να παίζει ούτως ώστε οι καταστροφικές οικονομικές και κοινωνικές συνέπειες των ακραίων φαινομένων στην καθημερινή ζωή των ανθρώπων να μετριάζονται.

Με ποιο τρόπο όμως μπορούν οι αναλογιστές να μελετήσουν τέτοιου είδους φαινόμενα που παρουσιάζονται σχετικά σπάνια και με πολύ μεγάλη έντασή; Το βασικότερο εργαλείο στα χέρια των αναλογιστών είναι η Θεωρία των Ακραίων Τιμών. Η συγκεκριμένη θεωρία δεν είναι μια ακόμα θεωρία για την μελέτη κατανομών με βαριά ουρά, όπως είναι οι κατανομές απώλειας, αλλά αποτελεί ένα πραγματικά αξιόπιστο εργαλείο για τη μελέτη ακραίων καταστροφικών κινδύνων, κινδύνων που παρουσιάζουν μεγάλες περιόδους επαναφοράς και μεγάλα επίπεδα επιστροφής.

Επίσης, πέρα από την παρουσίαση της Θεωρίας των Ακραίων Τιμών και του τρόπου που μπορεί να συμβάλει στη μελέτη και τιμολόγηση καταστροφικών κινδύνων, η συγκεκριμένη Θεωρία εφαρμόζεται σε ένα σετ πραγματικών δεδομένων που αφορούν αποζημιώσεις για καταστροφές από πλημμύρες που πραγματοποιήθηκαν στο λεκανοπέδιο της Αττικής.

Abstract

It is a common belief that nowadays the frequency and the intensity of extreme natural disasters are much more severe than it was in the past. According to the official records, worldwide, the economic losses from extreme natural disasters have become nine times more rigorous the last 40 years.

The role of Insurance Industry in mitigating the catastrophic economical and social consequences of such catastrophic phenomena is very crucial for the survival and the development of modern societies.

But in what way can the actuaries study such extreme phenomena? Extreme Value Theory is the most useful “tool” for the analysis of catastrophic disasters. What must be underlined is that Extreme Value Theory is not one more available theory for the study of heavy-tailed distributions. This theory is a really reliable “tool” for the study of extreme catastrophic disasters which are characterised by great return levels and great return periods.

Furthermore, in this study Extreme Value Theory is applied in a group of real data that concern extreme financial losses that have taken place in the area of Atica.

Περιεχόμενα

Κατάλογος Πινάκων.....	xvii
Κατάλογος Σχημάτων.....	xix
Κεφάλαιο 1 (Εισαγωγικές Έννοιες)	1
1.1 Εισαγωγή.....	1
1.2 Τρόποι Αντιμετώπιση.....	2
1.3 Η Θεωρία της Κατανομής των Ακραίων Τιμών.....	2
1.4 Προβλήματα στη Μελέτη των Ακραίων Φαινομένων.....	3
1.5 Ιστορική Αναδρομή.....	4
1.6 Κύρια Πεδία Εφαρμογή.....	5
1.7 Επίλογος.....	5
Κεφάλαιο 2 (Παρουσίαση της Θεωρίας των Ακραίων Τιμών)	7
2.1 Εισαγωγή.....	7
2.2 Μοντελοποίηση μέσω Κατανομών Ακραίων Τιμών.....	7
Η Μέθοδος των Μεγίστων.....	7
Οριακές Κατανομές Μεγίστων και το Θεώρημα των Fisher & Tippett.....	8
Οι Συναρτήσεις των Κατανομών των Ακραίων Τιμών.....	10
1) Παραμετροποίηση-α.....	10
2) Παραμετροποίηση-ξ ή Γενικευμένη Κατανομή Ακραίων Τιμών.....	13
Maximum Domain of Attraction.....	15
Ροπές.....	16
Η Μέθοδος του r-μεγαλύτερου Διατεταγμένου Δείγματος.....	19
2.3 Μοντελοποίηση μέσω των Κατανομών της Γενικευμένης Pareto.....	20
Η Μέθοδος των Υπερβαλουσών Τιμών πάνω από Υψηλά Κατώφλια.....	20
Οι Συναρτήσεις Κατανομών της Γενικευμένης Pareto.....	21
1) Παραμετροποίηση-α.....	21
2) Παραμετροποίηση-ξ ή Κατανομή Γενικευμένης Pareto.....	22
ROT-Stability.....	23
Οριακές Κατανομές και το Θεώρημα των Balkema-de Haan-Pickands.....	23
Προσαρμογή Ουράς της GPD.....	25
Ροπές.....	26
Τρόπος Επιλογής για το Κατώφλι u.....	27

1) Threshold Choice Plot.....	28
2) Mean Residual Life Plot.....	29
3) L-Moments Plot.....	30
4) Dispersion Index Plot.....	31
Κατανομή του Αριθμού των Υπερβαλουσών Τιμών.....	31
Η Μέθοδος της Σημειακής Ανέλιξης.....	33
2.4 Σχέση ανάμεσα σε GEV & GPD.....	34
2.5 Εκτιμητές Παραμέτρων.....	37
1) Εκτιμητές Μέγιστης Πιθανοφάνειας.....	37
2) Probability Weighted Moments.....	39
3) Pickands Estimator.....	39
4) Hill Estimator.....	40
5) Regular Variation Approach.....	40
6) Dekker-Elmahl-de Haan Estiamtor.....	40
2.6 Επίπεδο Επιστροφής για Περίοδο Επαναφοράς N-ετών.....	41
2.7 Διαγνωστικά Εργαλεία.....	42
1) Ιστόγραμμα για Ομαδοποιημένα Δεδομένα.....	42
2) Ιστόγραμμα για Διακριτά Δεδομένα.....	43
3) Scatterplot.....	43
4) Gumbel Plot.....	43
5) Sample Mean Excess Plot.....	44
6) QQ Plot & PP Plot.....	45
7) Return Level Plot.....	46
2.8 Επίλογος.....	47
Κεφάλαιο 3 (Η σπουδαιότητα της Ασφάλισης των Ακραίων Κινδύνων).....	49
3.1 Εισαγωγή.....	49
3.2 Ακραίοι Κίνδυνοι – Ακραίες Απώλειες.....	49
3.3 Η στάση της Διεθνούς Ασφ/κής Αγοράς Απέναντι στα Ακραία Φαινόμενα.....	53
3.4 Αντασφάλιση.....	54
Η Επίδραση των Υψηλών Απαιτήσεων στη Λήψη Αποφάσεων.....	56
Εκτίμηση του Καθαρού Ασφαλίστρου.....	57
Κατάτμιση Σύμφωνα με τη Μέγιστη Πιθανή Ζημιά.....	59

Η Διαδικασία Κινδύνου.....	62
3.5 Επίλογος.....	64
Κεφάλαιο 4 (Η εφαρμογή της Θεωρίας των Ακραίων Τιμών σε πραγματικά δεδομένα)...	65
4.1 Εισαγωγή.....	65
4.2 Παρουσίαση Δεδομένων.....	65
4.3 Διερευνητική Ανάλυση Δεδομένων.....	68
4.4 Επιλογή Κατάλληλου Κατωφλιού.....	71
4.5 Προσαρμογή της GPD.....	73
4.6 Μοντέλο Τιμολόγησης και Ευαισθησία Αποτελεσμάτων.....	78
Κεφάλαιο 5 (Γενικά Συμπεράσματα).....	83
5.1 Εισαγωγή.....	83
5.2 Γενικά Συμπεράσματα.....	83
5.3 Περιορισμοί Έρευνας.....	84
Βιβλιογραφία.....	86
Παράρτημα 1.....	91
Παράρτημα 2.....	93
Παράρτημα 3.....	94
Παράρτημα 4.....	95
Παράρτημα 5.....	101
Παράρτημα 6.....	102
Παράρτημα 7.....	104

ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΕΙΡΑΙΑ

Κατάλογος Πινάκων

4-1	Περιγραφικά Στατιστικά στοιχεία των δεδομένων	66
4-2	Οι εκτιμητές των παραμέτρων με τα αντίστοιχα σφάλματα εκτίμησης	76
4-3	Προσαρμογή της GPD για διάφορα κατώφλια με τις αντίστοιχες τιμές της παραμέτρου ξ , των 0.999 και 0.9999 ποσοστιαίων σημείων και των τιμών της P	79
5-1	Υποθετικά Μοντέλα με συν 1 και πλην 1 παρατήρηση πάνω από το κατώφλι u σε σχέση με το Πραγματικό Μοντέλο	84

ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΕΙΡΑΙΑ

Κατάλογος Σχημάτων

2-1	Η συμπεριφοράς της ουράς σύμφωνα με την τιμή της παραμέτρου σχήματος ξ	14
3-1	Διαχρονική εξέλιξη οικονομικών απωλειών από καταστροφικούς κινδύνους και αντίστοιχων ποσοστών ασφάλισης κατά των κινδύνων αυτών	50
3-2	Διαχρονική Εξέλιξη των απωλειών από πλημμύρες σε παγκόσμιο επίπεδο	51
3-3	Διαχρονική εξέλιξη του αριθμού των πλημμυροπαθών παγκοσμίως	52
4-1	Scatterplot δεδομένων (EUDMG)	67
4-2	Q-Q Plot έναντι της Εκθετικής Κατανομής και Sample Mean Excess Plot	69
4-3	Threshold Selection Plot (tcplot)	72
4-4	Mean Residual Life Plot (mrlplot)	73
4-5	Προσαρμογή της GPD για $u = 1,2,3$	74
4-6	Probability Plot, Quantile Plot, Return Level Plot και Density Plot για $u = 3$	75
4-7	Το profile log-likelihood 95% Διάστημα Εμπιστοσύνης για την παράμετρο ξ	77
4-8	Η εξάρτηση που υπάρχει ανάμεσα στην τιμή του κατωφλιού που επιλέγεται και στις αντίστοιχες τιμές της παραμέτρου ξ και των ποσοστιαίων σημείων	80

ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΕΙΡΑΙΑ

Κεφάλαιο 1

Εισαγωγικές Έννοιες

1.1 Εισαγωγή

Τις τελευταίες δεκαετίες όλο και πιο συχνά γίνεται λόγος για ακραία φυσικά φαινόμενα τα οποία ως επί το πλείστον αποδίδονται στις σοβαρές κλιματολογικές αλλαγές που έχει επιφέρει η ανθρώπινη δραστηριότητα στον πλανήτη.

Τι όμως ακριβώς σημαίνει ο όρος ακραία φαινόμενα ή ακραίες παρατηρήσεις; Στη διεθνή βιβλιογραφία ακραία θεωρούνται τα σπάνια ή ασυνήθιστα φαινόμενα, όπως σεισμοί, πλημμύρες, τυφώνες κ.α.. που συμβαίνουν με εξαιρετικά μεγάλη ένταση και προκαλούν πολύπλευρες και πολυποίκιλες επιπτώσεις στη ζωή των ανθρώπων.

Ειδικοί επιστήμονες όπως οι Kharin V.V. & Zwiers W.F. (2003), υποστηρίζουν ότι το θερμότερο κλιματολογικό μοντέλο που έχει δημιουργηθεί παγκοσμίως είναι πιο ευμετάβλητο και κατά συνέπεια σημαντικά πιο επιρρεπές στην δημιουργία ακραίων φαινομένων.

Ακόμα όμως και ανεξάρτητοι φορείς που ασχολούνται από διαφορετική σκοπιά με το θέμα καταλήγουν στα ίδια συμπεράσματα με τους ειδικούς επιστήμονες. Οι Chavez-Demoulin V. & Roehrl A., (2004) επισημαίνουν ότι μελέτη του οργανισμού αντασφαλιστών του Μονάχου αναφέρει ότι για πρώτη φορά στην ιστορία του πλανήτη οι κλιματολογικές αλλαγές είναι τόσο έντονες και σημαντικές που μπορεί να είναι πλέον και μη αναστρέψιμες. Ο Οργανισμός Αντασφαλιστών της Ελβετίας (2003), υποστηρίζει ότι η συχνότητα και η ένταση κάποιων φυσικών καταστροφών αναμένεται να αυξηθεί πέρα από τις κανονικές κυκλικές διακυμάνσεις.

Η σημαντικότερη παράμετρος αυτού του φαινομένου είναι οι αξιοσημείωτες επιπτώσεις που έχουν αυτά τα ακραία φαινόμενα στην καθημερινή ζωή εκατομμυρίων ανθρώπων. Οι Kovacs P. & Kunreuther H. (2001), και ο Οργανισμός Αντασφαλιστών της Ελβετίας (2003), υπογραμμίζουν ότι οι ακραίες φυσικές καταστροφές αποτελούν σημαντική απειλή για τις σημερινές κοινωνίες, με πολλές απώλειες σε ανθρώπινες ζωές και υλικές ζημιές. Μάλιστα σύμφωνα με στοιχεία διεθνών οργανισμών, όπως ο Ερυθρός Σταυρός

(2000), τη δεκαετία του 1990 περίπου 2.800 φυσικές καταστροφές προκάλεσαν πάνω από 500.000 θανάτους και επηρέασαν άμεσα τη ζωή 1,3 εκατομμυρίων ανθρώπων.

Όμως ποια αντανακλαστικά ή ποιους τρόπους αντιμετώπισης έχουν αναπτύξει οι σύγχρονες κοινωνίες για την αντιμετώπιση αυτών των προβλημάτων;

1.2 Τρόποι Αντιμετώπισης

Ο πιο διαδεδομένος και αποτελεσματικός τρόπος αντιμετώπισης τέτοιου είδους κινδύνων είναι η ασφάλιση. Επειδή οι κίνδυνοι αυτοί είναι ακραίοι και καταστροφικοί οι ειδικοί επιστήμονες των ασφαλιστικών εταιριών έχουν αναπτύξει σύγχρονα μαθηματικά εργαλεία προκειμένου να κάνουν αξιόπιστη αποτίμηση τέτοιων καταστροφικών απωλειών, ούτως ώστε να εξασφαλίζουν και την κάλυψη των πελατών τους αλλά και την βιωσιμότητα των ασφαλιστικών εταιριών.

Ποια είναι όμως αυτά τα μαθηματικά εργαλεία που δίνουν τη δυνατότητα στους επιστήμονες να μπορούν να αποτιμούν τη συχνότητα και την ένταση τέτοιου είδους καταστροφικών απωλειών;

1.3 Η Θεωρία των Ακραίων Τιμών

Το πλέον σύγχρονο μαθηματικό εργαλείο που χρησιμοποιείται διεθνώς για τη μελέτη καταστροφικών ζημιών είναι η θεωρία της κατανομής των ακραίων τιμών.

Ποια είναι όμως ακριβώς η χρησιμότητα της συγκεκριμένης θεωρίας; Παρά τη μεγάλη έρευνα που έχει γίνει τα τελευταία χρόνια για τη μελέτη ακραίων φαινομένων, δεν έχουν ακόμα αναπτυχθεί αξιόπιστες μέθοδοι πρόβλεψης για το πότε και που θα εμφανιστούν τέτοια φαινόμενα. Για παράδειγμα, δεν μπορεί να προβλεφθεί με βεβαιότητα κάποιος σεισμός, ενώ κάποια άλλα φαινόμενα όπως πλημμύρες ή ανεμοθύελλες μπορούν να προβλεφθούν στην καλύτερη περίπτωση μερικές μόλις μέρες πριν πραγματοποιηθούν. Όμως αυτή η μη-προβλεψιμότητα ή τυχαιότητα αυτών των φαινομένων είναι απαραίτητη προϋπόθεση προκειμένου να μπορούν τέτοιου είδους κίνδυνοι να είναι ασφαλίσιμοι.

Σύμφωνα όμως με τον Οργανισμό Αντασφαλιστών της Ελβετίας (2002), αυτό που μπορεί να γίνει, και που αποτελεί τη μεγάλη συμβολή της θεωρίας της κατανομής των

ακραίων τιμών στην αντιμετώπιση ακραίων φαινομένων μέσα από τη διαδικασία της αντ/ασφάλισης, είναι η πρόβλεψη, κατά μέσο όρο, της συχνότητας και της έντασης της εμφάνισης τέτοιων φαινομένων μέσα σε κάποια χρονική περίοδο.

Είναι όμως εύκολη μια τέτοια μελέτη; Ποια προβλήματα και ποιές δυσκολίες αντιμετωπίζει η εφαρμογή της συγκεκριμένης θεωρίας στην πράξη;

1.4 Προβλήματα στη Μελέτη των Ακραίων Φαινομένων

Οι Kovacs P. & Kunreuther H. (2001), υπογραμμίζουν ότι αυτά τα ακραία φαινόμενα, παρόλο που συμβαίνουν με μεγαλύτερη συχνότητα από ότι στο παρελθόν, εξακολουθούν να είναι σπάνια, άρα δεν υπάρχει μεγάλη πληθώρα διαθέσιμων στοιχείων, ενώ ο Οργανισμός Αντασφαλιστών της Ελβετίας (2003), συμπληρώνει ότι είναι αδύνατο να φτάσει κανείς σε αξιόλογες εκτιμήσεις έχοντας δεδομένα λίγων ετών.

Κάποια άλλα προβλήματα που αναφέρουν οι Kovacs P. & Kunreuther H. (2001) είναι τα διαφορετικά χαρακτηριστικά που έχουν τέτοια φαινόμενα ανάλογα με τη γεωγραφική περιοχή που συμβαίνουν καθώς και ότι κάποια από αυτά διαχρονικά μπορεί να γίνονται πιο επικίνδυνα, όπως οι πλημμύρες εξαιτίας της ανόδου της στάθμης της θάλασσας. Γενικά τα φαινόμενα αυτά δεν ακολουθούν κάποιο συγκεκριμένο κανόνα ως προς τη συχνότητα και την έντασή τους.

Η πρόβλεψη ακραίων ζημιών από φυσικά φαινόμενα είναι πολύ δύσκολη. Τα ακραία αυτά φαινόμενα έχουν πολύ μεγάλες περιόδους επιστροφής, δηλαδή μέσο χρόνο εντός του οποίου το φαινόμενο επαναλαμβάνεται με ίδια ή και μεγαλύτερη ένταση. Οι περίοδοι επιστροφής μπορούν να φτάνουν ακόμα και πολλές δεκαετίες ή και αιώνες. Αυτό σημαίνει ότι κοιτάζοντας τα δείγματα που έχουμε από το παρελθόν σπάνια παίρνουμε μια αντιπροσωπευτική εικόνα του πραγματικού κινδύνου. Μάλιστα όταν υπάρχει μακρά περίοδος επανεμφάνισης του φαινομένου τότε ο πραγματικός κίνδυνος συνήθως υποτιμάται.

Όλα τα παραπάνω αναφέρονται στην χρησιμοποίηση της θεωρίας για την αντιμετώπιση του προβλήματος της ασφάλισης ακραίων ζημιών. Ποια ήταν όμως τα αρχικά

προβλήματα που αντιμετώπισαν οι επιστήμονες και τους οδήγησαν στην ανάπτυξη μιας τέτοιας θεωρίας;

1.5 Ιστορική Αναδρομή

Οι Kotz S. & Nadarajah S. (2000), αναφέρουν ότι μια πρώτη ιδέα της θεωρίας των ακραίων τιμών μπορούμε να πούμε ότι δόθηκε από τον Nicolas Bernoulli το 1709 όταν ασχολήθηκε με το πρόβλημα της μέσης μεγαλύτερης απόστασης που μπορεί να έχουν n τυχαίως κατανομημένα σημεία πάνω σε μια ευθεία σταθερού μήκους t από την αρχή της ευθείας αυτής. Παρόλα αυτά η θεωρία των ακραίων τιμών άρχισε πιο ουσιαστικά να αναπτύσσεται στις αρχές του 19^{ου} αιώνα από την ανάγκη των αστρονόμων να αξιολογήσουν ή να απορρίψουν ακραίες παρατηρήσεις.

Σύμφωνα όμως με τους Kotz S. & Nadarajah S. (2000), η συστηματική ανάπτυξη της θεωρίας ξεκίνησε το 1922 με το άρθρο του Bortkiewicz ο οποίος ασχολήθηκε με την κατανομή της κύμανσης τυχαίων δειγμάτων που προέρχονταν από την κανονική κατανομή. Η σπουδαιότητα της δουλειάς του Bortkiewicz έγκειται στο γεγονός ότι το θέμα της κατανομής των ακραίων τιμών παρουσιάζοταν καθαρά για πρώτη φορά. Η προσπάθεια αυτή βρήκε πολλούς συνεχιστές όπως τους von Mises & Dodd που ασχολήθηκαν με την εκτίμηση της αναμενόμενης τιμής και της διαμέσου της κατανομής του Bortkiewicz, τον Frechet ο οποίος ασχολήθηκε με την ασυμπτωτική κατανομή των ακραίων τιμών, όπως και οι Fisher & Tippett, καθώς και άλλοι ερευνητές με πολύ αξιόλογα αποτελέσματα.

Την ανάπτυξη του θεωρητικού υπόβαθρου, που πραγματοποιήθηκε τις δεκαετίες του 20 και 30, ακολούθησε η δεκαετία του 40 με πληθώρα πρακτικών εφαρμογών της θεωρίας πάνω σε προβλήματα της καθημερινής ζωής όπως σεισμικές αναλύσεις, αναλύσεις βροχοπτώσεων, αντοχή υλικών, εκπομπή ραδιακτινοβολίας, με κύριους ερευνητές τους Gumbel, Weibull, Rantz, Potter κ.α.

Έκτοτε η θεωρία των ακραίων τιμών έχει συγκεντρώσει το ενδιαφέρον πολλών ερευνητών που προέρχονται από διαφορετικούς επιστημονικούς κλάδους λόγω των σημαντικών εφαρμογών της σε όλο και περισσότερους τομείς της ανθρώπινης δραστηριότητας

1.6 Κύρια Πεδία Εφαρμογής

Η θεωρία των ακραίων τιμών στη σημερινή εποχή χρησιμοποιείται σε πολλές και διαφορετικές περιοχές της σύγχρονης ζωής και επιστήμης.

Οι Kotz S. & Nadarajah S. (2000), αναφέρουν εφαρμογές της θεωρίας σε διαφορετικά πεδία, όπως πρόβλεψη αποτελεσμάτων σε ιπποδρομίες, ουρές σε supermarkets, πρόβλεψη πλημμύρων, ακραίες χιονοπτώσεις, μελέτες αξιοπιστίας για χρόνους μέχρι εμφάνιση αποτυχίας, θανατηφόρες επιδράσεις από τη μείωση του στρώματος του όζοντος, ύψος θαλάσσιων κυμάτων, μετεωρολογία, σεισμολογία, χρόνοι επιβίωσης μικροοργανισμών κ.α.

Ο James G. (2003), παραθέτει κάποιες εφαρμογές στην εξόρυξη μεταλλευμάτων για την εκτίμηση της ύπαρξης μεγάλων αποθεμάτων πολύτιμων λίθων, στην Ωκεανογραφία για την περιγραφή του επιπέδου της θάλασσας και στη μηχανική όπως είναι η κατανομή βάρους διαφόρων φορτίων, η μελέτη της αντοχής διαφόρων υλικών, η πρόβλεψη ακραίων κυκλοφοριακών φορτίων σε γέφυρες.

Επίσης, σημαντικές είναι οι σύγχρονες εφαρμογές της θεωρίας στα πεδία των χρηματοοικονομικών και των χρηματιστηριακών αγορών από τους Embrechts et al. (1997), στον τομέα του λειτουργικού κινδύνου από τους Chavez-Demoulin V. & Embrechts P. (2004), και στον τομέα της έρευνας αγοράς και Marketing από τους Dahan E. & Mendelson H. (2001).

Δύο νέες και πολύ ενδιαφέρουσες εφαρμογές της θεωρίας αφορά στην ασφάλιση τρομοκρατικών κινδύνων που παρουσιάζεται από τον Οργανισμό Αντασφαλιστών της Ελβετίας (2003), και κινδύνων από την ανάπτυξη του τομέα της ναυτοτεχνολογίας που επίσης παρουσιάζεται από τον Οργανισμό Αντασφαλιστών της Ελβετίας (2004).

1.7 Επίλογος

Μιλώντας κανείς γενικά θα μπορούσε να πει ότι καμία επιστημονική μέθοδος δεν είναι τέλεια, άρα ούτε και η θεωρία που σχετίζεται με την μελέτη των ακραίων φαινομένων. Όπως κάθε μοντέλο, έτσι και αυτό είναι απλώς μια προσπάθεια απεικόνισης της πραγματικότητας με όλες τις αδυναμίες που αυτό συνεπάγεται.

Η ουσία όμως βρίσκεται στην άποψη του Smith L.R. (2003), που αναφέρει ότι η θεωρία των ακραίων τιμών, παρόλο που δεν είναι τέλεια, προς το παρόν εξασφαλίζει την

καλύτερη χρήση δεδομένων ακραίων τιμών και σίγουρα μπορεί να επιτύχει περισσότερα από οποιαδήποτε άλλη μέθοδο, όπως η εμπειρική προσαρμογή καμπυλών. Επίσης, πολύ πρακτική η άποψη των Chavez-Demoulin V., & Roehrl A. (2004), που αναφέρουν ότι μπορεί να μην είναι τέλεια η συγκεκριμένη μέθοδος αλλά είναι πολύ ακριβό για μας να την αγνοούμε.

Έτσι λοιπόν στο επόμενο κεφάλαιο γίνεται μια λεπτομερής παρουσίαση της θεωρίας των ακραίων τιμών, των βασικών μοντέλων και των διαγνωστικών εργαλείων που έχουν αναπτυχθεί στη διεθνή βιβλιογραφία.

Κεφάλαιο 2

Παρουσίαση της Θεωρίας των Ακραίων Τιμών

2.1 Εισαγωγή

Στη διεθνή βιβλιογραφία έχουν αναπτυχθεί δυο κύριες μέθοδοι μοντελοποίησης ακραίων τιμών. Η πρώτη μέθοδος αναφέρεται στη μοντελοποίηση μέσω της Γενικευμένης Οικογένειας Κατανομών Ακραίων Τιμών (GEV), ενώ η δεύτερη μέθοδος αναφέρεται στη μοντελοποίηση μέσω της Οικογένειας Κατανομών της Γενικευμένης Pareto (GPD). Και οι δυο μέθοδοι παρουσιάζονται διεξοδικά στις επόμενες παραγράφους.

Επίσης, στο κεφάλαιο αυτό γίνεται λόγος για τους εκτιμητές των μεθόδων αυτών, για τα διαγνωστικά εργαλεία που υπάρχουν για τη μελέτη αυτών των κατανομών και για τη σχέση ανάμεσα σε GEV και GPD.

2.2 Μοντελοποίηση μέσω Κατανομών Ακραίων Τιμών

Σύμφωνα με αυτή τη μοντελοποίηση, που αναλύεται παρακάτω, η πραγματική κατανομή ενός μεγίστου αντικαθίσταται από μια συνάρτηση κατανομής ακραίας τιμής.

Η Μέθοδος των Μεγίστων (Maxima)

Έστω ότι έχουμε τις ανεξάρτητες και όμοια κατανεμημένες (iid) τυχαίες μεταβλητές X_1, X_2, X_3, \dots με κοινή συνάρτηση κατανομής F , δηλαδή $F(x) = \Pr(X_i \leq x)$. Έστω τώρα ότι τα δεδομένα y_i είναι μέγιστες τιμές, δηλαδή $y_i = \max\{x_{i,1}, \dots, x_{i,n}\}$, $i = 1, \dots, n$.

Ουσιαστικά αυτή η μέθοδος παίρνει τις μέγιστες τιμές από Μπλοκ δεδομένων και αποτελεί έναν από τους διαθέσιμους τρόπους επιλογής ανώτατων ακραίων τιμών από μια ομάδα δεδομένων. Η μέθοδος αυτή είναι γνωστή ως “ετήσια μέγιστα” ή “Μπλοκς” ή

“μέθοδος Gumbel” και ουσιαστικά εκφράζει τις μέγιστες τιμές που μπορεί κανείς να πάρει από ένα σετ δεδομένων μέσα σε κάποιο καθορισμένο χρονικό διάστημα.

Για τις iid τυχαίες μεταβλητές X_1, X_2, X_3, \dots , που ορίστηκαν παραπάνω με κοινή συνάρτηση κατανομής F , και για $M_n = \max\{X_1, X_2, \dots, X_n\}$ να είναι η n -th μέγιστη τιμή του δείγματος, μπορεί κανείς να υπολογίσει την συνάρτηση κατανομής των μέγιστων τιμών ως εξής:

$$\begin{aligned} F_{M_n}(x) &= \Pr\{M_n \leq x\} = \Pr\{\max_{i \leq n} X_i \leq x\} = \Pr\{X_1 \leq x, \dots, X_n \leq x\} \\ &= \Pr\{X_1 \leq x\} \cdot \dots \cdot \Pr\{X_n \leq x\} = F^n(x) \quad (1) \end{aligned}$$

Αν η συνθήκη iid δεν ισχύει, τότε και πάλι μια συνάρτηση κατανομής της παραπάνω μορφής μπορεί να είναι μια ακριβής προσέγγιση της πραγματικής κατανομής του μεγίστου. Για ανεξάρτητες, αλλά μη ισόνομες τυχαίες μεταβλητές X_j με συνάρτηση κατανομής F_j η παραπάνω σχέση μπορεί να σταθεί αν η $F^n(x)$ αντικατασταθεί από την $\prod_{j \leq n} F_j$. Μια συνάρτηση κατανομής της μορφής $F^n(x)$ μπορεί να προσαρμοστεί στην $\prod_{j \leq n} F_j$ αρκεί οι αποκλίσεις μεταξύ των F_j να μπορούν να παραλειφθούν.

Παρομοίως, αν υπάρχει μικρή εξάρτηση ανάμεσα στα δεδομένα, μια συνάρτηση κατανομής της μορφής $F^n(x)$ μπορεί να εξακολουθεί να αποτελεί μια καλή προσέγγιση της πραγματικής κατανομής του μεγίστου.

Οριακές Κατανομές Μεγίστων και το Θεώρημα των Fisher & Tippett

Από την σχέση (1) παρατηρούμε ότι καθώς όμως το n αυξάνεται το $M_n \rightarrow \infty$ και

$$\lim_{n \rightarrow \infty} F_{M_n}(x) = \begin{cases} 1, & F(x) = 1 \\ 0, & F(x) = 0 \end{cases} \quad (2)$$

που είναι μια εκφυλισμένη κατανομή.

Συνεπώς, αν υπάρχει κάποια κατανομή για την οποία ενδιαφερόμαστε πρέπει να την βρούμε ως μια οριακή κατανομή κάποιας ακολουθίας μετασχηματισμένων “μειωμένων” τιμών, όπως για παράδειγμα η $(a_n X'_n + b_n)$, όπου a_n και b_n μπορεί να εξαρτώνται από το n αλλά όχι από το x .

Για την απόκτηση λοιπόν μιας μη-εκφυλισμένης οριακής κατανομής (δηλαδή μιας κατανομής που δεν έχει όλη της τη μάζα σε ένα συγκεκριμένο σημείο όπως η (2), είναι απαραίτητο να “ελαττωθεί” η πραγματική μέγιστη τιμή μέσω ενός γραμμικού μετασχηματισμού, όπως για παράδειγμα

$$M_n^*(x) = \left(\frac{M_n - b_n}{a_n} \right),$$

με συντελεστές $a_n > 0, b_n$ που εξαρτώνται από το μέγεθος του δείγματος, τέτοιοι ώστε:

$$\Pr \left(\frac{M_n - b_n}{a_n} \leq x \right) = F^n(a_n x + b_n) \rightarrow H(x), \text{ καθώς } n \rightarrow \infty. \quad (3)$$

Σύμφωνα με τους Kotz S. & Nadarajah S. (2000), και τους Reiss D.R. & Thomas M. (2001), αν ισχύουν όλες οι παραπάνω συνθήκες τότε η ασυμπτωτική κατανομή της M_n^* , καθώς $n \rightarrow \infty$ αποδεικνύεται ότι ανήκει σε μια από τις τρεις οικογένειες κατανομών που είναι γνωστές ως κατανομές ακραίων τιμών, και δεδομένου ότι η H υπάρχει και είναι μη-εκφυλισμένη, τότε η $H(x)$ θα είναι μια από τις συναρτήσεις κατανομής ακραίων τιμών.

Ο James G. (2003), επισημαίνει ότι το παραπάνω αποτελεί το θεώρημα των Fisher & Tippett (1928) οι οποίοι το διατύπωσαν χωρίς ακριβή μαθηματική απόδειξη, η οποία δόθηκε αργότερα από τον Gnedenko (1943) και που ουσιαστικά διατυπώνει την άποψη ότι οι κατανομές των ακραίων τιμών αποκτήθηκαν ως οριακές κατανομές μέγιστων (ή ελάχιστων) τιμών τυχαίων δειγμάτων αυξανόμενου μεγέθους.

Ποιες είναι όμως αυτές οι συναρτήσεις κατανομών των ακραίων τιμών που προκύπτουν ως οριακές κατανομές μεγίστων;

Οι Συναρτήσεις Κατανομών των Ακραίων Τιμών

1) Παραμετροποίηση - α

Στην περίπτωση αυτή τα τρία υπομοντέλα της τυπικής κατανομής των ακραίων τιμών για τις διαφορετικές τιμές του α που παρουσιάζονται στην διεθνή βιβλιογραφία είναι τα εξής:

Gumbel: $H_0(x) = \exp(-e^{-x})$, για κάθε x

$$\text{Fréchet, } \alpha > 0: H_{1,\alpha}(x) = \begin{cases} 0, & , x < 0 \\ \exp(-x^{-\alpha}), & x \geq 0 \end{cases}$$

$$\text{Weibull, } \alpha < 0: H_{2,\alpha}(x) = \begin{cases} \exp(-(-x)^{-\alpha}), & x \leq 0 \\ 1, & , x > 0 \end{cases}$$

Η συνάρτηση κατανομής του Fréchet $H_{1,\alpha}(x)$ είναι 0 όταν $x < 0$ καθώς και η συνάρτηση κατανομής του Weibull είναι ίση με 1 όταν $x > 0$. Παρατηρούμε λοιπόν ότι κάθε διαφορετική τιμή της παραμέτρου α καθορίζει μια τυπική κατανομή ακραίων τιμών.

Οι αντίστοιχες συναρτήσεις πυκνότητας πιθανότητας $h = H'$ είναι οι εξής:

Gumbel: $h_0(x) = H_0(x)e^{-x}$, για κάθε x

Fréchet, $\alpha > 0: h_{1,\alpha}(x) = \alpha H_{1,\alpha}(x) x^{-(1+\alpha)}$, $x \geq 0$

Weibull $\alpha < 0: h_{2,\alpha}(x) = |a| H_{2,\alpha}(x) (-x)^{-(1+\alpha)}$, $x \leq 0$

Σύμφωνα με τους Finkenstädt B. & Rootzén H. (2004), οι παραπάνω πυκνότητες των ακραίων τιμών έχουν μια μόνο κορυφή, δηλαδή η πυκνότητα είναι αύξουσα (non-decreasing) αριστερά από κάποιο σημείο u και φθίνουσα (non-increasing) δεξιά του σημείου αυτού. Για

το λόγο αυτό το u ονομάζεται επικρατούσα τιμή. Οι πυκνότητες των Fréchet και Gumbel παρουσιάζουν δεξιά λοξότητα.

Οι Reiss D.R. & Thomas M. (2001), αναφέρουν ότι οι συναρτήσεις κατανομών Weibull προσφέρουν μια πλούσια οικογένεια μονοκόρυφων κατανομών. Οι πυκνότητες αυτής της οικογένειας κατανομών παρουσιάζουν αριστερή λοξότητα αν $\alpha > -3,6$. Αν το α είναι κοντά στο 0 τότε η καμπύλη της πυκνότητας φαίνεται συμμετρική ενώ παρουσιάζει δεξιά λοξότητα (όπως οι πυκνότητες των Fréchet και Gumbel), αν $\alpha < -3,6$.

Αν τώρα υποθέσουμε ότι έχουμε μια τυχαία μεταβλητή X με συνάρτηση κατανομής F , τότε η $\mu + \sigma X$ έχει συνάρτηση κατανομής $F_{\mu,\sigma}(x) = F((x - \mu)/\sigma)$, όπου μ και σ είναι παράμετροι θέσης και κλίμακας αντίστοιχα. Τα πλήρη στατιστικά μοντέλα απαιτούν και την προσθήκη των μεταβλητών αυτών και διαμορφώνονται ως εξής:

Gumbel:

$$H_{0,m,s}(x) = \exp[-e^{(x-\mu)/\sigma}], \text{ για κάθε } x$$

Fréchet, $\alpha > 0$:

$$H_{1,\alpha,m,s}(x) = \begin{cases} 0, & , x < m \\ \exp\left[-\left(\frac{x-m}{s}\right)^{-\alpha}\right], & , x \geq m \end{cases}$$

Weibull $\alpha < 0$:

$$H_{2,\alpha,m,s}(x) = \begin{cases} \exp\left[-\left(\frac{m-x}{s}\right)^{\alpha}\right], & , x \leq m \\ 0 & , x > m \end{cases}$$

$$\text{Δηλαδή, } H_{i,\alpha,m,s}(x) = H_{i,\alpha}\left(\frac{x-m}{s}\right), i = 1,2$$

Οι κατανομές των Fréchet, $H_{1,a,m,s}(x)$, και Weibull, $H_{2,a,m,s}(x)$, σχετίζονται με μια απλή αλλαγή πρόσημου.

Επίσης, αξίζει να επισημανθεί ότι οι κατανομές των Fréchet, $H_{1,a,m,s}(x)$, και Weibull, $H_{2,a,m,s}(x)$, μπορούν να μετασχηματισθούν στην κατανομή του Gumbel μέσω των μετασχηματισμών $Z = \log(X - \mu)$ και $Z = -\log(\mu - X)$ αντίστοιχα και για το λόγο αυτό η κατανομή του Gumbel, $H_{0,m,s}(x)$, είναι η πιο συχνά χρησιμοποιούμενη στη διεθνή βιβλιογραφία.

Η παράμετρος θέσης μ είναι το αριστερό άκρο της κατανομής Fréchet $H_{1,a,m,s}(x)$, και το δεξιό άκρο της κατανομής Weibull $H_{2,a,m,s}(x)$.

Οι αντίστοιχες συναρτήσεις πυκνότητας πιθανότητας $g = G'$ είναι οι εξής:

$$\text{Gumbel: } h_{0,m,s}(x) = \frac{1}{s} e^{-(x-m)/s} \exp(-e^{-(x-m)/s}), \text{ για κάθε } x$$

$$\text{Fréchet, } \alpha > 0: h_{1,a}(x) = \frac{1}{s} h_{1,a}\left(\frac{x-m}{s}\right), x \geq 0$$

$$\text{Weibull, } \alpha < 0: h_{2,a}(x) = \frac{1}{s} h_{2,a}\left(\frac{x-m}{s}\right) x \leq 0$$

Οι Kotz S. & Nadarajah S. (2000), αναφέρουν ότι ο όρος “κατανομές ακραίων τιμών” προσδίδεται στις παραπάνω κατανομές διότι μπορούν να θεωρηθούν ως οριακές κατανομές (καθώς το $n \rightarrow \infty$), της μέγιστης τιμής των n ανεξάρτητων τυχαίων μεταβλητών η κάθε μια από τις οποίες έχει την ίδια συνεχή κατανομή.

2) Παραμετροποίηση - ξ ή Γενικευμένη Κατανομή Ακραίων Τιμών

Οι παραπάνω τρεις οικογένειες κατανομών, όπως αναφέρουν και οι Reiss D.R. & Thomas M. (2001), μπορούν να ενωθούν και να σχηματίσουν μια ενιαία οικογένεια κατανομών γνωστή και ως Γενικευμένη Κατανομή Ακραίων Τιμών που αποδίδεται στον *von Mises* και έχει τύπο:

$$H_x(x) = \begin{cases} H_0(x) = \exp(-e^{-x}), \forall x \text{ και } x = 0 \\ \exp(-(1+xx)^{-1/x}), (1+xx) > 0, x \neq 0 \end{cases}$$

Εφαρμόζοντας ότι $(1 + \xi x)^{1/\xi} \rightarrow \exp(x)$ καθώς $\xi \rightarrow 0 \Rightarrow H_\xi(x) \rightarrow H_0(x)$ καθώς $\xi \rightarrow 0$.

όπου H_0 είναι η τυπική κατανομή του Gumbel

H_x είναι η κατανομή του Fréchet αν $\xi > 0$

H_x είναι η κατανομή του Weibull αν $\xi < 0$

Οι αντίστοιχες πυκνότητες είναι $h_0(x) = H_0(x)e^{-x}$, για κάθε x

$$h_x(x) = H_x(x)(1+xx)^{-(1+1/x)}, 1 + \xi x > 0, \xi \neq 0,$$

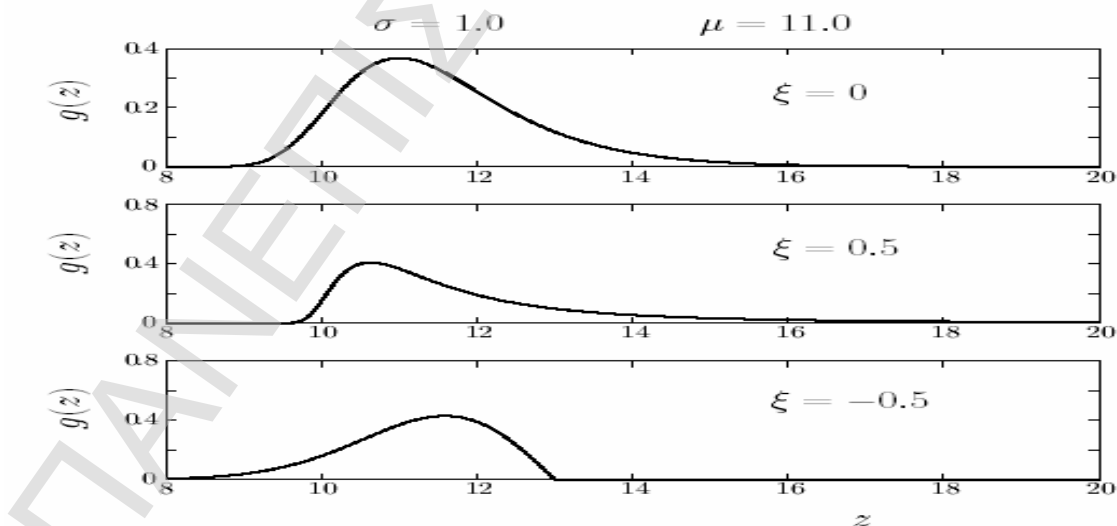
όπου $h_x(x) \rightarrow h_0(x)$, $\xi \rightarrow 0$

Στην περίπτωση τώρα όπου η $\mu + \sigma X$ έχει συνάρτηση κατανομής $F_{\mu,\sigma}(x) = F((x - \mu)/\sigma)$, όπου μ και σ είναι παράμετροι θέσης και κλίμακας αντίστοιχα, τότε οι τρεις τύποι κατανομών $H_{0,m,s}(x)$, $H_{1,a,m,s}(x)$ και $H_{2,a,m,s}(x)$ μπορούν και πάλι να παρουσιαστούν ως μέλη μιας γενικευμένης οικογένειας κατανομών με συνάρτηση κατανομής:

$$H(x; x, m, s) = \begin{cases} \exp\left\{-\left[1+x\frac{x-m}{s}\right]^{-1/x}\right\}, x \neq 0 \\ \exp\left\{-\exp\left[-\frac{x-m}{s}\right]\right\}, x = 0 \end{cases}$$

Ο παραπάνω τύπος ισχύει για τιμές του x που ικανοποιούν τη συνθήκη $\left[1+x\frac{x-m}{s}\right] > 0$, με $s > 0$ και m και ξ αυθαίρετα. Οι περιπτώσεις όπου $\xi = 0$, $\xi > 0$ και $\xi < 0$ μας δίνουν τις κατανομές των Gumbel, Fréchet και Weibull αντίστοιχα. Για το λόγο αυτό η παράμετρος σχήματος ξ αναφέρεται συχνά και ως δείκτης ακραίων τιμών, αφού η τιμή της καθορίζει και την αντίστοιχη κατανομή ακραίων τιμών που θα χρησιμοποιηθεί.

Επίσης, η τιμή της παραμέτρου ξ είναι ζωτικής σημασίας στον καθορισμό της συμπεριφοράς της ουράς της κατανομής. Όταν $\xi < 0$, τότε η κατανομή έχει ένα πεπερασμένο ανώτατο όριο που δίνεται από το $\mu - \sigma/\xi$. Όταν $\xi > 0$, η κατανομή έχει βαριά ουρά, δηλαδή προσεγγίζει αργά το άπειρο. Το παρακάτω σχήμα δείχνει τις τρεις διαφορετικές περιπτώσεις που μπορούν να δημιουργηθούν για τιμές $\mu = 11$, για την παράμετρο θέσης και $\sigma = 1$, για την παράμετρο κλίμακας.



Σχήμα 2-1: Η συμπεριφοράς της ουράς σύμφωνα με την τιμή της παραμέτρου σχήματος ξ .

Όπως φαίνεται και από το διάγραμμα η ουρά της κατανομής Gumbel, $\xi = 0$, δεν έχει πάνω όριο. Η βαριά ουρά της κατανομής Fréchet, $\xi > 0$, δεν έχει άνω όριο και παρουσιάζει μια αργή προσέγγιση στο άπειρο, ενώ η ουρά της κατανομής Weibull, $\xi < 0$, έχει άνω όριο.

Σύμφωνα με τους Kotz S. & Nadarajah S. (2000), η Γενικευμένη Κατανομή Ακραίων Τιμών έχει το πλεονέκτημα ότι μπορεί να εκτιμηθεί χωρίς να πρέπει εκ των προτέρων να έχει αποφασισθεί αν η παράμετρος σχήματος γ πρέπει να είναι θετική, αρνητική ή μηδέν, σε αντίθεση με τις κατανομές των ακραίων τιμών τύπου Gumbel, Fréchet ή Weibull.

Maximum Domain of Attraction

Όπως αναφέρει ο McNeil J.A. (1996), το θεώρημα των Fisher & Tippett είναι θεμελιώδες για την θεωρία των κατανομών των ακραίων τιμών και μπορεί να συγκριθεί σε σπουδαιότητα με το κεντρικό οριακό θεώρημα στη μελέτη των αθροισμάτων.

Ουσιαστικά το θεώρημα περιγράφει την οριακή συμπεριφορά κατάλληλα κανονικοποιημένων δειγματικών μεγίστων. Αν ισχύει η συνθήκη (3) λέμε ότι η F ανήκει στο Maximum Domain of Attraction της H και γράφουμε $F \in \text{MDA}(H)$.

Μάλιστα οι Fisher & Tippett (1928) έδειξαν ότι αν $F \in \text{MDA}(H) \Rightarrow H$ έχει τύπο H_x , για κάποιο ξ . Συνεπώς αν γνωρίζουμε ότι κατάλληλα κανονικοποιημένα μέγιστα συγκλίνουν σε κατανομή, τότε η οριακή κατανομή θα είναι μια κατανομή ακραίων τιμών για κάποιες τιμές των παραμέτρων ξ, μ, σ .

Ο Dalle Molle W.J. (2002), αναφέρει ότι ουσιαστικά το θεώρημα των maximum domains of attraction περιγράφει τις συνθήκες που εξασφαλίζουν ότι η ουρά μιας κατανομής θα ανήκει στο domain of attraction μιας κατανομής ακραίων τιμών αν και μόνο να ισχύει το θεώρημα των Fisher & Tippett.

Η κλάση των κατανομών F για τις οποίες η συνθήκη (3) ικανοποιείται είναι μεγάλη. Μια πληθώρα από ισοδύναμες συνθήκες μπορεί να εξαχθεί.

Ένα τέτοιο αποτέλεσμα είναι μια συνθήκη για την F για να ανήκει στο domain of attraction της κατανομής Fréchet που έχει βαριά ουρά ($H_x, \xi > 0$). Η συγκεκριμένη περίπτωση παρουσιάζει ενδιαφέρον διότι δεδομένα που αφορούν ασφαλιστικές ζημιές έχουν συνήθως βαριά ουρά.

Ο Gnedenko, έδειξε ότι για $\xi > 0$, $F \in \text{MDA}(H)$ αν και μόνο αν $1 - F(x) = x^{-1/\xi} L(x)$, για κάποια ελαφρώς μεταβαλλόμενη συνάρτηση $L(x)$.

Αυτό το αποτέλεσμα ουσιαστικά δείχνει ότι αν η ουρά της συνάρτησης κατανομής $F(x)$ εξασθενεί όπως μια δυναμοσυνάρτηση (power function), τότε η κατανομή ανήκει στο domain of attraction μιας κατανομής Fréchet. Η κλάση των κατανομών όπου η ουρά εξασθενεί όπως μια δυναμοσυνάρτηση είναι μεγάλη και περιλαμβάνει τις κατανομές Pareto, Burr, loggamma, Cauchy και t-distributions καθώς και διάφορα μικτά μοντέλα. Οι κατανομές αυτές είναι κατανομές με βαριά ουρά.

Κατανομές στο maximum domain of attraction μιας κατανομής Gumbel ($\text{MDA}(H_0)$) περιλαμβάνουν την Κανονική κατανομή, την εκθετική, την Γάμμα και την Λογαριθμοκανονική. Συγκεκριμένα η Λογαριθμοκανονική κατανομή έχει μια ελαφρώς βαριά ουρά και χρησιμοποιείται ευρέως ως κατανομή απωλειών. Παρόλα αυτά δεν έχει τόσο βαριά ουρά όπως οι κατανομές που ανήκουν στο $\text{MDA}(H_x)$ για $\xi > 0$.

Κατανομές στο domain of attraction of the Weibull ($H_x, \xi < 0$) είναι κατανομές με κοντή ουρά, όπως η ομοιόμορφη και η Βήτα. Αυτή η κλάση κατανομών δεν παρουσιάζει ιδιαίτερο ενδιαφέρον στις αντ/ασφαλιστικές εφαρμογές.

Ροπές

Η j -th $E(X^j)$ ροπή μιας Fréchet ή Weibull τυχαίας μεταβλητής X σαν μια συνάρτηση Γάμμα ως εξής:

$$\Gamma(\lambda) = \int_0^{\infty} x^{\lambda-1} e^{-x} dx, \quad \lambda > 0$$

Αν η X έχει συνάρτηση κατανομής του Gumbel και πυκνότητα $h = H'$, τότε:

$$m_{j,H} = E(X^j) = \int x^j dH(x) = \int x^j h(x) dx$$

Επίσης, $m_{j,H_{1,a}} = \Gamma(1 - j/a)$, αν $\alpha > j$

και

$$m_{j,H_{2,a}} = (-1)^j \Gamma(1 - j/a)$$

Όπως έχει ήδη επισημανθεί η πυκνότητα της κατανομής Fréchet έχει βαριά ουρά που είναι προσεγγιστικά ίση με $ax^{-(1+a)}$ και γι'αυτό η j th ροπή της κατανομής αυτής πάει στο άπειρο αν $\alpha \leq j$.

Για την ειδική περίπτωση που $j = 1$ έχουμε τις μέσες τιμές των κατανομών Fréchet και Weibull που είναι:

$$m_{1,H_{1,a}} = \Gamma(1 - 1/a), \text{ αν } \alpha > 1$$

$$m_{1,H_{2,a}} = -\Gamma(1 - 1/a)$$

Η πρώτη ροπή είναι άπειρη όταν $0 < \alpha \leq 1$

Για τις κεντρικές ροπές έχουμε ότι:

$$E((X - E(X))^j) = \int (x - m_1)^j dH(x) = \int (x - m_1)^j h(x) dx, \text{ όπου } m_1 = m_{1,H}$$

Για $j = 2$ προκύπτει η διακύμανση. Για τις κατανομές των Fréchet και Gumbel οι διακυμάνσεις προκύπτουν εύκολα από τη σχέση $\text{var} = m_2 - m_1^2$.

Έτσι, έχουμε ότι:

$$\text{var}_{H_{1,a}} = \Gamma(1 - 2/a) - \Gamma^2(1 - 1/a), \text{ αν } 1/\alpha < 1/2.$$

Για την περίπτωση της παραμετροποίησης ξ , που μας δίνει το ενιαίο μοντέλο, ο μέσος και η διακύμανση δίνονται από τις σχέσεις:

$$m_{H_x} = (\Gamma(1-x-1)/x, \text{ αν } \xi < 1 \text{ και } \text{var}_{G_x} = (\Gamma(1-2x) - \Gamma^2(1-x))/x^2,$$

αν $\xi < 1/2$ αντίστοιχα.

Η περίπτωση $\xi = 0$ ουσιαστικά περιλαμβάνεται στους παραπάνω τύπους παίρνοντας τα αντίστοιχα όρια με το γ να τείνει στο 0.

Έτσι έχουμε:

$$m_{H_0} = \lim_{x \rightarrow 0} m_{H_x} = \int_0^{\infty} (-\log x) e^{-x} dx = 1,$$

όπου $\lambda = 0,577216$, η σταθερά του Euler.

Αν κανονικοποιήσουμε τις κεντρικές ροπές μέσω της τυπικής απόκλισης $\sqrt{\text{var}}$ έχουμε:

$$E\left(\frac{X - EX}{\text{var}_G^{1/2}}\right)^j = \frac{E(X - EX)^j}{\text{var}_G^{j/2}},$$

όπου για $j=1,2$ οι κανονικοποιημένες κεντρικές ροπές είναι ίσες με το 0 και το 1 αντίστοιχα.

Για $j=3$ κανείς παίρνει τον συντελεστή ασυμμετρίας της κατανομής που είναι:

$$\text{skew}_G = \frac{E(X - EX)^3}{\text{var}_G^{3/2}}$$

Παρατηρούμε ότι οι κανονικοποιημένες κεντρικές ροπές είναι ανεξάρτητες από παραμέτρους θέσης και κλίμακας, όπως αναφέρουν και οι Reiss D.R. & Thomas M. (2001).

Η Μέθοδος του r-μεγαλύτερου Διατεταγμένου Δείγματος

Μια επέκταση της μεθόδου των Μεγίστων που δόθηκε στην αρχή του κεφαλαίου αποτελεί η Μέθοδος του r μεγαλύτερου Διατεταγμένου Δείγματος. Σύμφωνα με τη μέθοδο αυτή χρησιμοποιούμε τις $r, r > 1$, μεγαλύτερες παρατηρήσεις μιας συγκεκριμένης χρονικής περιόδου και όχι τη μεγαλύτερη παρατήρηση της περιόδου αυτής όπως συμβαίνει με τη μέθοδο των μεγίστων.

Η μέθοδος αυτή βασίζεται στο ότι η σχέση:

$$\Pr\left(\frac{M_n - b_n}{a_n} \leq x\right) = F^n(a_n x + b_n) \rightarrow H(x),$$

καθώς $n \rightarrow \infty$ (σχέση 3), μπορεί εύκολα να επεκταθεί στην από κοινού συνάρτηση των r μεγαλύτερων διατεταγμένων δειγμάτων καθώς το $n \rightarrow \infty$ για κάποιο δοσμένο $r > 1$.

Ένας προβληματισμός, όπως αναφέρουν οι Finkenstädt B. & Rootzén H. (2004), που συνδέεται με τα παραπάνω είναι ότι η μέθοδος αυτή είναι πιο επιρρεπής στο να αποκλίνει από τη συνθήκη iid, όταν για παράδειγμα υπάρχουν εποχικές κυμάνσεις στην κατανομή των δεδομένων ή αν υπάρχει εξάρτηση ανάμεσα στα δεδομένα.

Τα κύρια αποτελέσματα της μεθόδου αυτής έχουν ως εξής:

Αν $Y_{n,1} \geq Y_{n,2} \geq \dots \geq Y_{n,r}$ είναι τα r μεγαλύτερα διατεταγμένα δείγματα από iid δείγμα μεγέθους n και a_n και b_n είναι κανονικοποιητικές σταθερές της σχέσης 3, τότε :

$$\left(\frac{Y_{n,1} - b_n}{a_n}, \dots, \frac{Y_{n,r} - b_n}{a_n}\right)$$

συγκλίνει κατά κατανομή σε ένα οριακό τυχαίο διάνυσμα (X_1, \dots, X_r) του οποίου η πυκνότητα είναι:

$$h(x_1, \dots, x_r) = s^{-r} \exp\left\{-\left(1 + x \frac{x_x - m}{s}\right)^{-1/x} - \left(1 + \frac{1}{x}\right) \sum_{j=1}^r \log\left(1 + x \frac{x_j - m}{s}\right)\right\}$$

2.3 Μοντελοποίηση μέσω των Κατανομών της Γενικευμένης Pareto

Οι Corradin S. & Verbrigghe B. (2002), αναφέρουν ότι οι μοντέρνες μέθοδοι ανάλυσης των ακραίων τιμών βασίζονται σε τιμές που υπερβαίνουν υψηλά κατώφλια. Είναι δηλαδή μια μέθοδος που περιγράφει τη συμπεριφορά μεγάλων παρατηρήσεων που υπερβαίνουν κάποια υψηλά όρια. Ουσιαστικά η θεωρία αυτή εισάγει το ερώτημα ότι αν έχουμε μια ακραία παρατήρηση, πόσο ακραία μπορεί αυτή να είναι;

Η μέθοδος των Υπερβαλουσών Τιμών πάνω από Υψηλά Κατώφλια (Exceedances or Peaks Over Threshold)

Έστω x_i με συνάρτηση κατανομής F και έστω το κατώφλι u να είναι μικρότερο από το δεξιό τερματικό σημείο $x_0 = \sup\{x \in R : F(x) < 1\}$ της κατανομής F .

Επειδή το ζητούμενο είναι ένα αρκετά υψηλό κατώφλι, αυτό μπορεί να επιτευχθεί μόνο αν το κατώφλι u είναι πολύ κοντά στο δεξιό τερματικό σημείο x_0 της κατανομής. Τότε όμως $p = 1 - F(u)$ είναι πολύ μικρή και ο αριθμός k των υπερβαλλουσών παρατηρήσεων πάνω από το u μπορεί να θεωρηθεί ως μια τυχαία μεταβλητή από την κατανομή Poisson.

Όπως επισημαίνει και ο McNeil J. A. (1996), οι υπερβάσεις συμβαίνουν δοθέντος ότι οι παρατηρήσεις είναι μεγαλύτερες από το όριο u . Η δεσμευμένη συνάρτηση κατανομής F_u που λαμβάνεται ονομάζεται συνάρτηση κατανομής των υπερβαλουσών τιμών στο u . Αν X είναι η τυχαία μεταβλητή με συνάρτηση κατανομής F , τότε:

$$F_u(x) = \Pr(X - u \leq x | X > u) = \frac{\Pr(X + u \leq x, X > u)}{\Pr(X > u)} = \frac{F(x + u) - F(u)}{1 - F(u)}, \quad 0 \leq x < x_0 - u$$

Οι τυπικές συναρτήσεις κατανομών της γενικευμένης Pareto $G_{i,a}, G_x$ είναι οι παραμετρικές συναρτήσεις κατανομών για τις υπερβάλλουσες τιμές, δηλαδή τις τιμές που υπερβαίνουν κάποιο κατώφλι.

Οι Συναρτήσεις Κατανομών της Γενικευμένης Pareto

1) Παραμετροποίηση - α

Στην περίπτωση αυτή τα τρία υπομοντέλα της γενικευμένης Pareto είναι τα εξής:

$$\text{Εκθετική (GP0)} : G_0(x) = 1 - e^{-x}, x \geq 0$$

$$\text{Pareto (GP1), } \alpha > 0 : G_{1,\alpha}(x) = 1 - x^{-\alpha}, x \geq 1$$

$$\text{Beta (GP2), } \alpha < 0 : G_{2,\alpha}(x) = 1 - (-x)^{-\alpha}, -1 \leq x \leq 0$$

Η εκθετική συνάρτηση κατανομής G_0 είναι ίση με το 0 όταν $x < 0$, οι συναρτήσεις κατανομών της Pareto $G_{1,\alpha}$ είναι ίσες με το 0 όταν $x < 1$ και οι συναρτήσεις κατανομών της Beta $G_{2,\alpha}$ είναι ίσες με το 0 όταν $x < -1$ και ίσες με το 1 όταν $x > 0$.

Οι αντίστοιχες πυκνότητες είναι:

$$\text{Εκθετική (GP0): } g_0(x) = e^{-x}, x \geq 0$$

$$\text{Pareto (GP1), } \alpha > 0: g_{1,\alpha}(x) = \alpha x^{-(1+\alpha)}, x \geq 1$$

$$\text{Beta (GP2), } \alpha < 0: g_{2,\alpha}(x) = |a|(-x)^{-(1+\alpha)}, -1 \leq x < 0$$

Αν τώρα προστεθούν οι παράμετροι θέσεις και κλίμακας μ και $\sigma > 0$ μπορεί κανείς να πάρει τις πλήρεις οικογένειες κατανομών της γενικευμένης Pareto, όπου $G_{i,\alpha,m,s}(x) = G_{i,\alpha}((x-m)/s)$, $i = 1,2$, όπως αναφέρουν και οι Reiss D.R. & Thomas M. (2001).

2) Παραμετροποίηση- ξ ή Κατανομή Γενικευμένης Pareto

Θέτοντας $\xi = 1/a$ και επιλέγοντας κατάλληλες παραμέτρους θέσης και κλίμακας οι συναρτήσεις κατανομών της γενικευμένης Pareto είναι:

$$G_x(x) = \begin{cases} 1 - \exp(-x) & , x = 0 \\ 1 - (1 + \xi x)^{-1/\xi} & , x \neq 0 \end{cases},$$

όπου $\sigma > 0$ και $x > 0$ όταν $\xi > 0$ και $0 < x < 1/|\xi|$, αν $\xi < 0$.

Το αριστερό άκρο της G_x είναι το 0 για κάθε ξ , άρα η ξ - παραμετροποίηση έχει το πλεονέκτημα ότι η παράμετρος θέσης είναι πάντα το αριστερό άκρο της κατανομής, όπως επισημαίνουν οι Corradin S. & Verbrigghe B. (2001).

Οι Reiss D.R. & Thomas M. (2001), αναφέρουν ότι η Κατανομή της Γενικευμένης Pareto περιέχει τρεις άλλες κατανομές κάτω από τη συγκεκριμένη παραμετροποίηση που ουσιαστικά αντιστοιχούν και σε διαφορετικές συμπεριφορές της ουράς της κατανομής.

Σύμφωνα με τους McNeil J.A. & Saladin T. (1997), στην περίπτωση όπου $\xi > 0$ έχουμε την απλή κατανομή Pareto που αντιστοιχεί σε μια κατανομή με μακριά ουρά σύμφωνα με το ότι $\Pr(X > x) \sim x^{-a} l(x)$, για θετική σταθερά a . Η περίπτωση όπου $\xi < 0$ αναφέρεται σε κατανομές με πεπερασμένο άνω όριο, ενώ όταν έχουμε ότι $\xi = 0$ παίρνουμε την εκθετική κατανομή, δηλαδή αναφέρεται σε κατανομές που η ουρά τους μειώνεται εκθετικά όπως gamma, normal, Weibull, lognormal.

Οι αντίστοιχες πυκνότητες είναι:

$$g_0(x) = e^{-x}, x > 0, \text{ αν } \xi = 0$$

και

$$g_x(x) = (1 + \xi x)^{-(1 + 1/\xi)}, \text{ για } 0 < x, \text{ αν } \xi > 0 \text{ και } 0 < x < 1/|\xi|, \text{ αν } \xi < 0.$$

Στην περίπτωση όπου η $\mu + \sigma X$ έχει συνάρτηση κατανομής $F_{\mu,\sigma}(x) = F((x - m)/s)$, όπου μ και σ είναι παράμετροι θέσης και κλίμακας αντίστοιχα, τότε οι τρεις τύποι κατανομών $G_{0,m,s}(x)$, $G_{1,a,m,s}(x)$ και $G_{2,a,m,s}(x)$ μπορούν και πάλι να παρουσιαστούν ως μέλη μιας γενικευμένης οικογένειας κατανομών με συνάρτηση κατανομής:

$$G_{x,m,s}(x) = \begin{cases} 1 - \left(1 + x \frac{x-m}{s}\right)^{-1/x}, & x \neq 0 \\ 1 - \exp\left(-\frac{x-m}{s}\right), & x = 0 \end{cases}$$

όπου $\sigma > 0$ και $x \geq 0$ όταν $\xi \geq 0$ και $0 \leq x \leq -s/x$ όταν $\xi < 0$.

POT – Stability

Οι συναρτήσεις κατανομών της γενικευμένης Pareto είναι οι μόνες συνεχείς συναρτήσεις κατανομών για τις οποίες για μια συγκεκριμένη επιλογή σταθερών a_u και b_u έχουμε $F_u(b_u + a_u x) = F(x)$, όπου $F_u(x)$ είναι συνάρτηση κατανομής των υπερβαλουσών τιμών.

Οριακές Κατανομές και το Θεώρημα των Balkema – de Haan – Pickands

Όπως και στην περίπτωση της κλασσικής θεωρίας των ακραίων τιμών, έτσι και στην περίπτωση της Γενικευμένης Pareto, η παραμετρική μοντελοποίηση των συναρτήσεων κατανομών F_u των υπερβαλουσών τιμών βασίζεται σε κάποιο οριακό θεώρημα, το οποίο είναι γνωστό ως το θεώρημα των **Balkema – de Haan – Pickands**.

Για να γίνει αντιληπτή η σημασία του θεωρήματος θα πρέπει να έχει κανείς στο μυαλό του ότι αντικειμενικός σκοπός, στην περίπτωση της Γενικευμένης Pareto είναι η μοντελοποίηση των συναρτήσεων κατανομών F_u των υπερβαλουσών τιμών για υψηλά

κατώφλια και γι' αυτό πρέπει να θεωρείται ότι τα κατώφλια αυτά τείνουν στο δεξιό τερματικό σημείο της πραγματικής συνάρτησης κατανομής F .

Έστω λοιπόν ένα υψηλό κατώφλι u . Το ενδιαφέρον επικεντρώνεται για παρατηρήσεις που υπερβαίνουν αυτό το κατώφλι.

Έστω τώρα x_0 να είναι είτε το πεπερασμένο, είτε το άπειρο δεξιό τερματικό σημείο της κατανομής. Με άλλα λόγια $x_0 = \sup\{x \in \mathbb{R} : F(x) < 1\} \leq \infty$.

Όπως έχει ήδη αναφερθεί η συνάρτηση κατανομής των υπερβαλλουσών τιμών πάνω από το υψηλό κατώφλι u έχει ως εξής:

$$F_u(x) = \frac{F(x+u) - F(u)}{1 - F(u)}, \quad 0 \leq x < x_0 - u$$

Το θεώρημα των **Balkema – de Haan (1974), Pickands (1975)** δείχνει ότι κάτω από τις συνθήκες του Maximum Domain of Attraction, όπως αυτές δόθηκαν στην προηγούμενη ενότητα, η Γενικευμένη Pareto είναι η οριακή κατανομή της κατανομής των υπερβαλλουσών τιμών καθώς το κατώφλι u προσεγγίζει το δεξί τερματικό σημείο της κατανομής.

Αυτό σημαίνει ότι μπορούμε να βρούμε μια θετική μετρήσιμη συνάρτηση $S(u)$ τέτοια ώστε:

$$\lim_{u \rightarrow x_0} \sup_{0 \leq x < x_0 - u} |F_u(x) - G_{x,S(u)}(x)| = 0,$$

αν και μόνο αν $F \in MDA(G_x)$.

Το θεώρημα λοιπόν, όπως επισημαίνει και ο Dalle Molle W.J. (2002), ουσιαστικά αναφέρει ότι η Γενικευμένη Κατανομή Pareto (GPD) είναι η κατάλληλη προσέγγιση για την κατανομή των υπερβαλλουσών τιμών για ένα σημαντικά υψηλό κατώφλι u . Η GPD είναι η μόνη μη-εκφυλισμένη κατανομή που προσεγγίζει την κατανομή των υπερβαλλουσών τιμών πάνω από ένα σημαντικά υψηλό κατώφλι u καθώς το κατώφλι προσεγγίζει το ανώτατο τερματικό σημείο της κατανομής.

Παρόλο που οι *Balkema – de Haan – Pickands* διατύπωσαν το παραπάνω θεώρημα, σημαντική είναι και η συνεισφορά του Gnedenko (1943), ο οποίος απέδειξε ότι για επαρκώς υψηλά κατώφλια u , η κατανομή των υπερβαλουσών τιμών μπορεί να προσεγγιστεί από την $G_{\xi, \sigma}(x)$ για κάποιες τιμές των ξ και σ .

Ο McNeil J. A. (1997), υπογραμμίζει ότι η πρακτική εφαρμογή του παραπάνω αποτελέσματος είναι ότι επιτρέπει την προσπάθεια προσαρμογής της κατανομής της Γενικευμένης Pareto σε δεδομένα που υπερβαίνουν κάποιο υψηλό κατώφλι u . Με άλλα λόγια το παραπάνω θεώρημα και τα αποτελέσματα του δίνουν το θεωρητικό υπόβαθρο να αναμένουμε ότι αν επιλεγεί κάποιο αρκετά υψηλό κατώφλι, τα δεδομένα πάνω από το κατώφλι αυτό θα παρουσιάζουν τη συμπεριφορά μίας Γενικευμένη Pareto κατανομής.

Οι Reiss D.R. & Thomas M. (2001), αναφέρουν ότι η προσέγγιση αυτή αναπτύχθηκε από τους Davison (1984) και Davlson & Smith (1990) και το πρακτικό πρόβλημα που παρουσιάζει είναι η επιλογή του κατάλληλου κατωφλιού. Οι προβληματισμοί που συνδέονται με αυτό το θέμα αναλύονται παρακάτω.

Προσαρμογή Ουράς της GPD

Όπως ήδη αναφέρθηκε, η μεγάλη σημασία του παραπάνω θεωρήματος είναι ότι επιτρέπει την προσαρμογή μιας Γενικευμένης κατανομής Pareto στην υπό συνθήκη κατανομή των υπερβαλουσών τιμών που βρίσκονται πάνω από κάποιο υψηλό κατώφλι u .

Οι Reiss D.R. & Thomas M. (1996), έδειξαν ότι αν μπορεί να γίνει κάτι τέτοιο, τότε είναι δυνατό να προσαρμοστεί μια κατανομή Γενικευμένης Pareto και στη ουρά της πραγματικής κατανομής πάνω από το κατώφλι u .

Πιο συγκεκριμένα, για $x \geq u$, π.χ κάποια σημεία στην ουρά της κατανομής, ισχύει ότι

$$F(x) = \Pr(X \leq x) = (1 - \Pr\{X \leq u\})F_u(x) + \Pr\{X \leq u\}$$

Γνωρίζουμε ότι μπορούμε να προσεγγίσουμε την $F_u(x)$ από την $G_{x,u,s}(x)$ για κάποιο u υψηλό και ότι μπορούμε να προσεγγίσουμε την $\Pr(X \leq u)$ από τα δεδομένα μέσω της $F_n(u)$, δηλαδή μέσω της εμπειρικής συνάρτησης κατανομής στο u .

Συνεπώς για $x \geq u$ μπορούμε να προσεγγίσουμε την συνάρτηση κατανομής $F(x)$ μέσω της

$$\hat{F}(x) = (1 - F_n(u))G_{x,u,s}(x) + F_n(u) = G_{x,\tilde{m},\tilde{s}}(x)$$

Δηλαδή η $\hat{F}(x)$ είναι επίσης μια Γενικευμένη Pareto με την ίδια παράμετρο σχήματος ξ , αλλά με παράμετρο κλίμακας $\tilde{S} = s / (1 + xG_x^{-1}(F(u)))$ και παράμετρο θέσης $\tilde{m} = u - \tilde{S}G_x^{-1}(F(u))$.

Υπάρχει μια μοναδική σχέση ανάμεσα στα δυο ζευγάρια των παραμέτρων m, s και \tilde{m}, \tilde{S} που καθορίζεται από τις σχέσεις: $G_{x,\tilde{m},\tilde{S}}^{[u]} = G_{x,u,s}$ και $G_{x,\tilde{m},\tilde{S}}(u) = F(u)$.

Ροπές

Η j -th ροπή $m_{j,G}$ μιας Pareto και Beta συνάρτησης κατανομής μπορεί να υπολογιστεί με άμεσο τρόπο.

Δεδομένου ότι: $m_{j,G} = E(X^j) = \int x^j dG(x) = \int x^j g(x) dx$, όπου $g = G'$ είναι η πυκνότητα της G και X είναι μια τυχαία μεταβλητή κατανεμημένη σύμφωνα με την G , έχουμε ότι:

$$m_{j,G_{1,a}} = a / (a - j), \text{ αν } a > j$$

και

$$m_{j,G_{2,a}} = (-1)^j a / (a - j)$$

Η j -th ροπή μιας κατανομής Pareto, $G_{1,a}$, είναι άπειρη αν $a \leq j$.

Για την ειδική περίπτωση όπου $j=1$ κανείς παίρνει τις μέσες τιμές $m_G = m_{1,G}$ των συναρτήσεων κατανομών G των Pareto και Beta.

Άρα για $i = 1,2$ (δηλ. για τις κατανομές Pareto και Beta αντίστοιχα) 1^η ροπή είναι:

$$m_{j,G_{i,a}} = |a|/(a-1), \text{ αν } 1/a < 1, \text{ ενώ αν } 1/a < 1/2 \text{ οι αντίστοιχες διακυμάνσεις είναι:}$$
$$\text{var}_{G_{i,a}} = a/((a-1)^2(a-2)).$$

Οι μέσες τιμές και διακυμάνσεις των συναρτήσεων κατανομών G_x της Γενικευμένης Pareto είναι:

$$m_{G_x} = 1/(1-x), \text{ αν } x < 1$$

και

$$\text{var}_{G_x} = 1/((1-x)^2(1-2x)), \text{ αν } x < 1/2$$

Στην ειδική περίπτωση όπου $x = 0$ προκύπτει το γνωστό αποτέλεσμα της μέσης τιμής και διακύμανσης μιας τυπικής εκθετικής κατανομής που είναι μονάδα και για τα δυο.

Τρόπος Επιλογής για το Κατώφλι U

Πολύς λόγος έχει γίνει στη διεθνή βιβλιογραφία για την επιλογή του κατάλληλου κατωφλιού όταν εφαρμόζεται η μέθοδος POT.

Ο McNeil J.A. (1997), αναφέρει ότι η επιλογή του κατάλληλου κατωφλιού είναι ίσως το δυσκολότερο πρακτικό πρόβλημα που υπάρχει και που θέτει σε κίνδυνο τη μέθοδο που έχει εφαρμοστεί.

Σύμφωνα με τον James G. (2003), ένα υψηλό κατώφλι μειώνει τη μεροληψία καθώς αυτό ικανοποιεί τη σύγκλιση με τη θεωρία των ακραίων τιμών αλλά αυξάνει τη διακύμανση των εκτιμητών των παραμέτρων της GPD καθώς θα υπάρχουν λιγότερα δεδομένα για την εκτίμηση των παραμέτρων. Ένα χαμηλό κατώφλι από την άλλη μεριά λειτουργεί αντίθετα

καθώς έχει ως αποτέλεσμα μεγάλη μεροληψία αλλά μικρή διακύμανση των εκτιμητών αφού υπάρχουν περισσότερα δεδομένα που χρησιμοποιούνται για την εκτίμησή τους.

Ο Dalle Molle W.J. (2002), προσδιορίζει τη σπουδαιότητα της επιλογής του σωστού κατωφλιού στο γεγονός ότι η εκτίμηση των παραμέτρων που μας ενδιαφέρουν σχετίζεται με την επιλογή του κατωφλιού, καθώς στο ότι η τιμή του κατωφλιού καθορίζει την περιοχή της ουράς της κατανομής. Μάλιστα αναφέρει ότι το κριτήριο για την επιλογή του κατωφλιού θα πρέπει να είναι αυτό που εξασφαλίζει ότι η συνάρτηση κατανομής των τιμών που υπερβαίνουν το κατώφλι συγκλίνει οριακά σε κάποια από τις κατανομές των ακραίων τιμών.

Οι μέθοδοι που έχουν αναπτυχθεί στη διεθνή βιβλιογραφία για τον προσδιορισμό του ιδανικού κατωφλιού αναφέρονται κυρίως στη διερευνητική ανάλυση διαφόρων διαγραμμάτων που προσφέρουν στον ερευνητή μια οπτική τεκμηρίωση για το ποια θα πρέπει να είναι η τιμή του κατάλληλου κατωφλιού. Τα πιο γνωστά διαγράμματα είναι τα εξής:

1) *Threshold Choice Plot*

Η μέθοδος αυτή ουσιαστικά εξετάζει την σταθερότητα του υποδείγματος εκτίμησης των παραμέτρων, διότι όπως αναφέρουν και οι Corradin S. & Verbrigghe B. (2001), η ιδανική κατάσταση είναι το υπόδειγμα εκτίμησης των παραμέτρων να είναι σταθερό.

Έστω λοιπόν τυχαία μεταβλητή $X \sim GP(m, s, x)$. Έστω u να είναι ένα άλλο κατώφλι όπου $u > m$. Η τυχαία μεταβλητή $X|X > u$ ακολουθεί επίσης μια GPD με παραμέτρους $s_1 = s + x(u - m)$ και $x_1 = x$.

Έστω $s_+ = s_1 - x_1 u$. Με τη νέα αυτή παραμετροποίηση το s_+ είναι ανεξάρτητο του u . Επομένως οι εκτιμητές για το s_+ και x_1 είναι σταθεροί για κάθε $u > m$, αν m είναι ένα κατάλληλο κατώφλι για να ικανοποιείται η ασυμπτωτική προσέγγιση.

Το Threshold Choice Plot ορίζεται από τα σημεία: $\{(u, s_+): m_1 \leq x_{\max}\}$ και $\{(u, x_1): m_1 \leq x_{\max}\}$, όπου x_{\max} είναι η μέγιστη παρατήρηση.

Με άλλα λόγια, αν ικανοποιούνται οι συνθήκες του Θεωρήματος των Pickands-Balkema-de Haan, τότε για κάποιο κατώφλι u που είναι αρκετά υψηλό η παράμετρος σχήματος είναι σχεδόν σταθερή. Μάλιστα από τον εκτιμητή της παραμέτρου κλίμακας $S_1 = S + X(u - m)$ αν θέσουμε $S_+ = S_1 - X_1 u$, τότε και ο εκτιμητής της S_+ πρέπει να είναι σχεδόν σταθερός για κάποιο υψηλό κατώφλι u .

Όπως επισημαίνει και ο Han Z. (2003), ουσιαστικά η μέθοδος αυτή προτείνει την προσαρμογή διαφόρων μοντέλων σε έναν αριθμό διαφορετικών κατωφλιών ούτως ώστε να αποκτηθούν οι εκτιμητές \hat{X} και \hat{S}_+ . Στη συνέχεια επιλέγεται το κατώφλι u με τη μικρότερη τιμή για το οποίο οι εκτιμητές \hat{X} και \hat{S}_+ είναι σχεδόν σταθεροί.

2) Mean Residual Life Plot ή Mean Excess Plot

Όπως αναφέρει ο James G. (2003), το διάγραμμα αυτό στηρίζεται σε μια συνθήκη της GPD που αναφέρει ότι για $\xi < 1$, η μέση τιμή των υπερβάσεων πάνω από το κατώφλι u είναι μια γραμμική συνάρτηση του ύψους του κατωφλιού.

Ο Ribatet M. (2006), επισημαίνει ότι το Mean Residual Life Plot βασίζεται στη θεωρητική μέση τιμή της GPD.

Έστω τυχαία μεταβλητή $X \sim GP(m, S, \chi)$.

Τότε:

$$E[X] = m + \frac{S}{1 - \chi}, \text{ για } \chi < 1, \text{ ενώ για } \chi > 1 \text{ ο θεωρητικός μέσος πάει στο άπειρο.}$$

Στην πράξη αν η τυχαία μεταβλητή X αντιπροσωπεύει υπερβάσεις πάνω από κάποιο κατώφλι m_0 και αν η προσέγγιση από την GPD είναι καλή, τότε η αναμενόμενη τιμή των υπερβάσεων πάνω από το κατώφλι m_0 δίνεται από τη σχέση:

$$E[X - m_0 | X > m_0] = \frac{S_{m_0}}{1 - \chi}.$$

Όμως για κάθε νέο κατώφλι m_1 με $m_1 > m_0$ οι υπερβάλλουσες τιμές πάνω από το κατώφλι αυτό ακολουθούν επίσης μια GPD με παραμέτρους $s_1 = s_0 + x_0(m_1 - m_0)$ και $x_1 = x_0$.

Επομένως:

$$E[X - m_1 | X > m_1] = \frac{s_{m_1}}{1-x} = \frac{s_{m_0} + xm_1}{1-x}$$

Η συνάρτηση $E[X - m_1 | X > m_1]$ μπορεί εύκολα να εκτιμηθεί από την αντίστοιχη δειγματική συνάρτηση των υπερβολών που είναι:

$$e_n(m_1) = \frac{\sum_{i=1}^n (X_i - m_1)}{\sum_{i=1}^n I_{\{X_i - m_1\}}},$$

επομένως το συγκεκριμένο διάγραμμα δίνεται από τα σημεία $\{(m, e_n(m), m < x_{\max})\}$

Αρα λοιπόν και στο διάγραμμα αυτό η γραμμικότητα είναι το ζητούμενο, όπως υπογραμμίζουν και οι McNeil J.A. & Saladi T. (1998).

3) L-Moments Plot

Σύμφωνα με τον Ribatet M. (2006), οι L-Moments είναι περιγραφικά στατιστικά για κατανομές πιθανότητας και δείγματα δεδομένων. Παρέχουν μετρήσεις της θέσης, της διασποράς, της ασυμμετρίας, της κυρτότητας καθώς και άλλων παραμέτρων της κατανομής.

Για την GPD ισχύει η σχέση $t_4 = t_3 \frac{1+5t_3}{5+t_3}$, όπου t_4 είναι η L-κυρτότητα και t_3 η L-ασυμμετρία. Το L-Moments Plot δίνεται από τα σημεία $\{(t_{3,u}, t_{4,u}), u < x_{\max}\}$. Έχει παρατηρηθεί ότι το διάγραμμα αυτό έχει φτωχή απόδοση σε πραγματικά δεδομένα.

4) Dispersion Index Plot

Ο Ribatet M. (2006), επισημαίνει ότι το διάγραμμα αυτό είναι ιδιαίτερα χρήσιμο αν τα δεδομένα αναφέρονται σε κάποια χρονοσειρά.

Η Θεωρία των Ακραίων Τιμών αναφέρει ότι οι υπερβάλλουσες τιμές πάνω από κάποιο κατώφλι μπορούν να προσεγγιστούν μέσω μιας GPD και ότι ο αριθμός των υπερβάσεων πάνω από το κατώφλι ακολουθεί μια διαδικασία Poisson.

Έστω η τυχαία μεταβλητή $X \sim P(I)$, τότε:

$$\Pr[X = k] = e^{-I} \frac{I^k}{k!}, \quad k \in \mathbb{N}.$$

Από τα παραπάνω υπολογίζεται το στατιστικό $DI = \frac{s^2}{I}$, όπου s^2 είναι η ένταση της διαδικασίας Poisson και λ ο μέσος αριθμός των πραγματοποιήσεων σε ένα Μπλοκ δεδομένων, συνήθως είναι ένα έτος. Τα σημεία του διαγράμματος είναι $\{\mu, DI\}$.

Κατανομή του Αριθμού των Υπερβαλλουσών Τιμών

Μελετώντας μέχρι στιγμής τη μέθοδο των υπερβαλλουσών τιμών πάνω από κάποιο κατώφλι u έχουμε μιλήσει αποκλειστικά για το θέμα της κατανομής του μεγέθους ακραίων παρατηρήσεων, δηλαδή παρατηρήσεων που υπερβαίνουν κάποιο υψηλό όριο u .

Τι γίνεται όμως με την κατανομή του αριθμού των παρατηρήσεων που υπερβαίνουν το κατώφλι u ; Οι υπερβάλλουσες τιμές y_i πάνω από κάποιο όριο u είναι εκείνα τα x_i , με $x_i > u$, τα οποία λαμβάνονται είτε με την κανονική σειρά εμφάνισης τους είτε με οποιαδήποτε άλλη σειρά. Οι τιμές $y_i - u$ είναι τα ποσά πάνω από το κατώφλι u .

Ο αριθμός των υπερβαλλουσών τιμών πάνω από το όριο u συμβολίζεται με k ή και περιστασιακά με K για να δοθεί έμφαση στην τυχειότητα του αριθμού.

Δοθέντος των μεταβλητών X_1, X_2, \dots, X_n , μπορούμε να γράψουμε:

$$K = \sum_{i=1}^n I(X_i > u),$$

όπου η $I(X_i > u)$ είναι μια δείκτρια συνάρτηση με $I(X_i > u) = 1$, αν $X_i > u$ και $I(X_i > u) = 0$ διαφορετικά.

Αν τα X_i είναι τυχαίες μεταβλητές με κοινή συνάρτηση κατανομής F , τότε:

$$\Pr(K = k) = \binom{n}{k} p^k (1-p)^{n-k}, \quad k = 0, 1, \dots, n$$

που είναι η συνάρτηση πιθανότητας της Διωνυμικής Κατανομής με παραμέτρους $n, p = 1 - F(u)$, δηλαδή $K \sim B(n, p)$ και ο μέσος αριθμός των ποσοτήτων που υπερβαίνουν το όριο u είναι $E(K) = np = n(1 - F(u))$ και η οποία καθορίζει μια φθίνουσα συνάρτηση της μέσης τιμής.

Όμως η Διωνυμική κατανομή μπορεί να προσεγγιστεί από μια κατανομή Poisson και συνεπώς ο αριθμός K των υπερβάσεων πάνω από το όριο u να είναι μια Poisson τυχαία μεταβλητή.

Γνωρίζουμε ότι X_1, X_2, \dots, X_n είναι iid τυχαίες μεταβλητές οι οποίες λαμβάνουν τις τιμές 1 και 0 με πιθανότητες p και $1-p$ αντίστοιχα.

Γνωρίζουμε επίσης ότι:

$$\Pr(X_1 + X_2 + \dots + X_n = k) = \binom{n}{k} p^k (1-p)^{n-k} = B_{n,p}(k), \quad k = 0, 1, \dots, n,$$

δηλαδή ο συνολικός αριθμός των επιτυχιών ακολουθεί Διωνυμική Κατανομή, όπου np είναι ο αναμενόμενος αριθμός των επιτυχιών και $np(1-p)$ η διακύμανση.

Οι κατανομές Poisson μπορούν να προσαρμοστούν σε Διωνυμικές κατανομές κάτω από συγκεκριμένες συνθήκες. Αν $np \rightarrow \lambda$ καθώς το $n \rightarrow \infty$, τότε $B_{n,p}(k) \rightarrow P_\lambda(k)$,

$n \rightarrow \infty$, όπου $P_\lambda(k) = \frac{\lambda^k}{k!} e^{-\lambda}$, $k = 0, 1, 2, 3, \dots$, είναι η συνάρτηση πιθανότητας μιας κατανομής

Poisson με παράμετρο λ και $E(K) = \text{Var}(K) = \lambda$.

Είναι γνωστό ότι η ανισότητα $|B_{n,p}(A) - P_{np}(A)| \leq p$ ισχύει για κάθε ομάδα μη αρνητικών ακεραίων. Το δεξιό μέλος της ανισότητας δεν εξαρτάται από το n .

Συνεπώς, μια κατανομή Poisson μπορεί να προσαρμοστεί σε μια Διωνυμική Κατανομή όποτε το p είναι σημαντικά μικρό.

Η Μέθοδος της Σημειακής Ανέλιξης (Point Process Method)

Μέχρι τώρα τα μεγέθη και ο αριθμός των υπερβαλλουσών παρατηρήσεων πάνω από κάποιο κατώφλι u , εξετάστηκαν ως ξεχωριστές διαδικασίες. Η Μέθοδος Point Process Approach δίνει τη δυνατότητα αυτές οι δύο διαδικασίες να συνδυαστούν σε μια διαδικασία βασισμένη σε ένα διάγραμμα δυο κατευθύνσεων, του χρόνου και του μεγέθους των υπερβαλλουσών τιμών.

Οι Finkenstädt B. & Rootzén H. (2004) αναφέρουν ότι η ασυμπτωτική θεωρία των υπερβαλλουσών τιμών δείχνει ότι κάτω από κατάλληλη κανονικοποίηση αυτή η διαδικασία συμπεριφέρεται όπως μια μη-ομογενής διαδικασία Poisson.

Υποθέτουμε λοιπόν ότι η x έχει δυο διαστάσεις (t, y) , όπου t είναι ο χρόνος και $y \geq u$ είναι το μέγεθος της διαδικασίας και πεδίο ορισμού $D = [0, T] \times [u, \infty)$

Τότε η ένταση $I(t, y) = \frac{1}{s} \left(1 + x \frac{y - m}{s} \right)^{-1/x-1}$, ορίζεται όταν $\{1 + x(y - m)/s\} > 0$, διαφορετικά είναι 0.

Αν τώρα A είναι ένα σύνολο της μορφής $[t_1, t_2] \times [y, \infty)$, δηλαδή υποσύνολο του D και $N(A)$ καθορίζει τον αριθμό των σημείων στο A , τότε $N(A)$ ακολουθεί κατανομή Poisson με μέσο

$$\Lambda(A) = (t_2 - t_1) \left(1 + x \frac{y - m}{s} \right)^{-1/x},$$

δεδομένου ότι $\{1 + x(y - m)/s\} > 0$ και $y \geq u$

Για την προσαρμογή του μοντέλου σημειώνεται ότι αν παρατηρηθεί μια μη-ομογενής διαδικασία έντασης $I(t, y)$ με πεδίο ορισμού D και εάν $(T_1, Y_1), \dots, (T_n, Y_n)$ είναι τα N παρατηρηθέντα σημεία της διαδικασίας, τότε η από κοινού πυκνότητα είναι:

$$\prod_{i=1}^N I(T_i, Y_i) \exp \left\{ - \int_D I(t, y) dt dy \right\}$$

και μπορεί να χρησιμοποιηθεί ως συνάρτηση πιθανοφάνειας και να μεγιστοποιηθεί ως προς τις άγνωστες παραμέτρους της διαδικασίας.

Μια επέκταση της προσέγγισης αυτής επιτρέπει για μη στάσιμες χρονοσειρές οι παράμετροι μ, σ, γ να εξαρτώνται από το t και τότε:

$$I(t, y) = \frac{1}{s_t} \left(1 + x_t \frac{y - m_t}{s_t} \right)^{-1/x_t - 1}$$

2.4 Σχέση ανάμεσα σε GEV & GPD

O Menkens O. (2004), αναφέρει ότι η GEV H_x περιγράφει την οριακή κατανομή κανονικοποιημένων μεγίστων, ενώ η GPD G_x εμφανίζεται ως η οριακή κατανομή κλιμακούμενων υπερβαλλουσών τιμών πάνω από κάποιο υψηλό κατώφλι. Σε κάποιο μοντέλο που ο αριθμός των παρατηρήσεων που υπερβαίνουν κάποιο υψηλό κατώφλι είναι ακριβώς μια κατανομή Poisson και η κατανομή των υπερβαλλουσών τιμών πάνω από αυτό το κατώφλι είναι ακριβώς μια GPD, τότε η μέγιστη από αυτές τις υπερβάλλουσες τιμές έχει ακριβώς μια GEV.

Το παραπάνω συμπέρασμα, ότι δηλαδή ο αριθμός των υπερβάσεων πάνω από το κατώφλι u είναι τυχαίως και περιγράφεται από μια κατανομή Poisson, δίνει μια πολύ ενδιαφέρουσα σχέση ανάμεσα στη Γενικευμένη Κατανομή Ακραίων τιμών (GEV) και στη Γενικευμένη Κατανομή Pareto (GPD).

Δεδομένου ότι η τιμή του κατωφλιού είναι υψηλή, τότε ο αριθμός των υπερβαλλουσών τιμών μπορεί να προσεγγιστεί από μια κατανομή Poisson, δεδομένου ότι μια Διωνυμική κατανομή μπορεί να προσεγγιστεί από μια κατανομή Poisson όταν το p είναι

σημαντικά μικρό (πράγμα που εξασφαλίζεται από μια πολύ υψηλή τιμή του κατωφλιού u) και όταν τα ενδεχόμενα είναι ανεξάρτητα.

Οι Brodtkorb et al., (2000), αναφέρουν ότι “το μέγιστο, από ένα πλήθος ανεξάρτητων τυχαίων μεταβλητών από την GPD οι οποίες έχουν κατανομηθεί σύμφωνα με μια κατανομή Poisson, έχει κατανομή GEV”.

Έστω k να είναι μια τυχαία μεταβλητή κατανομημένη σύμφωνα με μια κατανομή Poisson με μέσο λ και το μέγιστο από k ανεξάρτητες GPD μεταβλητές τέτοιες ώστε $M_k = \max\{X_1, X_2, \dots, X_k\}$, άρα

$$\Pr(M_k \leq x) = \sum_{n=0}^{\infty} \Pr(k = n) \Pr(X_1 \leq x, X_2 \leq x, \dots, X_n \leq x)$$

Το πρώτο κομμάτι στη δεξιά πλευρά της εξίσωσης είναι η κατανομή Poisson και άρα έχουμε ότι

$$\Pr(M_k \leq x) = \sum_{n=0}^{\infty} e^{-1} \frac{1^n}{n!} \left[1 - \left(1 + x \frac{x}{\tilde{s}} \right)^{-1/x} \right]^n$$

το οποίο σύμφωνα με τους Brodtkorb et al., (2000) απλοποιείται στην παρακάτω έκφραση

$$\Pr(M_k \leq x) = \exp \left\{ - \left[1 + x \left(\frac{x - m}{s} \right) \right]^{-1/x} \right\}$$

Η παραπάνω έκφραση είναι μια GEV κατανομή. Οι Brodtkorb et al., (2000), αναφέρει ότι οι παράμετροι θέσης και κλίμακας αυτής της GEV κατανομής δίνονται από τις ακόλουθες σχέσεις

$$s = \tilde{s} l_u^x \text{ και } m = u + \frac{(s - \tilde{s})}{x}$$

όπου \tilde{S} είναι η παράμετρος κλίμακας από την GPD, I_u είναι ο μέσος αριθμός των υπερβάσεων πάνω από το όριο u μέσα στο προκαθορισμένο χρονικό διάστημα και προέρχονται από μια κατανομή Poisson και γ είναι η παράμετρος σχήματος και η οποία μπορεί να ταιριάζει είτε στη GEV είτε στη GPD ανάλογα με το μοντέλο που επιλέγεται.

Ο James G. (2003), αναφέρει ότι η εκτίμηση των παραμέτρων της GPD συνήθως εκτιμούν τις σ και ξ , γι' αυτό η μόνη ποσότητα που παραμένει άγνωστη είναι ο μέσος αριθμός των υπερβαλλουσών τιμών μ της κατανομής Poisson. Αυτή μπορεί να εκτιμηθεί ως ο μέσος αριθμός των υπερβάσεων πάνω από το κατώφλι u από τα παρατηρούμενα δεδομένα, με μια προσαρμογή για την επιθυμητή χρονική περίοδο. Αυτή η σχέση χρησιμοποιείται για να εκτιμήσει μια GEV από το εκτιμημένο GPD μοντέλο.

Οι Reiss R.D. & Thomas M. (2001) σχολιάζουν ότι υπάρχει μια απλή σχέση που συνδέει τις συναρτήσεις κατανομών της γενικευμένης Pareto με τις αντίστοιχες συναρτήσεις κατανομών των ακραίων τιμών και είναι η εξής: $W(x) = 1 + \log G(x)$, αν $\log G(x) > -1$, μεταξύ της συνάρτησης κατανομής της Γενικευμένης Pareto και της συνάρτησης κατανομής των Ακραίων Τιμών.

Βασιζόμενος στην παραπάνω σχέση ο Dalle Molle W.J. (2002), διαπιστώνει τις εξής σχέσεις ανάμεσα στα τρία υπομοντέλα των GEV και GPD:

1) Η εκθετική συνάρτηση κατανομής αντιστοιχεί στην κατανομή Gumbel αφού

$$G_0(x) = 1 + \log H_0(x) = 1 + \log(\exp(-e^{-x})) = 1 - e^{-x}, \text{ για } x \geq 0$$

2) Η συνάρτηση κατανομής της Pareto αντιστοιχεί στην συνάρτηση κατανομής Frechet. Η κατανομή Pareto έχει μεγάλη ιστορία στα αναλογιστικά μαθηματικά ως κατανομή μεγάλων απωλειών αφού έχει βαριά ουρά. Συνδέεται με την κατανομή Frechet ως εξής

$$G_{1,a}(x) = 1 + \log H_{1,a}(x) = 1 + \log(\exp(-x^{-a})) = 1 - x^{-a}, \text{ για } x \geq 1, a > 0$$

3) Η συνάρτηση κατανομής της Beta αντιστοιχεί στην συνάρτηση κατανομής της Weibull αφού

$$G_{2,a}(x) = 1 + \log H_{2,a}(x) = 1 + \log(\exp(-(-x)^{-a})) = 1 - (-x)^{-a}, \text{ για } -1 \leq x \leq 0, a < 0$$

2.5 ΕΚΤΙΜΗΤΕΣ ΠΑΡΑΜΕΤΡΩΝ

Παρακάτω παρατίθενται οι πιο σημαντικοί και ευρέως χρησιμοποιούμενοι στη διεθνή βιβλιογραφία. Για την εκτίμηση των παραμέτρων υποθέτουμε ότι X_i είναι iid τυχαίες μεταβλητές από μια κατανομή Ακραίων τιμών G_q , όπου $\theta = (\xi, \mu, \sigma)$

1) **Εκτιμητές Μέγιστης Πιθανοφάνειας** (Κατάλληλοι κυρίως για $\xi > -1/2$)

- Για GEV

Η συνάρτηση πιθανοφάνειας δίνεται από τη σχέση

$L(q; X) = \prod_{i=1}^n h_q(X_i)$, και λογαριθμίζοντας τη συνάρτηση πιθανοφάνειας έχουμε ότι:

$$l(q; X) = \log(L(q; X)) = -N \log s - \left(\frac{1}{x} + 1\right) \sum_i \log\left(1 + x \frac{X_i - m}{s}\right) - \sum_i \left(1 + x \frac{X_i - m}{s}\right)^{-1/x},$$

δοθέντος ότι $1 + x(X_i - m)/s > 0$, για κάθε i .

Ο MLE εκτιμητής για το θ είναι $\hat{q} = \arg \max_{q \in \Theta} l(q; X)$, όπου $\Theta = \{(\xi, \mu, \sigma): \xi, \mu \in \mathbb{R}, \sigma > 0\}$

Παραγωγίζοντας ως προς μ και σ και θέτοντας τις εξισώσεις ίσες με 0 και λύνοντας ως προς σ και μ έχουμε τους ζητούμενους εκτιμητές.

Οι καλές ιδιότητες του MLE υπάρχουν μόνο όταν $\xi > -1/2$, ενώ στην περίπτωση μικρών δειγμάτων οι ιδιότητες του MLE γίνονται αμφισβητήσιμες.

- Για GPD

Υποθέτοντας ότι (X_i) είναι iid από μια GPD κατανομή F_q , με $\theta = (\xi, \sigma)$, η αντίστοιχη λογαριθμική συνάρτηση πιθανοφάνειας δίνεται από τη σχέση:

$$l((\mathbf{x}, \mathbf{s}'); X) = -n \ln(\mathbf{s}) - \left(\frac{1}{\mathbf{x}} + 1 \right) \sum_{i=1}^n \ln \left(1 + \frac{\mathbf{x}}{\mathbf{s}} X_i \right)$$

Παραγωγίζοντας πάλι ως προς ξ και σ και θέτοντας τις εξισώσεις ίσες με 0 έχουμε ότι:

$$\frac{1}{\mathbf{x}^2} \sum_{i=1}^n \ln \left(1 + \frac{\mathbf{x}}{\mathbf{s}} X_i \right) - \left(\frac{1}{\mathbf{x}} + 1 \right) \sum_{i=1}^n \frac{\mathbf{x} X_i}{\mathbf{s} + \mathbf{x} X_i} = 0$$

$$n - \left(\frac{1}{\mathbf{x}} + 1 \right) \sum_{i=1}^n \frac{\mathbf{x} X_i}{\mathbf{s} + \mathbf{x} X_i} = 0$$

και ισχύουν οι ίδιοι προβληματισμοί που ήδη συζητήθηκαν και για την περίπτωση του MLE για GEV.

Όμως είναι πιο ρεαλιστικό να υποθέσουμε ότι η υπερβάλλουσες τιμές Y_1, \dots, Y_{N_u} ακολουθούν GPD, με την προϋπόθεση ότι N_u είναι ανεξάρτητες από τις Y_i . Για την επίλυση της αντίστοιχης συνάρτησης πιθανοφάνειας αλλάζουμε τις παραμέτρους (ξ, σ) με τις (ξ, τ) , όπου $\tau = -\xi/\sigma$.

Αυτό δίνει τον ακόλουθο εκτιμητή

$$\hat{x}_{N_u} = \hat{x}_{N_u}(t) = \frac{1}{N_u} \sum_{i=1}^{N_u} \ln(1 - t Y_i)$$

όπου το τ δίνεται από τη σχέση

$$\frac{1}{t} + \frac{1}{N_u} \left(\frac{1}{\hat{x}_{N_u}(t)} + 1 \right) \sum_{i=1}^{N_u} \frac{Y_i}{1-tY_i} = 0$$

2) Probability-Weighted Moments

Η τεχνική της μεθόδου των ροπών βασίζεται στον υπολογισμό των ροπών του μοντέλου μέσω των εμπειρικών ροπών.

Έστω $q = (x, s)$, και $w_r(q) = E[XG_q^x(X)]$, $r = 0, 1, 2, \dots$

Όπου $w_0 = \frac{s}{1-x}$ και $w_1 = \frac{s}{4-2x}$, οπότε $x = \frac{w_0 - 4w_1}{w_0 - 2w_1}$ και $s = \frac{2w_0w_1}{w_0 - 2w_1}$

Η εκτίμηση των παραμέτρων επιτυγχάνεται με την αντικατάσταση των ροπών του μοντέλου $w_r(q)$ με τις εμπειρικές ροπές που δίνονται από τη σχέση:

$$\hat{w}_r(q) = \frac{1}{n} \sum_{j=1}^n X_{(j)} G_q^x(X_{(j)}), \text{ για } r = 1, 2, \dots,$$

όπου $X_{(1)}, X_{(2)}, \dots, X_{(n)}$ είναι το διατεταγμένο δείγμα.

3) Pickands Estimator, για $\xi \in R$

$$\hat{x}_{k,n} = \frac{1}{\ln(2)} \ln \left(\frac{X_{k,n} - X_{2k,n}}{X_{2k,n} - X_{4k,n}} \right)$$

Όπως σχολιάζει και ο Menkens O. (2004), ο παραπάνω εκτιμητής περιλαμβάνει μια σειρά από upper order statistics που αυξάνει καθώς το n αυξάνει. Ουσιαστικά το πρόβλημα είναι να βρεθεί η βέλτιστη επιλογή του k . Αυτό που συνήθως γίνεται είναι να χρησιμοποιείται η βοήθεια του διαγράμματος Pickand $\{(k, \hat{x}_{k,n}) : k = 1, \dots, n\}$.

Η δισοθητική επιλογή του k είναι εκεί που το διάγραμμα Pickand είναι περίπου επίπεδο.

4) Hill Estimator, για $\xi > 0$

Ο εκτιμητής Hill που είναι ένας εκτιμητής αναλογίας ροπών για την παράμετρο του δείκτη της ουράς, είναι ο αντεστραμμένος μέσος όρος των λογαριθμικών αναλογιών του διατεταγμένου δείγματος των δεδομένων. Χρησιμοποιείται για να προσεγγίσει την παράμετρο σχήματος.

$$\hat{\chi}_{k,n} = \frac{1}{k} \sum_{j=1}^k \ln(X_{j,n}) - \ln(X_{k,n})$$

Ο ρυθμός σύγκλισης του Hill εκτιμητή μπορεί να είναι πολύ αργός. Συνήθως χρησιμοποιείται το διαγράμματος Hill $\{(k, \hat{\chi}_{k,n}) : k = 2, \dots, n\}$. Η δισοθητική επιλογή του k είναι εκεί που το διάγραμμα Pickand είναι περίπου επίπεδο. Ο εκτιμητής Hill είναι πολύ ευαίσθητος αναφορικά με αλληλεξαρτήσεις μέσα στο δείγμα.

5) Regular Variation Approach, για $\xi > 0$

$$\hat{\chi}_{k,n} = \frac{1}{k} \sum_{j=1}^k \ln(X_{j,n}) - \ln(X_{k+1,n})$$

Ο Menkens O. (2004), επισημαίνει ότι ο εκτιμητής βασίζεται στην ίδια βάση με τον εκτιμητή Pickand παρόλο που μοιάζει πολύ με τον Hill εκτιμητή.

6) Dekker-Elnmahl-de Haan Estimator, για $\xi \in R$

$$\hat{\chi}_{k,n} = 1 + H_{k,n}^{(1)} + \frac{1}{2} \left(\frac{(H_{k,n}^{(1)})^2}{H_{k,n}^{(2)}} - 1 \right)^{-1}, \text{ με } H_{k,n}^{(1)} = \frac{1}{k} \sum_{j=1}^k \ln(X_{j,n}) - \ln(X_{k+1,n})$$

και

$$H_{k,n}^{(2)} = \frac{1}{k} \sum_{j=1}^k (\ln(X_{j,n}) - \ln(X_{k+1,n}))^2$$

Ο Dekker-Elnmahl-de Haan Estimator συχνά καλείται και εκτιμητής ροπής του ξ διότι οι $H_{k,n}^{(i)}$ μπορεί να ερμηνευθούν και ως εμπειρικές ροπές.

2.6 Επίπεδο Επιστροφής για Περίοδο Επαναφοράς N-ετών

Το επίπεδο επιστροφής είναι το επίπεδο της ζημιάς που αναμένεται να ξεπεραστεί κατά μέσο όρο μια φορά στα N-έτη, ενώ η περίοδος επαναφοράς αναφέρεται στο χρόνο αναμονής (συνήθως μετριέται σε έτη), μέχρι ένα συγκεκριμένο επίπεδο επιστροφής ξεπεραστεί, όπως αναφέρει ο Han Z. (2003).

Όπως έχει ήδη αναφερθεί ο αριθμός των υπερβαλουσών τιμών πάνω από κάποιο κατώφλι u για ένα συγκεκριμένο χρονικό διάστημα ακολουθεί Διωνυμική Κατανομή. Αν τώρα μια GPD με παραμέτρους ξ και σ αποτελεί ένα καλό μοντέλο για τις υπερβάλλουσες τιμές, τότε για $x > 0$:

$$\Pr[X > x | X > u] = \left(1 + \xi \left(\frac{x - m}{\sigma} \right) \right)^{-1/\xi}$$

Αν $p_u = \Pr(X > u)$, τότε το επίπεδο επιστροφής x_m των m -παρατηρήσεων είναι η λύση της εξίσωσης $\Pr(X > x_m) = \frac{1}{m}$.

Όμως

$$\Pr(X > x_m) = \Pr[X > x_m | X > u] \Pr(X > u)$$

$$\Rightarrow \left(1 + \xi \left(\frac{x - m}{\sigma} \right) \right)^{-1/\xi} p_u = \frac{1}{m}$$

και τελικά $x_m = m + \frac{S}{X} [(mp_u)^x - 1]$, δεδομένου ότι $\xi \neq 0$ και m αρκετά μεγάλο ώστε να εξασφαλίζει ότι $x_m > u$.

Στην περίπτωση που $\xi=0$, τότε $x_m = m + S \ln(mp_u)$

Έτσι για μια περίοδο επιστροφής N -ετών, το επίπεδο επιστροφής θα είναι:

$$x_N = \begin{cases} m + \frac{S}{X} [(Nn_y p_u)^x - 1], & x \neq 0 \\ m + S \ln(Nn_y p_u), & x = 0 \end{cases}$$

2.7 Διαγνωστικά Εργαλεία

Είναι σημαντικό να υπάρχουν κάποιες διαγνωστικές διεργασίες προκειμένου να μπορεί να διαπιστωθεί πόσο καλά τα εκάστοτε μοντέλα προσαρμόζονται στα δεδομένα.

1) Ιστογράμματα για Ομαδοποιημένα Δεδομένα

Τα ιστογράμματα είναι ένα σημαντικό εργαλείο στα χέρια των ερευνητών κυρίως εξαιτίας της απλότητας τους στην απεικόνιση των δεδομένων ακόμα και όταν αυτά είναι συνεχή.

Το βασικότερο μειονέκτημα τους είναι ότι πρέπει κανείς να επιλέξει την θέση του πλέγματος των γραμμών.

Ας υποθέσουμε ότι n_j είναι η συχνότητα των δεδομένων στο διάστημα $(t_j, t_{j+1}]$. Παίρνοντας την παράγωγο της δειγματικής συνάρτησης κατανομής για ομαδοποιημένα

δεδομένα, που είναι $F_n(x) = \hat{F}_n(t_j) + \frac{(x-t_j)n_j}{t_{j+1}-t_j}$, για $t_j < x \leq t_{j+1}$, έχουμε την πυκνότητα

$$f_n(x) = \frac{n(j)}{n(t_{j+1}-t_j)}, \quad t_j < x \leq t_{j+1}$$

2) Ιστογράμματα για Διακριτά Δεδομένα

Στην περίπτωση αυτή ένα δειγματικό ιστόγραμμα δίνεται από την $p_n(j) = n(j)/n$, όπου $n(j)$ είναι ο αριθμός των δεδομένων x_1, x_2, \dots, x_n ίσος με τον ακέραιο j . Και αναλογικά με τη σχέση $\hat{F}_n(x) \approx F(x)$ έχουμε ότι $p_n(j) \approx P\{j\}$, όπου $P\{j\}$ είναι η υποκείμενη διακριτή κατανομή των x_i .

3) Διάγραμμα Διασποράς (Scatterplot)

Τα Scatterplots είναι πολύ χρήσιμα διαγραμματικά εργαλεία και λόγω της ευκολίας κατασκευής τους, αλλά και λόγω της ευκλείας κατανοήσεως τους.

Σημεία $(i, x(i))$ ή γενικότερα $(t(i), x(i))$ για $1 \leq i \leq n$ παρουσιάζονται σε διάγραμμα scatterplot.

Τέτοιου είδους διαγράμματα είναι ιδιαίτερα χρήσιμα ειδικά σε δεδομένα χρονοσειρών διότι δίνουν μια πολύ χρήσιμη απεικόνιση της κατανομής των δεδομένων στο χρόνο. Έτσι ο ερευνητής έχει τη δυνατότητα να δει με μια ματιά πότε για παράδειγμα πραγματοποιήθηκαν οι μεγαλύτερες απώλειες και πόσο πιο ακραίες είναι αυτές οι παρατηρήσεις σε σχέση με τις υπόλοιπες.

4) Gumbel Plot

Είναι η παλαιότερη μέθοδος εξέτασης της προσαρμογής ετησίων (ή και άλλων μεγίστων) στην κατανομή Gumbel. Αν τα ετήσια μέγιστα μιας σειράς N ετών είναι Y_1, \dots, Y_N τα οποία διατάσσονται ως $Y_{1:N} \leq \dots \leq Y_{N:N}$, τότε κατασκευάζεται ένα διάγραμμα με $Y_{i:N}$, για $1 \leq i \leq N$ και $x_{i:N}$, όπου $x_{i:N} = -\log(-\log p_{i:N})$, με $p_{i:N}$ να είναι το i -th σημείο του διαγράμματος το οποίο συνήθως λαμβάνεται ως το σημείο $(i-1/2)/N$.

Μια ευθεία γραμμή μαρτυρεί καλή προσαρμογή των δεδομένων στην κατανομή Gumbel. Καμπύλη με θετική ή αρνητική κλίση ίσως να προσδιορίζει μια κατανομή Frechet ή Weibull αντίστοιχα, όπως αναφέρει ο Smith L. R. (2003). Η μέθοδος είναι επίσης χρήσιμη για την ανίχνευση outliers.

Διαγράμματα αυτού του είδους χρησιμοποιούνταν ευρέως πριν εισαχθούν αυτόματες μέθοδοι όπως maximum likelihood, τόσο για εκτιμήσεις όσο και για έλεγχο των μοντέλων.

Στις μέρες μας δεν είναι πια τόσο σημαντική μέθοδος αλλά σαν διαγνωστική μέθοδος εξακολουθεί να είναι χρήσιμη.

5) *Sample Mean Excess Plot*

Η ιδέα για το συγκεκριμένο διάγραμμα εισήχθη από τους Davidson και Smith και είναι κάτι ανάλογο με το διάγραμμα Gumbel για δεδομένα που υπερβαίνουν κάποιο κατώφλι, με την έννοια ότι είναι ένα διαγνωστικό εργαλείο που χρησιμοποιείται πριν την προσαρμογή κάποιου μοντέλου και γι' αυτό μπορεί να χρησιμοποιηθεί ως οδηγός για την επιλογή του κατάλληλου κατωφλιού.

Όπως έχει ήδη αναφερθεί το mean excess plot περιγράφεται από τα σημεία $\{(u, e_N(u)), X_{N:N} < u < X_{1:N}\}$, όπου $X_{1:n}$ και $X_{n:n}$ είναι το πρώτο και το n – οστό διατεταγμένο στατιστικό αντίστοιχα, ενώ η

$$e_n(u) = \frac{\sum_{i=1}^n (X_i - u)^+}{\sum_{i=1}^n I_{\{X_i > u\}}}$$

είναι η δειγματική συνάρτηση των μέσων υπερβαλουσών τιμών πάνω από το κατώφλι u .

Οι McNeil J.A. & Saladi T. (1998), σχολιάζουν ότι αν το διάγραμμα φαίνεται να ακολουθεί μια ευθεία γραμμή με θετική κλίση πάνω από κάποιο κατώφλι u , τότε αυτό αποτελεί μια πρώτη ένδειξη για την τιμή του κατωφλιού που θα πρέπει να επιλεγεί.

Όμως το Sample Mean Excess Plot δίνει πολύ χρήσιμες πληροφορίες και για την κατανομή των παρατηρήσεων πάνω από το κατώφλι u αλλά και για την τιμή της παραμέτρου σχήματος ξ . Για παράδειγμα, αν το διάγραμμα φαίνεται να ακολουθεί μια ευθεία γραμμή με θετική κλίση πάνω από κάποιο κατώφλι u , τότε αυτό αποτελεί μια καλή ένδειξη ότι οι υπερβάλλουσες τιμές πάνω από αυτό το κατώφλι ακολουθούν μια GPD με θετική παράμετρο σχήματος ξ .

Άρα η κλίση του διαγράμματος δίνει μια πολύ γρήγορη εκτίμηση για το ξ αφού ένα αύξων διάγραμμα σημαίνει $\xi > 0$, ένα φθίνων διάγραμμα σημαίνει $\xi < 0$ και ένα διάγραμμα με σχετικά σταθερή κλίση σημαίνει ότι το γ είναι κοντά στο μηδέν.

Μια δυσκολία της συγκεκριμένης μεθόδου που επισημαίνει τόσο ο Smith (2003), όσο και ο McNeil A. (1996), είναι ότι το sample mean excess plot παρουσιάζει πολύ υψηλή

μεταβλητότητα ειδικά σε υψηλά κατώφλια, οπότε ίσως είναι δύσκολο για κάποιον να αποφασίσει αν η παρατηρούμενη απόκλιση από την γραμμικότητα οφείλεται πράγματι στην ακαταλληλότητα της GPD ή είναι απλώς δειγματική μεταβλητότητα.

6) Q – Q Plot και P – P Plot

Τα διαγράμματα Q – Q Plot και P – P Plot είναι πολύ χρήσιμα διαγνωστικά εργαλεία προκειμένου να διαπιστώσει κανείς αν το θεωρητικό μοντέλο \tilde{F} είναι μια καλή προσέγγιση της πραγματικής συνάρτησης κατανομής F που εκτιμάται από την \hat{F} .

α) P – P Plot

Έστω ένα διατεταγμένο δείγμα $x_{(1)} \leq x_{(2)} \leq \dots \leq x_{(n)}$ ανεξαρτήτων παρατηρήσεων από έναν πληθυσμό με εκτιμημένη συνάρτηση κατανομής \hat{F} . Τότε το P-P Plot θα δίνεται από τα

$$\text{σημεία } \left\{ \left(\hat{F}(x_{(i)}), \frac{i}{n+1} \right) : i = 1, \dots, n \right\}.$$

Ο Han Z. (2003), αναφέρει ότι αν η \hat{F} είναι μια καλή εκτίμηση της πληθυσμιακής κατανομής F , τότε τα σημεία του διαγράμματος πρέπει να βρίσκονται πολύ κοντά στη διαγώνιο. Ουσιαστική απόκλιση από τη διαγώνιο αποτελεί ένδειξη κακής προσαρμογής της \hat{F} στο μοντέλο.

β) Q – Q Plot

Έστω ένα διατεταγμένο δείγμα $x_{(1)} \leq x_{(2)} \leq \dots \leq x_{(n)}$ ανεξαρτήτων παρατηρήσεων από έναν πληθυσμό με εκτιμημένη συνάρτηση κατανομής \hat{F} . Τότε το Q-Q Plot θα δίνεται από τα

$$\text{σημεία } \left\{ \left(\hat{F}^{-1}\left(\frac{i}{n+1}\right), x_{(i)} \right) : i = 1, \dots, n \right\}.$$

Το ποσοστιαίο σημεία $x_{(i)}$ είναι το αντίστοιχο $\frac{i}{n+1}$ ποσοστιαίο σημείο της πληθυσμιακής συνάρτησης κατανομής και η εκτίμηση του είναι η $\hat{F}^{-1}\left(\frac{i}{n+1}\right)$.

Αν η \hat{F} είναι μια λογική εκτίμηση, τότε το quantile plot πρέπει να είναι σχεδόν γραμμικό. Αυτό θα ισχύει αν τα δεδομένα προέρχονται από ένα γραμμικό μετασχηματισμό της κατανομής.

Εκτροπές παρατηρήσεις μπορούν πολύ εύκολα να ανιχνευθούν στο Q-Q Plot. Επειδή όμως το αντικείμενο μελέτης της Θεωρίας των Ακραίων τιμών είναι το πάνω μέρος της ουράς της κατανομής, πρέπει να είναι κανείς πολύ προσεχτικός σχετικά με τα σημεία στην ουρά της κατανομής που παρουσιάζουν ουσιαστική απόκλιση από το μοντέλο. Μάλιστα, όπως σχολιάζει ο Han Z. (2003), από τη στιγμή που η παράμετρος ξ καθορίζει το πόσο βαριά είναι η ουρά της κατανομής από το συγκεκριμένο διάγραμμα μπορούν να εξαχθούν συμπεράσματα για την ουρά της κατανομής.

Γενικά θα μπορούσε να ειπωθεί ότι μια υπερεκτίμηση της παραμέτρου ξ θα οδηγήσει σε μια κοίλη απόκλιση από την ευθεία, ενώ μια υποεκτίμηση της παραμέτρου ξ θα οδηγήσει σε μια κυρτή απόκλιση.

7) Return Level Plot

Έστω (X_i) μια ακολουθία ισόνομων και ανεξάρτητων τυχαίων μεταβλητών με συνεχή συνάρτηση κατανομής F και u ένα δοσμένο κατώφλι.

Έστω η ακολουθία $(I_{(X_i > u)})$ ανεξάρτητων και ισόνομων τυχαίων μεταβλητών από μια κατανομή Bernoulli με πιθανότητα επιτυχίας $p = 1 - F(u)$.

Τότε η $E[L(u)] = \frac{1}{p}$ είναι η περίοδος επαναφοράς του ενδεχομένου $(X_i > u)$, όπου

$L(u) = \min\{i \geq 1 : X_i > u\}$ είναι η χρονική στιγμή που πραγματοποιείται η πρώτη υπέρβαση του κατωφλιού u , και η $L(u)$ ακολουθεί Γεωμετρική Κατανομή με παράμετρο p .

Έστω τώρα (X_i) μια ακολουθία ισόνομων και ανεξάρτητων τυχαίων μεταβλητών με συνεχή συνάρτηση κατανομής F . Αν $F(z_p) = 1 - p$, τότε z_p θα είναι το επίπεδο επιστροφής για την περίοδο επαναφοράς $\frac{1}{p}$.

Όμως $z_p = m - \frac{s}{x} \{1 - [-\ln(1-p)]^{-x}\}$, για $x \neq 0$

και

$$z_p = m - s \ln[-\ln(1-p)], \text{ για } x = 0$$

Έστω $y_p = -\ln(1-p)$, τότε το return level plot ορίζεται από τα σημεία: $(z_p, -\ln y_p)$.

Αν $x = 0$, τότε το διάγραμμα είναι γραμμικό. Αν $x > 0$, τότε το διάγραμμα είναι κυρτό με ασυμπτωτικό όριο $m - \frac{s}{x}$ καθώς το $p \rightarrow 0$. Αν $x < 0$, τότε το διάγραμμα είναι κοίλο και χωρίς όρια καθώς $p \rightarrow 0$.

2.8 Επίλογος

Μετά τη διεξοδική παρουσίαση της θεωρίας των ακραίων τιμών καθώς και κάποιων βασικών θεμάτων που σχετίζονται άμεσα με την εφαρμογή της θεωρίας αυτής, η ανάλυση μπορεί να περάσει στο στάδιο της συσχέτισης της Θεωρίας των Ακραίων Τιμών με ασφαλιστικά χαρτοφυλάκια ακραίων κινδύνων.

ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΕΙΡΑΙΑ

Κεφάλαιο 3

Η σπουδαιότητα της Ασφάλισης των Ακραίων Κινδύνων

3.1 Εισαγωγή

Έχοντας παρουσιάσει διεξοδικά την Θεωρία των Ακραίων Τιμών στην προηγούμενη ενότητα, το κεφάλαιο αυτό στοχεύει στο να συνδέσει την προαναφερθείσα θεωρία με την ασφαλιστική αγορά με τελικό σκοπό την ανάλυση δεδομένων ακραίων φυσικών καταστροφών που ακολουθεί στο επόμενο κεφάλαιο. Ουσιαστικά λοιπόν το κεφάλαιο αυτό αποτελεί το συνδυαστικό κρίκο μεταξύ της Θεωρία των Ακραίων Τιμών και της εφαρμογής που θα ακολουθήσει στο 4^ο κεφάλαιο.

Για το λόγο αυτό στη συγκεκριμένη ενότητα γίνεται αναφορά στη σημασία της ασφάλισης καταστροφικών κινδύνων που συνεπάγονται και ακραίες απώλειες, καθώς και στη στάση της Διεθνούς και της Ελληνικής Ασφαλιστικής Αγοράς απέναντι στους κινδύνους αυτούς. Ιδιαίτερη μνεία γίνεται για ακραίες απώλειες λόγω πλημμύρων διότι αυτή είναι και η φύση των δεδομένων που θα αναλυθούν παρακάτω. Επιπλέον, παρουσιάζεται ο ρόλος της Αντασφάλισης στην κάλυψη καταστροφικών κινδύνων, τα βασικά μοντέλα Αντασφάλισης που υπάρχουν στη διεθνή αγορά, καθώς επίσης και κάποια τεχνικά στοιχεία που αφορούν την τιμολόγηση τέτοιου είδους κινδύνων.

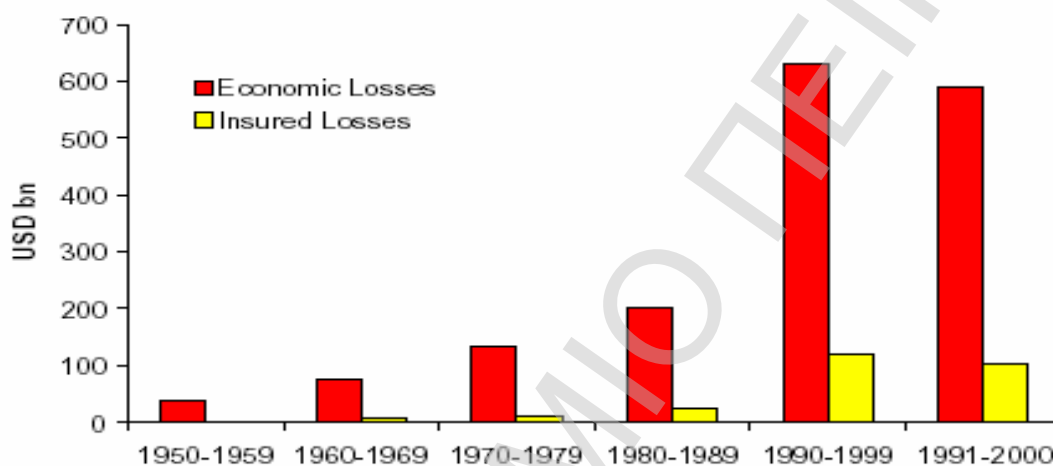
3.2 Ακραίοι Κίνδυνοι – Ακραίες Απώλειες

Σύμφωνα με τους Chavez-Demoulin V. & Roehrl A. (2004), τα ακραία φαινόμενα οδηγούν και σε ακραία κόστη. Οι Guin J. & Saxena V. (2001), επισημαίνουν ότι οι οικονομικές ζημιές από ακραίες φυσικές καταστροφές έχουν εννιαπλασιαστεί τα τελευταία 40 χρόνια. Ο Dlugolecki A.D. (2001), αναφέρει ότι δεν υπάρχει πλέον χώρος για αδιαφορία προς τα ακραία φαινόμενα και τις επιπτώσεις τους.

Σύμφωνα με στοιχεία του Ερυθρού Σταυρού (2000), τη δεκαετία του 60 οι απώλειες ήταν 73 δισεκατομμύρια δολάρια, τη δεκαετία του 80 οι ζημιές ανήλθαν σε 204 δις. δολάρια,

ενώ τη δεκαετία του 90 οι οικονομικές απώλειες παγκοσμίως από φυσικές καταστροφές ξεπέρασαν τα 700 δις. δολάρια.

Το παρακάτω διάγραμμα της εταιρίας αντασφαλιστών του Μονάχου που δείχνει τη διαχρονική εξέλιξη των ζημιών λόγω φυσικών καταστροφών είναι ιδιαίτερα αποκαλυπτικό.



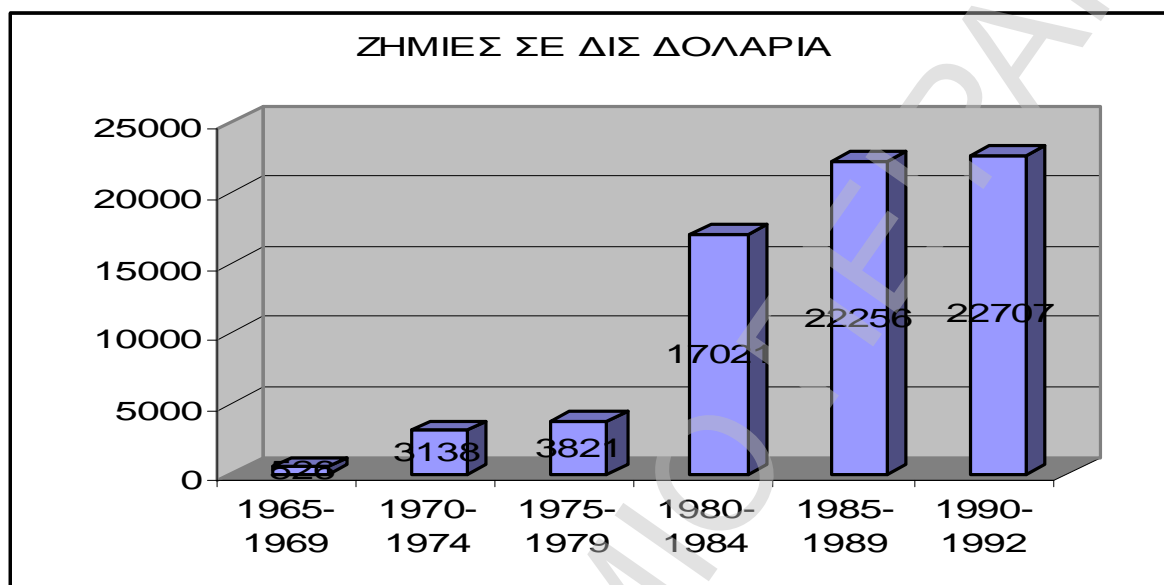
Σχήμα 3-1. Διαχρονική εξέλιξη οικονομικών απωλειών από καταστροφικούς κινδύνους και αντίστοιχων ποσοστών ασφάλισης κατά των κινδύνων αυτών.

(Πηγή: Guin J. & Saxena V. 2001)

Από το 1950 μέχρι σήμερα υπάρχει μια πολύ σημαντική αύξηση των οικονομικών απωλειών εξαιτίας ακραίων φυσικών φαινομένων. Τα μεγέθη αυτά απεικονίζονται από τις μπάρες με το κόκκινο χρώμα. Όμως το παραπάνω διάγραμμα παρουσιάζει και τα αντίστοιχα μεγέθη των ζημιών που έχουν ασφαλιστεί. Παρόλο που και τα μεγέθη των ασφαλισμένων κινδύνων αυξάνονται διαχρονικά, εξακολουθούν να παρουσιάζουν σημαντική υστέρηση σε σχέση με τα αντίστοιχα μεγέθη των οικονομικών απωλειών που παρατηρούνται κάθε φορά που παρουσιάζεται κάποιο ακραίο φαινόμενο.

Αξιοσημείωτες είναι και οι οικονομικές απώλειες που προέρχονται από πλημμύρες. Όπως φαίνεται και από το παρακάτω διάγραμμα το παγκόσμιο μέγεθος των ζημιών που

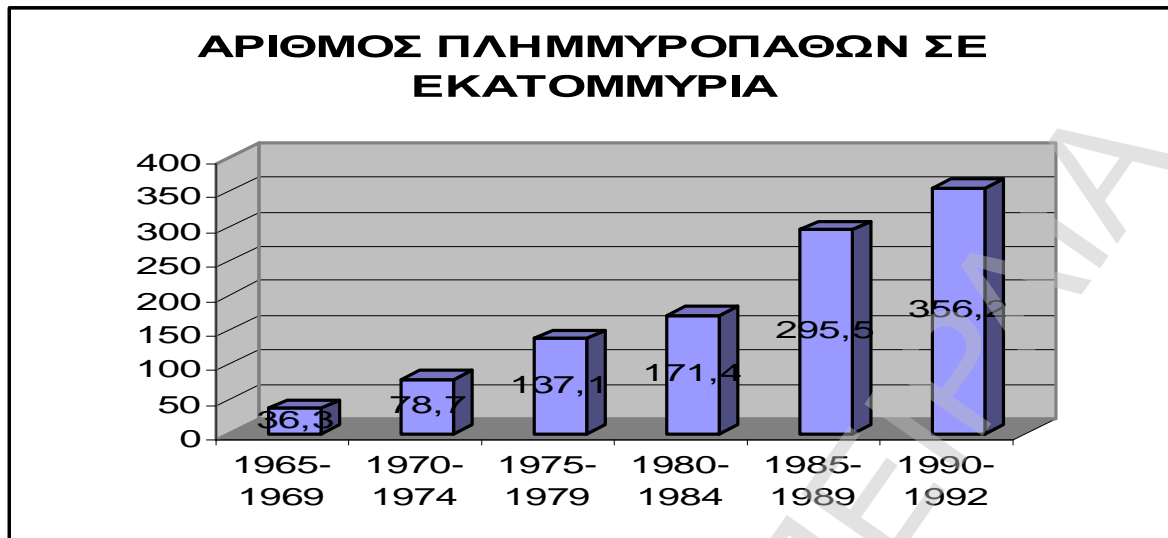
προέρχονται από πλημμύρες από 526 δις. Δολάρια που ήταν το 1965 αυξήθηκε σε 22.707 δις δολάρια το 1992. Μέσα δηλαδή σε λιγότερο από 30 χρόνια η ένταση των οικονομικών απωλειών λόγω πλημμύρων έγινε πάνω 43 φορές μεγαλύτερη σε παγκόσμιο επίπεδο.



Σχήμα 3-2. Διαχρονική Εξέλιξη των απωλειών από πλημμύρες σε παγκόσμιο επίπεδο

(Πηγή: Ένωση Ασφαλιστικών Εταιριών Ελλάδος, 1997)

Οι επιπτώσεις όμως των φυσικών καταστροφών δεν έχουν μόνο οικονομικές αλλά και κοινωνικές προεκτάσεις. Είναι χαρακτηριστικό, όπως φαίνεται και στο παρακάτω διάγραμμα, ότι ο αριθμός των ανθρώπων σε παγκόσμιο επίπεδο που υπέστησαν τις συνέπειες ακραίων καταστροφών λόγω πλημμύρων αυξήθηκαν από 36,3 εκατομμύρια που ήταν το 1965 σε 356,2 εκατομμύρια το 1992. Με άλλα λόγια παρατηρείτε μια δραματική αύξηση και στον αριθμό των πλημμυροπαθών διαχρονικά.



Σχήμα 3-3. Διαχρονική εξέλιξη του αριθμού των πλημμυροπαθών παγκοσμίως.

(Πηγή: Ένωση Ασφαλιστικών Εταιριών Ελλάδος, 1997)

Όμως και στην Ελλάδα και κυρίως στην περιοχή της Αττικής οι ζημιές από πλημμύρες είναι πάρα πολύ σημαντικές. Αξίζει να σημειωθεί, σύμφωνα με τα στοιχεία της Ένωσης Ελληνικών Ασφαλιστικών Εταιριών (1997), το 1977 στις περιοχές Περιστέρη και Αιγάλεω και το 1993 στα νότια προάστια του λεκανοπέδιου, οι υλικές ζημιές ανήλθαν σε 1,5 δις. Δραχμές για κάθε ανά από τα έτη αυτά. Επίσης το 1994 στα Άνω Λιόσια και στις γύρω περιοχές πραγματοποιήθηκαν ζημιές της τάξης του 0,5 δις δραχμές.

Πέρα όμως από το σημαντικό μέγεθος των οικονομικών απωλειών που παρατηρείται στον Ελλαδικό χώρο από την εκδήλωση ακραίων φαινομένων, το πιο ανησυχητικό είναι το χαμηλό επίπεδο ασφάλισης τέτοιων κινδύνων. Είναι χαρακτηριστικό ότι μόλις το 6% του συνόλου των κτιρίων στην Ελλάδα έχουν ασφαλιστεί για τον κίνδυνο της πλημμύρας.

Από όλα τα παραπάνω είναι προφανές ότι ο θεσμός της ασφάλισης έχει να παίξει έναν ιδιαίτερα σημαντικό ρόλο προκειμένου τα ακραία φαινόμενα να μην σημαίνουν και ακραίες απώλειες για τους πολίτες. Η σπουδαιότητα της ασφάλισης καταστροφικών κινδύνων έγκειται στο γεγονός ότι μετριάξει τις οικονομικές, κοινωνικές και άλλες επιπτώσεις που συνεπάγεται η πραγματοποίηση τέτοιων κινδύνων.

Πως όμως αντιμετωπίζει η Διεθνής και η Ελληνική Αντ/ασφαλιστική Αγορά τέτοιου είδους ακραία και ιδιαίτερα ζημιογόνα φαινόμενα;

3.3 Η στάση της Διεθνούς Ασφαλιστικής Αγοράς απέναντι στα Ακραία Φαινόμενα

Σύμφωνα με τον Boulaton A. (2003), η στάση της ασφαλιστικής αγοράς είναι κοινή απέναντι στα ακραία καταστροφικά φαινόμενα και γίνεται ακόμα πιο ομοιογενής όσο προχωράει η παγκοσμιοποίηση των εθνικών οικονομιών.

Έτσι, μετά από την πραγματοποίηση κάποιου καταστροφικού γεγονότος είναι σύνηθες το φαινόμενο οι αντ/ασφαλιστικές εταιρίες που κάλυπταν τέτοιους κινδύνους, στη συνέχεια να ανακοινώνουν ότι δεν ανανεώνουν τα συμβόλαια των πελατών τους και να εγκαταλείπουν τη συγκεκριμένη αγορά. Το αποτέλεσμα είναι να δημιουργούνται πολλά προβλήματα σε διάφορους κλάδους της οικονομίας, όπως στην αγορά και χρηματοδότηση ακινήτων, στον κατασκευαστικό κλάδο κ.α

Από την άλλη μεριά, οι αντ/ασφαλιστικές εταιρίες που παραμένουν στην αγορά προχωρούν σε σημαντικές αυξήσεις ασφαλίσεων που συνδέονται με την κάλυψη καταστροφικών κινδύνων. Μάλιστα, στην ασφαλιστική αγορά της Αμερικής μετά από γεγονότα όπως ο τυφώνας Andrew ή το τρομοκρατικό κτύπημα της 11^{ης} Σεπτεμβρίου παρατηρήθηκε αύξηση των ασφαλίσεων από τις εταιρίες που παρέμειναν στην αγορά που έφτασε και το 400%.

Μετά από μερικό καιρό κυβερνητικοί μηχανισμοί επεμβαίνουν με διάφορους τρόπους για να ομαλοποιήσουν την αγορά.

Όσο προχωράει ο καιρός και παρόμοια γεγονότα δεν συμβαίνουν οι εταιρίες που είχαν αποχωρήσει από το συγκεκριμένο κομμάτι της αγοράς επανέρχονται και σταδιακά τα ασφάλιστρα, τουλάχιστον αυτά που αφορούν φυσικές καταστροφές, επανέρχονται σχεδόν στα ίδια επίπεδα πριν συμβεί το καταστροφικό γεγονός. Στα καταστροφικά τρομοκρατικά κτυπήματα η μείωση των ασφαλίσεων είναι πολύ πιο μετριοπαθής και η επανείσοδος των εταιριών που είχαν φύγει πολύ πιο αργή.

Αυτή η περιγραφή για το πώς λειτουργούν οι αντ/ασφαλιστικές αγορές στα ακραία φαινόμενα δείχνει συνοπτικά ότι από τη μια μεριά οι τιμές τείνουν να αυξάνονται απότομα και μετά να μειώνονται σταδιακά καθώς κανένα ακραίο γεγονός δεν συμβαίνει πια, και από την άλλη εταιρίες αφήνουν την αγορά άμεσα και επιστρέφουν σταδιακά καθώς κανένα ακραίο γεγονός δεν συμβαίνει πια. Γενικά, παρατηρείται οι εταιρίες να εγκαταλείπουν την αγορά όταν οι τιμές είναι ψηλά και να επανέρχονται όταν οι τιμές πέφτουν.

3.4 ΑΝΤΑΣΦΑΛΙΣΗ

Η σχέση αντασφάλισης και καταστροφικών κινδύνων είναι προφανής. Οι εξαιρετικά μεγάλες απαιτήσεις που υπερβαίνουν κάποιο πάρα πολύ υψηλό όριο και που είναι συνέπεια κάποιου καταστροφικού γεγονότος καλύπτονται από τον αντασφαλιστή.

Υπάρχουν τέσσερις διαφορετικές μορφές αντασφάλισης που είναι οι εξής:

1) *Αντασφάλιση Υπερβάλλοντος Ποσού Ζημιάς (Excess Loss):*

Όταν ο αντασφαλιστής πληρώνει το υπερβάλλον ποσό από κάποιο συγκεκριμένο και προκαθορισμένο όριο για ατομικά μεγέθη ζημιών.

Στην πιο απλή της μορφή η ασφαλιστική εταιρία καταβάλλει τις αποζημιώσεις για οποιοδήποτε ποσό έως κάποιο όριο d , ενώ το υπερβάλλον ποσό ζημιάς (αν υπάρχει) το καταβάλλει ο αντασφαλιστής.

Έτσι λοιπόν αν X_i είναι το μέγεθος της i ζημιάς, Y_i είναι το ποσό της αποζημίωσης που καταβάλλει η πρωτασφαλίστρια εταιρία και Z_i είναι το ποσό της αποζημίωσης που καταβάλλει η αντασφαλίστρια εταιρία έχουμε ότι:

$$Y_i = \begin{cases} X_i, & X_i \leq d \\ d, & X_i > d \end{cases}$$

και

$$Z_i = \begin{cases} 0, & X_i \leq d \\ X_i - d, & X_i > d \end{cases}$$

2) *Αναλογική Αντασφάλιση:*

Ο ασφαλιστής καταβάλλει ως αποζημίωση ένα συγκεκριμένο ποσοστό της ζημιάς ανεξάρτητα από το συνολικό μέγεθός της. Άρα αν X_i είναι το μέγεθος της i ζημιάς, Y_i είναι το ποσό της αποζημίωσης που καταβάλλει η πρωτασφαλίστρια εταιρία και Z_i είναι το ποσό της αποζημίωσης που καταβάλλει η αντασφαλίστρια εταιρία έχουμε ότι: $Y_i = aX_i$ και $Z_i = (1 - a)X_i$

3) *Αντασφάλιση stop-loss:*

Ο αντασφαλιστής καλύπτει το ποσό που υπερβαίνει κάποιο όριο για το συνολικό ποσό των απαιτήσεων του χαρτοφυλακίου του ασφαλιστή.

4) *Αντασφάλιση ECOMOR:*

Είναι μια τροποποίηση της αντασφάλισης υπερβάλλοντος ποσού ζημιάς με το n-οστό μεγαλύτερο ατομικό μέγεθος απαίτησης να λαμβάνεται ως τυχαίο όριο, γι' αυτό και ο αντασφαλιστής πληρώνει μόνο τα ποσά που υπερβαίνουν το n-οστό μεγαλύτερο μέγεθος απαίτησης.

Στην παρούσα εργασία θα ασχοληθούμε κυρίως με κινδύνους σε σχέση με την αντασφάλιση υπερβάλλοντος ποσού ζημιάς. Οι άλλες μορφές αντασφάλισης παρουσιάζουν μικρότερο πρακτικό ενδιαφέρον σε σχέση με το αντικείμενο που πραγματεύεται η παρούσα εργασία.

Όταν αναφερόμαστε σε τιμολόγηση για ασφαλιστικές και αντασφαλιστικές εταιρίες ενδιαφερόμαστε στο να έχουμε καλά πιθανοθεωρητικά μοντέλα για τις κατανομές του αριθμού και του ποσού των ζημιών που πραγματοποιήθηκαν μέσα σε κάποιο χρόνο. Την τελευταία δεκαετία η αντασφαλιστική βιομηχανία υπέφερε από καταστροφικά γεγονότα των οποίων το μέγεθος ήταν πέρα από κάθε πρόβλεψη για απώλειες που είχαν γίνει από τα διαθέσιμα μοντέλα. Γι' αυτό το λόγο μια πραγματική κατανόηση της στατιστικής μοντελοποίησης των ακραίων φαινομένων απέκτησε μεγάλο ενδιαφέρον ανάμεσα στους αναλογιστές των αντ/ασφαλιστικών εταιριών.

Ο Embrechts et al (1997), αναφέρει ότι η θεωρία των ακραίων τιμών έχει έναν πολύ σημαντικό ρόλο να παίζει στην τιμολόγηση αντ/ασφαλιστικών συμβολαίων. Αυτό που αποτελεί κύριο στόχο των αναλογιστών είναι η δημιουργία καλών μοντέλων πρόβλεψης για τους καταστροφικούς κινδύνους προκειμένου να υπολογίζονται ασφάλιστρα που να μην είναι ούτε πολύ υψηλά (δηλαδή απαγορευτικά για τον αγοραστή), αλλά ούτε και πολύ χαμηλά (ούτως ώστε να θέτουν σε κίνδυνο την εταιρία), σύμφωνα και με τη Mata A. (2000).

Ο αντ/ασφαλιστής προκειμένου να μπορέσει να ικανοποιήσει τις οικονομικές απαιτήσεις των πελατών του όταν ένα ακραίο γεγονός πραγματοποιηθεί πρέπει να είναι σε θέση να μπορεί να αποτιμήσει σωστά τον κίνδυνο που υπάρχει. Δηλαδή θα πρέπει να

μπορέσει να εκτιμήσει με όσο πιο ακριβή τρόπο γίνεται τη σχέση ενδεχόμενης ζημιάς και συχνότητας εμφάνισης.

Η Επίδραση των Υψηλών Απαιτήσεων στη Λήψη Αποφάσεων

1) Αριθμός Απαιτήσεων και Συνολικό Ποσό Απαιτήσεων

Έστω $S_n = \sum_{i \leq n} X_i$ το συνολικό (αθροιστικό) ποσό απαιτήσεων των πρώτων n τυχαίων απαιτήσεων X_1, X_2, X_3, \dots . Τότε, το συνολικό ποσό των απαιτήσεων για ένα συγκεκριμένο χρονικό διάστημα θα είναι $S_N = \sum_{i \leq N} X_i$, όπου N είναι ο αριθμός των απαιτήσεων που πραγματοποιούνται μέσα στο συγκεκριμένο χρονικό διάστημα και είναι και αυτό μια τυχαία μεταβλητή.

Το αναμενόμενο ποσό των συνολικών αποζημιώσεων $E(S_n)$ είναι στην ουσία το καθαρό ασφάλιστρο.

2) Η Συνάρτηση Κατανομής των Συνολικών Απαιτήσεων

Σύμφωνα με το ομογενές μοντέλο κινδύνου (homogeneous risk model) τα μεγέθη των απαιτήσεων X_1, X_2, X_3, \dots , είναι iid τυχαίες μεταβλητές με κοινή συνάρτηση κατανομής F . Επιπροσθέτως, ο αριθμός των απαιτήσεων N και η ακολουθία X_1, X_2, X_3, \dots είναι ανεξάρτητα.

Άρα η συνάρτηση κατανομής των συνολικών απαιτήσεων θα είναι:

$$\Pr\{S_N \leq x\} = \sum_{n=0}^{\infty} \Pr\{S_N \leq x, N = n\} = \sum_{n=0}^{\infty} \Pr\{N = n\} F^{n*}(x),$$

όπου $S_0 = 0$, $F^{0*}(x) = I(0 \leq x)$ και $F^{n*}(x) = \Pr(S_n \leq x)$ είναι η n -οστή συνέλιξη της F .

Επίσης, το καθαρό ασφάλιστρο είναι, όπως έχει ήδη αναφερθεί, το αναμενόμενο ποσό των συνολικών αποζημιώσεων $E(S_n)$.

Δεδομένου όμως ότι X είναι τυχαία μεταβλητή με ίδια κατανομή με την X_i και ότι οι τυχαίες μεταβλητές S_n και $I(N = n)$ είναι ανεξάρτητες, έχουμε ότι:

$$\begin{aligned} E(S_n) &= \sum_{n=0}^{\infty} E(S_n I(N = n)) = \sum_{n=1}^{\infty} E(S_n) E(I(N = n)) \\ &= E(X) \sum_{n=1}^{\infty} n \Pr(N = n) = E(X) E(N) \end{aligned}$$

και

$$V(S_N) = V(X)E(N) + (EX)^2 V(N)$$

Εκτίμηση του Καθαρού Ασφαλίστρου

1) Μη – Παραμετρική Εκτίμηση του Καθαρού Ασφαλίστρου

Υποθέτουμε ότι οι χρόνοι άφιξης των απαιτήσεων T_1, T_2, T_3, \dots αποτελούν μια διαδικασία παραγωγής απαιτήσεων:

$$N(t) = \sum_{i=1}^{\infty} I(T_i \leq t), t \geq 0$$

με σταθερό ρυθμό άφιξης που είναι:

$$E(N(t_2) - N(t_1)) = (t_2 - t_1)E(N(1))$$

Η εκτίμηση του καθαρού ασφαλίστρου $E(S_n)$ για την επόμενη περίοδο θα βασίζεται στο συνολικό ποσό των απαιτήσεων $S_{N(T)}$ των περασμένων T – ετών. Κάτω από τις συνθήκες του ομογενούς μοντέλου κινδύνου και κάτω από τη σχέση $E(S_n) = E(X)E(N)$, έχουμε ότι $E(S_n) = E(N(1))E(X)$ και κάτω από τη συνθήκη $E(N(t_2) - N(t_1)) = (t_2 - t_1)E(N(1))$ έχουμε τελικά ότι

$$E(S_{N(T)}) = T E(N(1))E(X)$$

Άρα ο λόγος $S_{N(T)}/T$ είναι ένας αμερόληπτος εκτιμητής για το καθαρό ασφαλίστρο $E(S_n)$.

2) Παραμετρική Εκτίμηση του Καθαρού Ασφαλιστρού

Υποθέτουμε ότι οι χρόνοι άφιξης και τα μεγέθη των ζημιών ικανοποιούν τις συνθήκες μιας διαδικασίας Poisson (λ, F). Παρατηρούμε ότι το λ είναι ο μέσος αριθμός απαιτήσεων σε μια περίοδο μοναδιαίου μήκους και F είναι η συνάρτηση κατανομής των τιμών που υπερβαίνουν το όριο u .

Υποθέτουμε ότι η $F = W_{g,u,s}$ είναι μια Γενικευμένη Pareto. Το μέσο μέγεθος των απαιτήσεων $E(X)$ είναι ο μέσος της συνάρτησης κατανομής της Γενικευμένης Pareto (για την γ - παραμετροποίηση). Το καθαρό ασφαλιστρού για την επόμενη περίοδο είναι $E(S_N) = I m w_{g,u,s}$, όπου m είναι ο μέσος μιας συνάρτησης κατανομής F .

Συνεπώς, εκτιμώντας κανείς τις παραμέτρους λ, γ, σ παίρνει μια εκτίμηση του καθαρού ασφαλιστρού.

Επίσης, $I_{N(T)} = N(T)/T$ είναι ένας εκτιμητής του $I = E(N)$. Αν $g_{N(T)}$ και $s_{N(T)}$ είναι εκτιμητές των παραμέτρων g και s αντίστοιχα, τότε ο μέσος

$$m_{N(T)} = u + \frac{s_{N(T)}}{1 + g_{N(T)}}$$

της συνάρτησης κατανομής $W_{g_{N(T)},u,s_{N(T)}}$ μιας Γενικευμένης Pareto είναι ένας εκτιμητής του μέσου μεγέθους των απαιτήσεων $E(X) = m w_{g,u,s}$. Για το λόγο αυτό $I_{N(T)} m_{N(T)}$ είναι ένας εκτιμητής του καθαρού ασφαλιστρού $E(S_N)$.

3) Ο Συνδυασμός της Παραμετρικής και της Μη - Παραμετρικής Προσέγγισης

Οι δύο αυτές προσεγγίσεις μπορούν να συνδυαστούν για να εκτιμηθεί η συνάρτηση κατανομής του μεγέθους των απαιτήσεων:

- πάνω από ένα καθορισμένο όριο u με παραμετρική προσέγγιση
- κάτω από το όριο u με ένα μη - παραμετρικό τρόπο

Έστω $\hat{F}_n(x;0)$ η δειγματική συνάρτηση κατανομής η οποία βασίζεται στα μεγέθη των απαιτήσεων x_i . Έστω $G_{x,\hat{m},\hat{s}}$ η συνάρτηση κατανομής μιας Γενικευμένης Pareto προσαρμοσμένης στο πάνω άκρο της ουράς της δειγματικής συνάρτησης κατανομής $\hat{F}_n(x;0)$.

Και οι δύο συναρτήσεις κατανομών μπορούν να ενωθούν ομαλά σε ένα ενιαίο σύνολο αφού $G_{x,\tilde{m},\tilde{s}} = \hat{F}_n(x;u)$.

Η συνάρτηση κατανομής $\hat{F}_n(x;x) I(x \leq u) + W_{g,\tilde{m},\tilde{s}} I(x > u)$ είναι μια μη-παραμετρική – παραμετρική εκτίμηση της συνάρτησης κατανομής του μεγέθους των απαιτήσεων.

Αυτή η προσέγγιση πρέπει να προτιμάται και στις περιπτώσεις που η παραμετρική υπόθεση είναι βάσιμη μόνο για κάποιο όριο μεγαλύτερο από το προκαθορισμένο από τη συνθήκη Excess of Loss(XL).

Κατάτμηση Σύμφωνα με την Πιθανή Μέγιστη Ζημιά

1) Μέσο Μέγεθος Ζημιών σε Συνάρτηση με την Πιθανή Μέγιστη Ζημιά

Η δυνητική ζημιά δεν είναι ομογενής για όλους τους κινδύνους ενός χαρτοφυλακίου αλλά εξαρτάται από την υποκείμενη έκθεση του ατομικού κινδύνου του συμβολαίου. Μια πάρα πολύ σημαντική έννοια είναι αυτή της Μέγιστης Πιθανής Ζημιάς ενός ατομικού κινδύνου που στην ουσία περιγράφει το μέγιστο μέγεθος μιας ατομικής απαίτησης η οποία καλύπτεται από το συμβόλαιο. Η αξία αυτή μπορεί να θεωρηθεί ως το άνω άκρο της κατανομής του ατομικού μεγέθους της απαίτησης.

Γι' αυτό πέρα από τη μεταβλητή X του μεγέθους της απαίτησης για το ατομικό συμβόλαιο, υπάρχει η μέγιστη πιθανή ζημιά. Αυτό που μας ενδιαφέρει είναι η υπό συνθήκη μέση τιμή $E(X|z)$ του X δοθέντος του $Z = z$, όπου z είναι η μέγιστη πιθανή ζημιά για κάποιο συμβόλαιο. Ο σκοπός είναι να εκτιμηθεί το υπό συνθήκη μέσο μέγεθος απαίτησης $E(X|z)$ σαν συνάρτηση της μέγιστης πιθανής ζημιάς z .

Τα επόμενα βήματα είναι να εκτιμηθούν ο μέσος αριθμός απαιτήσεων και το μέσο συνολικό ποσό των απαιτήσεων σε συνάρτηση με τη μέγιστη πιθανή ζημιά.

2) Εκτίμηση του μέσου μεγέθους απαιτήσεων για μια ομάδα μέγιστων πιθανών ζημιών χρησιμοποιώντας διαβαθμίσεις απαιτήσεων

Η i -th ομάδα μέγιστης πιθανής ζημιάς καθορίζεται από αυτούς τους κινδύνους με μέγιστες πιθανές ζημιές ανάμεσα στα όρια p_i και p_{i+1} , όπου

$$u = p_0 < p_1 < \dots < p_{m-1} < p_m = \infty$$

είναι μια προκαθορισμένη διαμέριση.

Γράφουμε n_i για τον αριθμό των απαιτήσεων και $x_{i,j}$ για τα μεγέθη των απαιτήσεων που ανήκουν στην i -th ομάδα μέγιστης πιθανής ζημιάς. Η μέγιστη πιθανή ζημιά που χαρακτηρίζει το $x_{i,j}$ δηλώνεται ως $z_{i,j}$.

Επίσης, $n^{-1} \sum_{j \leq n_i} x_{i,j}$ είναι ένας εκτιμητής του μέσου μεγέθους της απαίτησης μέσα στην i -th ομάδα μέγιστης πιθανής ζημιάς.

Προκειμένου να είναι τα αποτελέσματα ανάμεσα σε διαφορετικές ομάδες μέγιστης πιθανής ζημιάς συγκρίσιμα εισάγουμε τις διαβαθμίσεις $x_{i,j} / z_{i,j}$ και τις εμπειρικές μέσες βαθμίδες $d_i = n^{-1} \sum_{j \leq n_i} x_{i,j} / z_{i,j}$.

Το μέσο μέγεθος της απαίτησης μέσα στην i -th ομάδα μπορεί να εκτιμηθεί από τη σχέση $m_i = q_i d_i$, όπου $q_i = (p_i + p_{i+1}) / 2$ είναι τα κέντρα των διαστημάτων μέγιστης πιθανής ζημιάς. Το πλεονέκτημα της χρησιμοποίησης των εμπειρικών μέσων βαθμίδων d_i για την εκτίμηση των μέσων μεγεθών των ζημιών είναι ότι κανείς μπορεί να εξομαλύνει την απόκλιση αυτών των τιμών χρησιμοποιώντας π.χ πολυωνυμική συνάρτηση ελαχίστων τετραγώνων.

Χρησιμοποιώντας κανείς τις εξομαλυσμένες μέσες βαθμίδες d'_i από την ευθεία ελαχίστων τετραγώνων παίρνει την εκτίμηση $m'_i = q_i d'_i$ του μέσου μεγέθους της απαίτησης στην i -th ομάδα πιθανής μέγιστης ζημιάς.

3) Σχετικές Συχνότητες για Τμηματοποιημένα Χαρτοφυλάκια και Εκτίμηση του Καθαρού Ασφαλίστρου για Ομάδες Πιθανής Μέγιστης Ζημιάς

Η σχετική συχνότητα απαίτησης είναι $f_i = n_i / r_i$, όπου n_i είναι ο αριθμός των απαιτήσεων και r_i είναι ο αριθμός των κινδύνων που ανήκουν στην i -th ομάδα πιθανής μέγιστης ζημιάς.

Έστω τώρα ότι m'_i είναι και πάλι ένας εκτιμητής του μέσου μεγέθους της απαίτησης στην i -th ομάδα πιθανής μέγιστης ζημιάς και f'_i είναι οι εξομαλυσμένες σχετικές συχνότητες.

Τότε, πέρα από το $\sum_{j=1}^{n_i} x_{i,j}$ κανείς παίρνει μέσα από την $m'_i f'_i r_i$ έναν εκτιμητή του καθαρού ασφαλίστρου για την i -th ομάδα πιθανής μέγιστης ζημιάς. Ως επακόλουθο, m'_i θα αντικατασταθεί από παραμετρικούς εκτιμητές του μέσου μεγέθους της απαίτησης. Αυτοί οι εκτιμητές βασίζονται είτε σε βαθμίδες απαιτήσεων είτε σε πραγματικά μεγέθη ζημιών. Η εξομάλυνση από τις γραμμές ελαχίστων τετραγώνων γίνεται για τις παραμέτρους των κατανομών της Γενικευμένης Pareto.

4) Η Παραμετρική Προσέγγιση

Το μέσο μέγεθος απαίτησης στην i -th ομάδα πιθανής μέγιστης ζημιάς μπορεί να εκτιμηθεί με έναν παραμετρικό τρόπο.

Καθορίζεται ένα μοντέλο Γενικευμένης Pareto για τις βαθμίδες των απαιτήσεων ή για τα πραγματικά μεγέθη των απαιτήσεων μέσα στην i -th ομάδα πιθανής μέγιστης ζημιάς και χρησιμοποιούμε τον μέσο της εκτιμημένης κατανομής της Γενικευμένης Pareto για την εκτίμηση της μέσης βαθμίδας απαίτησης ή του μέσου μεγέθους απαίτησης.

5) Η Παραμετρική Κοντινότερη Γειτονική Προσέγγιση

Το μειονέκτημα της προηγούμενης μεθόδου είναι ότι για κάποιες από τις ομάδες πιθανής μέγιστης ζημιάς η παραμετρική εκτίμηση πρέπει να βασίζεται σε πάρα πολύ μικρό αριθμό απαιτήσεων.

Έχει ήδη αναφερθεί ότι το μέσο μέγεθος της απαίτησης μέσα σε μια ομάδα πιθανής μέγιστης ζημιάς σχετίζεται με της υπό συνθήκη προσδοκία $E(X|z)$ του μεγέθους της απαίτησης X , με τη δέσμευση της πιθανής μέγιστης ζημιάς ότι $Z = z$.

Η εκτίμηση της υπό συνθήκη προσδοκία $E(X|z)$ μπορεί επίσης να βασίζεται στα μεγέθη των απαιτήσεων y_j , $j = 1, \dots, k$, που αναφέρεται στις k πιθανές μέγιστες ζημιές z_j που είναι κοντύτερα στο z .

Τότε $m_k = k^{-1} \sum_{j \leq k} y_j$ είναι ένας εμπειρικός εκτιμητής του δεσμευμένου μέσου μεγέθους απαίτησης $E(X|z)$.

Παρομοίως η ίδια εκτίμηση μπορεί να γίνει για ένα μοντέλο Γενικευμένης Pareto. Για να γίνει αυτό πρέπει πρώτα να υπολογιστούν οι εκτιμήσεις $\hat{x}_k(z)$ και $\hat{s}_k(z)$ των παραμέτρων shape και κλίμακας $x(z)$ και $s(z)$ βασιζόμενοι στα y_j του μοντέλου της Γενικευμένης Pareto και χρησιμοποιούμε τον μέσο της Γενικευμένης Pareto κατανομής $W_{\hat{g}_k(z), \mu, \hat{s}_k(z)}$, ως ένας εκτιμητής του $E(X|z)$.

Η Διαδικασία Κινδύνου

1) Καθαρά και Συνολικά Ασφάλιστρα

Το καθαρό ασφάλιστρο στο χρόνο t είναι η προσδοκώμενη τιμή $E(S(t))$ των συνολικών απαιτήσεων $S(t)$. Υποθέτοντας ανεξαρτησία για τα μεγέθη των απαιτήσεων X_1, X_2, X_3, \dots καθώς επίσης και για τη διαδικασία άφιξης των απαιτήσεων T_1, T_2, T_3, \dots , ισχύει η ανεξαρτησία ανάμεσα στα μεγέθη των απαιτήσεων X_1, X_2, X_3, \dots και τον αριθμό των απαιτήσεων $N(t)$.

Επίσης, υποθέτουμε ότι τα μεγέθη των απαιτήσεων X_i έχουν την ίδια κατανομή. Το καθαρό ασφάλιστρο μπορεί να γραφεί ως:

$$E(S(t)) = E(X)E(N(t)).$$

Αν η διαδικασία του αριθμού των απαιτήσεων $N(t)$, $t \geq 0$ είναι μια ομογενής διαδικασία Poisson έντασης λ , τότε: $E(S(t)) = E(X)It$.

Αν $N(t)$, $t \geq 0$ είναι μια διαδικασία Polya – Lundberg, τότε ένας αντίστοιχος τύπος ισχύει με το λ να αντικαθίσταται από το a .

Έστω τώρα ότι το συνολικό ασφάλιστρο $c(t)$, δηλαδή το συνολικό ποσό που είναι να πληρωθεί από τους ασφαλισμένους για να αντισταθμίσουν τις μελλοντικές απώλειες, είναι ένα πολλαπλάσιο του καθαρού ασφάλιστρου και μπορεί να γραφεί ως:

$c(t) = (1 + p) E(S(t))$, όπου p είναι μια σταθερά που ονομάζεται συντελεστής ασφαλείας (safety loading).

2) Η Διαδικασία του Κινδύνου

Ο τύπος που δίνει τη διαδικασία του κινδύνου είναι ο εξής:

$$U(t) = s + E(S(t)) + b(t) - S(t), \quad t \geq 0$$

όπου:

$s = U(0) \geq 0$ είναι το αρχικό απόθεμα (κεφάλαιο) του ασφαλιστή για κάποιο συγκεκριμένο χαρτοφυλάκιο.

$E(S(t))$ είναι το καθαρό ασφάλιστρο μέχρι κάποιο χρονικό σημείο t .

$b(t)$ είναι η συνάρτηση ασφαλείας (μέση υπεραξία) στο χρόνο t που ορίζεται ως:

- Τη διαφορά $pE(S(t))$ ανάμεσα στο συνολικό και στο καθαρό ασφάλιστρο
- Συν τα τοκομερίδια από το συνολικό κεφάλαιο
- Πλην έξοδα, φόρους, μερίσματα κ.α

$S(t) = S_{N(t)} = \sum_{i \leq N(t)} X_i$ είναι το συνολικό ποσό απαιτήσεων μέχρι το χρόνο t .

Ο μέσος της μεταβλητής του κεφαλαίου $U(t)$ είναι $E(U(t)) = s + b(t)$ και για το λόγο αυτό το $b(t)$ είναι η υπεραξία που μπορεί να προστεθεί στο απόθεμα (κεφάλαιο). Είναι φανερό ότι η συνάρτηση ασφαλείας $b(t)$ είναι μη-αρνητική.

3.5 Επίλογος

Μετά την παρουσίαση της Θεωρίας των Ακραίων Τιμών, που ήταν το αντικείμενο του 2^{ου} Κεφαλαίου, και την παρουσίαση της Ασφάλισης ως του αποτελεσματικότερου τρόπου αντιμετώπισης ακραίων κινδύνων, που ήταν το αντικείμενο του παρόντος Κεφαλαίου, η ανάλυση μπορεί να προχωρήσει στο στάδιο της επεξεργασίας δεδομένων καταστροφικών κινδύνων κάποιου ασφαλιστικού χαρτοφυλακίου με τη χρήση της Θεωρίας των Ακραίων Τιμών.

ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΕΙΡΑΙΑΣ

Κεφάλαιο 4

Η εφαρμογή της Θεωρίας των Ακραίων Τιμών σε πραγματικά δεδομένα

4.1 Εισαγωγή

Στο προηγούμενο κεφάλαιο έγινε λόγος για τη σπουδαιότητα της ασφάλισης ακραίων και καταστροφικών κινδύνων προκειμένου να μετριάζονται οι οικονομικές και κοινωνικές τους συνέπειες. Η θεωρία των ακραίων τιμών έχει αποδειχθεί ένα πολύ χρήσιμο εργαλείο στα χέρια των αναλογιστών προκειμένου να εκτιμηθούν και να τιμολογηθούν τέτοιου είδους κίνδυνοι.

Σκοπός λοιπόν του κεφαλαίου αυτού είναι η ανάλυση, μέσω της θεωρίας των ακραίων τιμών, δεδομένων που προέρχονται από την Ελληνική Ασφαλιστική αγορά και αφορούν αποζημιώσεις που έχουν δοθεί από κάποια εταιρεία τα έτη 2002-04 για καταστροφές λόγω πλημμύρων στην περιοχή της Αττικής.

Πώς όμως ορίζεται ο όρος πλημμύρα στον ασφαλιστικό κλάδο; Η Ένωση Ασφαλιστικών Εταιριών Ελλάδος (1997), ορίζει την πλημμύρα ως: "...κατακλυσμό από θαλάσσια ύδατα, την λόγω φυσικών αιτιών υπερχειλίση ή παρέκκλιση από τα συνήθη κανάλια απορροής των φυσικών ή τεχνητών υδατορευμάτων, την διάρρηξη ή υπερχειλίση του δημοσίου δικτύου παροχής ύδατος και την οιαδήποτε άλλη εισροή ή συσσώρευση υδάτων προερχομένων από χώρους εκτός ασφαλιζόμενων κτιρίων ή κτιρίων που περιέχουν τα ασφαλιζόμενα περιουσιακά στοιχεία."

Προτού επιχειρηθεί η οποιαδήποτε ανάλυση είναι απαραίτητο να γίνει μια αναλυτικότερη παρουσίαση των δεδομένων που θα χρησιμοποιηθούν.

4.2 Παρουσίαση Δεδομένων

Αρχικά είναι χρήσιμο να διευκρινιστεί ότι τα δεδομένα των τριών ετών αναλύονται σαν ένα ενιαίο σύνολο δεδομένων και όχι ξεχωριστά για κάθε έτος. Η κατάτμηση τους σε

τρεις ομάδες δεδομένων, ξεχωριστά για το 2002, 2003 και 2004, δημιουργεί πολλά προβλήματα στην εφαρμογή της θεωρίας των ακραίων τιμών και στην εξαγωγή αξιόπιστων αποτελεσμάτων καθώς στην περίπτωση αυτή ο αριθμός των δεδομένων σε κάθε ομάδα είναι αρκετά μικρός. Μια τέτοιου είδους αντιμετώπιση των δεδομένων κρίθηκε σκόπιμη από τον ερευνητή διότι ενώ δημιουργεί τις κατάλληλες προϋποθέσεις για την εφαρμογή της συγκεκριμένης θεωρίας, από την άλλη μεριά δεν αλλοιώνει το σκοπό που εξυπηρετεί μια τέτοιου είδους εργασία.

Πρέπει επίσης να επισημανθεί ότι τα δεδομένα έχουν προσαρμοστεί έτσι ώστε να εκφράζονται σε χιλιάδες ευρώ. Επιπλέον, επειδή τα δεδομένα εκφράζουν τα μεγέθη των αποζημιώσεων τριών διαφορετικών ετών είναι απαραίτητη η προσαρμογή τους στον πληθωρισμό. Τα μεγέθη των αποζημιώσεων αναπροσαρμόστηκαν με βάση ένα μέσο επίπεδο πληθωρισμού της τάξης του 3% για κάθε έτος και έτσι τα δεδομένα που τελικά θα χρησιμοποιηθούν εκφράζονται όλα σε τιμές του 2004, ούτως ώστε τα μεγέθη να είναι συγκρίσιμα.

Η επεξεργασία των δεδομένων έγινε με τη βοήθεια του στατιστικού προγράμματος R. Στο Παράρτημα 1 της εργασίας (σελ.91), παρουσιάζονται αναλυτικά τόσο τα ίδια τα δεδομένα όσο και οι εντολές που χρησιμοποιήθηκαν για την εισαγωγή τους στο συγκεκριμένο πρόγραμμα. Τα δεδομένα παρουσιάζονται σε δύο στήλες με τίτλους “OBS” και “EUDMG”, που ουσιαστικά μεταφράζονται σε “Παρατηρήσεις”, και δίνει τον αριθμό των παρατηρήσεων και “Ζημίες σε Ευρώ”, και δίνει τα αντίστοιχα μεγέθη σε χιλιάδες ευρώ.

Μετά από όλες αυτές τις απαραίτητες διευκρινήσεις προκειμένου να γίνει κατανοητό το είδος το δεδομένων, καθώς επίσης το πλαίσιο και ο τρόπος με τον οποίο θα αναλυθούν, δίνονται κάποια απλά περιγραφικά στατιστικά στοιχεία.

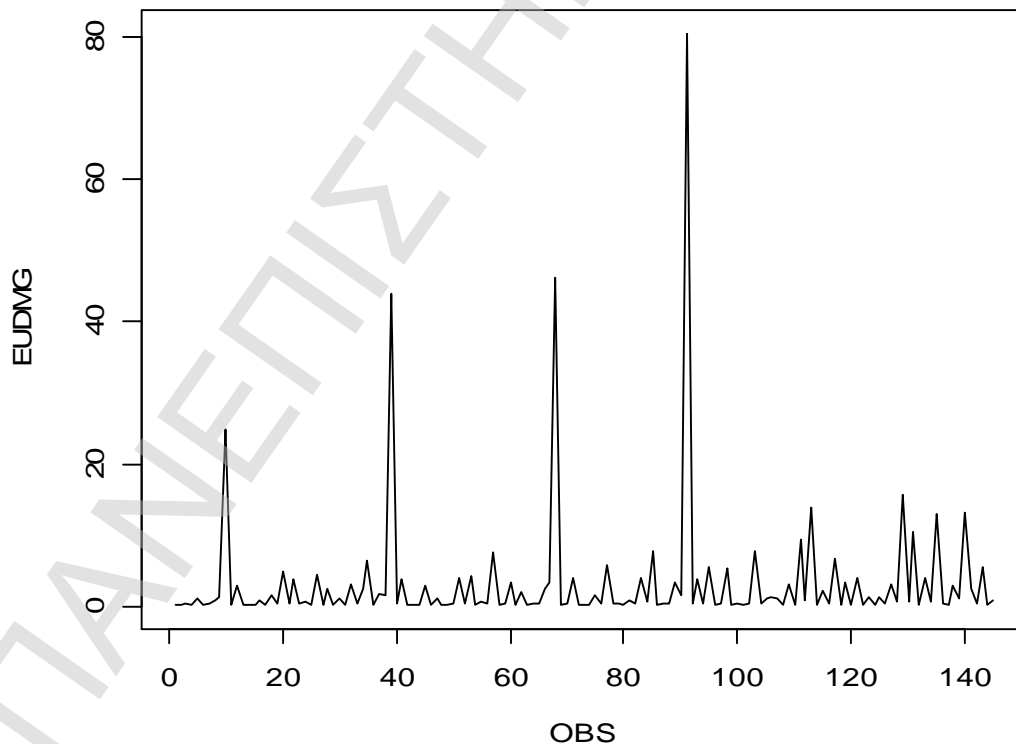
Πίνακας 4-1. Περιγραφικά Στατιστικά στοιχεία των δεδομένων

EUDMG (σε χιλ. ευρώ)			
N	145.000000	Median	0.530000
Mean	3.236421	Q3	3.090000
Std.Dev	8.879416	Max	80.504000
Min	0.127000	Missing values	0.000000
Q1	0.288000		

Όπως προκύπτει, τα δεδομένα αφορούν συνολικά 145 αποζημιώσεις που δόθηκαν από την εταιρεία την τριετία 2002-04 που ξεκινούν από 127 ευρώ (min) και φτάνουν σχεδόν μέχρι τις 81.000 ευρώ (max), ενώ δεν υπάρχουν ελλείπουσες παρατηρήσεις (missing values).

Πέρα από τα παραπάνω στοιχεία δίνονται ακόμα η τυπική απόκλιση (Std.Dev) και η διάμεσος (median) των παρατηρήσεων καθώς επίσης και οι τιμές του πρώτου (Q1) και του τρίτου (Q3) τεταρτημορίου. Αυτό που αξίζει να σχολιασθεί είναι η τιμή του τρίτου τεταρτημορίου που αντιπροσωπεύει την τιμή της μεταβλητής (EUDMG), που είναι τέτοια ώστε το 75% των τιμών των δεδομένων να είναι μικρότερες ή ίσες με αυτήν. Παρατηρείται λοιπόν ότι το 75% των τιμών της μεταβλητής είναι μικρότερο ή ίσο με 3.000 ευρώ τη στιγμή που η μέγιστη παρατήρηση ανέρχεται σχεδόν σε 81.000 ευρώ. Είναι λοιπόν φανερό πόσο ακραία είναι μια τέτοια τιμή για αυτό το σύνολο των δεδομένων.

Επιπλέον παρακάτω δίνεται και το Scatterplot των δεδομένων, προκειμένου ο αναγνώστης να μπορεί να έχει και μια οπτική παρουσίαση της κατανομής των δεδομένων και συνεπώς μια συνολικότερη εικόνα.



Σχήμα 4-1. Scatterplot δεδομένων (EUDMG)

Στον x – άξονα παρουσιάζεται ο αριθμός των παρατηρήσεων, ενώ στον y – άξονα παρουσιάζονται τα μεγέθη των ζημιών σε χιλιάδες ευρώ. Στο παραπάνω διάγραμμα φαίνονται παραστατικά οι ακραίες ζημιές και πόσο πραγματικά πιο ακραίο είναι το μέγεθος τους σε σχέση με το μέγεθος των περισσότερων ζημιών.

Μετά από την παρουσίαση των δεδομένων μαζί με κάποια απαραίτητα περιγραφικά στατιστικά στοιχεία η έρευνα μπορεί να προχωρήσει σε μια πρώτη διερευνητική ανάλυση των δεδομένων.

4.3 Διερευνητική Ανάλυση Δεδομένων

Από τη διερευνητική αυτή ανάλυση θα προκύψει αν όντως η Γενικευμένη Pareto (GPD) είναι η κατάλληλη κατανομή για την περιγραφή της ουράς της κατανομής των δεδομένων.

Αρχικά δίνεται το QQ-plot έναντι της Εκθετικής Κατανομής. Το συγκεκριμένο διάγραμμα είναι ένα πολύ χρήσιμο εργαλείο διότι ουσιαστικά εξετάζει οπτικά την υπόθεση ότι τα δεδομένα προέρχονται από μια εκθετική κατανομή ή κατεπέκταση από μια κατανομή με μεσαίου μεγέθους ουρά.

Για την κατασκευή αυτού του διαγράμματος, στον x – άξονα τοποθετούνται τα ποσοστιαία σημεία της εμπειρικής συνάρτησης κατανομή ενώ στον άξονα των y τοποθετούνται τα ποσοστιαία σημεία της εκθετικής συνάρτησης κατανομής. Με άλλα λόγια,

$$\left\{ \left(X_{k,n}, G_{0,1}^{-1} \left(\frac{n-k+1}{n+1} \right) \right), k = 1, \dots, n \right\},$$

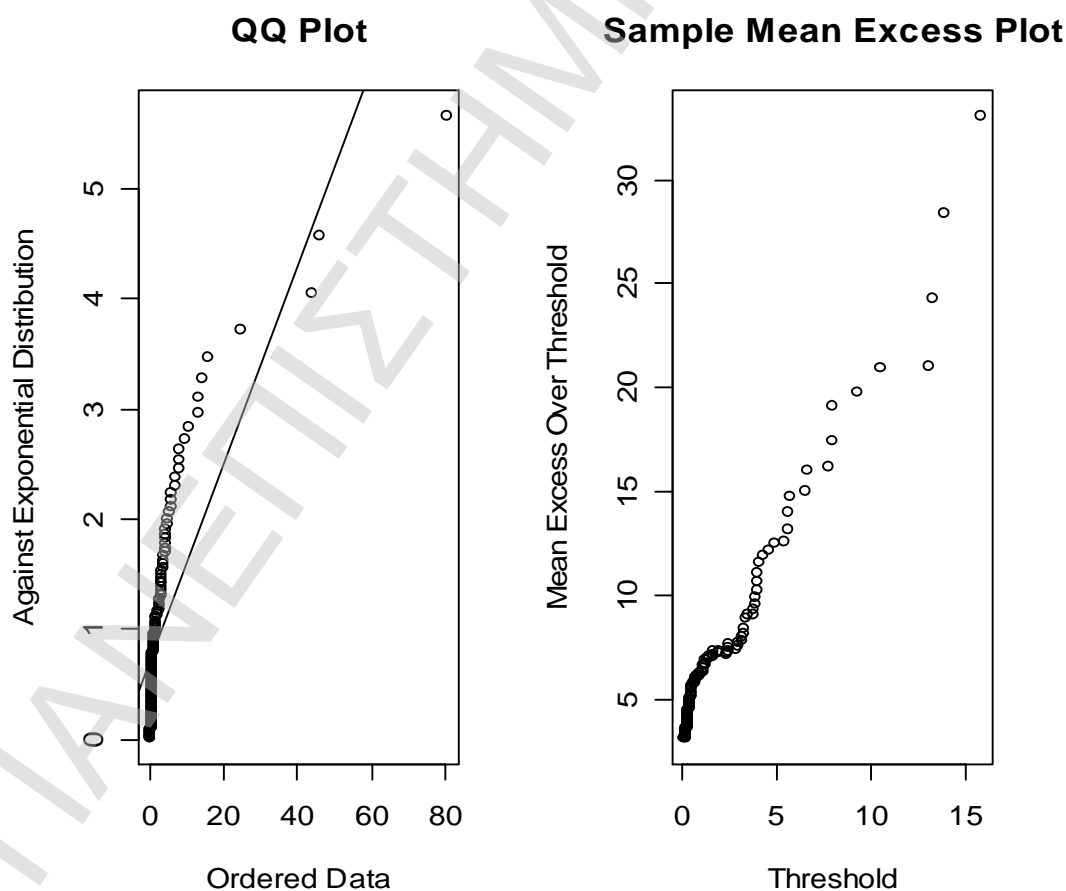
όπου $X_{k,n}$ είναι το k – οστό διατεταγμένο στατιστικό και $G_{0,1}^{-1}$ είναι η αντίστροφη συνάρτηση κατανομής της εκθετικής κατανομής (που είναι μια ειδική περίπτωση της GPD).

Τα σημεία πρέπει να βρίσκονται προσεγγιστικά πάνω σε μια ευθεία γραμμή αν τα δεδομένα προέρχονται από ένα iid δείγμα μια Εκθετικής Κατανομής. Μια κοίλη απόκλιση από την ιδεατή ευθεία γραμμή αποτελεί ένδειξη ότι τα δεδομένα προέρχονται από κάποια

κατανομή με πιο βαριά ουρά, ενώ μια κυρτή απόκλιση από την ιδεατή ευθεία γραμμή αποτελεί ένδειξη ότι τα δεδομένα προέρχονται από κάποια κατανομή με πιο κοντή ουρά.

Στο συγκεκριμένο διάγραμμα είναι φανερό ότι υπάρχει μια κοίλη απόκλιση των σημείων από την ευθεία, πράγμα που αποτελεί μια πρώτη ένδειξη ότι χρειάζεται μια κατανομή με πιο βαριά ουρά, όπως η GPD, για την περιγραφή των δεδομένων.

Θα πρέπει όμως να διευκρινιστεί ότι κάποιες φορές ακόμα και δεδομένα που προέρχονται από την εκθετική κατανομή παρουσιάζουν αποκλίσεις από μια τυπική εκθετική συμπεριφορά. Γενικά θα μπορούσε να ειπωθεί ότι όσο μεγαλύτερος είναι ο αριθμός των δεδομένων, τόσο πιο ασφαλή είναι τα συμπεράσματα που προκύπτουν από το QQ-plot. Στην προκειμένη περίπτωση ο αριθμός των 145 δεδομένων δίνει μια σχετική ασφάλεια στο συμπέρασμα ότι τα δεδομένα ακολουθούν μια κατανομή με βαριά ουρά, αλλά σίγουρα μια περαιτέρω διερεύνηση των δεδομένων θα οδηγήσει σε ασφαλέστερα συμπεράσματα.



Σχήμα 4-2. Q-Q Plot έναντι της Εκθετικής Κατανομής και Sample Mean Excess Plot

Το εργαλείο που θα χρησιμοποιηθεί προκειμένου να γίνει μια πιο εμπειριστατωμένη ανάλυση των δεδομένων είναι το Sample Mean Excess Plot, όπως αυτό έχει περιγραφεί στο Κεφάλαιο 2 της εργασίας. Το Sample Mean Excess Plot είναι ουσιαστικά η γραφική απεικόνιση της δειγματικής συνάρτησης των μέσων υπερβαλουσών τιμών πάνω από κάποιο κατώφλι u .

Στο παραπάνω διάγραμμα έχουν παραληφθεί τέσσερα επιπλέον σημεία στα οποία παρατηρήθηκε μια ακανόνιστη συμπεριφορά.

Αν τα σημεία δείχνουν μια ανοδική τάση, τότε αυτό αποτελεί μια ένδειξη συμπεριφοράς βαριάς ουράς. Δεδομένα που είναι εκθετικά κατανομημένα θα έδιναν μια σχεδόν οριζόντια γραμμή, ενώ δεδομένα από μια κατανομή με κοντή ουρά θα είχαν καθοδική τάση. Πιο συγκεκριμένα, αν το γράφημα των σημείων φαίνεται να ακολουθεί προσεγγιστικά μια ευθεία γραμμή με θετική κλίση πάνω από κάποια τιμή του u , τότε αυτό είναι μια ένδειξη ότι τα δεδομένα ακολουθούν μια κατανομή Γενικευμένης Pareto (GPD), με παράμετρο σχήματος που έχει θετική τιμή στην περιοχή της ουράς πάνω από την τιμή του κατωφλιού u .

Από το διάγραμμα είναι φανερό ότι τα σημεία παρουσιάζουν μια καθαρά ανοδική τάση, πράγμα που δείχνει συμπεριφορά κατανομής με βαριά ουρά. Το συμπέρασμα αυτό είναι σε απόλυτη συμφωνία με το συμπέρασμα που προέκυψε από το QQ-plot πράγμα που δίνει μεγαλύτερη ακόμα ασφάλεια ότι τα δεδομένα προέρχονται από κάποια κατανομή με βαριά ουρά.

Επιπλέον, στο παραπάνω διάγραμμα φαίνεται ότι για τιμές κατωφλιών πάνω από 2 ή 3 περίπου, τα σημεία βρίσκονται προσεγγιστικά πάνω σε μια ευθεία σε πάρα πολύ ικανοποιητικό βαθμό. Αυτό λοιπόν αποτελεί ένδειξη ότι πάνω από κατώφλι με τιμή 2 ή 3 η GPD θα δώσει μια πολύ καλή προσαρμογή στα δεδομένα με θετική παράμετρο σχήματος. Αυτό φυσικά αποτελεί απλά μια ένδειξη και θα χρειαστεί επιπλέον διερεύνηση για την επιλογή του κατάλληλου κατωφλιού. Γενικά πάντως ίσως η GPD να μπορεί να προσαρμοστεί σχετικά καλά στο σύνολο των δεδομένων μιας και ολόκληρο το διάγραμμα των σημείων δεν δείχνει σημαντική απόκλιση από την ευθεία.

Στο Παράρτημα 2 (σελ. 93) δίνονται οι εντολές που χρησιμοποιήθηκαν για την κατασκευή των Q-Q Plot και Sample Mean Excess Plot.

4.4 Επιλογή Κατάλληλου Κατωφλιού

Όπως έχει ήδη αναφερθεί, ένα από τα δυσκολότερα πρακτικά προβλήματα, ίσως το πιο δύσκολο, στην εφαρμογή της μεθόδου των υπερβαλουσών τιμών (POT) πάνω από κάποιο κατώφλι u , είναι η επιλογή του κατάλληλου κατωφλιού πάνω από το οποίο η Κατανομή της Γενικευμένης Pareto δίνει καλή προσαρμογή στα δεδομένα. Όπως αναφέρει και ο Menkens O.,(2004), το κατώφλι u πρέπει ουσιαστικά να επιλεγεί με τέτοιο τρόπο που να εξασφαλίζει τη χρυσή τομή ανάμεσα στη μεροληψία και το σφάλμα των εκτιμητών.

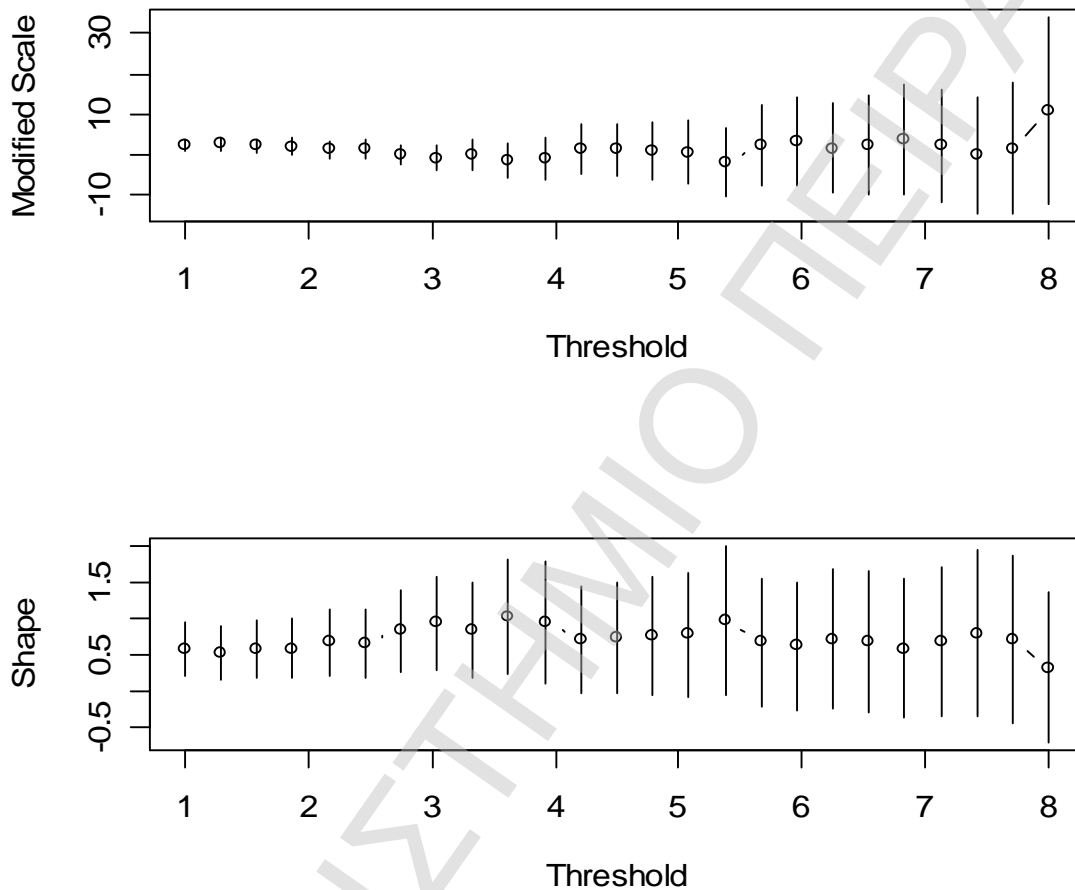
Στο Κεφάλαιο 2 αναφέρθηκε ότι οι πιο γνωστές μέθοδοι που έχουν αναπτυχθεί στη διεθνή βιβλιογραφία για την επιλογή του κατάλληλου κατωφλιού είναι οι εξής: Threshold Choice plot (tcplot), Mean Residual Life plot (mrlplot), L-Moments plot (lmomplot), Dispersion Index plot (diplot). Όλες οι παραπάνω μέθοδοι χρησιμοποιούν την ασυμπτωτική προσέγγιση που ορίζεται από τη συνάρτηση $\Pr[X < y | X > u] \rightarrow H(y)$, καθώς το $u \rightarrow x_0$. Με άλλα λόγια επιλέγεται κάποιο κατώφλι για το οποίο η ασυμπτωτική κατανομή H είναι μια καλή προσέγγιση.

Οι μέθοδοι που θα χρησιμοποιηθούν στην παρούσα εργασία είναι το Threshold Choice plot (tcplot) και το Mean Residual Life plot (mrlplot). Το L-Moments plot (lmomplot) δεν θα χρησιμοποιηθεί διότι έχει φτωχή προσαρμογή στα συγκεκριμένα δεδομένα και έτσι μπορεί να οδηγήσει σε λανθασμένα συμπεράσματα, ενώ το Dispersion Index plot (diplot) έχει κυρίως εφαρμογή σε δεδομένα χρονολογικών σειρών.

Για την κατασκευή των διαγραμμάτων μέσα από το πρόγραμμα R επιλέχθηκε το εύρος των κατωφλιών να κυμαίνεται από το 1 μέχρι το 8. Πρακτικά όμως ακόμα και κατώφλια από το 5 και πάνω δεν είναι πολύ καλό να χρησιμοποιηθούν, ακόμα και αν φαίνονται κατάλληλα διότι ο αριθμός των παρατηρήσεων είναι πάρα πολύ μικρός και μια τέτοια επιλογή αυξάνει πολύ τη διακύμανση των εκτιμητών. Από την άλλη μεριά κατώφλια κάτω από το 1 είναι πάρα πολύ μικρά (περιλαμβάνουν σχεδόν το 50% των παρατηρήσεων), και θέτουν σε κίνδυνο την ασυμπτωτική συνθήκη που απαιτεί υψηλά κατώφλια.

Από τα παρακάτω διαγράμματα (tcplot) φαίνεται ότι μια επιλογή κατωφλιού κοντά στο 3 ή 4 είναι η ενδεδειγμένη. Επειδή όπως αναφέρθηκε και στη αρχή του παρόντος

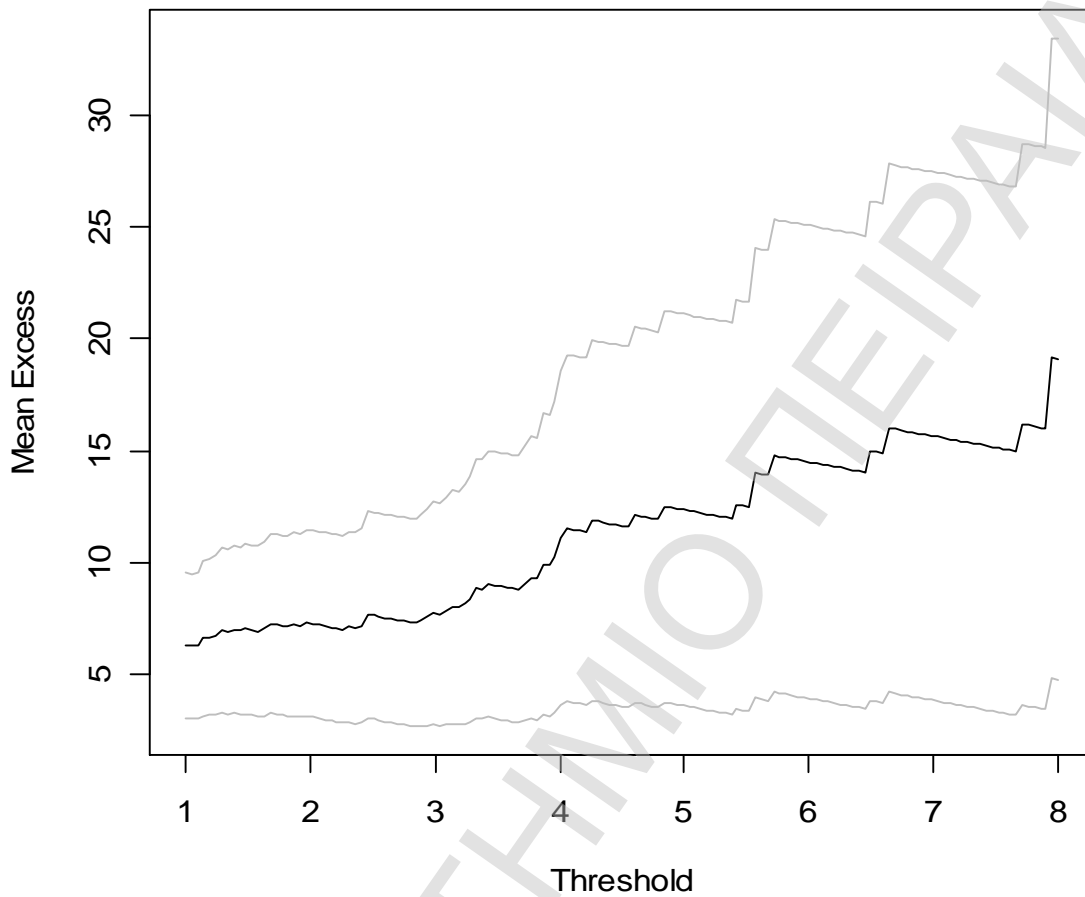
κεφαλαίου ο συνολικός αριθμός των δεδομένων δεν είναι πάρα πολύ μεγάλος, βαρύνουσα σημασία στην τελική επιλογή του κατωφλιού πέρα από τα διαγράμματα θα έχει και ο αριθμός των δεδομένων που υπάρχει πάνω από κάποιο κατώφλι.



Σχήμα 4-3. *Threshold Selection Plot (tcplot)*

Μετά την κατασκευή του tcplot ενδιαφέρον παρουσιάζει και η εξέταση του mrlplot. Και στην περίπτωση αυτή το εύρος των κατωφλιών που επιλέχθηκε να εξεταστεί είναι από 1 έως 8. Παρόλο που και σε αυτό το διάγραμμα κατώφλια της τάξης του 3 ή 4 φαίνονται ότι είναι συνολικά η καλύτερη επιλογή, ίσως η τιμή 4 να παρουσιάζεται ελαφρώς καλύτερη μιας και το γράφημα ομαλοποιείται σε μεγάλο βαθμό από την τιμή αυτή και πάνω.

Mean Residual Life Plot



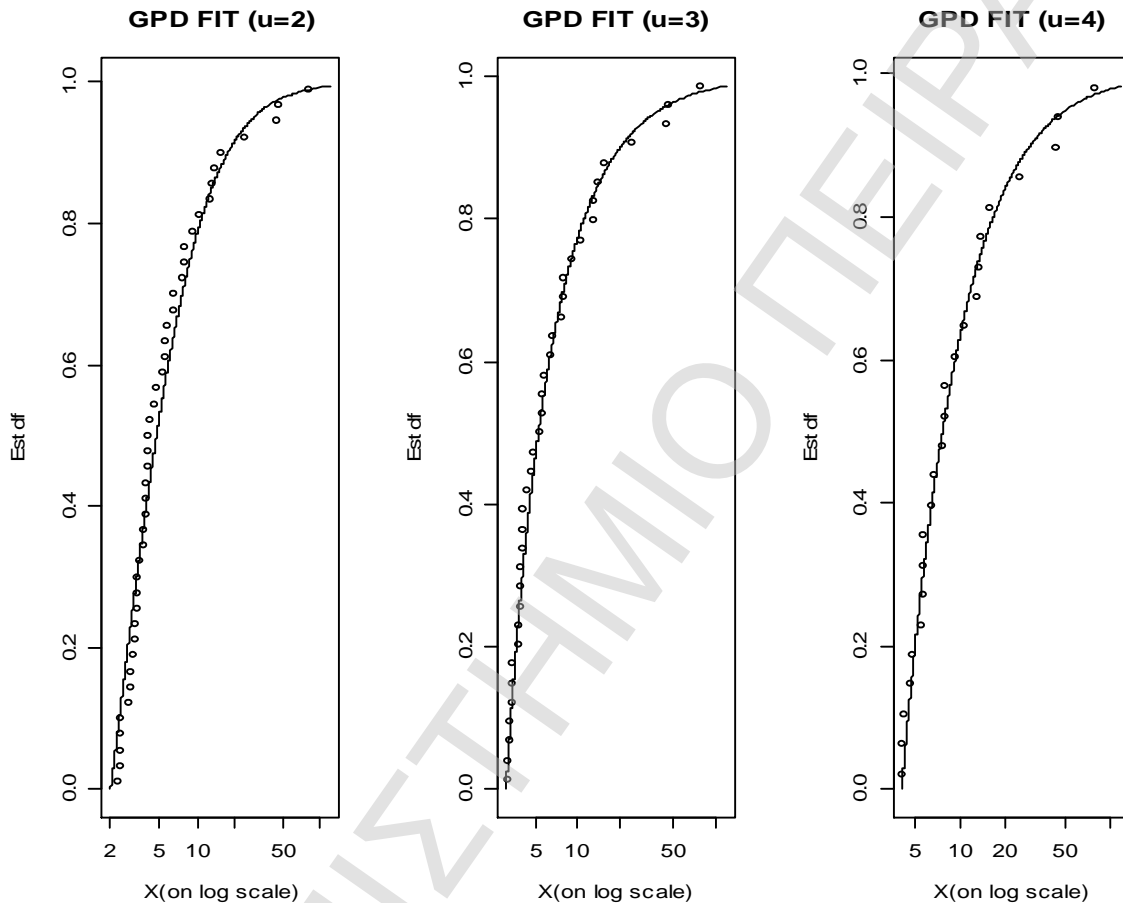
Σχήμα 4-4. Mean Residual Life Plot (mrlplot)

Στο Παράρτημα 2 (σελ.93) δίνονται οι εντολές που χρησιμοποιήθηκαν για την κατασκευή των Threshold Choice Plot (tcplot) και Mean Residual Life Plot (mrlplot).

4.5 Προσαρμογή της GPD

Επειδή από την ανάλυση που προηγήθηκε η εικόνα δεν είναι τελείως ξεκάθαρη για το πιο κατώφλι είναι το καταλληλότερο, θα εξεταστεί η προσαρμογή της GPD για $u = 2, 3$ και 4 . Και στις τρεις περιπτώσεις η τιμή της παραμέτρου θέσεως είναι η τιμή του κατωφλιού ($\mu = u$). Στη συνέχεια αφού εκτιμηθούν οι παράμετροι κλίμακας και σχήματος με τη μέθοδο της

μέγιστης πιθανοφάνειας γίνεται το διάγραμμα της προσαρμογής της GPD στην εμπειρική συνάρτηση κατανομής των υπερβαλουσών τιμών. Οι εκτιμητές αυτοί είναι εκτιμητές της υπό συνθήκη κατανομής των απωλειών, δοθέντος ότι οι απώλειες υπερβαίνουν το κατώφλι u .



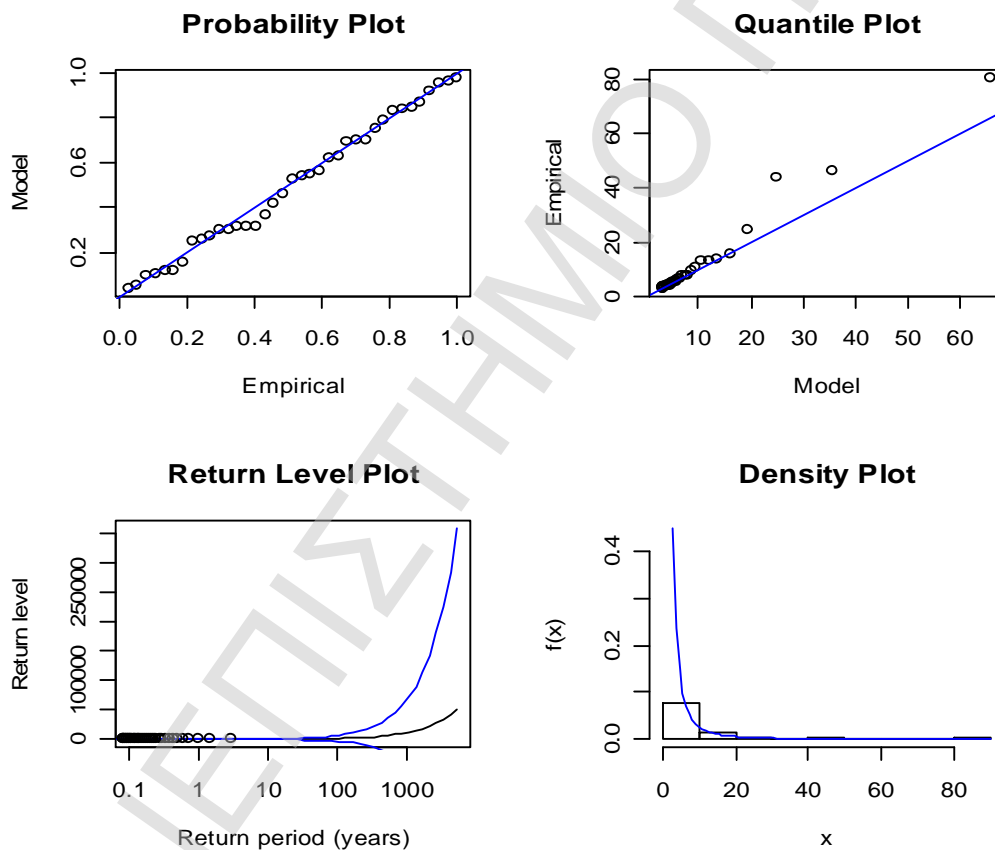
Σχήμα 4-5. Για $u=2$ η GPD έχει προσαρμοστεί σε 45 παρατηρήσεις ενώ οι εκτιμητές των παραμέτρων είναι $\xi=0.604$, $\mu=2.0$, $\sigma=3.13$. Για $u=3$ η GPD έχει προσαρμοστεί σε 37 παρατηρήσεις ενώ οι εκτιμητές των παραμέτρων είναι $\xi=0.90$, $\mu=3.0$, $\sigma=2.23$. Για $u=4$ η GPD έχει προσαρμοστεί σε 24 παρατηρήσεις ενώ οι εκτιμητές των παραμέτρων είναι $\xi=0.805$, $\mu=4.0$, $\sigma=3.8$.

Σχολιάζοντας κανείς και τα τρία διαγράμματα θα μπορούσε να πει ότι η προσαρμογή της GPD για $u=3$ παρουσιάζεται καθαρά ως η καλύτερη από τις τρεις περιπτώσεις μιας και τα σημεία της εμπειρικής συνάρτησης κατανομής συνολικά αλλά κυρίως στην περιοχή της ουράς που μας ενδιαφέρει περισσότερο, βρίσκονται πιο κοντά στην καμπύλη της GPD, μαρτυρώντας καλύτερη προσαρμογή. Επιπλέον αξίζει να σχολιασθεί ότι η περίπτωση $u=4$

ακόμα και αν παρουσιαζόταν να έχει την καλύτερη προσαρμογή ίσως να ήταν παρακινδυνευμένο να επιλεγεί μιας και ο αριθμός των παρατηρήσεων πάνω από το κατώφλι αυτό είναι σημαντικά μικρός (24 μόλις παρατηρήσεις).

Στο Παράρτημα 3 (σελ.94) παρουσιάζονται οι εντολές που χρησιμοποιήθηκαν για την κατασκευή των παραπάνω διαγραμμάτων.

Μετά την επιλογή του κατάλληλου κατώφλιου, ακολουθεί η εξέταση κάποιων διαγνωστικών διαγραμμάτων για την προσαρμογή της GPD πάνω από το κατώφλι $u = 3$.



Σχήμα 4-6. Probability Plot, Quantile Plot, Return Level Plot και Density Plot για $u = 3$

Το probability plot φανερώνει μια πάρα πολύ καλή προσαρμογή των δεδομένων, καθώς συνολικά τα σημεία δεν αποκλίνουν σχεδόν καθόλου από την ευθεία. Επίσης, και το quantile plot δείχνει μια αρκετά καλή προσαρμογή παρόλο που κάποια σημεία αποκλίνουν

κάπως από την ευθεία. Το return level plot δίνει μια ιδέα για το αναμενόμενο επίπεδο επιστροφής για κάθε περίοδο επιστροφής.

Στο Παράρτημα 5 (σελ.101) δίνονται κάποιες πληροφορίες που αυτόματα το πρόγραμμα παρέχει για την προσαρμογή της GPD πάνω από το κατώφλι u .

Αναφορικά με τους εκτιμητές πρέπει να επισημανθεί ότι η μέθοδος που υιοθετείται στην παρούσα εργασία είναι αυτή της Μέγιστης Πιθανοφάνειας. Οι Hoskin & Wallis (1987) αναφέρουν ότι για δεδομένα που προέρχονται από μια GPD με παράμετρο σχήματος $0 \leq \xi \leq 0.4$ και ειδικά για μικρά δείγματα, η μέθοδος *rwmm* πλεονεκτεί έναντι της μεθόδου της μέγιστης πιθανοφάνειας αφού οι *rwmm* εκτιμητές παρουσιάζουν μικρότερο τετραγωνικό σφάλμα. Παρόλα αυτά όσο το δείγμα μεγαλώνει οι διαφορές αυτές γίνονται λιγότερο σημαντικές. Από την άλλη μεριά, οι Rootzen & Tajvidi (1996) αναφέρουν ότι για δεδομένα με βαριά ουρά με $\xi \geq 0.5$ η μέθοδος *rwmm* δίνει σοβαρά μεροληπτικούς εκτιμητές των παραμέτρων, ενώ η μέθοδος της μέγιστης πιθανοφάνειας δίνει συνεπείς εκτιμητές. Επειδή λοιπόν και στην παρούσα εργασία αναλύονται ασφαλιστικά δεδομένα με βαριά ουρά προτιμάται η μέθοδος της μέγιστης πιθανοφάνειας.

Παρόλα αυτά στον παρακάτω πίνακα δίνονται οι εκτιμήσεις των παραμέτρων με όλες τις μεθόδους που είναι διαθέσιμες από το πρόγραμμα R (για $u = 3$), ούτως ώστε να μπορεί να γίνει κάποια σύγκριση. Δίνονται επίσης και τα σφάλματα των εκτιμήσεων σε όσες περιπτώσεις αυτά είναι διαθέσιμα από το πρόγραμμα.

Πίνακας 4-2. Οι εκτιμητές των παραμέτρων με τα αντίστοιχα σφάλματα εκτίμησης

Method	Scale	Std Error	Shape	Std Error
<i>mle</i>	2.2324	0.7311	0.8980	0.3196
<i>mom</i>	4.8414	3.7743	0.3737	0.2560
<i>Pickands'</i>	2.665	-	0.708	-
<i>med</i>	2.6596	-	0.7144	-
<i>mdpd</i>	2.1633	-	0.9629	-
<i>rwmb</i>	2.7587	0.4839	0.6431	-
<i>rwmu</i>	2.5888	0.4776	0.6651	-

Στο Παράρτημα 4 (σελ. 95), δίνονται οι εντολές που χρησιμοποιήθηκαν για τον υπολογισμό των εκτιμητών που παρουσιάζονται στον παραπάνω πίνακα, καθώς και όλες οι πληροφορίες που δίνει αυτόματα το πακέτο για κάθε μέθοδο εκτίμησης.

Έτσι λοιπόν η μέθοδος της μέγιστης πιθανοφάνειας δίνει εκτιμητές $(\hat{S}, \hat{x}) = (2.2324, 0.8980)$, με log-likelihood περίπου -100 και πίνακα διακυμάνσεων – συνδιακυμάνσεων:

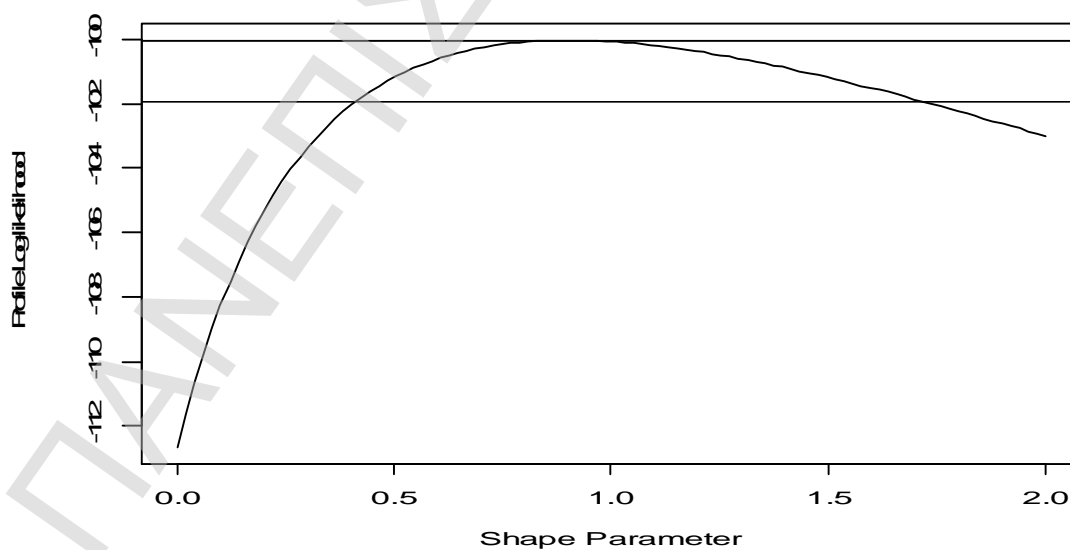
Variance-Covariance Matrix

Parameter covariance:

	[1]	[2]
[1,]	0.5366081	-0.1248071
[2,]	-0.1248071	0.1017811

Ένα 95% Διάστημα Εμπιστοσύνης για την ξ είναι: $0.8980 \pm 1.96 \times 0.3196 = [0.2715, 1.5244]$, το οποίο αποτελεί ισχυρή ένδειξη ότι $\xi > 0$, αφού το 0 δεν περιλαμβάνεται μέσα στο διάστημα εμπιστοσύνης.

Η μέθοδος profile log-likelihood δίνει 95% Διάστημα Εμπιστοσύνης για την παράμετρο ξ που είναι $[0.40926, 1.71933]$.



Σχήμα 4-7. Το profile log-likelihood 95% Διάστημα Εμπιστοσύνης για την παράμετρο ξ .

Το διάστημα εμπιστοσύνης αυτό και δεν διαφέρει σημαντικά από το προηγούμενο και ενισχύει ακόμα περισσότερο την ένδειξη ότι $\xi > 0$.

Τα αποτελέσματα αυτά έρχονται να επιβεβαιώσουν την αρχική ένδειξη που υπήρχε από την ανάλυση του Sample Mean Excess Plot, ότι τα δεδομένα ακολουθούν κάποια κατανομή με θετική παράμετρο σχήματος ξ . Επίσης, δικαιώνουν την επιλογή των εκτιμητών μέγιστης πιθανοφάνειας ως την καλύτερη ως την καλύτερη μέθοδο εκτίμησης των παραμέτρων αφού τα δεδομένα ακολουθούν κατανομή με βαριά ουρά αλλά και ξ σχεδόν μεγαλύτερο από 0.5.

4.6 Μοντέλο Τιμολόγησης & Ευαισθησία Αποτελεσμάτων ανάλογα με την επιλογή του Κατωφλιού u

Έχει ήδη αναφερθεί ότι το μοντέλο ανασφάλισης που θα εξεταστεί είναι το μοντέλο υπερβάλλοντος ποσού ζημιάς όπου οι αποζημιώσεις που καταβάλει ο ανασφαλιστής για κάποιο όριο ίδιας κράτησης d είναι:

$$Z_i = \begin{cases} 0, & X_i \leq d \\ X_i - d, & X_i > d \end{cases}$$

Αν υποθέσουμε ότι υπάρχει κάποιο όριο R μέχρι του οποίου ο ανασφαλιστής καλύπτει τη ζημία τότε:

$$Z_i = \begin{cases} 0, & X_i \leq d \\ X_i - d, & d < X_i < R \\ X_i - R, & X_i \geq R \end{cases}$$

Όμως στην περίπτωση αυτή η τιμή του ασφαλιστρού δίνεται από τη σχέση:

$$P = \int_d^R (x-r) f_X(x) dx + (R-d)(1-F_X(R))$$

όπου $f_X(x) = dF_X(x)/dx$ είναι η συνάρτηση πυκνότητας των απωλειών.

Επιλέγοντας ένα υψηλό κατώφλι u ($< d$) και προσαρμόζοντας το μοντέλο της Γενικευμένης Pareto στις υπερβάλλουσες τιμές, μπορούμε να εκτιμήσουμε την $F_X(x)$ για $x > u$ από τη σχέση $F_X(x) = (1 - F_n(u))G_{x,u,S}(x) + F_n(u)$, όπου \hat{x}, \hat{s} είναι οι εκτιμητές μέγιστης πιθανοφάνειας των παραμέτρων και $F_n(u) = n^{-1} \sum_{i=1}^n I_{\{X_i < u\}}$ είναι ένας εκτιμητής της $\Pr(X < u)$ βασισμένος στην εμπειρική συνάρτηση κατανομής των δεδομένων.

Αφού το d είναι το ποσό από το οποίο και πάνω δημιουργείται υποχρέωση για τον αντασφαλιστή, τότε σίγουρα πρέπει να είναι αρκετά υψηλό. Έτσι για $d = 30$, που ικανοποιεί και την προϋπόθεση $u < d$, και $R = 80$, που είναι ένα ποσό πολύ κοντά στην μεγαλύτερη απώλεια που υπάρχει στα δεδομένα, το P γίνεται:

$$P = \int_{30}^{80} (x - 30) f_X(x) dx + (R - d)(1 - F_X(R))$$

Στον παρακάτω πίνακα δίνονται αναλυτικά οι τιμές του P που προέκυψαν από την προσαρμογή της GPD για διάφορα κατώφλια $u = 0.9, 1, 2, 3$ και 4 .

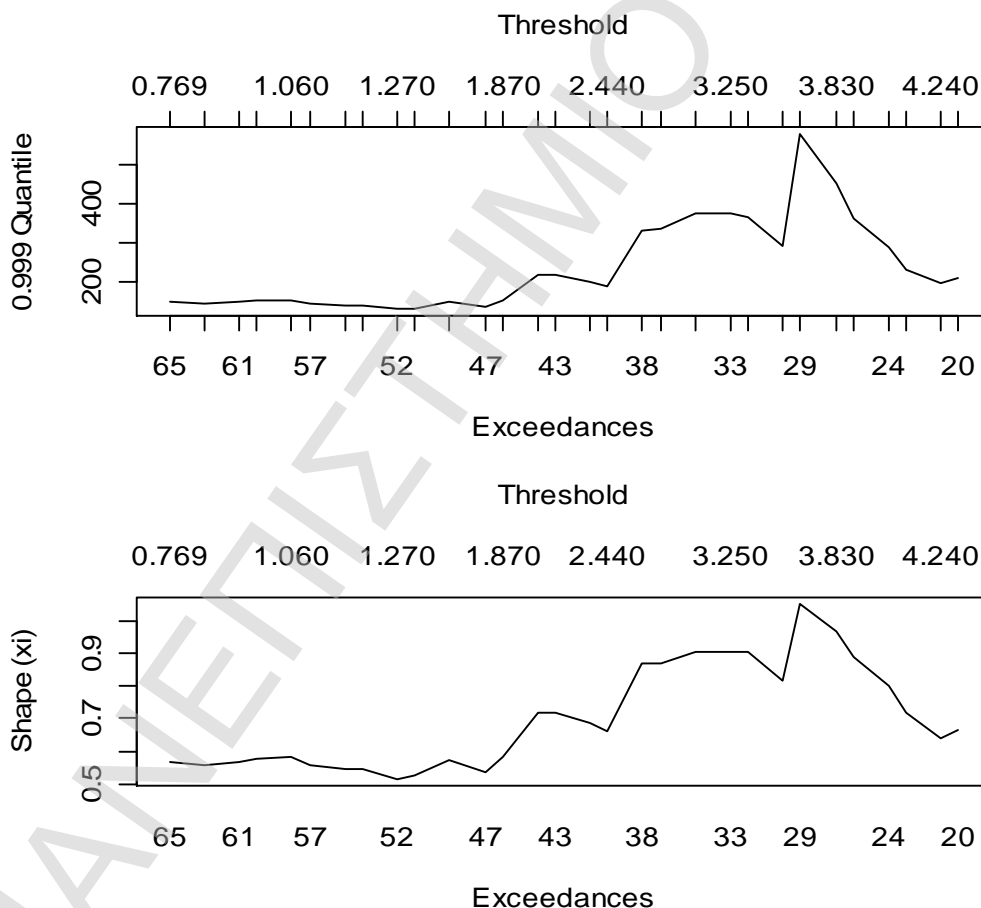
Πίνακας 4-3. Προσαρμογή της GPD για διάφορα κατώφλια με τις αντίστοιχες τιμές της παραμέτρου ξ , των 0.999 και 0.9999 ποσοστιαίων σημείων και των τιμών της P .

Μοντέλο	Κατώφλι	Υπερβαλ. Τιμές	ξ	s.e	.999	.9999	P
GPD	0.9	63	0.59	0.19	160	634	0.34
GPD	1.0	60	0.57	0.19	153	584	0.32
GPD	2.0	45	0.60	0.21	163	664	0.33
GPD	3.0	37	0.89	0.31	365	-	0.46
GPD	4.0	24	0.80	0.40	287	-	0.43

Μάλιστα στον πίνακα αυτό δίνονται εκτός από τις τιμές του P και οι τιμές των εκτιμήσεων της παραμέτρου σχήματος ξ , καθώς επίσης και οι τιμές των εκτιμήσεων των 0.999 και 0.9999 ποσοστιαίων σημείων για τα αντίστοιχα κατώφλια.

Ο λόγος που παρουσιάζονται αυτές οι πληροφορίες είναι για να φανεί η εξάρτηση που υπάρχει ανάμεσα στην τιμή του κατωφλιού που επιλέγεται και στις αντίστοιχες τιμές της παραμέτρου ξ , των ποσοστιαίων σημείων και της τιμής P που προκύπτουν. Αυτό είναι φανερό και από τον πίνακα, αλλά και από τα παρακάτω διαγράμματα.

Στο Παράρτημα 7 (σελ.104) παρουσιάζονται οι εντολές που χρησιμοποιήθηκαν για την κατασκευή αυτών των διαγραμμάτων.



Σχήμα 4-8. Η εξάρτηση που υπάρχει ανάμεσα στην τιμή του κατωφλιού που επιλέγεται και στις αντίστοιχες τιμές της παραμέτρου ξ και των ποσοστιαίων σημείων.

Συγκεκριμένα, και για την παράμετρο ξ αλλά και για τα ποσοστιαία σημεία οι διαφορές αρχίζουν να γίνονται πολύ έντονες από το κατώφλι $u = 2$ και πάνω. Μάλιστα, τα δύο διαγράμματα παρουσιάζουν εντελώς παρόμοια συμπεριφορά.

Γενικά μπορεί να παρατηρήσει κανείς ότι όσο πιο μεγάλη είναι η τιμή του ξ , τόσο πιο βαριά είναι η ουρά της κατανομής και τόσο πιο υψηλές είναι οι τιμές των ασφαλιστρών P και των ποσοστιαίων σημείων που προκύπτουν.

Η απάντηση λοιπόν στο ερώτημα πιο κατώφλι είναι καλύτερο εξαρτάται από το σκοπό που εξυπηρετούν τα εξαγόμενα αποτελέσματα. Για παράδειγμα στα συγκεκριμένα δεδομένα μπορεί να επιλεγεί κατώφλι $u = 1$ που είναι μικρότερο από το 3, που είχε φανεί ως η καλύτερη επιλογή. Το μοντέλο της GPD μπορεί να μην προσαρμόζεται τόσο καλά σε αυτό το χαμηλότερο κατώφλι, όπως συμβαίνει για $u = 3$, αλλά μπορεί να κριθεί ότι είναι καλύτερο διότι λόγοι έντονου ανταγωνισμού επιβάλουν ένα χαμηλό ασφάλιστρο. Από την άλλη μεριά, μπορεί να θεωρηθεί ασφαλέστερο η τιμή του ασφαλιστρου να υπολογιστεί με βάση ένα υψηλότερο κατώφλι.

ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΕΙΡΑΙΑ

Κεφάλαιο 5

Γενικά Συμπεράσματα

5.1 Εισαγωγή

Ο σκοπός του κεφαλαίου αυτού είναι διπλός. Πρώτα απ' όλα στοχεύει στο να παρουσιάσει συνοπτικά τα κυριότερα αποτελέσματα από την ανάλυση των δεδομένων που πραγματοποιήθηκε στο προηγούμενο κεφάλαιο.

Επίσης, επειδή καμία μέθοδος ανάλυσης δεδομένων δεν είναι τέλεια στο κεφάλαιο αυτό θα παρουσιαστούν όλοι οι περιορισμοί που προέκυψαν από την εφαρμογή της μεθόδου και την ανάλυση των συγκεκριμένων δεδομένων που χρησιμοποιήθηκαν.

5.2 Γενικά Συμπεράσματα

Το κυριότερο συμπέρασμα που προκύπτει από την παραπάνω εφαρμογή της Θεωρίας των Ακραίων Τιμών σε πραγματικά δεδομένα είναι η επιβεβαίωση της άποψης του McNeil A. (1997), ότι η επιλογή του κατάλληλου κατωφλιού είναι ίσως το δυσκολότερο πρακτικό πρόβλημα που υπάρχει και που θέτει σε κίνδυνο τη μέθοδο που έχει εφαρμοστεί.

Η συμπερασματολογία στην περίπτωση αυτή είναι άμεσα συνυφασμένη με την επιλογή του κατωφλιού. Όμως, τα συμπεράσματα που εξάγονται εξαρτώνται και από τις τιμές που έχουν οι μεγαλύτερες σε μέγεθος παρατηρήσεις των δεδομένων που αναλύονται.

Επιστρέφοντας στο μοντέλο που είχε επιλέγει ως καλύτερο κατά τη διάρκεια της ανάλυσης και που αντιστοιχούσε σε κατώφλι με τιμή 3, παρατηρεί κανείς ότι το 0.999 ποσοστιαίο σημείο έχει τιμή 365. Όπως φαίνεται από το δεύτερο μοντέλο του παρακάτω πίνακα, απαλείφοντας την μεγαλύτερη παρατήρηση (80.504) από το σετ των δεδομένων (και έχοντας τώρα 36 υπερβάλλουσες τιμές πάνω από το κατώφλι $u = 3$) το 0.999 ποσοστιαίο σημείο μειώνεται σε 174. Η παράμετρος σχήματος ξ μειώνεται από 0.89 σε 0.73. Επίσης, σημαντική είναι και η μείωση της τιμής P του ασφαλίστρου.

Πίνακας 5-1. Υποθετικά Μοντέλα με συν 1 και πλην 1 παρατήρηση πάνω από το κατώφλι u σε σχέση με το Πραγματικό μοντέλο.

Υποθετικά Μοντέλα							
Μοντέλο	Κατώφλι	Υπερβαλ. Τιμές	ξ	s.e	.999	.9999	P
GPD	3.0	38	1.14	0.34	1.076	-	0.74
GPD	3.0	36	0.73	0.30	174	-	0.28

Από την άλλη μεριά, αν στο μοντέλο που προαναφέρθηκε ($u = 3$) εισαχθεί μια νέα παρατήρηση, μεγαλύτερη από τις ήδη υπάρχουσες, πάλι φαίνεται να επηρεάζονται οι εκτιμήσεις της ξ , των ποσοστιαίων σημείων και του P.

Πιο συγκεκριμένα στο μοντέλο εισήχθη μια νέα παρατήρηση με τιμή 261.300 που είναι κατά πολύ μεγαλύτερη από τις ήδη υπάρχουσες. Στην περίπτωση αυτή η τιμή της παραμέτρου ξ αλλά κυρίως η τιμή του 0.999 ποσοστιαίου σημείου αυξάνονται σε σημαντικό βαθμό. Εξίσου αξιόλογη είναι και η αύξηση στην τιμή του ασφαλιστηρίου συμβολαίου.

Στο Παράρτημα 6 (σελ. 102) παρουσιάζονται οι εντολές για τις εκτιμήσεις των ποσοστιαίων σημείων των Υποθετικών Μοντέλων.

5.3 Περιορισμοί Έρευνας

Ο πρώτος περιορισμός της έρευνας αυτής σχετίζεται με την ποιότητα των διαθέσιμων δεδομένων. Μιλώντας για ποιότητα δεδομένων θα πρέπει να γίνει λόγος τόσο για το πλήθος των δεδομένων, όσο και για χρονική έκταση που καλύπτουν τα δεδομένα.

Η έλλειψη λοιπόν μακροχρόνιων δεδομένων περιόρισε σημαντικά τις δυνατές μεθόδους ανάλυσης που ήταν δυνατόν να χρησιμοποιηθούν στην παρούσα εργασία. Είναι φανερό ότι σε δεδομένα τριών μόνο ετών, έστω και αν αυτά αναλύονταν ξεχωριστά για κάθε έτος, δεν μπορεί να γίνει ανάλυση μέσω της GEV. Επιπλέον, δεν είναι δυνατή ούτε η εφαρμογή των Μεθόδων του r -μεγαλύτερου διατεταγμένου δείγματος, ούτε και της μεθόδου Point Process. Άρα ο μόνος τρόπος να αναλυθούν τα δεδομένα ήταν να ομαδοποιηθούν και

να αναλυθούν συνολικά μέσω της μεθόδου των Υπερβαλουσών Τιμών, της μόνης μεθόδου που μπορούσε να εφαρμοστεί.

Η έλλειψη μεγάλου αριθμού δεδομένων καθιστά τη συμπερασματολογία της έρευνας σχετικά με την ουρά της κατανομής πιο αβέβαιη αφού τα δεδομένα που τελικά χρησιμοποιούνται για την εξαγωγή συμπερασμάτων είναι λίγα. Γενικά έχει παρατηρηθεί ότι ένα δείγμα 1000 παρατηρήσεων είναι επαρκές για την εξαγωγή συμπερασμάτων για τη μέση τιμή της κατανομής με τη χρήση του κεντρικού οριακού θεωρήματος. Επιπλέον, όπως φάνηκε από την ανάλυση, η συμπερασματολογία επηρεάζεται σε σημαντικό βαθμό από την τιμή των μεγαλύτερων παρατηρήσεων καθώς η εισαγωγή νέων ακραίων απωλειών στα δεδομένα επηρεάζει τα εξαγόμενα συμπεράσματα.

Ένας δεύτερο πεδίο αμφισβήτησης στην συγκεκριμένη έρευνα μπορεί να είναι η κατάλληλη επιλογή του ύψους του κατωφλιού u . Παρόλο που έγινε προσπάθεια να ληφθούν υπόψη όλες οι παράμετροι που σχετίζονται με την επιλογή του ιδανικού κατωφλιού, το κατώφλι που τελικά επιλέχθηκε μπορεί να είναι πολύ υψηλό οπότε να υπάρχει μεγάλη διακύμανση των εκτιμητών ή πιο μικρό από όσο θα έπρεπε οπότε υπάρχει μεγάλη μεροληψία. Όπως προέκυψε και από τη συγκεκριμένη μελέτη, τα εξαγόμενα συμπεράσματα επηρεάζονται από την επιλογή του κατωφλιού.

Όμως, ακόμα και αν τα δεδομένα της έρευνας καθώς και τα χρησιμοποιούμενα μοντέλα είναι πάρα πολύ καλά πάντα θα υπάρχουν σφάλματα στην εκτίμηση των παραμέτρων. Μάλιστα, αλλαγές στις τιμές των παραμέτρων, ειδικά της παραμέτρου ξ , επηρεάζουν και τη συμπερασματολογία.

Επιπλέον τα δεδομένα που χρησιμοποιήθηκαν θεωρήθηκαν ως iid τυχαίες μεταβλητές. Παρόλα αυτά στην πράξη μπορεί να υπάρχουν διαφόρων ειδών εξαρτήσεις ανάμεσα στα δεδομένα.

ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΕΙΡΑΙΑ

ΒΙΒΛΙΟΓΡΑΦΙΑ

Ελληνική

Ένωση Ασφ/κών Εταιριών Ελλάδος, 1997, Ανάλυση Κινδύνου Πλημμύρας στο Λεκανοπαίδιο Αττικής.

Ξένα

Boulatov A., 2003, The Response of Catastrophe Insurance Markets to Extreme Events: A Real Option Approach.

Brodtkorb, P. A., Johannesson, P., Lindgren, G., Rychlik, I., Rydén, J., Sjö, E., and Sköold, M. (2000), WAFO—A Matlab Toolbox for Analysis of Random Waves and Loads, Lund's Institute of Technology, Department of Mathematical Statistics.

Chavez-Demoulin V., Roehrl A., 2004, Extreme Value Theory can save your neck, Department of Mathematics, University of Zurich.

Chavez-Demoulin V. & Embrechts P., 2004, Advanced Extremal Models for Operational Risk, Department of Mathematics, University of Zurich.

Corradin S., 2002, Economic Risk Capital and Reinsurance: an Extreme Value Theory's Application to Fire Claims of an Insurance Company, RAS Spa.

Corradin S. & Verbrigghe B., 2001, Economic Risk Capital and Reinsurance: an Application to Fire Claims of an Insurance Company.

Dahan E., Mendelson H., 2001, An Extreme-Value Model of Concept Testing, Management Science, Vol. 47, No. 1, pp. 102–116.

Dalle Molle W.J., 2002, Modelling Catastrophic Loss Events Using Extreme Value Theory – Fundamental Results and Applications in Finance and Insurance.

Dlugolecki, A. D., 2001, “Climate Change and the Financial Services Industry”, Insurance Economics, The Geneva Association, pp.18-20.

Embrechts et.al, 1997, Modelling Extremal Events for Insurance and Finance, Springer - Verlag.

Finkenstädt B., Rootzén H., 2004, Extreme Values in Telecommunications, Finance and the Environment, Monographs on Statistics and Applied Probability 99, Chapman and Hall/CRC Press LLC, pp.7, 8, 13, 14, 15.

Guin J., Saxena V., Extreme Losses from Natural Disasters - Earthquakes, Tropical Cyclones and Extratropical Cyclones, Applied Insurance Research Inc., 101 Huntington Ave, Boston, MA 02199.

Han Z., 2003, Doctoral Thesis, Actuarial Modelling of Extremal Events Using Transformed Generalized Extreme Value Distributions and Generalized Pareto Distributions, Ohio State University.

International Federation of Red Cross and Red Crescent Societies, 2000, World Disasters Report, Michael Day and Jonathon Walter editors, Geneva, Switzerland.

James G., 2003, Analysis of Traffic Load Effects on Railway Bridges, Doctoral Thesis, Structural Engineering Division, Royal Institute of Technology, SE-100 44 Stockholm, Sweden, pp. 68, 69, 70, 71.

Kharin V. V. & Zwiers W. F., 2003, Estimating extremes in transient climate change simulations, Canadian Centre for Climate Modelling and Analysis.

Kotz S., Nadarajah S., 2000, Extreme Value Distributions: Theory and Applications, Imperial College Press, pp.2, 3, 4, 5.

Kovacs P., Kunreuther H., 2001, Managing Catastrophic Risk: Lessons from Canada, ICLR Research, Paper Series – No. 13.

Mata A., 2000, General Insurance Convention, Parameter Uncertainty for Extreme Value Distributions.

McNeil J. A., 1996, Estimating the Tails of Loss Severity Distributions using Extreme Value Theory, Department of Mathematics, ETH Zentrum.

McNeil J.A., 1997, Estimating the tail of loss severity distributions using extreme value theory, Department of Mathematics, ETH Zentrum.

McNeil J. A. & Saladin T., 1997, The Peaks over Thresholds Methods for Estimating High Quantiles of Loss Distributions.

McNeil J. A. & Saladi T., 1998 Developing Scenarios for Future Extreme Losses Using the POT Model.

Menkens O., 2004, Extreme Value Theory, CF901 - Week 3/4, Department of Economics, University of Essex.

Reiss D. R., Thomas M., 1996, “Statistical Analysis of Extreme Values”, Documentation for XTREMES software package.

Reiss D. R., Thomas M., 2001, Statistical Analysis of Extreme Values with Applications to Insurance, Finance, Hydrology and other Fields, Birkhäuser Verlag, Second Edition, pp.8, 9, 10, 12, 15, 16, 17, 18, 19 – 21, 23, 24, 25, 26, 27, 28, 29 – 30.

Ribatet M., 2006, A User's Guide to the POT Package (Version 1.1), University of Qu'ebec, Department of Hydrological Statistic.

Swiss Re, 2003, Natural Catastrophes and Reinsurance.

Swiss Re, 2002, Natural Catastrophes – game or dice?

Swiss Re, 2003, Terrorism Risk in Property Insurance and their Insurability after 11 September 2001.

Swiss Re, 2004, Nanotechnology: Small Matter, Many Unknowns.

Smith L. R., 2003, Statistics of Extremes with applications in Environment, Insurance and Finance.

ΠΑΡΑΡΤΗΜΑ 1 (Εισαγωγή και Παρουσίαση Δεδομένων στο R)

Με κόκκινο χρώμα παρουσιάζονται οι εντολές ενώ με μαύρο χρώμα παρουσιάζονται τα εξαγόμενα από το πρόγραμμα R αποτελέσματα.

A) Εισαγωγή Δεδομένων στο Πρόγραμμα:

```
> library(extRemes)
```

```
>"Adjustedr1"<-structure(list(OBS=c(1,2,3,4,5,6,7,8,9,10,
11,12,13,14,15,16,17,18,19,20,21,22,23,24,25,26,27,28,29,
30,31,32,33,34,35,36,37,38,39,40,41,42,43,44,45,46,47,48,
49,50,51,52,53,54,55,56,57,58,59,60,61,62,63,64,65,66,67,
68,69,70,71,72,73,74,75,76,77,78,79,80,81,82,83,84,85,86,
87,88,89,90,91,92,93,94,95,96,97,98,99,100,101,102,103,104,
105,106,107,108,109,110,111,112,113,114,115,116,117,118,119,
120,121,122,123,124,125,126,127,128,129,130,131,132,133,134,
135,136,137,138,139,140,141,142,143,144,145),EUDMG=c(0.162,
0.148,0.530,0.209,1.126,0.234,0.352,0.815,1.235,24.762,0.307,
2.905,0.299,0.136,0.200,0.769,0.219,1.448,0.369,4.836,0.399,
3.828,0.467,0.677,0.137,4.550,0.307,2.440,0.269,1.099,0.262,
3.127,0.332,2.434,6.484,0.268,1.872,1.627,43.921,0.319,3.721,
0.275,0.213,0.271,2.967,0.253,1.000,0.164,0.285,0.472,4.017,
0.395,4.238,0.200,0.659,0.363,7.693,0.165,0.371,3.297,0.263,
1.952,0.169,0.350,0.319,2.418,3.255,46.070,0.158,0.480,4.000,
0.158,0.127,0.306,1.648,0.425,5.715,0.331,0.357,0.170,0.824,
0.332,3.925,0.570,7.926,0.133,0.335,0.328,3.399,1.641,80.504,
0.381,3.764,0.444,5.550,0.138,0.413,5.385,0.141,0.515,0.141,
0.488,7.913,0.324,1.099,1.246,1.176,0.308,3.226,0.231,9.281,
0.933,13.842,0.317,2.308,0.488,6.594,0.285,3.285,0.288,3.927,
0.213,1.270,0.213,1.359,0.472,3.090,0.657,15.730,0.563,10.447,
0.275,4.005,0.668,13.002,0.518,0.272,2.870,1.058,13.189,2.363,
0.425,5.562,0.206,0.936)),.Names=c("OBS","EUDMG"),
row.names=c("1","2","3","4","5","6","7","8","9","10","11","12",
"13","14","15","16","17","18","19","20","21","22","23","24","25",
"26","27","28","29","30","31","32","33","34","35","36","37","38",
"39","40","41","42","43","44","45","46","47","48","49","50","51",
"52","53","54","55","56","57","58","59","60","61","62","63","64",
"65","66","67","68","69","70","71","72","73","74","75","76","77",
"78","79","80","81","82","83","84","85","86","87","88","89","90",
"91","92","93","94","95","96","97","98","99","100","101","102",
"103","104","105","106","107","108","109","110","111","112","113",
"114","115","116","117","118","119","120","121","122","123","124",
"125","126","127","128","129","130","131","132","133","134","135",
"136","137","138","139","140","141","142","143","144","145"),
class="data.frame")
```

B) Παρουσίαση Δεδομένων:

>Adjustedr1

>\$data

OBS	EUDMG	OBS	EUDMG	OBS	EUDMG	OBS	EUDMG
1	0.162	46	0.253	91	80.504	136	0.518
2	0.148	47	1.000	92	0.381	137	0.272
3	0.530	48	0.164	93	3.764	138	2.870
4	0.209	49	0.285	94	0.444	139	1.058
5	1.126	50	0.472	95	5.550	140	13.189
6	0.234	51	4.017	96	0.138	141	2.363
7	0.352	52	0.395	97	0.413	142	0.425
8	0.815	53	4.238	98	5.385	143	5.562
9	1.235	54	0.200	99	0.141	144	0.206
10	24.762	55	0.659	100	0.515	145	0.936
11	0.307	56	0.363	101	0.141		
12	2.905	57	7.693	102	0.488		
13	0.299	58	0.165	103	7.913		
14	0.136	59	0.371	104	0.324		
15	0.200	60	3.297	105	1.099		
16	0.769	61	0.263	106	1.246		
17	0.219	62	1.952	107	1.176		
18	1.448	63	0.169	108	0.308		
19	0.369	64	0.350	109	3.226		
20	4.836	65	0.319	110	0.231		
21	0.399	66	2.418	111	9.281		
22	3.828	67	3.255	112	0.933		
23	0.467	68	46.070	113	13.842		
24	0.677	69	0.158	114	0.317		
25	0.137	70	0.480	115	2.308		
26	4.550	71	4.000	116	0.488		
27	0.307	72	0.158	117	6.594		
28	2.440	73	0.127	118	0.285		
29	0.269	74	0.306	119	3.285		
30	1.099	75	1.648	120	0.288		
31	0.262	76	0.425	121	3.927		
32	3.127	77	5.715	122	0.213		
33	0.332	78	0.331	123	1.270		
34	2.434	79	0.357	124	0.213		
35	6.484	80	0.170	125	1.359		
36	0.268	81	0.824	126	0.472		
37	1.872	82	0.332	127	3.090		
38	1.627	83	3.925	128	0.657		
39	43.921	84	0.570	129	15.730		
40	0.319	85	7.926	130	0.563		
41	3.721	86	0.133	131	10.447		
42	0.275	87	0.335	132	0.275		
43	0.213	88	0.328	133	4.005		
44	0.271	89	3.399	134	0.668		
45	2.967	90	1.641	135	13.002		

\$name

[1] "Adjustedr1.dat"

ΠΑΡΑΡΤΗΜΑ 2 (QQ-Plot και Sample Mean Excess Plot)

A) Εντολές που χρησιμοποιήθηκαν για την κατασκευή των Q-Q Plot και Sample Mean Excess Plot.

```
>library(evir)
>par(mfrow=c(1,2))
> qqplot(adjusted1,labels=FALSE)
> title(xlab="Ordered Data",ylab="Against Exponential Distribution",main="QQ Plot")
> meplot(adjusted1,omit=4,labels=FALSE)
> title(xlab="Threshold",ylab="Mean Excess Over Threshold",main="Sample Mean Excess Plot")
```

B) Εντολές που χρησιμοποιήθηκαν για την κατασκευή των Threshold Choice Plot (tcplot) και Mean Residual Life Plot (mrlplot)

```
>library(POT)
>par(mfrow=c(2,1))
> tcplot(adjusted1,u.range=c(1.000,8.000))
>par(mfrow=c(1,1))
> mrlplot(adjusted1,u.range=c(1.000,8.000))
```

ΠΑΡΑΡΤΗΜΑ 3 (Προσαρμογή της GPD για $u = 2$, $u = 3$, $u = 4$)

Με κόκκινο χρώμα παρουσιάζονται οι εντολές, ενώ με μπλε χρώμα οι εκτιμήσεις των παραμέτρων για $u = 1, 2, 3$ όπως τις έδωσε το πρόγραμμα R.

```
>library(evir)
```

```
>par(mfrow=c(1,3))
```

```
> out2<-gpd(adjusted1,2.000)
```

```
> out3<-gpd(adjusted1,3.000)
```

```
> out4<-gpd(adjusted1,4.000)
```

```
> plot(out2,labels=FALSE)
```

```
"threshold = 2  xi = 0.604  scale = 3.13  location = 2"
```

```
> title(xlab="X(on log scale)",ylab="Est df",main="GPD FIT (u=2)")
```

```
> plot(out3,labels=FALSE)
```

```
"threshold = 3  xi = 0.9  scale = 2.23  location = 3"
```

```
> title(xlab="X(on log scale)",ylab="Est df",main="GPD FIT (u=3)")
```

```
> plot(out4,labels=FALSE)
```

```
"threshold = 4  xi = 0.805  scale = 3.8  location = 4"
```

```
> title(xlab="X(on log scale)",ylab="Est df",main="GPD FIT (u=4)")
```

ΠΑΡΑΡΤΗΜΑ 4 (Εκτιμητές Παραμέτρων)

Με κόκκινο χρώμα παρουσιάζονται οι εντολές ενώ με μαύρο χρώμα παρουσιάζονται τα εξαγόμενα, από το πρόγραμμα R, αποτελέσματα.

1) Method of Moments

```
mom<-fitgpd(adjusted1,3.000,"moments")
```

```
> mom
```

```
Estimator: moments
```

```
Varying Threshold: FALSE
```

```
Threshold Call: 3
```

```
Number Above: 37
```

```
Proportion Above: 0.2552
```

```
Estimates
```

```
scale  shape  
4.8414  0.3737
```

```
Standard Error Type:
```

```
Standard Errors
```

```
scale  shape  
3.7743  0.2560
```

```
Asymptotic Variance Covariance
```

```
scale  shape  
scale  14.24570 -1.01250  
shape -1.01250  0.06551
```

```
Correlation
```

```
scale  shape  
scale  1.000 -1.048  
shape -1.048  1.000
```

```
Optimization Information
```

```
Convergence: NA
```

```
Function Evaluations: NA
```

```
Message: Asymptotic theory assumptions for standard error may not be fulfilled!
```

2) Maximum Likelihood Estimators

```
mle<-fitgpd(adjusted1,3.000,"mle")
```

```
> mle
```

Estimator: MLE

Deviance: 200.0446

AIC: 204.0446

Varying Threshold: FALSE

Threshold Call: 3

Number Above: 37

Proportion Above: 0.2552

Estimates

scale shape

2.2324 0.8980

Standard Error Type: observed

Standard Errors

scale shape

0.7311 0.3196

Asymptotic Variance Covariance

scale shape

scale 0.5345 -0.1248

shape -0.1248 0.1021

Optimization Information

Convergence: successful

Function Evaluations: 24

Gradient Evaluations: 12

3) Pickands' Estimators

```
pickands<-fitgpd(adjusted1,3.000,"pickands")  
> pickands
```

Estimator: pickands

Varying Threshold: FALSE

Threshold Call: 3

Number Above: 37

Proportion Above: 0.2552

Estimates

scale shape

2.665 0.708

Optimization Information

Convergence: NA

Function Evaluations: NA

Message: Estimates are valid

4) Median Estimators

```
med<-fitgpd(adjusted1,3.000,"med")  
> med
```

Estimator: MEDIANS

Varying Threshold: FALSE

Threshold Call: 3

Number Above: 37

Proportion Above: 0.2552

Estimates

scale shape

2.6596 0.7144

Optimization Information

Convergence: successful

Function Evaluations: 39

5) Mean Power Density Divergence

```
mdpd<-fitgpd(adjusted1,3.000,"mdpd")  
> mdpd
```

Estimator: MDPD

Varying Threshold: FALSE

Threshold Call: 3

Number Above: 37

Proportion Above: 0.2552

Estimates

scale	shape
2.1633	0.9629

Optimization Information

Convergence: successful

Function Evaluations: 36

Gradient Evaluations: 31

6) Biased Probability Weighted Moments

```
pwmb<-fitgpd(adjusted1,3.000,"pwmb")
```

```
Warning message: NaNs produced in: sqrt(diag(var.cov))
```

```
> pwmb
```

Estimator: PWMB

Varying Threshold: FALSE

Threshold Call: 3

Number Above: 37
Proportion Above: 0.2552

Estimates
scale shape
2.7587 0.6431

Standard Error Type:
Standard Errors
scale shape
0.4839 NaN

Asymptotic Variance Covariance
scale shape
scale 0.23417 0.10406
shape 0.10406 -0.04287

Correlation
scale shape
scale 1 NaN
shape NaN 1

Optimization Information
Convergence: NA
Function Evaluations: NA

Message: PWMB

7) Unbiased Probability Weighted Moments

```
pwmu<-fitgpd(adjusted1,3.000,"pwmu")
```

```
Warning message: NaNs produced in: sqrt(diag(var.cov))
```

```
> pwmu
```

Estimator: PWMU
Varying Threshold: FALSE

Threshold Call: 3
Number Above: 37
Proportion Above: 0.2552

Estimates

scale	shape
2.5888	0.6651

Standard Error Type:

Standard Errors

scale	shape
0.4776	NaN

Asymptotic Variance Covariance

	scale	shape
scale	0.22807	0.08766
shape	0.08766	-0.03568

Correlation

	scale	shape
scale	1	NaN
shape	NaN	1

Optimization Information

Convergence: NA

Function Evaluations: NA

Message: Asymptotic theory assumptions for standard error may not be fulfilled!

ΠΑΡΑΡΤΗΜΑ 5 (Προσαρμογή Μοντέλου GPD για $u = 3$)

"Threshold = 3"

[1] "Number of exceedances of threshold = 37"

[1] "Exceedance rate (per year) = 93.201724137931"

[1] "Maximum Likelihood Estimates:"

	MLE	Std. Err.
Scale (sigma):	2.2389622	0.7325354
Shape (xi):	0.8980095	0.3190314

[1] "Negative log-likelihood: 100.057549830801"

Parameter covariance:

	[1]	[2]
[1,]	0.5366081	-0.1248071
[2,]	-0.1248071	0.1017811

[1] "Convergence code (see help file for optim): 0"

NULL

Model name: gpd.fit1

ΠΑΡΑΡΤΗΜΑ 6 (Εκτιμήσεις Ποσοστιαίων Σημείων)

Με κόκκινο χρώμα παρουσιάζονται οι εντολές, ενώ με μπλε χρώμα οι προειδοποιήσεις όταν το πρόγραμμα R δεν μπορεί να δώσει εκτιμητές.

A) Πραγματικά Μοντέλα (Πίνακας 4-3)

```
>library(evir)
```

```
> out1<-gpd(adjusted1,0.9)
```

```
> tp1<-tailplot(out1)
```

```
> out2<-gpd(adjusted1,1.000)
```

```
> tp2<-tailplot(out2)
```

```
> out3<-gpd(adjusted1,2.000)
```

```
> tp3<-tailplot(out3)
```

```
> out4<-gpd(adjusted1,3.000)
```

```
> tp4<-tailplot(out4)
```

```
> out5<-gpd(adjusted1,4.000)
```

```
> tp5<-tailplot(out5)
```

```
> gpd.q(tp1,0.999)
```

```
> gpd.q(tp2,0.999)
```

```
> gpd.q(tp3,0.999)
```

```
> gpd.q(tp4,0.999)
```

```
> gpd.q(tp5,0.999)
```

```
> gpd.q(tp1,0.9999)
```

```
> gpd.q(tp2,0.9999)
```

```
> gpd.q(tp3,0.9999)
```

```
> gpd.q(tp4,0.9999)
```

```
Lower CI Estimate Upper CI
```

```
Inf 2894.463 -Inf
```

```
Warning messages:
```

```
1: no non-missing arguments to min; returning Inf
```

```
2: no non-missing arguments to max; returning -Inf
```

```
> gpd.q(tp5,0.9999)
```

```
Lower CI Estimate Upper CI
```

```
Inf 1840.203 -Inf
```

```
Warning messages:
```

```
1: no non-missing arguments to min; returning Inf
```

```
2: no non-missing arguments to max; returning -Inf
```

B) Υποθετικά Μοντέλα (Πίνακας 5-1)

```
>library(evir)
```

```
>outf<-gpd(full,3.000)
```

```
> tpf<-tailplot(outf)
```

```
> gpd.q(tpf,0.999)
```

```
> gpd.q(tpf,0.9999)
```

```
Lower CI Estimate Upper CI
```

```
Inf 14931.58 -Inf
```

```
Warning messages:
```

```
1: no non-missing arguments to min; returning Inf
```

```
2: no non-missing arguments to max; returning -Inf
```

```
> outr2<-gpd(adjusted2,3.000)
```

```
> tpr2<-tailplot(outr2)
```

```
> gpd.q(tpr2,0.999)
```

```
> gpd.q(tpr2,0.9999)
```

```
Lower CI Estimate Upper CI
```

```
Inf 933.9292 -Inf
```

```
Warning messages:
```

```
1: no non-missing arguments to min; returning Inf
```

```
2: no non-missing arguments to max; returning -Inf
```

ΠΑΡΑΡΤΗΜΑ 7 (Κατασκευή Διαγραμμάτων Συσχέτισης της Παραμέτρου ξ και των Ποσοστιαίων Σημείων με την τιμή του Κατωφλιού u)

```
>library(evir)
```

```
>par(mfrow=c(2,1))
```

```
>quant(adjusted1,p=0.999,models=30,start=20,end=65,reverse=TRUE,ci=0,auto.scale=TRUE  
)
```

```
>shape(adjusted1,models=30,start=20,end=65,reverse=TRUE,ci=0,auto.scale=TRUE)
```

ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΕΙΡΑΙΑ