



# «Ανθεκτικοί Εκτιμητές της Μεθόδου Chain Ladder»

---

Τμήμα Στατιστικής και Ασφαλιστικής Επιστήμης  
Μεταπτυχιακό Πρόγραμμα Σπουδών στην Αναλογιστική  
Επιστήμη και Διοικητική Κινδύνου

**ΓΡΗΓΟΡΙΑΔΟΥ ΒΑΣΙΛΙΚΗ**  
**ΜΑΕ 09018**  
**5/12/2011**

# Ευχαριστίες

---

Αρχικά θα ήθελα να ευχαριστήσω τον καθηγητή μου κ. Γεώργιο Πιτσέλη για την στήριξη και την βοήθειά του καθόλη τη διάρκεια της εκπόνησης αυτής της διπλωματικής εργασίας και κυρίως για την ευκαιρία που μου έδωσε να παρουσιάσω ένα μέρος αυτής στο 15<sup>ο</sup> Διεθνές Συνέδριο Insurance, Mathematics and Economics που πραγματοποιήθηκε 14-17 Ιουνίου 2011 στη Τεργέστη της Ιταλίας. Θεωρώ αυτή την εμπειρία ως το σημαντικότερο εφόδιο της ακαδημαϊκής μου πορείας, καθώς είχα τη δυνατότητα να παρουσιάσω τη δουλειά μου μπροστά σε ένα επιστημονικό κοινό και να ξεπεράσω τυχόν φόβους μου.

Ευχαριστώ, επίσης, τους καθηγητές μου κ. Σπυρίδων Βρόντο και κ. Γεώργιο Ψαρράκο που συμμετείχαν στη τριμελή επιτροπή.

Επίσης, θα ήθελα να ευχαριστήσω την οικογένεια και τους φίλους μου για την αμέριστη συμπαράσταση υλική και ψυχολογική που μου προσφέρουν πάντα. Τους χρεωστάω ότι έχω καταφέρει μέχρι τώρα.

# ΠΕΡΙΕΧΟΜΕΝΑ

ΕΥΧΑΡΙΣΤΙΕΣ .....	2
ΠΕΡΙΕΧΟΜΕΝΑ .....	3
ΕΙΣΑΓΩΓΗ .....	6
<b>ΚΕΦΑΛΑΙΟ 1<sup>ο</sup>: ΣΤΑΤΙΣΤΙΚΑ ΜΟΝΤΕΛΑ .....</b>	<b>10</b>
1.1 ΓΕΝΙΚΟ ΓΡΑΜΜΙΚΟ ΜΟΝΤΕΛΟ .....	10
1.1.1 Εκτιμητής Ελαχίστων Τετραγώνων και Ιδιότητες .....	10
1.1.2 Εκτιμώμενες από το μοντέλο τιμές και ιδιότητες .....	11
1.1.3 Ανάλυση Διασποράς για την παλινδρόμηση .....	12
1.2 ΜΟΝΤΕΛΟ ΑΝΑΛΥΣΗΣ ΔΙΑΚΥΜΑΝΣΗΣ 2 ΠΑΡΑΓΟΝΤΩΝ (TWO-WAY MODEL OF ANOVA) .....	13
<b>ΚΕΦΑΛΑΙΟ 2<sup>ο</sup>: ΜΕΤΡΑ ΑΝΘΕΚΤΙΚΟΤΗΤΑΣ .....</b>	<b>16</b>
2.1 ΣΥΝΑΡΤΗΣΗ ΕΠΙΔΡΑΣΗΣ (INFLUENCE FUNCTION, IF) .....	17
2.2 ΣΗΜΕΙΟ ΚΑΤΑΡΡΕΥΣΗΣ (BREAKDOWN POINT, BP) .....	19
2.3 ΕΥΑΙΣΘΗΣΙΑ ΓΕΝΙΚΟΥ ΣΦΑΛΜΑΤΟΣ (GROSS ERROR SENSITIVITY) .....	20
<b>ΚΕΦΑΛΑΙΟ 3<sup>ο</sup>: ΑΝΘΕΚΤΙΚΟΙ ΕΚΤΙΜΗΤΕΣ .....</b>	<b>21</b>
3.1 Μ-ΕΚΤΙΜΗΤΕΣ .....	21
3.1.1 Μονοδιάστατη Περίπτωση .....	21
3.1.2 Πολυπαραμετρική Περίπτωση .....	23
3.1.3 Μοντέλο Γραμμικής Παλινδρόμησης .....	23
3.1.4 Αλγόριθμος Μ-εκτιμητή (Huber & Dutter, 1974) .....	28
3.2 L-ΕΚΤΙΜΗΤΕΣ .....	29
3.2.1 Παλινδρόμηση Ελαχίστων Απολύτων Αποκλίσεων (Least Absolute Deviations, LAD) .....	30
3.2.2 Παλινδρόμηση Ελαχίστων Διαμέσων Τετραγώνων (Least Median Squares, LMS) .....	30
3.2.3 Παλινδρόμηση Ελαχίστων Περικεκομμένων Τετραγώνων (Least Trimmed Squares, LTS) .....	31
<b>ΚΕΦΑΛΑΙΟ 4<sup>ο</sup>: ΕΚΤΙΜΗΣΗ ΑΠΟΘΕΜΑΤΩΝ ΖΗΜΙΩΝ (CLAIMS RESERVING) .....</b>	<b>35</b>
4.1 ΜΕΘΟΔΟΣ CHAIN LADDER .....	37
4.2 ΛΟΓΑΡΙΘΜΙΚΟ ΜΟΝΤΕΛΟ ΠΑΛΙΝΔΡΟΜΗΣΗΣ (VERALL, 1991) .....	39
4.3 ΣΥΝΤΟΜΗ ΕΦΑΡΜΟΓΗ ΛΟΓΑΡΙΘΜΟΚΑΝΟΝΙΚΟΥ ΜΟΝΤΕΛΟΥ .....	42
4.3.1 Εκτίμηση μελλοντικών ατομικών ζημιών .....	44
4.4 ΑΝΘΕΚΤΙΚΟ ΛΟΓΑΡΙΘΜΟΚΑΝΟΝΙΚΟ ΜΟΝΤΕΛΟ ΠΑΛΙΝΔΡΟΜΗΣΗΣ .....	45
Βήμα 1 <sup>ο</sup> : Υπολογισμός ανθεκτικής εκτιμήτριας του $\beta$ .....	45
Βήμα 2 <sup>ο</sup> : Υπολογισμός ανθεκτικού πίνακα συνδιακύμανσης του $\beta$ .....	46
Βήμα 3 <sup>ο</sup> : Υπολογισμός ανθεκτικής Μ-εκτίμησης του $\tilde{\theta}_{ij}$ .....	47

Βήμα 4 <sup>ο</sup> : Υπολογισμός ανθεκτικής Μ-εκτίμησης του $\bar{\theta}_{ij}$ .....	47
4.5 IBNR ΖΗΜΙΕΣ ΚΑΙ ΜΟΝΤΕΛΟ ΑΝΑΛΥΣΗΣ ΔΙΑΚΥΜΑΝΣΗΣ ΔΥΟ ΠΑΡΑΓΟΝΤΩΝ (KREMER,1982).....	48
4.6 ΑΝΘΕΚΤΙΚΗ ΑΝΑΛΥΣΗ ΔΙΑΚΥΜΑΝΣΗΣ ΔΥΟ ΠΑΡΑΓΟΝΤΩΝ .....	50
4.6.1 Εφαρμογή του LAD Εκτιμητή στο Γενικό Μοντέλο Ανθεκτικής ANOVA .....	51
<b>ΚΕΦΑΛΑΙΟ 5<sup>ο</sup>: ΕΦΑΡΜΟΓΗ.....</b>	<b>53</b>
5.1 ΕΦΑΡΜΟΓΗ ΛΟΓΑΡΙΘΜΙΚΟΥ ΜΟΝΤΕΛΟΥ ΠΑΛΙΝΔΡΟΜΗΣΗΣ (0 OUTLIERS) .....	53
5.1.1 Εκτιμήσεις Παραμέτρων.....	55
5.2 ΑΝΘΕΚΤΙΚΕΣ ΜΕΘΟΔΟΙ (0 OUTLIERS) .....	57
5.2.1 Παλινδρόμηση με χρήση Μ-Εκτιμητή (0 outliers) .....	58
5.2.2 Παλινδρόμηση Ελαχίστων Απολύτων Αποκλίσεων - Least Absolute Deviation (0 outliers)..	59
5.2.3 Κλασσική & Ανθεκτική Ανάλυση Διακύμανσης (ANOVA) (0 outliers).....	60
<b>ΔΙΑΓΝΩΣΤΙΚΑ ΓΡΑΦΗΜΑΤΑ .....</b>	<b>62</b>
5.2.4 Παλινδρόμηση Ελαχίστων Περικεκομμένων Τετραγώνων (0 outliers) .....	64
5.2.5 Παλινδρόμηση Ελαχίστων Διαμέσων Τετραγώνων (Least Trimmed Squares, LMS) (0 outliers)	66
5.3 ΕΦΑΡΜΟΓΗ ΛΟΓΑΡΙΘΜΙΚΟΥ ΜΟΝΤΕΛΟΥ ΠΑΛΙΝΔΡΟΜΗΣΗΣ (1 OUTLIER ΣΤΗ ΘΕΣΗ (4,4) ΤΟΥ ΤΡΙΓΩΝΟΥ	67
ΕΞΕΛΙΞΗΣ ΖΗΜΙΩΝ).....	67
5.3.1 Εκτιμήσεις παραμέτρων .....	69
5.4 ΑΝΘΕΚΤΙΚΕΣ ΜΕΘΟΔΟΙ (1 OUTLIER).....	70
5.4.1 Παλινδρόμηση με χρήση Μ-Εκτιμητή (1 outlier).....	70
5.4.2 Παλινδρόμηση Ελαχίστων Απολύτων Αποκλίσεων - Least Absolute Deviation (1 outlier)..	71
5.4.3 Κλασσική & Ανθεκτική ANOVA (1 outlier) .....	72
<b>ΔΙΑΓΝΩΣΤΙΚΑ ΓΡΑΦΗΜΑΤΑ .....</b>	<b>74</b>
5.4.4 Παλινδρόμηση Ελαχίστων Περικεκομμένων Τετραγώνων -Least Trimmed Squares (1 outlier)	76
5.4.5 Παλινδρόμηση Ελαχίστων Διαμέσων Τετραγώνων (Least Median Squares, LMS) (1 outlier)	77
5.5 ΕΦΑΡΜΟΓΗ ΛΟΓΑΡΙΘΜΙΚΟΥ ΜΟΝΤΕΛΟΥ ΠΑΛΙΝΔΡΟΜΗΣΗΣ (2 OUTLIERS ΣΤΙΣ ΘΕΣΕΙΣ (4,4) ΚΑΙ (1,2)) .....	78
5.5.1 Εκτιμήσεις παραμέτρων .....	79
5.6 ΑΝΘΕΚΤΙΚΕΣ ΜΕΘΟΔΟΙ (2 OUTLIERS) .....	80
5.6.1 Παλινδρόμηση με χρήση Μ-Εκτιμητή (2 outliers) .....	81
5.6.2 Παλινδρόμηση Ελαχίστων Απολύτων Αποκλίσεων (Least Absolute Deviation, LAD) (2	82
outliers) .....	82
5.6.3 Κλασσική & Ανθεκτική ANOVA (2 outliers).....	83
<b>ΔΙΑΓΝΩΣΤΙΚΑ ΓΡΑΦΗΜΑΤΑ .....</b>	<b>84</b>
5.6.4 Παλινδρόμηση Ελαχίστων Περικεκομμένων Τετραγώνων (Least Trimmed Squares, LTS) (2	86
outliers) .....	86

5.6.5	Παλινδρόμηση Ελαχίστων Διαμέσων Τετραγώνων ( <i>Least Median Squares, LMS</i> ) (2 outliers)	87
5.7	ΣΥΓΚΡΙΣΗ ΕΚΚΡΕΜΩΝ ΑΠΟΘΕΜΑΤΩΝ ΚΑΘΕ ΜΕΘΟΔΟΥ ΓΙΑ 0, 1, 2 OUTLIERS	88
5.7.1	Λογαριθμικό Μοντέλο Παλινδρόμησης	88
5.7.2	Παλινδρόμηση με χρήση M-Εκτιμητή	89
5.7.3	Παλινδρόμηση Ελαχίστων Απολύτων Αποκλίσεων ( <i>LAD</i> )	90
5.7.4	Ανθεκτική Ανάλυση Διασποράς ( <i>Robust ANOVA</i> )	90
5.7.5	Παλινδρόμηση Ελαχίστων Περικεκομμένων Τετραγώνων ( <i>LTS</i> )	91
5.7.6	Παλινδρόμηση ελαχίστων Διαμέσων Τετραγώνων ( <i>LMS</i> )	91
<b>ΚΕΦΑΛΑΙΟ 6<sup>ο</sup>: ΣΥΓΚΡΙΣΗ ΑΠΟΘΕΜΑΤΟ- ΠΟΙΗΣΗΣ ΚΛΑΣΣΙΚΩΝ &amp; ΑΝΘΕΚΤΙΚΩΝ ΜΕΘΟΔΩΝ</b>		<b>93</b>
6.1	ΣΥΓΚΡΙΣΗ ΑΠΟΘΕΜΑΤΟΠΟΙΗΣΗΣ ΚΛΑΣΣΙΚΩΝ & ΑΝΘΕΚΤΙΚΩΝ ΜΕΘΟΔΩΝ (0 OUTLIERS)	94
6.2	ΣΥΓΚΡΙΣΗ ΑΠΟΘΕΜΑΤΟΠΟΙΗΣΗΣ ΚΛΑΣΣΙΚΩΝ & ΑΝΘΕΚΤΙΚΩΝ ΜΕΘΟΔΩΝ (1 OUTLIER)	96
6.3	ΣΥΓΚΡΙΣΗ ΑΠΟΘΕΜΑΤΟΠΟΙΗΣΗΣ ΚΛΑΣΣΙΚΩΝ & ΑΝΘΕΚΤΙΚΩΝ ΜΕΘΟΔΩΝ (2 OUTLIERS)	98
6.4	ΣΥΓΚΡΙΣΗ ΜΕΘΟΔΩΝ ΠΡΙΝ ΚΑΙ ΜΕΤΑ ΤΗΝ ΠΡΟΣΘΗΚΗ ΈΚΤΟΠΩΝ ΠΑΡΑΤΗΡΗΣΕΩΝ	100
<b>ΒΙΒΛΙΟΓΡΑΦΙΑ</b>		<b>102</b>
<b>ΠΑΡΑΡΤΗΜΑ</b>		<b>104</b>
A <sub>1</sub>	ΣΥΜΠΛΗΡΩΜΕΝΑ ΤΡΙΓΩΝΑ ΕΞΕΛΙΞΗΣ ΖΗΜΙΩΝ ΜΕ ΛΟΓΑΡΙΘΜΟΥΣ ΑΤΟΜΙΚΩΝ ΖΗΜΙΩΝ (0 OUTLIERS)	104
A <sub>2</sub>	ΣΥΜΠΛΗΡΩΜΕΝΑ ΤΡΙΓΩΝΑ ΕΞΕΛΙΞΗΣ ΖΗΜΙΩΝ ΜΕ ΛΟΓΑΡΙΘΜΟΥΣ ΑΤΟΜΙΚΩΝ ΖΗΜΙΩΝ (1 OUTLIER)	106
A <sub>3</sub>	ΣΥΜΠΛΗΡΩΜΕΝΑ ΤΡΙΓΩΝΑ ΕΞΕΛΙΞΗΣ ΖΗΜΙΩΝ ΜΕ ΛΟΓΑΡΙΘΜΟΥΣ ΑΤΟΜΙΚΩΝ ΖΗΜΙΩΝ (2 OUTLIERS)	108
B	ΔΙΑΚΥΜΑΝΣΕΙΣ ΤΩΝ $\hat{Y}_{ij}$	110
B <sub>1</sub>	Διακυμάνσεις των $\hat{Y}_{ij}$ (0 outliers)	110
B <sub>2</sub>	Διακυμάνσεις των $\hat{Y}_{ij}$ (1 outlier)	111
B <sub>3</sub>	Διακυμάνσεις των $\hat{Y}_{ij}$ (2 outliers)	112

# Εισαγωγή

---

Η διαδικασία εκτίμησης του απαιτούμενου αποθεματικού για την αποπληρωμή αποζημιώσεων - loss reserving - συνιστά μια από τις σημαντικότερες διεργασίες μέσα στα πλαίσια μιας ασφαλιστικής εταιρείας. Ο ασφαλιστής προκειμένου να είναι σε θέση να ανταπεξέλθει στις μελλοντικές του υποχρεώσεις, οι οποίες προκύπτουν από συμβόλαια που είτε ισχύουν σήμερα είτε έχουν συναφθεί κατά το παρελθόν, είναι απαραίτητο να κατέχει αποθέματα εκκρεμών ζημιών.

Η κρισιμότητα της διαδικασίας αυτής συνίσταται στο γεγονός ότι αφενός η ημερομηνία είσπραξης των ασφαλίσεων και η ημερομηνία πληρωμής της αποζημίωσης δεν συμπίπτουν χρονικά και αφετέρου υπάρχει μια καθυστέρηση ανάμεσα στη χρονική στιγμή επέλευσης ενός ζημιογόνου ενδεχομένου και την ημερομηνία διακανονισμού και αποπληρωμής αυτού. Συνεπώς, η ευημερία αλλά και η φερεγγυότητα της ασφαλιστικής εταιρείας είναι άμεσα συνυφασμένες με την διαδικασία του loss reserving.

Στην αναλογιστική βιβλιογραφία υπάρχουν αρκετά άρθρα τα οποία περιγράφουν διάφορες μεθόδους αποθεματοποίησης, μεταξύ αυτών είναι: Zehnwirth (1985), Verall (1990, 1991a, 1991b), Renshaw (1989).

Ο Kremer (1997) ενσωμάτωσε τις ιδέες της ανθεκτικής στατιστικής σε διάφορες τεχνικές αποθεματοποίησης ζημιών, εκτελώντας πρακτικές εφαρμογές και χρησιμοποιώντας τη μέθοδο lag factor (ή link-ratio).

Οι Verdonck et al. (2009) δημιούργησαν ένα απλό παράδειγμα με σκοπό να εξετάσουν πως η μέθοδος chain ladder συμπεριφέρεται στη προσθήκη έκτοπων παρατηρήσεων, οι οποίες ενδέχεται να οδηγήσουν σε λανθασμένες εκτιμήσεις αποθεμάτων και επίσης εισήγαγαν μία τεχνική ανίχνευσης έκτοπων παρατηρήσεων. Για να λύσουν το πρόβλημα των έκτοπων παρατηρήσεων αντικατέστησαν τη μέση τιμή με τη διάμεσο, η οποία αποτελεί μία πιο ανθεκτική εκτίμηση.

Οι Verdonck και Debruyne (2010) ανέλυσαν την επίδραση της νόθευσης (contamination) των δεδομένων υπολογίζοντας συναρτήσεις επίδρασης στο πλαίσιο των γενικευμένων γραμμικών μοντέλων για τη μέθοδο chain ladder.



Οι Busse et al. (2010) σχεδίασαν ένα φίλτρο για τις έκτοπες παρατηρήσεις και παρουσίασαν μία ανθεκτική εκδοχή του εκτιμητή διακύμανσης κατά Mack.

Στην παρούσα εργασία θα αναλύσουμε και θα συγκρίνουμε τα μοντέλα των κλασικών και ανθεκτικών μεθόδων αποθεματοποίησης. Συγκεκριμένα, στο Κεφάλαιο 1 περιγράφονται σε γενικές γραμμές τα στατιστικά μοντέλα που θα χρησιμοποιήσουμε, δηλαδή το γενικό γραμμικό μοντέλο παλινδρόμησης και το μοντέλο ανάλυσης διακύμανσης δύο παραγόντων. Σε αυτά βασίζονται τα αντίστοιχα μοντέλα για την αποθεματοποίηση ζημιών, δηλαδή το μοντέλο της λογαριθμικής παλινδρόμησης, που μελετήθηκε μεταξύ άλλων από τον Verall (1991), και της ανάλυσης διακύμανσης δύο παραγόντων για IBNR (incurred but not reported) ζημιές που παρουσιάστηκε από τον Kremer (1982) τα οποία θα αναπτύξουμε αργότερα.

Στο Κεφάλαιο 2 θα αναλύσουμε την έννοια της ανθεκτικής στατιστικής. Πολλές φορές σε ένα σύνολο δεδομένων γενικών ασφαλίσεων παρουσιάζονται έκτοπες παρατηρήσεις (outliers). Οι ακραίες αυτές τιμές μπορούν να συμβούν κατά τύχη σε οποιαδήποτε κατανομή αλλά είναι συχνά ενδεικτικές είτε ενός σφάλματος μέτρησης είτε ότι ο πληθυσμός προέρχεται από μία κατανομή με βαριά ουρά. Στη πρώτη περίπτωση κάποιος έχει τη δυνατότητα να τα απορρίψει ή να χρησιμοποιήσει στατιστικά εργαλεία ανθεκτικά σε ακραίες τιμές, ενώ στη δεύτερη περίπτωση πρέπει κανείς να είναι πολύ προσεκτικός με τη χρήση εργαλείων ή πρακτικών που θεωρούν δεδομένη μια κανονική κατανομή. Δυστυχώς, όταν υπάρχουν ακραίες τιμές στα δεδομένα, οι κλασικές γραμμικές μέθοδοι έχουν συχνά πολύ χαμηλές επιδόσεις οπότε η ανθεκτική στατιστική (Robust Statistics) επιδιώκει να παρέχει μεθόδους που δεν επηρεάζονται από ακραίες τιμές ή άλλες μικρές αποκλίσεις από τις υποθέσεις του μοντέλου. Στη συνέχεια του συγκεκριμένου κεφαλαίου θα περιγράψουμε περιληπτικά τα διάφορα **μέτρα ανθεκτικότητας** που έχουν εισαχθεί προκειμένου να προσδιοριστεί ο βαθμός ανθεκτικότητας μιας μεθόδου. Τα πιο γνωστά είναι η συνάρτηση επίδρασης (Influence Function- IF), το σημείο κατάρρευσης (Breakdown Point- BP) και η ευαισθησία γενικού σφάλματος (Gross Error Sensitivity- GES).

Στο Κεφάλαιο 3 θα αναλύσουμε εκτενώς τους **ανθεκτικούς εκτιμητές** τους οποίους και θα χρησιμοποιήσουμε στο αριθμητικό παράδειγμα. Οι εκτιμητές αυτοί είναι οι M-εκτιμητές, οι εκτιμητές Ελαχίστων Διαμέσων Τετραγώνων (Least Median Squares - LMS), οι εκτιμητές Ελαχίστων Περικεκομμένων Τετραγώνων (Least

Trimmed Squares - LTS) και οι εκτιμητές Ελαχίστων Απολύτων Αποκλίσεων (Least Absolute Deviation - LAD). Στο σημείο αυτό αξίζει να τονίσουμε ότι σκοπός της παρούσας εργασίας δεν είναι η λεπτομερής ανάλυση των ανθεκτικών εκτιμητών, για τους οποίους υπάρχει εκτενής βιβλιογραφία. Στη πράξη, τους χρησιμοποιούμε σαν στατιστικά εργαλεία για την εκτίμηση των εκκρεμών ζημιών και των απαιτούμενων αποθεματικών των ασφαλιστικών εταιρειών.

Στο Κεφάλαιο 4 θα αναπτύξουμε εκτενώς μία από τις σημαντικότερες στατιστικές μεθόδους εκτίμησης αποθεματικών, μέσω της πρόβλεψης εκκρεμών ζημιών από παρελθοντικές ζημιές, την Chain Ladder ή Μέθοδος Τριγώνου Εξέλιξης Ζημιών στην οποία είναι εμφανής η ιστορική εξέλιξη της ζημιάς. Επίσης, θα παρουσιάσουμε το μοντέλο πολλαπλής λογαριθμικής παλινδρόμησης (multiple loglinear regression), της ανάλυσης διακύμανσης 2 παραγόντων (two-way ANOVA) για IBNR ζημιές και τα αντίστοιχα ανθεκτικά τους μοντέλα, τα οποία θα χρησιμοποιήσουμε για να εκτιμήσουμε τις παραμέτρους των μοντέλων μας και κατ' επέκταση τις μελλοντικές ζημιές και τα προβλεπόμενα αποθεματικά. Επίσης, στο Κεφάλαιο 4 συμπεριλαμβάνεται μία σύντομη εφαρμογή για ένα 3×3 τρίγωνο εξέλιξης ζημιών στην οποία θα περιγράψουμε επακριβώς τα απαιτούμενα βήματα για την εκτίμηση των παραμέτρων και την πρόβλεψη των εκκρεμών ζημιών.

Στο Κεφάλαιο 5 γίνεται εφαρμογή των κλασσικών και ανθεκτικών μεθόδων στα δεδομένα μας. Αρχικά, θα παρουσιάσουμε τα δεδομένα (τρίγωνο εξέλιξης ζημιών) των Taylor and Ashe (1983), με βάση τα οποία θα τις εφαρμόσουμε. Το τρίγωνο ζημιών που θα μελετήσουμε αποτελείται από 10 έτη ατυχήματος και 10 έτη εξέλιξης, απεικονίζει ατομικές (όχι αθροιστικές) ζημιές και έχει ήδη χρησιμοποιηθεί από πολλούς ερευνητές. Στη συνέχεια, ακολουθούν τα συμπληρωμένα τρίγωνα εξέλιξης ζημιών και διάφορα διαγνωστικά γραφήματα καταλοίπων για όλες τις εξεταζόμενες μεθόδους. Η διαδικασία αυτή επαναλαμβάνεται μετά την προσθήκη μίας και δύο ακραίων τιμών στα δεδομένα μας στις θέσεις (4,4) και (1,2) του τριγώνου εξέλιξης ζημιών. Το κεφάλαιο αυτό ολοκληρώνεται με μία σύγκριση υπό τη μορφή γραφημάτων των χρησιμοποιούμενων μεθόδων για μία, δύο και καμία έκτοπη παρατήρηση στα δεδομένα μας.

Στο Κεφάλαιο 6 εκθέτονται οι εκτιμήσεις των τελικών ζημιών, των απαιτούμενων εκκρεμών αποθεματικών και του συνολικού αποθεματικού για κάθε μέθοδο. Η



διαδικασία αυτή επαναλαμβάνεται μετά την προσθήκη μίας και δύο έκτοπων παρατηρήσεων στα δεδομένα μας στις θέσεις (4,4) και (1,2) ώστε να διευκρινιστεί η επιρροή αυτών των έκτοπων παρατηρήσεων στη πρόβλεψη των αποθεματικών. Τέλος, η παρούσα εργασία ολοκληρώνεται με μία σύγκριση όλων των μεθόδων, κλασσικών και ανθεκτικών, πριν και μετά την προσθήκη έκτοπων παρατηρήσεων και εξάγονται τα συμπεράσματα μας.

Στο σημείο αυτό αξίζει να αναφέρουμε ότι οι υπολογισμοί και οι εκτιμήσεις των κλασσικών και ανθεκτικών μεθόδων παλινδρόμησης και ανάλυσης διακύμανσης, πραγματοποιήθηκαν με τη βοήθεια των στατιστικών πακέτων S-Plus και R, καθώς επίσης και του Excel.

Μέρος αυτής της εργασίας παρουσιάστηκε στο 15<sup>ο</sup> Διεθνές Συνέδριο, Insurance Mathematics & Economics που πραγματοποιήθηκε 14-17 Ιουνίου 2011 στη Τεργέστη της Ιταλίας. Από την ελληνική βιβλιογραφία χρησιμοποιήθηκαν οι διπλωματικές εργασίες των Χ. Ζαχαροπούλου (Ανίχνευση Έκτοπων Μετρήσεων, Διαγνωστική και Ανθεκτική Εκτίμηση: Ανασκόπηση, 2004) και Δ. Τσατσαρώνη (Δυναμικές Μέθοδοι Εκτίμησης Αποθέματος Ζημιών Γενικών Ασφαλίσεων, 2010).

# Κεφάλαιο 1<sup>ο</sup>: Στατιστικά Μοντέλα

## 1.1 Γενικό Γραμμικό Μοντέλο

Υποθέτουμε ότι έχουμε ένα υπό εξέταση μοντέλο που μπορεί να γραφεί στη μορφή

$$Y = X\beta + e, \quad (1.1)$$

όπου  $Y$ :  $(n \times 1)$  διάνυσμα παρατηρήσεων,

$X$ :  $(n \times p)$  πίνακας σχεδιασμού γνωστής μορφής,

$\beta$ :  $(p \times 1)$  διάνυσμα παραμέτρων,

$e$ :  $(n \times 1)$  διάνυσμα τυχαίων σφαλμάτων με μέση τιμή  $E(e) = 0_{(n \times 1)}$ ,

$Var(e) = \sigma^2 I_{(n \times 1)}$  και  $Cov(e_i, e_j) = 0$  για κάθε  $i, j$  με  $i \neq j$ ,

οπότε οι τιμές του διανύσματος παρατηρήσεων  $Y$  είναι οι εξαρτημένες μεταβλητές και  $X$  είναι ο πίνακας σχεδιασμού που περιέχει τις ανεξάρτητες/επεξηγηματικές μεταβλητές.

### 1.1.1 Εκτιμητής Ελαχίστων Τετραγώνων και Ιδιότητες

Η μέθοδος Ελαχίστων Τετραγώνων έχει ως στόχο την εύρεση εκτιμητριών για το διάνυσμα των παραμέτρων  $\beta$  έτσι ώστε να ελαχιστοποιείται το άθροισμα των τετραγώνων των καταλοίπων

$$\begin{aligned} SSE(\beta) &= \sum_{i=1}^n e_i^2 = e^T e = (Y - X\beta)^T (Y - X\beta) \\ &= Y Y^T - Y^T X \beta - (X\beta)^T Y + (X\beta)^T (X\beta) \\ &= Y Y^T - 2\beta^T X^T Y + \beta^T X^T X \beta \end{aligned}$$

Τελικά, ο εκτιμητής ελαχίστων τετραγώνων,  $\hat{\beta}$ , υπολογίζεται παίρνοντας την παράγωγο του ως προς  $\beta$  και θέτοντας την ίση με το 0, δηλαδή

$$\frac{\partial SSE(\beta)}{\partial \beta} = 0$$

$$\Rightarrow -2X^T\Upsilon + \beta^T X^T \frac{\partial}{\partial \beta} X\beta + \frac{\partial}{\partial \beta} \beta^T X^T = 0$$

$$\Rightarrow -2X^T\Upsilon - 2X^T X\hat{\beta} = 0$$

$$\Rightarrow (X^T X)\hat{\beta} = \beta^T \Upsilon$$

$$\Rightarrow (X^T X)^{-1} X^T X\hat{\beta} = (X^T X)^{-1} X^T \Upsilon$$

$$\Rightarrow \hat{\beta} = (X^T X)^{-1} X^T \Upsilon. \quad (1.2)$$

Ο εκτιμητής  $\hat{\beta}$  είναι γραμμικός αμερόληπτος εκτιμητής ελάχιστης διασποράς για το  $\beta$ . Συνεπώς, αν τα σφάλματα κατανέμονται ανεξάρτητα με μέση τιμή 0 και σταθερή διασπορά τότε ο εκτιμητής ελαχίστων τετραγώνων αποτελεί τον πιο αποτελεσματικό (efficient) γραμμικό εκτιμητή του  $\beta$ .

Η μέση τιμή του εκτιμητή  $\hat{\beta}$  είναι

$$E(\hat{\beta}) = E[(X^T X)^{-1} X^T \Upsilon] = (X^T X)^{-1} X^T E(\Upsilon) = (X^T X)^{-1} X^T X\beta = \beta.$$

Η διασπορά του εκτιμητή  $\hat{\beta}$  δίνεται από τη σχέση

$$\begin{aligned} Cov(\hat{\beta}) &= Cov[(X^T X)^{-1} X^T \Upsilon] \\ &= (X^T X)^{-1} X^T Cov(\Upsilon) X (X^T X)^{-1} \\ &= \sigma^2 (X^T X)^{-1} X^T X (X^T X)^{-1} \\ &= \sigma^2 (X^T X)^{-1}. \end{aligned}$$

### 1.1.2 Εκτιμώμενες από το μοντέλο τιμές και ιδιότητες

Δοθέντων των τιμών  $(X_1, \dots, X_p)$  οι οποίες συμβολίζονται με  $X_f = (X_{f_1}, \dots, X_{f_p})^T$  λαμβάνουμε την πρόβλεψη των εκτιμώμενων τιμών,  $\hat{Y}_f$ , ως εξής

$$\hat{Y}_f = X_f^T \hat{\beta},$$

όπου  $X_f$  ο μελλοντικός πίνακας σχεδιασμού και ισχύουν τα εξής

$$E(\hat{Y}_f) = E(X_f^T \hat{\beta}) = X_f^T E(\hat{\beta}) = X_f^T \beta = E(Y_f),$$

και

$$\begin{aligned}
\text{Var}(\hat{Y}_f) &= E(X_f^T \hat{\beta} - X_f^T \beta)^2 \\
&= E(X_f^T \hat{\beta} - X_f^T \beta)(X_f^T \hat{\beta} - X_f^T \beta)^T \\
&= E[X_f^T (\hat{\beta} - \beta)(\hat{\beta} - \beta)^T X_f] \\
&= X_f^T \text{Var}(\hat{\beta}) X_f \\
&= \sigma^2 X_f^T (X_f^T X)^{-1} X_f.
\end{aligned}$$

Μετά την προσαρμογή του μοντέλου (1.1) στα δεδομένα μας, το άθροισμα των τετραγώνων των σφαλμάτων  $e_i = Y_i - \hat{Y}_i$  είναι ελάχιστο και το συμβολίζουμε με  $SSE$ . Ισχύει

$$SSE(\hat{\beta}) = e^T e = (Y - \hat{Y})^T (Y - \hat{Y}) = (Y - X\hat{\beta})^T (Y - X\hat{\beta}) = \hat{Y}^T Y - \hat{\beta}^T X^T Y,$$

και αντικαθιστώντας την τιμή του  $\hat{\beta}$ , από την (1.2) έχουμε

$$SSE(\hat{\beta}) = Y^T (I_n - X(X^T X)^{-1} X^T) Y,$$

όπου  $I_n$  είναι ένας  $n \times n$  μοναδιαίος πίνακας.

### 1.1.3 Ανάλυση Διασποράς για την παλινδρόμηση

Η ανάλυση διασποράς (ANOVA) εξετάζει τη σχέση της εξαρτημένης μεταβλητής με την ανεξάρτητη, υπολογίζοντας στην ουσία το αν η μεταβλητότητα των τιμών της εξαρτημένης μεταβλητής  $Y$  εξηγείται από την ανεξάρτητη μεταβλητή  $X$ . Η δειγματική διασπορά των παρατηρήσεων  $Y_i$  αποδεικνύεται ότι χωρίζεται σε δύο αθροίσματα

$$\sum_{i=1}^n (Y_i - \bar{Y})^2 = \sum_{i=1}^n (Y_i - \hat{Y}_i)^2 + \sum_{i=1}^n (\hat{Y}_i - \bar{Y})^2 \Rightarrow$$

$$SST = SSE + SSR.$$

Το  $SST$  εκφράζει τη συνολική παρατηρούμενη μεταβλητότητα των  $Y_i$ , το  $SSR$  εκφράζει τη μεταβλητότητα των προσαρμοσμένων τιμών και το  $SSE$  εκφράζει τη μεταβλητότητα των  $Y_i$  σε σχέση με τις αντίστοιχες προσαρμοσμένες τιμές. Η μεταβλητότητα του  $SSR$  ερμηνεύεται από το μοντέλο σε αντίθεση με αυτήν του  $SSE$ .

Η ανάλυση της διασποράς για το απλό γραμμικό μοντέλο μπορεί να παρουσιασθεί ως εξής

Πηγή Μεταβλητότητας	Βαθμοί Ελευθερίας	Άθροισμα Τετραγώνων	Μέσο Τετράγωνο	Λόγος F
Παλινδρόμηση	1	$SSR = \sum_{i=0}^n (\hat{Y}_i - \bar{Y})^2$	$MSR = \frac{SSR}{1}$	$F = \frac{MSR}{MSE}$
Υπόλοιπα	n - 2	$SSE = \sum_{i=1}^n (Y_i - \hat{Y}_i)^2$	$MSE = S^2 = \frac{SSE}{n - 2}$	
Ολική	n - 1	$SST = \sum_{i=1}^n (Y_i - \bar{Y})^2$		

Πίνακας 1.1: Πίνακας Ανάλυσης Διασποράς (ANOVA Table)

## 1.2 Μοντέλο Ανάλυσης Διακύμανσης 2 παραγόντων (Two-way Model of ANOVA)

Υποθέτουμε ότι στο πρόβλημα που μας ενδιαφέρει υπεισέρχονται δύο παράγοντες, A και B, που εμφανίζονται με p και q στάθμες αντίστοιχα. Έστω, επίσης, ότι σε κάθε στάθμη του παράγοντα A έγιναν (r \* q) παρατηρήσεις, από r για κάθε στάθμη του B. Σχηματικά το πρόβλημα περιγράφεται από τον παρακάτω πίνακα

Παράγοντας A	Παράγοντας B				
		$B_1$	$B_2$	...	$B_q$
$A_1$		$Y_{111}, \dots, Y_{11r}$	$Y_{121}, \dots, Y_{12r}$	...	$Y_{1q1}, \dots, Y_{1qr}$
$A_2$		$Y_{211}, \dots, Y_{21r}$	$Y_{221}, \dots, Y_{22r}$	...	$Y_{2q1}, \dots, Y_{2qr}$
...		...	...	...	...
$A_p$		$Y_{p11}, \dots, Y_{p1r}$	$Y_{p21}, \dots, Y_{p2r}$	...	$Y_{pq1}, \dots, Y_{pqr}$

Πίνακας 1.2: Δύο παράγοντες με r παρατηρήσεις ανά κελί

Σ' ένα τέτοιο πρόβλημα η τιμή της εξαρτημένης μεταβλητής Y μπορεί να εξαρτάται ή όχι από τους παράγοντες A και B. Θεωρούμε τις (p \* r) παρατηρήσεις που έχουν γίνει στη στάθμη  $B_1$  και αγνοώντας τις άλλες στάθμες υπολογίζουμε τις κύριες επιδράσεις του παράγοντα A (όταν ο παράγοντας B είναι στη στάθμη  $B_1$ ). Αφού επαναλάβουμε την ίδια διαδικασία για όλες τις στάθμες του παράγοντα B, ενδέχεται σε ορισμένες από αυτές οι κύριες επιδράσεις του A να είναι πολύ διαφορετικές αφενός μεταξύ τους και αφετέρου με τις συνολικές επιδράσεις που



εκτιμήθηκαν. Στη περίπτωση αυτή, οι παράγοντες A και B αλληλεπιδρούν στο κελί  $(i, j)$  και συμβολίζουμε την αλληλεπίδραση αυτή με  $(\alpha\beta)_{ij}$ . Ένα κατάλληλο μοντέλο για τη περίπτωση αυτή είναι

$$Y_{ijk} = \mu + \alpha_i + b_j + (\alpha\beta)_{ij} + e_{ijk}.$$

Αν υποθέσουμε ότι δεν υπάρχει αλληλεπίδραση μεταξύ των παραγόντων A και B, τότε το μοντέλο γίνεται

$$Y_{ij} = \mu + \alpha_i + b_j + e_{ij}, \quad i = 1, 2, \dots, p \text{ και } j = 1, 2, \dots, q$$

όπου:  $Y_{ij}$ : η τιμή της μεταβλητής Y στη στάθμη του παράγοντα A και στη στάθμη του παράγοντα B,

$\mu$ : ο μέσος όρος,

$\alpha_i$ : η επίδραση της στάθμης του παράγοντα A στη διαμόρφωση των  $Y_{ij}$ ,

$b_j$ : επίδραση της στάθμης του παράγοντα B στη διαμόρφωση των  $Y_{ij}$ ,

$e_{ij}$ : σφάλμα κατά την εκτίμηση των  $Y_{ij}$ .

Ο πίνακας διακυμάνσεων-συνδιακυμάνσεων (ANOVA Table) του μοντέλου ανάλυσης διακύμανσης για δύο παράγοντες χωρίς αλληλεπίδραση είναι ο εξής

Πηγή Μεταβλητότητας	Άθροισμα Τετραγώνων (SS)	Βαθμοί Ελευθερίας	Μέσο Άθροισμα Τετραγώνων (MS)	Λόγος F
Παράγοντας A	$SSA = q \sum_{i=1}^p (\bar{Y}_i - \bar{Y})^2$	$p - 1$	$MSA = \frac{SSA}{p - 1}$	$F_A = \frac{MSA}{MSE}$
Παράγοντας B	$SSB = p \sum_{j=1}^q (\bar{Y}_j - \bar{Y})^2$	$q - 1$	$MSB = \frac{SSB}{q - 1}$	$F_B = \frac{MSB}{MSE}$
Σφάλμα	$SSE = SST - SSA - SSB$	$(p - 1)(q - 1)$	$MSE = \frac{SSE}{(p - 1)(q - 1)}$	
Συνολική	$SST = \sum_{i=1}^p \sum_{j=1}^q (Y_{ij} - \bar{Y})^2$	$pq - 1$		

Πίνακας 1.3: Πίνακας Ανάλυσης Διασποράς (ANOVA Table)

όπου η συνολική μεταβλητότητα (SST) είναι ένα άθροισμα μεταβολών που οφείλεται στη δράση του παράγοντα A (SSA), στη δράση του παράγοντα B (SSB) και σε μη ελεγχόμενους παράγοντες -υπόλοιπη μεταβλητότητα ή σφάλμα- (SSE). Δηλαδή

$$SST = SSA + SSB + SSE.$$

Στη συνέχεια ακολουθούν οι έλεγχοι για την σημαντικότητα των παραγόντων A και B σε επίπεδο σημαντικότητας  $\alpha\%$ .

Έλεγχοι Σημαντικότητας Παραγόντων	Έλεγχος Υποθέσεων	Στατιστικό	Περιοχή Απόρριψης $H_0$
Παράγοντας A	$H_{0A}$ : ο παράγοντας A δεν επιδρά vs $H_{1A}$ : ο παράγοντας A επιδρά	$F_A = \frac{MSA}{MSE}$	$F_A > F_{(p-1),(p-1)(q-1);\alpha}$
Παράγοντας B	$H_{0B}$ : ο παράγοντας B δεν επιδρά vs $H_{1B}$ : ο παράγοντας B επιδρά	$F_B = \frac{MSB}{MSE}$	$F_B > F_{(q-1),(p-1)(q-1);\alpha}$

Πίνακας 1.4: Έλεγχοι Σημαντικότητας Παραγόντων A και B

Αν  $F_A > F_{(p-1),(p-1)(q-1);\alpha}$  τότε σε επίπεδο σημαντικότητας  $\alpha\%$  απορρίπτουμε τη μηδενική υπόθεση ότι ο παράγοντας A δεν επιδρά στα δεδομένα και συνεπώς συμπεραίνουμε ότι είναι σημαντικός.

## Κεφάλαιο 2<sup>ο</sup>: Μέτρα Ανθεκτικότητας

Πολλές φορές σε ένα σύνολο δεδομένων παρουσιάζονται έκτοπες παρατηρήσεις (outliers). Σύμφωνα με τον Hawkins (1980):

*«Έκτοπη είναι μία παρατήρηση που αποκλίνει τόσο πολύ από τις άλλες παρατηρήσεις, ώστε να μας εξεγείρει τις υποψίες ότι δημιουργήθηκε από διαφορετικό μηχανισμό.»*

Οι συχνότερες αιτίες εμφάνισης έκτοπων παρατηρήσεων είναι συνήθως λάθη αντιγραφής και εναλλαγής τιμών, παροδικές επιδράσεις, απρόσεκτη παρατήρηση, παραπονημένες ερωτήσεις και ελλιπείς απαντήσεις. Οφείλονται, δηλαδή, σε κάποιο μηχανικό σφάλμα, αλλαγή σε μία συμπεριφορά ενός συστήματος ή σε ανθρώπινα λάθη. Η απόφαση σχετικά με το αν μία παρατήρηση αποτελεί outlier λαμβάνεται με βάση διάφορα κριτήρια και τη κρίση του ερευνητή (βλέπε Rousseeuw & Leroy (1987)).

Οι ακραίες αυτές τιμές μπορούν να συμβούν κατά τύχη σε οποιαδήποτε κατανομή αλλά είναι συχνά ενδεικτικές είτε ενός σφάλματος μέτρησης είτε ότι ο πληθυσμός προέρχεται από μία κατανομή με βαριά ουρά. Στη πρώτη περίπτωση κάποιος έχει τη δυνατότητα να τα απορρίψει ή να χρησιμοποιήσει στατιστικά εργαλεία ανθεκτικά σε ακραίες τιμές, ενώ στη δεύτερη περίπτωση πρέπει κανείς να είναι πολύ προσεκτικός με τη χρήση εργαλείων ή πρακτικών που θεωρούν δεδομένη μια κανονική κατανομή. Δυστυχώς, όταν υπάρχουν ακραίες τιμές στα δεδομένα, οι κλασσικές γραμμικές μέθοδοι έχουν συχνά πολύ χαμηλές επιδόσεις οπότε η ανθεκτική στατιστική (Robust statistics) επιδιώκει να παρέχει μεθόδους που δεν επηρεάζονται από ακραίες τιμές ή άλλες μικρές αποκλίσεις από τις υποθέσεις του μοντέλου.

Υποθέτουμε ότι έχουμε ένα μονοδιάστατο δείγμα  $(X_1, X_2, \dots, X_n)$  που αποτελείται από ανεξάρτητα και ομοιόμορφα κατανεμημένες παρατηρήσεις με συνάρτηση κατανομής  $F_\theta$ , η οποία ανήκει σε μία παραμετρική οικογένεια κατανομών  $\{F_\theta; \theta \in \Omega\}$ .

Στην κλασσική στατιστική υποθέτουμε ότι τα  $X_i$  είναι κατανεμημένα ακριβώς όπως η  $F_\theta$ , συνεπώς αυτό που απομένει είναι να εκτιμήσουμε την παράμετρο  $\theta$  με βάση τα δεδομένα μας. Πολύ συχνά οι υποθέσεις της κλασσικής στατιστικής, όπως

για παράδειγμα η κανονικότητα, η γραμμικότητα και η ανεξαρτησία, αποτελούν απλά προσεγγίσεις της πραγματικότητας. Τις περισσότερες φορές εφαρμόζουμε τις κατάλληλες τεχνικές κάτω από αυτές τις υποθέσεις, αλλά ακόμα και η μικρότερη απόκλιση από το μοντέλο ενδέχεται να επηρεάσει τα αποτελέσματα σε μεγάλο βαθμό.

Αντιθέτως, στην ανθεκτική στατιστική αντιλαμβανόμαστε ότι το μοντέλο  $\{F_\theta; \theta \in \Omega\}$  δεν είναι τίποτα άλλο παρά μία εξιδανικευμένη προσέγγιση της πραγματικότητας. Επίσης, σκοπός της ανθεκτικής στατιστικής είναι να περιγράψει το μοντέλο που εφαρμόζεται καλύτερα στον κύριο όγκο των δεδομένων, να αναγνωρίσει τις έκτοπες τιμές (outliers) και τέλος να προειδοποιήσει για τυχόν σημεία μόχλευσης (leverage points).

Με σκοπό να κατανοήσουμε γιατί ορισμένοι εκτιμητές έχουν κάποιες συγκεκριμένες ιδιότητες, είναι απαραίτητο να μελετήσουμε τα διάφορα μέτρα ανθεκτικότητας και τα εργαλεία που τα συνοδεύουν.

## 2.1 Συνάρτηση Επίδρασης (Influence Function, IF)

Θεωρούμε  $F_n$  την εμπειρική κατανομή του δείγματος  $(X_1, X_2, \dots, X_n)$ , η οποία δίνεται από τη σχέση

$$F_n = \frac{1}{n} \sum_{i=1}^n \Delta_{x_i},$$

όπου  $\Delta_x$  έχει μάζα πιθανότητας 1 στο  $x$ . Για την εκτίμηση του  $\theta$  θεωρούμε τις στατιστικές συναρτήσεις,  $T_n$ , της μορφής:

$$T_n = T_n(X_1, X_2, \dots, X_n) = T_n(F_n).$$

Ένας εκτιμητής μπορεί, με την ευρύτερη έννοια, να θεωρηθεί σαν μία ακολουθία στατιστικών συναρτήσεων  $\{T_n; n \geq 1\}$ , μία για κάθε πιθανό μέγεθος δείγματος  $n$ . Η συνάρτηση επίδρασης περιγράφει την επίδραση που έχει στον εκτιμητή μια πολύ μικρή αλλοίωση σε ένα σημείο  $x$  του δείγματος μας, μετρώντας την ασυμπτωτική μεροληψία που προκαλείται.

**Ορισμός 1.** Η Συνάρτηση Επίδρασης (Influence Function, IF) του εκτιμητή  $T$  με κατανομή  $F$  δίνεται από τη σχέση

$$IF(x; T, F) = \lim_{t \rightarrow 0} \frac{T[(1-t)F + t\Delta_x] - T(F)}{t} \quad (2.1).$$

Αν στην παραπάνω σχέση για τη συνάρτησης επίδρασης αντικαταστήσουμε το  $F$  με το  $F_{n-1}$ , όπου  $F_n$  η εμπειρική κατανομή εκτίμησης του  $F$  προερχόμενη από το δείγμα  $(X_1, X_2, \dots, X_n)$  και θέσουμε  $t = \frac{1}{n}$ , τότε η  $IF(x; T, F_{n-1})$  μετράει περίπου  $n$  φορές την αλλαγή που προκαλείται στον εκτιμητή  $T$  και οφείλεται στην επιπλέον παρατήρηση στο  $x$ , όταν ο  $T$  εφαρμόζεται σε ένα μεγάλο δείγμα παρατηρήσεων μεγέθους  $(n-1)$  (βλέπε *Hampel et al. (1986)*, σελ. 84). Όταν μελετάμε πεπερασμένα δείγματα δεν χρησιμοποιούμε  $T(F)$  αλλά  $T(F_n)$ .

Αν αντικαταστήσουμε μία παρατήρηση με μία έκτοπη παρατήρηση  $X$  (outlier), τότε

$$T(X_1, X_2, \dots, X_n, X) \cong T(F_n) + \frac{1}{n} IF(x; T, F).$$

Αν αντικαταστήσουμε ένα μικρό ποσοστό  $t = \frac{m}{n}$  παρατηρήσεων με έκτοπες παρατηρήσεις, τότε

$$T\left(X_1, X_2, \dots, X_n, \underbrace{X, \dots, X}_{m \text{ φορές}}\right) \cong T(F_n) + \frac{m}{n} IF(x; T, F).$$

Έστω μια κατανομή  $G$ , η οποία είναι «κοντά» στην  $F$ , τότε το von Mises ανάπτυγμα 1<sup>ης</sup> τάξης του  $T$  στην  $F$ , εκτιμώμενο για  $G$  δίνεται από τη σχέση

$$T(G) = T(F) + \int IF(x; T, F) d(G - F)(x) + R_n,$$

όπου ο όρος  $R_n$  αντιστοιχεί στο υπόλοιπο (remainder).

Στο σημείο αυτό θα επικεντρωθούμε στην σχέση μεταξύ της συνάρτησης επίδρασης και της ασυμπτωτικής διασποράς (βλέπε *Hampel et al. (1986)*, σελ. 85). Όταν οι παρατηρήσεις  $X_i$  είναι ανεξάρτητα και ισόνομα κατανεμημένες σύμφωνα με την κατανομή  $F$ , τότε σύμφωνα με το θεώρημα Glivenko-Cantelli, η εμπειρική συνάρτηση κατανομής  $F_n$  θα τείνει στην  $F$ . Συνεπώς, στον παραπάνω τύπο μπορούμε



να αντικαταστήσουμε το  $G$  με το  $F_n$ , για αρκετά μεγάλα μεγέθη δειγμάτων. Ας υποθέσουμε επίσης ότι  $T_n(X_1, X_2, \dots, X_n) = T_n(F_n)$  μπορεί να προσεγγιστεί ικανοποιητικά από το  $T(F_n)$ . Κάνοντας χρήση της σχέσης

$$\int IF(x; T, F) dF(x) = 0,$$

λαμβάνουμε:

$$T_n(F_n) = T(F) + \int IF(x; T, F) dF_n(x) + R_n.$$

Υπολογίζοντας το ολοκλήρωμα ως προς  $F_n$  καταλήγουμε στην παρακάτω σχέση

$$\sqrt{n}(T_n - T(F)) \approx \frac{1}{\sqrt{n}} \sum_{i=1}^n IF(X_i; T, F) + R_n.$$

Σύμφωνα με το Κεντρικό Οριακό Θεώρημα, ο πρώτος όρος στο δεξί μέλος της εξίσωσης είναι ασυμπτωτικά κανονικός. Επίσης, στις περισσότερες περιπτώσεις ο όρος  $R_n$  (remainder) γίνεται αμελητέος για  $n \rightarrow \infty$ , με αποτέλεσμα και ο εκτιμητής  $T_n$  να είναι ασυμπτωτικά κανονικός. Αυτό σημαίνει ότι

$$\left( \sqrt{n}(T_n - T(F)) \right) \rightarrow N(0, V(T, F)),$$

όπου η ασυμπτωτική διασπορά  $V(T, F)$  ορίζεται ως εξής

$$V(T, F) = \int IF(x; T, F)^2 dF(x) \quad (2.2).$$

## 2.2 Σημείο Κατάρρευσης (Breakdown Point, BP)

Το σημείο κατάρρευσης μας εξασφαλίζει ένα πρόχειρο άνω φράγμα του ποσοστού των έκτοπων παρατηρήσεων (outliers) για το οποίο μια τέτοια γραμμική προσέγγιση μπορεί να είναι χρήσιμη (βλέπε *Rousseeuw & Leroy (1987)*). Μετράει, δηλαδή, πόση αλλοίωση μπορεί να αντέξει ο εκτιμητής, υπολογίζοντας το μέγιστο ποσοστό αλλοίωσης που διατηρεί τον εκτιμητή υπό έλεγχο, ενώ ένα μεγαλύτερο ποσοστό αλλοίωσης προκαλεί κατάρρευση του εκτιμητή.

**Ορισμός 2.** Το σημείο κατάρρευσης (BP) μιας εκτιμήτριας  $T$  για ένα παρατηρούμενο δείγμα  $(X_1, X_2, \dots, X_n)$  ορίζεται ως εξής

$$\varepsilon_n^*(T, X) := \min \left[ \frac{m}{n}; \text{bias}(m, T, X) < \infty \right] \quad (2.3),$$

όπου  $\text{bias}(m, T, X)$  είναι η μέγιστη μεροληψία (bias) που μπορεί να προκληθεί από μία νόθευση (contamination) των δεδομένων (παρουσία έκτοπων παρατηρήσεων) και  $m$  είναι ο αριθμός των αρχικών παρατηρήσεων που αντικαταστάθηκαν από τυχαίες (αυθαίρετες) τιμές. Με άλλα λόγια το  $\varepsilon_n^*(T, X)$  είναι το μικρότερο ποσοστό νόθευσης που μπορεί να προκαλέσει την εκτιμήτρια  $T$  να πάρει τυχαίες τιμές μακριά από  $T_n$ . Επίσης, το BP συνήθως δεν εξαρτάται από την τιμή των παρατηρήσεων του δείγματος  $X$ , ενώ εξαρτάται ελάχιστα από τον αριθμό των παρατηρήσεων του δείγματος,  $n$ .

### 2.3 Ευαισθησία Γενικού Σφάλματος (Gross Error Sensitivity)

Η ευαισθησία γενικού σφάλματος περιγράφει την μέγιστη επίδραση που προκαλεί στην τιμή του εκτιμητή μια μικρή αλλοίωση της κατανομής και θα πρέπει να είναι πεπερασμένη.

**Ορισμός 3.** Η ευαισθησία γενικού σφάλματος μιας εκτιμήτριας  $T$  για δοσμένη κατανομή  $F$ , ορίζεται ως εξής

$$\gamma^*(T, F) := \sup |IF(x; T, F)| \quad (2.4),$$

όπου το supremum ορίζεται για όλα τα  $x$ , για τα οποία υπάρχει η συνάρτηση επίδρασης  $IF(x; T, F)$ . Τυπικά, ο καθορισμός ενός ορίου για το  $\gamma^*$  αποτελεί το πρώτο βήμα για να κάνουμε μια εκτιμήτρια πιο ανθεκτική και αυτό συχνά έρχεται σε σύγκρουση με τον στόχο της ασυμπτωτικής αποδοτικότητας (βλέπε Hampel et al. (1986), σελ. 87).

# Κεφάλαιο 3<sup>ο</sup>: Ανθεκτικοί Εκτιμητές

## 3.1 Μ-Εκτιμητές

Στην ανθεκτική στατιστική, οι Μ-εκτιμητές αποτελούν μια γενικευμένη κλάση εκτιμητών που παράγονται από την ελαχιστοποίηση του αθροίσματος διαφόρων συναρτήσεων.

Η ονομασία «Μ-Εκτιμήτρια» (βλέπε *Huber, 1981, σελ.43*) προέρχεται από την έννοια της «γενικευμένης μεγίστης πιθανοφάνειας», αφού αποτελούν γενίκευση των εκτιμητών μεγίστης πιθανοφάνειας. Επίσης, συνιστούν από τις αρχικές προσπάθειες συνδυασμού της αποτελεσματικότητας της μεθόδου ελαχίστων τετραγώνων και της ανθεκτικότητας των εκτιμητών ελαχίστων απολύτων αποκλίσεων, μέθοδοι οι οποίες και οι δύο αποτελούν ειδικές περιπτώσεις Μ-εκτιμητών.

Παρά το γεγονός ότι έχουν καλύτερη εφαρμογή όταν η αλλοίωση εντοπίζεται κυρίως στην κατεύθυνση των  $y_i$ , οι Μ-εκτιμητές θεωρούνται «ευαίσθητοι» σε δεδομένα που περιλαμβάνουν αποκλίνοσες  $x$  τιμές. Συνεπώς, το σημείο κατάρρευσης (Breakdown Point) των Μ-εκτιμητών είναι 0%, εξαιτίας του ότι δεν είναι ανθεκτικοί στα σημεία μόχλευσης (βλέπε *Rousseeuw και Leroy 1987, σελ.145*).

Με σκοπό να εξερευνήσουμε τις διάφορες πτυχές της ανθεκτικής στατιστικής, θα ξεκινήσουμε με το πιο απλό πλαίσιο, της εκτίμησης, δηλαδή, μίας μονοδιάστατης παραμέτρου με βάση τη συλλογή μονοδιάστατων παρατηρήσεων. Το πιο κοινό μονοδιάστατο πρόβλημα είναι η εκτίμηση μιας παραμέτρου θέσης. Στη συνέχεια, θα ασχοληθούμε με πολυδιάστατα προβλήματα όπως η πολυμεταβλητή συνδιασπορά-θέσης, η ανάλυση παλινδρόμησης και ο έλεγχος γραμμικών μοντέλων.

### 3.1.1 Μονοδιάστατη Περίπτωση

Στην μονοδιάστατη περίπτωση, λοιπόν, ο Μ-εκτιμητής  $T_n(X_1, X_2, \dots, X_n)$  για μια αυθαίρετη αντικειμενική συνάρτηση  $\rho$  και το δείγμα  $(X_1, X_2, \dots, X_n)$  είναι η τιμή του  $\beta$  που ελαχιστοποιεί τη σχέση

$$\sum_{i=1}^n \rho(X_i; T_n) \quad (3.1),$$

όπου  $\rho$  μία συνάρτηση με παράγωγο ως προς  $\beta$ ,  $\psi$ , η οποία θα είναι

$$\psi(x, \beta) = \frac{\partial}{\partial \beta} \rho(x, \beta).$$

Μπορούμε εναλλακτικά να υπολογίσουμε με μεγαλύτερη ευκολία τον M-εκτιμητή, σαν την τιμή του  $T_n$  που ικανοποιεί τη σχέση

$$\sum_{i=1}^n \psi(X_i; T_n) = 0 \quad (3.2).$$

Κάτω από αυτές τις υποθέσεις ένας M-εκτιμητής μπορεί να είναι  $\psi$ -τύπου – και είναι η λύση της εξίσωσης (3.2) – ή  $\rho$ -τύπου αν είναι η τιμή που ελαχιστοποιεί τη συνάρτηση (3.1).

Έστω  $G_n$  η εμπειρική αθροιστική συνάρτηση που προέρχεται από το δείγμα, τότε η λύση  $T_n$  της εξίσωσης (3.2) μπορεί επίσης να γραφεί σαν  $T(G_n)$  και  $T$  το συναρτησοειδές που δίνεται από τη σχέση

$$\int \psi(x, T(G)) dG(x) = 0 \quad (3.3),$$

για όλες τις κατανομές  $G$  για τις οποίες ορίζεται το ολοκλήρωμα. Στο σημείο αυτό, θα αντικαταστήσουμε στη σχέση (3.3) το  $G$  με  $F_{t,x} = (1-t)F + t\Delta_x$  και στη συνέχεια θα παραγωγίσουμε ως προς  $t$ . Οπότε παίρνουμε

$$\int \psi(y, T(F)) d(\Delta_x - F) + \int \frac{\partial}{\partial \beta} [\psi(y; \beta)]_{T(F)} dF(x) \frac{\partial}{\partial t} [T(F_{t,x})]_{t=0} = 0.$$

Κάνοντας χρήση των σχέσεων (2.1) και (3.3) παίρνουμε τη συνάρτηση επίδρασης (Influence Function) ενός M-εκτιμητή

$$IF(x; \psi, F) = \frac{\psi[x; T(F)]}{-\int (\partial/\partial \beta)[\psi(y; \beta)]_{T(F)} dF(y)} \quad (3.4),$$

με τη προϋπόθεση ότι ο παρανομαστής είναι διάφορος του μηδενός.

Η ασυμπτωτική διασπορά (Asymptotic Variance) ενός M-εκτιμητή είναι

$$V(T, F) = \int IF(x; F, T)^2 dF(x)$$

$$= \frac{\int \psi^2[x; T(F)] dF(x)}{[\int (\partial/\partial\beta)[\psi(y; \beta)]_{T(F)} dF(y)]^2}$$

### 3.1.2 Πολυπαραμετρική Περίπτωση

Οι Μ-εκτιμητές για την πολυπαραμετρική περίπτωση (multiparameter case) ορίζονται ακριβώς όπως εκείνοι για μία παράμετρο. Κάθε Μ-εκτιμητής που ορίζεται από μία παραγωγίσιμη συνάρτηση  $\rho$  αντιστοιχεί σε έναν εκτιμητή που ορίζεται από μία  $\psi$  συνάρτηση για την οποία ισχύει:  $\psi(x, \beta) = \frac{\partial}{\partial\beta} \rho(x, \beta)$ , υπό την προϋπόθεση ότι η τελευταία υπάρχει και είναι μοναδική. Οι Μ-εκτιμητές που ορίζονται από μία  $\psi$  συνάρτηση αποτελούν μία πιο ευρεία και πιο ευέλικτη κλάση, αφού οι  $\psi^{(j)}(x, \beta)$  δεν χρειάζεται να είναι μερικές παράγωγοι κάποιας συνάρτησης του  $\beta$  (βλέπε *Hampel et al (1986), σελ.230*). Για τον λόγο αυτό ορίζουμε τους Μ-εκτιμητές με τη βοήθεια  $\psi$  συναρτήσεων.

Η συνάρτηση επίδρασης παράγεται όπως η σχέση (3.4) και είναι

$$IF(x; T, F) = M(\psi, F)^{-1} \psi(x, T(F)) \quad (3.5),$$

όπου  $M$  ένας  $p \times p$  πίνακας για τον οποίο ισχύει

$$M(\psi, F)^{-1} := - \int \left[ \frac{\partial}{\partial\beta} \psi(x, \beta) \right]_{T(F)} dF(x).$$

Η ασυμπτωτική διασπορά ενός Μ-Εκτιμητή στην πολυδιάστατη περίπτωση δίνεται από τη σχέση

$$V(T, F) = M(\psi, F)^{-1} Q(\psi, F) M(\psi, F)^{-T},$$

όπου

$$Q(\psi, F) := \int \psi(x, T(F)) \psi(x, T(F))^T dF(x).$$

### 3.1.3 Μοντέλο Γραμμικής Παλινδρόμησης

Το μοντέλο της γραμμικής παλινδρόμησης αποτελεί ένα από τα πιο διαδεδομένα εργαλεία της στατιστικής ανάλυσης, ενώ η μέθοδος ελαχίστων τετραγώνων είναι η πιο γνωστή τεχνική εκτίμησης αυτού του μοντέλου. Όμως παρόλη τη μαθηματική απλότητα του, ο εκτιμητής ελαχίστων τετραγώνων (LSE) διακατέχεται από έλλειψη



ανθεκτικότητας, αφού μία και μοναδική έκτοπη παρατήρηση (απόμακρη) μπορεί να έχει σημαντική αρνητική επίδραση στην εκτίμηση. Σκοπός μας εδώ είναι να διερευνήσουμε τις ανθεκτικές ιδιότητες μίας ευρύτερης κλάσης εκτιμητών παλινδρόμησης στην οποία ανήκει και ο LSE. Οπότε, υποθέτουμε το παρακάτω γραμμικό μοντέλο

$$y_i = x_i^T \beta + e_i \quad i = 1, 2, \dots, n,$$

όπου,  $y_i$  η  $i^{\text{η}}$  παρατήρηση,

$x_i$  η  $i^{\text{η}}$  γραμμή (γραμμένη σαν διάνυσμα στήλης) του πίνακα σχεδιασμού  $X_{n \times p}$ ,

$\beta \in \theta \subset \mathbb{R}^p$  είναι ένα  $p$ -διάνυσμα αγνώστων παραμέτρων ( $p \geq 1$ ),

$e_i$  το  $i^{\text{ο}}$  σφάλμα.

Έστω ότι  $K(x)$  η συνάρτηση κατανομής του  $x_i$  και  $k(x)$  η συνάρτηση πυκνότητας πιθανότητας του. Επίσης, με  $f_{\theta}(x_i, y_i)$  δηλώνουμε την από κοινού συνάρτηση πυκνότητας των  $x_i$  και  $y_i$ , η οποία θα είναι

$$f_{\beta}(x, y) = \sigma^{-1} g\left(\frac{(y - x^T \beta)}{\sigma}\right) k(x),$$

και με  $F_{\beta}(x, y)$  η αντίστοιχη από κοινού συνάρτηση κατανομής.

Έτσι, για παράδειγμα, ένας εκτιμητής ελαχίστων τετραγώνων ( $T_n^{LS}$ ) θα είναι η τιμή του  $\beta$  που ελαχιστοποιεί τη σχέση

$$\Gamma(\beta) = \sum_{i=1}^n \left(\frac{(y_i - x_i^T \beta)}{\sigma}\right)^2,$$

δηλαδή ο εκτιμητής  $T_n^{LS}$  ορίζεται ως εξής

$$\Gamma(T^{LS}) = \min\{\Gamma(\beta) | \beta \in \theta\}.$$

Το 1973 ο Huber επέκτεινε τα αποτελέσματα του για την ανθεκτική εκτίμηση μίας παραμέτρου θέσης στην περίπτωση της γραμμικής παλινδρόμησης (βλέπε Huber, 1973a). Ουσιαστικά πρότεινε να υπολογίσει σταθμισμένες εκτιμήσεις ελαχίστων τετραγώνων με βάρη (redefined iteratively) της μορφής

$$w_i = \min \{1, c/|r_i|\},$$

όπου  $r_i$  είναι το  $i$  κατάλοιπο και  $c$  μία θετική σταθερά. Συνεπώς, τα βάρη δεν είναι σταθερά, αλλά εξαρτώνται από την εκτίμηση. Γενικότερα, ο Huber πρότεινε  $M$ -εκτιμητές  $T_n$  που ορίζονται ως εξής

$$\Gamma(T_n) = \min\{\Gamma(\beta) | \beta \in \Theta\},$$

όπου

$$\Gamma(\beta) = \sum_{i=1}^n \rho\left(\frac{(y_i - x_i^T \beta)}{\sigma}\right).$$

Για μία παραγωγίσιμη συνάρτηση  $\rho$  και  $\sigma$  σταθερό. Αν η συνάρτηση  $\rho$  έχει παράγωγο  $\frac{\partial}{\partial r} \rho(r) = \psi(r)$ , τότε ο  $T_n$  ικανοποιεί το παρακάτω σύστημα των εξισώσεων

$$\sum_{i=1}^n \psi\left(\frac{(y_i - x_i^T T_n)}{\sigma}\right) x_i = 0 \quad (3.6).$$

Στην πράξη, συνήθως πρέπει να εκτιμήσουμε την παράμετρο κλίμακας  $\sigma$  μαζί με την παράμετρο  $\beta$ . Ένας πιθανός τρόπος να το κάνουμε αυτό είναι να ελαχιστοποιήσουμε, με δεδομένο το  $A$ , τη σχέση

$$\sum_{i=1}^n \left[ \rho\left(\frac{(y_i - x_i^T \beta)}{\sigma}\right) + A \right] \sigma \quad (3.7).$$

Παραγωγίζοντας ως προς  $\beta$  (για σταθερό  $\sigma$ ) και  $\sigma$  (για σταθερό  $\beta$ ) και θέτοντας ίσο με μηδέν, παίρνουμε αντίστοιχα τις παρακάτω εξισώσεις

$$\sum_{i=1}^n \psi\left(\frac{(y_i - x_i^T T_n)}{\hat{\sigma}}\right) x_i = 0,$$

$$\sum_{i=1}^n \chi\left(\frac{(y_i - x_i^T T_n)}{\hat{\sigma}}\right) = 0,$$

όπου  $\psi(r) = \rho'(r)$  και  $\chi(r) = r\psi(r) - \rho(r) - A$ .

Ακολουθώντας Hampel et al. (1986) παραθέτουμε τις ανθεκτικές ιδιότητες του Huber Εκτιμητή,  $T^{Hu}$ , θα υπολογίσουμε την συνάρτηση επίδρασης με συνάρτηση κατανομής  $F_\beta(x, y)$  και πυκνότητα (αγνοώντας την παράμετρο κλίμακας,  $\sigma$ )

$$f_\beta(x, y) = \varphi(y - x^T \beta) k(x).$$

Χρησιμοποιώντας τη σχέση (3.5) με  $F = F_\beta$  και  $\psi(x, \beta) = \psi_c(y - x^T \beta)$  παίρνουμε την παρακάτω σχέση για τη συνάρτηση επίδρασης του Huber Εκτιμητή

$$IF(x, y; T^{Hu}, F_\beta) = \psi_c(y - x^T \beta) M^{-1} x \quad (3.8),$$

όπου:

$$M = (E\psi'_c)(Exx^T) = \left( \int \psi'_c(r) d\Phi(r) \right) \left( \int xx^T dK(x) \right).$$

Εισάγοντας την παράμετρο κλίμακας  $\sigma$  στη σχέση (3.8), η ανθεκτική συνάρτηση επίδρασης γίνεται

$$IF(x, y; T^{Hu}, F_\beta) = (Exx^T)^{-1} x \left( \frac{\psi(r/\sigma) \sigma}{E[\psi'(r/\sigma)]} \right).$$

Κάτω από τις κατάλληλες συνθήκες κανονικότητας (regularity), το  $T(F_n)$  δέχεται την παρακάτω ασυμπτωτική προσέγγιση, το ανάπτυγμα von Mises,

$$T(F_n) = T(F) + n^{-1} \sum_{i=1}^n IF(x, y; T^{Hu}, F_\beta) + R_n,$$

όπου ο όρος  $R_n$  αντιστοιχεί στο υπόλοιπο (remainder). Ειδικότερα

$$E\{IF[x, y; T^{Hu}, F_\beta]\} = 0, \text{ το οποίο συνεπάγεται ότι, } E[T(F_n)] \simeq T(F).$$

Η ασυμπτωτική διασπορά δίνεται από τη σχέση

$$V(T, F) = E\{IF[x, y; T^{Hu}, F_\beta]\}IF[x, y; T^{Hu}, F_\beta].$$

Η εμπειρική συνάρτηση επίδρασης, κάτω από την προσέγγιση του Hampel, ορίζεται ως εξής

$$IF[x, y; T^{Hu}, F_n] = n(X^T X)^{-1} \frac{\psi(r_i^{Hu}/\sigma) \sigma}{\frac{1}{n} \sum_{i=1}^n \psi'(r_i^{Hu}/\sigma)} x_i,$$

όπου ο όρος  $\psi(r_i^{Hu}/\sigma)$  αναφέρεται στη βιβλιογραφία σαν κατάλοιπα κατά Winsor κάτω από μία  $\psi$  συνάρτηση και μία παράμετρο κλίμακας  $\sigma$  (βλέπε Huber, 1981).

Βλέπουμε ότι η συνάρτηση επίδρασης της σχέσης (3.8) εξαρτάται από το  $y$  μόνο μέσω των καταλοίπων  $r := y - x^T \beta$ . Μπορούμε, λοιπόν, να ορίσουμε την IF σαν γινόμενο δύο παραγόντων, της επίδρασης των καταλοίπων (κλίμακας) (-scalar-influence of the residual, IR) και της επίδρασης της θέσης (-vector valued- influence of position in factor space, IP), ως εξής

$$\begin{aligned} IF(x, x^T \theta + r; T^{Hu}, F_\beta) &= IT(x, r; T^{Hu}, F_\beta) \\ &= IR(r; T^{Hu}, \Phi) IP(x; T^{Hu}, K), \end{aligned}$$

όπου

$$IR(r; T^{Hu}, \Phi) = \psi_c(r) / (E\psi'_c),$$

$$IP(x; T^{Hu}, K) = (Exx^T)^{-1} x.$$

Η παραπάνω παραγοντοποίηση είναι μοναδική αν ορίσουμε ως IR τη συνάρτηση επίδρασης του αντίστοιχου M-εκτιμητή θέσης (βλέπε Hampel et al., 1986, σελ.313). Ως IT ορίσαμε την συνολική επίδραση σε έναν εκτιμητή για τη συνάρτηση κατανομής  $F_\beta$ . Η επίδραση των καταλοίπων  $IR(r; T^{Hu}, \Phi)$  είναι φραγμένη. Το γεγονός αυτό αποτελεί μία βελτίωση σε σχέση με τον εκτιμητή Ελαχίστων Τετραγώνων από την άποψη της ανθεκτικότητας. Αντίθετα, η επίδραση της θέσης  $IP(x; T^{Hu}, K)$  είναι μη φραγμένη, που σημαίνει ότι αν προστεθεί μία έκτοπη παρατήρηση (outlier) στα δεδομένα μας, αυτή θα καθορίσει σχεδόν εντελώς την προσαρμογή (βλέπε Hampel, 1973a). Σε αυτή την περίπτωση ο Huber εκτιμητής και όλοι οι εκτιμητές που ορίζονται από τη σχέση (3.6), συμπεριλαμβανομένου και του L1-εκτιμητή, είναι το πρώτο βήμα για να κάνουμε έναν εκτιμητή παλινδρόμησης ανθεκτικό.

### 3.1.4 Αλγόριθμος M-εκτιμητή (Huber & Dutter, 1974)

Ο τυπικός αλγόριθμος για την ελαχιστοποίηση της σχέσης (3.7) έχει σαν πλεονέκτημα ότι είναι απλός, παρόλο τον αργό ρυθμό σύγκλισης του. Υποθέτουμε ένα επίπεδο ανεκτικότητας  $\varepsilon > 0$  και τιμές έναρξης,  $\sigma^{(0)}$ . Αν οι προβλεπόμενες (θεωρητικές) τιμές  $x_i^T \beta$  είναι γραμμικές, μπορούμε να χρησιμοποιήσουμε τον LSE σαν τιμή έναρξης. Ο αλγόριθμος, όπως τον διατύπωσαν οι Huber & Dutter (1974), είναι ο εξής

1.  $m := 0$ , παίρνουμε σαν τιμή έναρξης του  $\beta^{(0)}$  την τιμή του LSE.
2. Υπολογίζουμε τα κατάλοιπα  $\Delta_i$

$$\Delta_i := y_i - x_i^T \beta^{(m)} \quad i = 1, 2, \dots, n.$$

3. Υπολογίζουμε μία καινούργια τιμή για το  $\sigma$

$$(\sigma^{(m+1)})^2 := \frac{1}{A} \sum \chi \left( \frac{\Delta_i}{\sigma^{(m)}} \right) (\sigma^{(m)})^2,$$

όπου A: διορθωτικός παράγοντας για να πετύχουμε αμεροληψία.

4. Εκτιμούμε τα κατάλοιπα κατά Winsor

$$\Delta_i^* := \psi \left( \frac{\Delta_i}{\sigma^{(m+1)}} \right) \sigma^{(m+1)}.$$

5. Υπολογίζουμε τις μερικές παραγώγους

$$c_{ik} := \frac{d}{d\beta} (x_i^T \beta).$$

6. Λύνουμε:  $\sum (\Delta_i - \sum_k c_{ik} \hat{t}_k)^2 = \min \stackrel{i.e}{\Rightarrow}$  λύνουμε ως προς  $\hat{t}$ :  $C^T C \hat{t} = C^T \Delta$ .

7. Θέτουμε  $\beta^{(m+1)} := \beta^{(m)} + q \hat{t}$ , όπου  $0 < q < 2$  είναι ένας αυθαίρετος σταθερός παράγοντας.

8. Αν  $\|\hat{t}\| < \varepsilon$ , πηγαίνουμε στο βήμα 10, δηλαδή η επανάληψη σταματάει εφόσον η τυπική απόκλιση των εκτιμητών μεταβληθεί κατά λιγότερο από  $\varepsilon$  φορές.

9. Εναλλακτικά, θέτουμε  $m := m + 1$ , πάω στο βήμα 2.



10. Εκτίμηση  $\beta$  από  $\beta^{(m+1)}$ ,  $var(x_i)$  από  $(\sigma^{(m+1)})^2$ ,  $cov(\beta)$  από  $K^2 \frac{1}{n-p} \sum \Delta_i^2$ ,

όπου  $K$  είναι ένας διορθωτικός παράγοντας (βλέπε Huber (1974), σελ.168),

$$K = \frac{1 + \frac{p}{n} \frac{var\psi'}{(ave\psi')^2}}{ave\psi'}$$

Οι ποσότητες  $ave\psi'$  και  $var\psi'$  δίνονται από τις εξής σχέσεις

$$ave\psi' = \frac{1}{n} \sum \psi' \left( \frac{\Delta_i}{\sigma^{(m+1)}} \right) = \mu,$$

$$var\psi' = \frac{1}{n} \sum \left[ \psi' \left( \frac{\Delta_i}{\sigma^{(m+1)}} \right) - \mu \right]^2,$$

όπου  $\psi'(u) = \frac{d}{du} \psi(u)$ .

### 3.2 L-Εκτιμητές

Κάθε εκτιμητής που υπολογίζεται από έναν γραμμικό συνδυασμό στατιστικών στοιχείων μπορεί να χαρακτηριστεί ως L-εκτιμητής. Ο πρώτος L-εκτιμητής που προτάθηκε και είναι κάπως πιο ανθεκτικός από τον εκτιμητή ελαχίστων τετραγώνων είναι η παλινδρόμηση ελαχίστων απολύτων αποκλίσεων (least absolute deviations, LAD). Επίσης γνωστός και ως  $L_1$  Παλινδρόμηση επειδή ελαχιστοποιεί το « $L_1 - norm$ » (δηλαδή το άθροισμα απολύτων αποκλίσεων), ο LAD εκτιμητής αποτελεί την απλούστερη και παλαιότερη προσέγγιση της ανθεκτικής στατιστικής. Σε αυτόν τον ορισμό των L-εκτιμητών ταιριάζει επίσης η παλινδρόμηση ελαχίστων τετραγώνων, στην οποία μερικές φορές αναφερόμαστε σαν  $L_2$  Παλινδρόμηση επειδή ελαχιστοποιεί το « $L_2 - norm$ » (δηλαδή το άθροισμα τετραγώνων αποκλίσεων). Άλλοι γνωστοί L-εκτιμητές τους οποίους θα αναλύσουμε εκτενώς είναι ο εκτιμητής ελαχίστων διαμέσων τετραγώνων (least median squares, LMS) και ο εκτιμητής ελαχίστων περικεκομμένων τετραγώνων (least trimmed squares, LTS). Για λεπτομέρειες ο αναγνώστης μπορεί να απευθυνθεί στους Huber (1981), Hampel et al. (1986) και Rousseeuw & Leroy (1987).

### 3.2.1 Παλινδρόμηση Ελαχίστων Απολύτων Αποκλίσεων (Least Absolute Deviations, LAD)

Η LAD παλινδρόμηση είναι πολύ ανθεκτική σε δεδομένα με ασυνήθιστες τιμές  $y$ . Οι εκτιμήσεις βρίσκονται ελαχιστοποιώντας το άθροισμα των απολύτων τιμών των καταλοίπων, ως εξής

$$\min \sum_{i=1}^n |e_i| = \min \sum_{i=1}^n \left| y_i - \sum x_{ij} \beta_j \right|.$$

Επίσης, μπορεί να γενικευθεί σαν  $\alpha$ -παλινδρόμηση ποσοστιαίου σημείου ( $\alpha$ -regression quantile) και η συνάρτηση που θα πρέπει να ελαχιστοποιηθεί θα έχει την εξής μορφή

$$\sum_{i=1}^n \rho_a(e_i),$$

όπου αν  $\alpha$  είναι ένα ποσοστιαίο σημείο, τότε

$$\rho_a(e_i) = \begin{cases} \alpha e_i, & \text{αν } e_i \geq 0 \\ (1 - \alpha) e_i, & \text{αν } e_i < 0 \end{cases}.$$

Παρά το γεγονός ότι ο LAD εκτιμητής επηρεάζεται λιγότερο σε σχέση με τον LSE από ακραίες τιμές στην κατεύθυνση του  $y$ , αποτυγχάνει να αντιμετωπίσει τα αρνητικά σημεία μόχλευσης (βλέπε *Mosteller and Tukey, 1977*), δηλαδή την αλλοίωση στην κατεύθυνση των  $x$ . Συνεπώς, το σημείο κατάρρευσης του LAD εκτιμητή είναι 0. Επίσης, οι LAD εκτιμητές έχουν χαμηλή αποδοτικότητα. Ο συνδυασμός χαμηλής αποδοτικότητας και χαμηλού σημείου κατάρρευσης καθιστά την  $L_1$  Παλινδρόμηση λιγότερο ελκυστική σε σχέση με άλλες μεθόδους ανθεκτικής στατιστικής.

### 3.2.2 Παλινδρόμηση Ελαχίστων Διαμέσων Τετραγώνων (Least Median Squares, LMS)

Η παλινδρόμηση ελαχίστων διαμέσων τετραγώνων (LMS) προτάθηκε για πρώτη φορά από τον *Rousseeuw (1984)* και αντικαθιστά το άθροισμα των τετραγωνισμένων καταλοίπων που χαρακτηρίζει τον LSE με την αποδεδειγμένα πιο ανθεκτική διάμεσο τους. Οι εκτιμήσεις, λοιπόν, βρίσκονται ως εξής

$$\min \text{Med}(y_i - \sum x_{ij} \beta_j)^2 = \min \text{Med}(e_i^2) \quad (3.9),$$

όπου  $Med$  η διάμεσος.

#### Ιδιότητες της Μεθόδου Ελαχίστων Διαμέσων Τετραγώνων:

- Υπάρχει πάντα μία λύση για την εξίσωση (3.9).
- Αντικαθιστώντας το άθροισμα με τη διάμεσο προκύπτει ένας εκτιμητής ανθεκτικός σε ακραίες τιμές. Αν ο αριθμός των παραμέτρων της παλινδρόμησης  $p$  είναι μεγαλύτερος της μονάδας, δηλαδή  $p > 1$ , τότε το σημείο κατάρρευσης του εκτιμητή LMS είναι 50% και ορίζεται από τον τύπο

$$\varepsilon_n^*(T, Z) = ([n/2] - p + 2)/n.$$

- Παρά το γεγονός ότι επιτυγχάνεται ένα ιδανικό σημείο κατάρρευσης της τάξης του 50%, ο LMS έχει σημαντικές ελλείψεις που περιορίζουν τη χρήση του. Έχει στην καλύτερη περίπτωση μία σχετική αποτελεσματικότητα, περίπου 37%, όταν τα σφάλματα είναι κανονικά καταναμημένα και η τιμή σύγκλισης της μεθόδου είναι  $n^{-1/3}$  (βλέπε *Rousseeuw και Croux, 1993*). Συνεπώς, ο εκτιμητής LMS δεν είναι ασυμπτωτικά κανονικός. Ο αργός αυτός ρυθμός σύγκλισης του εκτιμητή θα μπορούσε να βελτιωθεί χρησιμοποιώντας διαφορετική αντικειμενική συνάρτηση (βλέπε *Rousseeuw, 1984*).
- Στερείται μίας καλά ορισμένης συνάρτησης επίδρασης, ωστόσο μπορούμε να πάρουμε μία ιδέα σχετικά με τις ανθεκτικές του ιδιότητες, κατασκευάζοντας τις καμπύλες ευαισθησίας του (βλέπε *Andrews et. al., 1972*).
- Ο εκτιμητής LMS ικανοποιεί και τις τρεις ιδιότητες ισοδυναμίας: ίσης μεταβολής ως προς την παλινδρόμηση, ίσης μεταβολής ως προς την κλίμακα και ως προς τη θέση και την κλίμακα.

### **3.2.3 Παλινδρόμηση Ελαχίστων Περικεκομμένων Τετραγώνων (Least Trimmed Squares, LTS)**

Μία άλλη μέθοδος που αναπτύχθηκε από τον *Rousseeuw (1984)* είναι η παλινδρόμηση ελαχίστων περικεκομμένων τετραγώνων (LTS). Συνιστά έναν εναλλακτικό τρόπο βελτίωσης του αργού ρυθμού σύγκλισης του εκτιμητή LMS και επιτυγχάνεται με την χρήση διαφορετικής αντικειμενικής συνάρτησης. Έτσι, αντί να προσθέσουμε όλα τα τετραγωνισμένα κατάλοιπα (όπως στην μέθοδο LSE), μπορούμε

να περιοριστούμε σε ένα περικεκομμένο άθροισμα τετραγώνων. Αρχικά, λοιπόν, τα διατεταγμένα τετραγωνισμένα κατάλοιπα από το μικρότερο στο μεγαλύτερο είναι

$$(r^2)_1 \leq (r^2)_2 \leq \dots \leq (r^2)_n.$$

Στη συνέχεια αρκεί να προσθέσουμε μόνο τα πρώτα  $q$  από αυτά τα τετραγωνισμένα κατάλοιπα, οπότε ο εκτιμητής LTS δίνεται από τη σχέση

$$\min \sum_{i=1}^q e_i^2 \quad (3.10).$$

Σαν επέκταση του περικομμένου μέσου (trimmed mean)- ο οποίος προκύπτει από τον μέσο των παρατηρήσεων που απομένουν αφού περικόψουμε το άνω και κάτω 25% των διατεταγμένων παρατηρήσεων- η παλινδρόμηση ΕΠΤ ελαχιστοποιεί το άθροισμα των περικεκομμένων τετραγωνισμένων καταλοίπων.

Όπως αναφέρθηκε και πρωτίτερα,  $q$  είναι ο αριθμός των παρατηρήσεων που απομένουν μετά την περικοπή του δείγματος κατά  $\alpha\%$  και συμπεριλαμβάνονται στον υπολογισμό του εκτιμητή. Τότε

$$q = [n(1 - \alpha) + 1].$$

- Για  $\alpha = 50\%$ , (δηλαδή αν περικόψουμε τις παρατηρήσεις μας κατά 50%), τότε

$$q = [n/2] + 1,$$

και ο εκτιμητής LTS μπορεί να αντισταθεί στην επίδραση έκτοπων παρατηρήσεων κατά 50% όπως και ο εκτιμητής LMS.

- Αν  $q$  είναι ίσο με

$$q = [n/2] + [(p + 1)/2],$$

ο εκτιμητής LTS επιτυγχάνει το μέγιστο δυνατό σημείο κατάρρευσης, το οποίο θα ορίζεται από τη σχέση

$$\varepsilon_n^*(T, Z) = ([n - p]/2 + 1)/n.$$

#### Ιδιότητες της Μεθόδου Ελαχίστων Περικεκομμένων Τετραγώνων:

- Υπάρχει πάντα μία λύση για την εξίσωση (3.10).

- Σε αντίθεση με τον αργό ρυθμό σύγκλισης του εκτιμητή LMS, η τιμή σύγκλισης της μεθόδου LTS είναι  $n^{-1/2}$ , με την ίδια ασυμπτωτική αποτελεσματικότητα (στην κανονική κατανομή) με τον M-Εκτιμητή που ορίζεται από την συνάρτηση

$$\psi(x) = \begin{cases} x, & |x| \leq \Phi^{-1}(1 - \alpha/2) \\ 0, & \text{διαφορετικά} \end{cases},$$

και αναφέρεται ως Huber-type skipped mean. Η συνάρτηση επίδρασης τέτοιων M-Εκτιμητών με συμμετρική κατανομή F δίνεται από τη σχέση

$$IF(y; T, F) = \frac{\psi(y)}{\int \psi' dF} = \frac{\psi(y)}{-\int \psi(z) f'(z) dz},$$

με ασυμπτωτική διασπορά

$$V(\psi, F) = \frac{\int \psi^2 dF}{(\int \psi' dF)^2} \quad (3.11).$$

Αντίστοιχα, η συνάρτηση επίδρασης του εκτιμητή LTS δίνεται από τον τύπο (βλέπε Rousseeuw και Leroy 1987, σελ. 191)

$$IF(y; LTS, \Phi) = \frac{y l_{[-q, q]}(y)}{2\Phi(q) - 1 - 2q\Phi(q)},$$

ή εναλλακτικά:

$$IF(y; LTS, \Phi) = \begin{cases} 14.021y, & \text{αν } |y| \leq q \\ 0, & \text{διαφορετικά} \end{cases},$$

με  $q = \Phi^{-1}(0.75) = 0.6745$ .

Από την (3.11), η ασυμπτωτική διασπορά του εκτιμητή LTS θα είναι

$$V(LTS, \Phi) = 14,021.$$

- Ο εκτιμητής LTS ικανοποιεί και τις τρεις ιδιότητες ισοδυναμίας: ίσης μεταβολής ως προς την παλινδρόμηση, ίσης μεταβολής ως προς την κλίμακα και ως προς τη θέση και την κλίμακα.
- Παρά την υψηλή του ανθεκτικότητα, που οφείλεται στο υψηλό σημείο κατάρρευσης που πετυχαίνει, η μέθοδος LTS δεν κρίνεται αρκετά

αποτελεσματική όταν στο δείγμα μας υπάρχουν περισσότερες έκτοπες παρατηρήσεις σε σχέση με τις περικεκομμένες.

- Το κυριότερο μειονέκτημα της μεθόδου LTS έγκειται στο γεγονός ότι η αντικειμενική της συνάρτηση απαιτεί την ταξινόμηση των τετραγωνισμένων καταλοίπων.

ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΕΡΑΙΑ



# Κεφάλαιο 4<sup>ο</sup>: Εκτίμηση Αποθεμάτων Ζημιών (Claims Reserving)

Όπως έχουμε αναφέρει, κύρια υποχρέωση μίας ασφαλιστικής εταιρείας είναι η εκτίμηση των μελλοντικών της αποζημιώσεων και κατ' επέκταση ο υπολογισμός ενός αποθεματικού για την κάλυψη των εκκρεμών υποχρεώσεων της προς τους ασφαλισμένους πελάτες.

Οι απαιτήσεις/ζημιές που μελετάμε είναι αυτές που είναι γνωστό ότι υπάρχουν, αλλά το τελικό μέγεθος τους είναι άγνωστο κατά τη στιγμή που πρέπει να καθοριστούν τα αποθέματα. Στη βιβλιογραφία αναφέρονται σαν τις απαιτήσεις που συνέβησαν αλλά δεν δηλώθηκαν εξολοκλήρου (Incurred But Not Reported Completely, IBNRC). Οι απαιτήσεις στις οποίες επικεντρωνόμαστε είναι αυτές που παίρνουν μήνες ή και χρόνια για να εμφανιστούν, ανάλογα με την πολυπλοκότητα των ζημιών.

Η καθυστέρηση της πληρωμής μπορεί να οφείλεται, για παράδειγμα, λόγω πολυετών νομικών διαδικασιών ή δυσκολιών στον καθορισμό του μεγέθους των αξιώσεων. Ως εκ τούτου οι ασφαλιστές θα πρέπει να έχουν δημιουργήσει αποθεματικό που θα τους δώσει τη δυνατότητα να πληρώσουν τις εκκρεμείς ζημιές και να ανταποκριθούν στις απαιτήσεις που θα προκύψουν στο μέλλον.

Ας υποθέσουμε, λοιπόν, ότι έχουμε το ακόλουθο σύνολο ατομικών ζημιών

$$\{P_{ij}, i = 1, 2, \dots, n; j = 1, 2, \dots, n\},$$

όπου  $i$  αναφέρεται στη γραμμή και δείχνει το έτος ατυχήματος ή το εν λόγω ασφαλιστικό έτος, ενώ το  $j$  αναφέρεται στη στήλη και απεικονίζει το έτος καθυστέρησης ή εξέλιξης.

Στην περίπτωση που δεν είναι ευδιάκριτοι οι δείκτες  $i$  και  $j$ , θα βάζουμε ανάμεσα τους ένα “,”.

Οι ατομικές αυτές ζημιές αναπαριστώνται στο παρακάτω τρίγωνο εξέλιξης ζημιών (Πίνακας 4.1).

Έτος Ατυχήματος ( $i$ )	Έτος Εξέλιξης ( $j$ )						
	1	2	...	$k$	...	$n-1$	$n$
1	$P_{11}$	$P_{12}$	...	$P_{1k}$	...	$P_{1,n-1}$	$P_{1n}$
2	$P_{21}$	$P_{22}$	...	$P_{2k}$	...	$P_{2,n-1}$	
...	...	...	...	...	...		
$i$	$P_{i1}$	$P_{i2}$	...	$P_{ik}$			
...	...	...	...				
$n-1$	$P_{n-1,1}$	$P_{n-1,2}$					
$n$	$P_{n1}$						

Πίνακας 4.1: Τρίγωνο Εξέλιξης Ζημιών

Οι απαιτήσεις που βρίσκονται πάνω στη διαγώνιο (για τις οποίες ισχύει:  $i + j - 1 = c$ ) αναπαριστούν τις ζημιές που έγιναν το ημερολογιακό έτος  $c$ . Οι τεχνικές αποθεματοποίησης μπορούν εύκολα να επεκταθούν για εξαμηνιαία, τριμηνιαία ή και μηνιαία στοιχεία, αλλά οι περισσότερες ασφαλιστικές εταιρείες χρησιμοποιούν ετήσια στοιχεία. Στις περισσότερες περιπτώσεις δεν έχει σημασία αν χρησιμοποιούνται ατομικά ή αθροιστικά δεδομένα (απαιτήσεις). Συχνά χρειάζονται και οι δύο μορφές και η μετάβαση από τη μία στην άλλη είναι πολύ εύκολη. Οι αθροιστικές ζημιές ορίζονται ως οι τιμές των  $C_{ij}$  και για  $i + j \leq n + 1$  είναι γνωστές στον ερευνητή. Το ζητούμενο είναι να υπολογίσουμε τις εκκρεμείς ζημιές για τις οποίες  $i + j > n + 1$  και ιδιαίτερα τα μέγιστα ποσά αποζημιώσεων  $C_{in}$  για κάθε έτος ατυχήματος  $i = 1, 2, \dots, n$ . Τα εκκρεμή αποθεματικά ζημιών για κάθε έτος ατυχήματος  $i$  θα είναι

$$R_i = C_{in} - C_{i,n-i+1}, \quad 1 \leq i \leq n,$$

υπό την προϋπόθεση ότι  $C_{i,n-i+1}$  έχουν ήδη πληρωθεί. Η ασφαλιστική εταιρεία, βέβαια, ενδιαφέρεται περισσότερο για το συνολικό απαιτούμενο αποθεματικό  $R$ , που αντιστοιχεί στο άθροισμα των  $R_i$  αποθεματικών, για  $i = 1, 2, \dots, n$ . Τα δεδομένα μας έχουν τη μορφή ενός τριγώνου εξέλιξης ζημιών, όπως θα δούμε παρακάτω.

Σκοπός της αποθεματοποίησης ζημιών (claims reserving) είναι να προβλέπει ζημιές που θα πληρωθούν σε μελλοντικά ημερολογιακά έτη. Σκοπός μας είναι να συμπληρώσουμε το τρίγωνο εξέλιξης ζημιών ώστε να έχει τη μορφή τετραγώνου,

επειδή το άθροισμα των ζημιών που θα βρίσκονται στο κάτω δεξί μέρος του τριγώνου στο συνολικό αποθεματικό  $R$  που θα πρέπει να πληρωθεί στο μέλλον. Υπάρχουν διάφορες μέθοδοι εκτίμησης αποθεματικών ζημιών που βασίζονται σε διαφορετικές υποθέσεις και έχουν διαφορετικές αξιώσεις. Στο επόμενο κεφάλαιο θα περιγράψουμε μία από τις παλαιότερες και πλέον δημοφιλείς μεθόδους εκτίμησης εκκρεμών αποθεμάτων ζημιών, την Chain Ladder μέθοδο.

#### 4.1 Μέθοδος Chain Ladder

Η Chain Ladder μέθοδος, CL (ή ελληνιστί μέθοδος τριγώνου εξέλιξης ζημιών), λόγω της ευκολίας στην εφαρμογή της στα δεδομένα των ασφαλιστικών εταιριών, είναι μία από τις πιο γνωστές μεθόδους πρόβλεψης αποθεματικών. Αποτελεί μία στατιστική μέθοδο εκτίμησης εκκρεμών ζημιών κατά την οποία ο σταθμισμένος μέσος όρος των προηγούμενων απαιτήσεων προβάλλεται στο μέλλον. Η πρόβλεψη αυτή βασίζεται στην αναλογία μεταξύ των παρελθοντικών σωρευτικών ζημιών, πληρωθεισών ή εκκρεμών, για την εκτίμηση των μελλοντικών εκκρεμών απαιτήσεων (για τα διαδοχικά έτη εξέλιξης). Απαιτεί το νωρίτερο έτος ατυχήματος να είναι πλήρως απορροές ή τουλάχιστον ότι το τελικό αποτέλεσμα για το συγκεκριμένο έτος να μπορεί να εκτιμηθεί με βεβαιότητα. Συνεπώς, η CL μέθοδος βασίζεται στην ανάλυση της ιστορικότητας των αποζημιώσεων.

Στη μέθοδο αυτή τα δεδομένα μας εισάγονται σε ένα τρίγωνο εξέλιξης ζημιών (run-off triangle), όπως τον Πίνακα 4.1, του οποίου κάθε γραμμή παριστά το έτος ατυχήματος (accident year) και κάθε στήλη το έτος εξέλιξης (development year). Σαν έτος ατυχήματος θεωρούμε το έτος εκείνο κατά το οποίο πραγματοποιήθηκε το ζημιογόνο γεγονός, ενώ σαν έτος εξέλιξης τον αριθμό των ετών που μεσολάβησαν μέχρι την αποπληρωμή της ζημιάς. Επίσης, κάθε διαγώνιος του τριγώνου αποτελεί ένα ημερολογιακό έτος (calendar year). Οι γραμμές του τριγώνου, εκτός από το έτος ατυχήματος, θα μπορούσαν να παριστάνουν και το έτος αναγγελίας της ζημιάς (reporting year) ή το έτος σύναψης του ασφαλιστήριου συμβολαίου (policy year).

Τα δεδομένα του τριγώνου εξέλιξης ζημιών μπορεί να είναι ατομικές (incremental) ή σωρευτικές (cumulative) ζημιές. Η κλασική Chain Ladder μέθοδος επεξεργάζεται αθροιστικές ζημιές για τον υπολογισμό των συντελεστών εξέλιξης, ως το πηλίκο δύο διαδοχικών ποσών ζημιάς, και στη συνέχεια αυτοί πολλαπλασιαζόμενοι με τις

σφαιρικές ζημιές οδηγούν στην εκτίμηση των μελλοντικών εκκρεμών ζημιών. Οι σφαιρικές ζημιές προκύπτουν από τις ατομικές ως εξής

$$C_{i,k} = \sum_{j=1}^k P_{ij}.$$

Στη συνέχεια υπολογίζουμε τους συντελεστές εξέλιξης (development factors), ως το πηλίκο δύο διαδοχικών ποσών ζημιών (Mack, 1993), δηλαδή

$$\hat{f}_k = \frac{\sum_{i=1}^{n-k} C_{i,k+1}}{\sum_{i=1}^{n-k} C_{i,k}}, \quad \text{για } 1 \leq k \leq n-1 \quad (4.1),$$

οπότε οι εκτιμώμενες τελικές αθροιστικές αποζημιώσεις θα είναι

$$\hat{C}_{i,n} = C_{i,n+1-i} \hat{f}_{n+1-i} \dots \hat{f}_{n-1} \quad (4.2).$$

Συνεπώς, από τον τύπο (4.1) υπολογίζουμε τους σταθμισμένους συντελεστές εξέλιξης ανά έτος ατυχήματος

$\hat{f}_1$	$\hat{f}_2$	...	$\hat{f}_{n-1}$
-------------	-------------	-----	-----------------

Τελικά παίρνουμε τον Πίνακα (4.2) με τις εκτιμώμενες αθροιστικές εκκρεμείς ζημιές και τα αντίστοιχα αποθεματικά ζημιών για κάθε έτος ατυχήματος.

Έτος Ατυχήματος (i)	Έτος Εξέλιξης j							Πληρωθείσες Ζημιές	Απόθεμα Ζημιών
	1	2	...	k	...	n-1	n		
1							$C_{1,n}$	$C_{1,n}$	$\hat{R}_1$
2						$C_{2,n-1}$	$\hat{C}_{2,n}$	$C_{2,n-1}$	$\hat{R}_2$
...					...	...	...	...	...
i				$C_{i,k}$	...	$\hat{C}_{i,n-1}$	$\hat{C}_{i,n}$	$C_{i,k}$	$\hat{R}_i$
...			...	...	...	...	...	...	...
n-1		$C_{n-1,2}$	...	$\hat{C}_{n-1,k}$	...	$\hat{C}_{n-1,n-1}$	$\hat{C}_{n-1,n}$	$C_{n-1,2}$	$\hat{R}_{n-1}$
n	$C_{n,1}$	$\hat{C}_{n,2}$	...	$\hat{C}_{n,k}$	...	$\hat{C}_{n,n-1}$	$\hat{C}_{n,n}$	$C_{n,1}$	$\hat{R}_n$

Πίνακας 4.2: Εκτιμώμενες Αθροιστικές Ζημιές και Αποθεματικά

όπου η στήλη n αντιστοιχεί στις εκτιμήσεις των τελικών ζημιών.

Το αντίστοιχο προβλεπόμενο αποθεματικό για κάθε έτος ατυχήματος  $i$ ,  $R_i$  θα είναι

$$\hat{R}_i = C_{i,n+1-i}(\hat{f}_{n+1-i} \dots \hat{f}_{n-1} - 1) \quad (4.3).$$

Αναλυτικά, για να εκτιμήσουμε τις αθροιστικές εκκρεμείς αποζημιώσεις πολλαπλασιάζουμε τις σωρευτικές πληρωθείσες ζημιές με τους αντίστοιχους συντελεστές εξέλιξης, για παράδειγμα

$$\hat{C}_{n,2} = C_{n,1} \cdot \hat{f}_1,$$

$$\hat{C}_{2,n} = C_{2,n-1} \cdot \hat{f}_{n-1}.$$

Για να υπολογίσουμε το απόθεμα ζημιών για κάθε έτος ατυχήματος  $i$ , θα πρέπει να αφαιρέσουμε τις πληρωθείσες ζημιές από τις εκτιμήσεις των τελικών ζημιών, δηλαδή

$$\hat{R}_1 = C_{1,n} - C_{1,n} = 0,$$

$$\hat{R}_2 = \hat{C}_{2,n} - C_{2,n-1},$$

...

$$\hat{R}_{n-1} = \hat{C}_{n-1,n} - C_{n-1,2},$$

$$\hat{R}_n = \hat{C}_{n,n} - C_{n,1}.$$

Το συνολικό απόθεμα  $\hat{R}$ , που θα πρέπει να έχει μία ασφαλιστική εταιρεία προκειμένου να ανταποκριθεί στις μελλοντικές της υποχρεώσεις προς τους πελάτες της, είναι το άθροισμα των παραπάνω εκτιμώμενων αποθεματικών  $\hat{R}_i$ , επομένως είναι

$$\hat{R} = \hat{R}_1 + \hat{R}_2 + \dots + \hat{R}_n.$$

## 4.2 Λογαριθμικό Μοντέλο Παλινδρόμησης (Verall, 1991)

Αν υποθέσουμε ότι

- $A_i$ : οι τελικές σωρευτικές ζημιές για το έτος ατυχήματος  $i$ ,
- $B_j$ : το ποσοστό των τελικών ζημιών που πληρώθηκαν κατά το έτος εξέλιξης  $j$ ,
- $P_{ij}$ : οι ατομικές (μη σωρευτικές) πληρωθείσες ζημιές για το έτος ατυχήματος  $i$  που πληρώθηκαν κατά το έτος εξέλιξης  $j$ .

Το μοντέλο της Chain Ladder μεθόδου μπορεί, δηλαδή, να περιγραφεί από τις παρακάτω σχέσεις

$$P_{ij} = A_i * B_j, \quad \text{για } i, j \text{ από } 1 \text{ έως } n \quad (4.4),$$

υπό την προϋπόθεση ότι

$$\sum B_j = 1 \quad (4.5).$$

Παίρνοντας λογαρίθμους για την σχέση (4.4) οδηγούμαστε στην παρακάτω γραμμική εξίσωση

$$\ln(P_{ij}) = \ln(A_i * B_j) = \ln(A_i) + \ln(B_j).$$

Ενώ αντίστοιχα παίρνοντας λογαρίθμους για την σχέση (4.5) δεν οδηγούμαστε σε γραμμική εξίσωση αφού

$$\ln\left(\sum B_j\right) \neq \sum \ln(B_j)$$

Επομένως, δεν θα λάβουμε υπόψη μας τη σχέση (4.5) και θα προσπαθήσουμε να εκτιμήσουμε τις παραμέτρους μας από τη γραμμική σχέση (4.4), την οποία και θα πρέπει να αναπροσαρμόσουμε προκειμένου να πάρουμε ακριβείς εκτιμήσεις εισάγοντας και έναν όρο σφάλματος στο μοντέλο μας. Υποθέτοντας, λοιπόν, τα παρακάτω

- $a_i = \ln(A_i)$ ,
- $b_j = \ln(B_j)$ ,
- $e_{ij}$  κάποιο σφάλμα που υποθέτουμε ότι ακολουθεί κανονική κατανομή με μέσο 0 (δηλαδή  $e_{ij} \sim N(0, I\sigma^2)$ ).

Το αναπροσαρμοσμένο μοντέλο μας είναι

$$\ln(P_{ij}) = Y_{ij} = a_i + b_j + e_{ij} \quad (4.6).$$

Η πιο γενικά

$$Y = X\beta + e,$$

όπου  $Y$ :  $(n \times 1)$  διάνυσμα λογαριθμισμένων ζημιών,



$X$ : ( $n \times p$ ) πίνακας σχεδιασμού,

$\beta$ : ( $p \times 1$ ) διάνυσμα παραμέτρων,

$e$ : ( $n \times 1$ ) διάνυσμα τυχαίων κανονικών σφαλμάτων,

$\sigma^2 > 0$ : άγνωστη παράμετρος διακύμανσης.

Έχουμε, λοιπόν, υποθέσει ότι  $Y_{ij} = \ln(P_{ij})$ . Τότε οι  $Y_{ij}$ , με  $i = 1, 2, \dots, n; j = 1, 2, \dots, n - i + 1$ , είναι ανεξάρτητες, κανονικά κατανοημένες τ.μ. για τις οποίες ισχύει

$$E(Y_{ij}) = x_{ij}^T \beta \text{ και } \text{Var}(Y_{ij}) = \sigma^2,$$

όπου  $x_{ij}^T$  είναι ένα διάνυσμα γραμμής με  $p$  επεξηγηματικές μεταβλητές (σταθερές) και  $\beta$  ένα ( $p \times 1$ ) διάνυσμα αγνώστων συντελεστών παλινδρόμησης. Σε μορφή πινάκων το παραπάνω μοντέλο γράφεται ως εξής

$$E(Y) = X\beta \text{ και } \text{Var}(Y_{ij}) = \sigma^2.$$

Αν υποθέσουμε ότι  $E(P_{ij}) = \theta_{ij}$ , τότε η αναμενόμενη τιμή του  $\theta_{ij}$  είναι

$$\theta_{ij} = \exp\left(x_{ij}^T \beta + \frac{1}{2} \sigma^2\right),$$

και ο εκτιμητής μεγίστης πιθανοφάνειας του  $\theta_{ij}$  είναι

$$\hat{\theta}_{ij} = \exp\left(x_{ij}^T \hat{\beta} + \frac{1}{2} \hat{\sigma}^2\right),$$

όπου  $\hat{\beta}$  είναι μία εκτιμήτρια του  $\beta$  και  $\hat{\sigma}^2$  μία εκτίμηση του  $\sigma^2$

$$\hat{\beta} = (X^T X)^{-1} X^T Y \quad (4.7).$$

Μία αμερόληπτη εκτιμήτρια  $\hat{s}^2$  του  $\hat{\sigma}^2$  είναι

$$\hat{s}^2 = \frac{1}{(n - p)} (Y - X\hat{\beta})^T (Y - X\hat{\beta}) \quad (4.8).$$

Συνεπώς μία αμερόληπτη εκτίμηση για το  $\theta_{ij}$  (βλέπε Verall, 1991), είναι

$$\tilde{\theta}_{ij} = \exp(x_{ij}^T \hat{\beta}) g_m \left[ \frac{1}{2} (1 - x_{ij}^T (X^T X)^{-1} x_{ij}) s^2 \right] \quad (4.9),$$

όπου

$$g_m(t) = \sum_{k=0}^{\infty} \frac{m^k (m+2k)}{m(m+2) \dots (m+2k)} \frac{t^k}{k!},$$

η συνάρτηση που εισήγαγε ο Finney (1941) με σκοπό να διορθώσει τη μεροληψία και  $m$  είναι οι βαθμοί ελευθερίας που σχετίζονται με το  $\hat{\sigma}^2$  (στη δική μας περίπτωση  $m = n - 1$ ). Επίσης έχουμε

$$\text{Cov}(\hat{\beta}) = (X^T X)^{-1} \sigma^2 \quad \text{και} \quad \text{Cov}(x_{ij}^T \hat{\beta}) = x_{ij}^T (X^T X)^{-1} x_{ij} \sigma^2.$$

Η διασπορά της αμερόληπτης εκτίμησης του  $\theta_{ij}$  είναι  $\tau_{ij}^2$ . Μία αμερόληπτη εκτιμήτρια του  $\tau_{ij}^2$  (βλέπε Verall, 1991, σελ.79), είναι

$$\begin{aligned} \tilde{\tau}_{ij}^2 = \exp(2x_{ij}^T \hat{\beta}) & \left[ g_m \left( \frac{1}{2} (1 - x_{ij}^T (X^T X)^{-1} x_{ij}) s^2 \right) \right. \\ & \left. - g_m \left( (1 - 2x_{ij}^T (X^T X)^{-1} x_{ij}) s^2 \right) \right] \quad (4.10). \end{aligned}$$

### 4.3 Σύντομη Εφαρμογή Λογαριθμοκανονικού Μοντέλου

Υποθέτουμε ένα τρίγωνο εξέλιξης ζημιών που περιλαμβάνει τις ατομικές ζημιές  $P_{ij}$ , για 3 έτη ατυχήματος και 3 έτη εξέλιξης.

Έτος Ατυχήματος ( $i$ )	Έτος Εξέλιξης ( $j$ )		
	1	2	3
1	$P_{11}$	$P_{12}$	$P_{13}$
2	$P_{21}$	$P_{21}$	
3	$P_{31}$		

Πίνακας 4.3: Τρίγωνο Εξέλιξης Ζημιών

Έστω  $\theta_{ij} = E(P_{ij})$ . Το ζητούμενο είναι η εκτίμηση των  $\theta_{ij}$  και των τυπικών αποκλίσεων αυτών των εκτιμήσεων. Ειδικότερα, θέλουμε να εκτιμήσουμε τις αναμενόμενες εκκρεμείς ζημιές  $\{\theta_{ij}, i = 2, 3; j = 3 - i + 2\}$ . Επίσης πρέπει να

ληφθούν υπόψη οι συνολικές εκτιμώμενες ζημιές κάθε γραμμής αφού ουσιαστικά πρόκειται για τις συνολικές ζημιές κάθε έτους εργασίας.

Υποθέτουμε, επίσης, ότι οι ατομικές ζημιές  $P_{ij}$  είναι ανεξάρτητες και λογαριθμοκανονικά κατανομημένες, δηλαδή ότι οι λογάριθμοι τους,  $Y_{ij} = \ln(P_{ij})$ , είναι ανεξάρτητες, κανονικά κατανομημένες ζημιές για τις οποίες ισχύει

$$E(Y_{ij}) = X_{ij}\beta \quad \text{και} \quad \text{Var}(Y_{ij}) = \sigma^2,$$

όπου  $X_{ij}$  είναι ένα διάνυσμα γραμμής με  $p$  επεξηγηματικές μεταβλητές (σταθερές) και  $\beta$  ένα  $(p \times 1)$  διάνυσμα αγνώστων συντελεστών παλινδρόμησης.

Το γραμμικό μοντέλο που θα χρησιμοποιήσουμε για την πρόβλεψη των εκκρεμών ζημιών είναι

$$E(Y) = X\beta.$$

Οι παράμετροι που θα πρέπει να εκτιμήσουμε χρησιμοποιώντας πολλαπλή γραμμική παλινδρόμηση (multiple linear regression) μεταξύ των ετών εξέλιξης με τη βοήθεια πινάκων, είναι συνολικά 5 και είναι οι εξής

$$(\alpha_1, \alpha_2, \alpha_3, b_2, b_3).$$

Οι υποθέσεις του μοντέλου μας θα είναι

- $Y$ :  $(6 \times 1)$  διάνυσμα λογαρίθμων ατομικών ζημιών,
- $X$ :  $(6 \times 5)$  πίνακας σχεδιασμού που αποτελείται από 0 ή 1,
- $\beta$ :  $(5 \times 1)$  διάνυσμα των παραμέτρων προς εκτίμηση,
- $e$ :  $(6 \times 1)$  διάνυσμα σφαλμάτων,
- $b_1 = 0$ .

Στον Πίνακα 4.4 αναπαριστώνται αναλυτικά οι λογάριθμοι των ατομικών ζημιών ( $Y_{ij} = \ln(P_{ij})$ ) και ο πίνακας σχεδιασμού  $X$  που περιλαμβάνει τις παραμέτρους του μοντέλου μας και καθορίζει τη θέση των ζημιών στο τρίγωνο εξέλιξης ανάλογα με την τιμή που αναγράφει (0 ή 1).

<i>i</i>	<i>j</i>	<i>Y<sub>ij</sub></i>	Πίνακας Σχεδιασμού <i>X</i>				
			$\alpha_1$	$\alpha_2$	$\alpha_3$	$b_2$	$b_3$
1	1	$Y_{11}$	1	0	0	0	0
1	2	$Y_{12}$	1	0	0	1	0
1	3	$Y_{13}$	1	0	0	0	1
2	1	$Y_{21}$	0	1	0	0	0
2	2	$Y_{22}$	0	1	0	1	0
3	1	$Y_{31}$	0	0	1	0	0

Πίνακας 4.4: Λογάριθμοι Ατομικών Ζημιών ( $Y_{ij}$ ) και Πίνακας Σχεδιασμού  $X$

Συνεπώς κάθε γραμμή αντιστοιχεί σε μία ατομική ζημιά του τριγώνου εξέλιξης καθώς και στη θέση που κατέχει σε αυτό ανάλογα με την τιμή των παραμέτρων στον πίνακα σχεδιασμού. Για παράδειγμα η 3<sup>η</sup> γραμμή αντιπροσωπεύει τη ζημιά που πραγματοποιήθηκε το 1<sup>ο</sup> έτος ατυχήματος και πληρώθηκε μετά από 3 έτη και υπολογίζεται σαν το άθροισμα των παραμέτρων  $\alpha_1$  και  $b_3$  ( $Y_{13} = \alpha_1 + b_3$ ).

#### 4.3.1 Εκτίμηση μελλοντικών ατομικών ζημιών

Στη συνέχεια θα εκτιμήσουμε τις ατομικές ζημιές ( $P_{ij}$ ) από τις υπολογισμένες τιμές των  $\ln(P_{ij})$ , χρησιμοποιώντας την προσέγγιση των *Zehnwirth* και *Renshaw* (1985), που αργότερα χρησιμοποιήθηκε και από τον *Christofides* (1990). Οι εκτιμήσεις που παράγονται από αυτή την προσέγγιση δεν είναι απαραίτητως αμερόληπτες ειδικότερα σε περιπτώσεις που αφορούν μικρό αριθμό δεδομένων.

Οι μελλοντικές (εκτιμώμενες) ατομικές ζημιές  $\hat{P}_{ij}$  υπολογίζονται από τις εκτιμήσεις των εκτιμώμενων λογαρίθμων  $\hat{Y}_{ij}$  ως εξής

$$\hat{P}_{ij} = \exp(\hat{Y}_{ij} + 0.5Var(\hat{Y}_{ij})) \quad (4.11).$$

Τα τυπικά σφάλματα υπολογίζονται ως εξής:

$$s.e.(\hat{P}_{ij}) = \hat{P}_{ij} \sqrt{\exp(Var(\hat{Y}_{ij}))^{-1}} \quad (4.12).$$

Για τον υπολογισμό των διακυμάνσεων των  $\hat{Y}_{ij}$  απαιτείται η κατασκευή του πίνακα διακυμάνσεων - συνδιακυμάνσεων (variance-covariance matrix), ο οποίος δίνεται από τη σχέση

$$\sigma^2 X_f (X^T X)^{-1} X_f^T,$$

όπου  $\sigma^2$ : η διακύμανση του μοντέλου που εξαρτάται από τα δεδομένα,

$X_f$ : ο μελλοντικός πίνακας σχεδιασμού,

$X_f^T$ : ο ανάστροφος του μελλοντικού πίνακα σχεδιασμού  $X_f$ .

<i>i</i>	<i>j</i>	Μελλοντικός Πίνακας Σχεδιασμού $X_f$				
		$\alpha_1$	$\alpha_2$	$\alpha_3$	$b_2$	$b_3$
2	3	0	1	0	0	1
3	2	0	0	1	1	0
3	3	0	0	1	0	1

Πίνακας 4.5: Μελλοντικός Πίνακας Σχεδιασμού

Οπότε, σχεδιάζουμε τον παραπάνω πίνακα  $X_f (X^T X)^{-1} X_f^T$  και στη συνέχεια πολλαπλασιάζουμε με την αντίστοιχη διακύμανση  $\sigma^2$  του εκάστωτε μοντέλου. Οι διακυμάνσεις των μελλοντικών τιμών του μοντέλου μας είναι το άθροισμα των διαγωνίων του πίνακα διακυμάνσεων- συνδιακυμάνσεων συν την διακύμανση του μοντέλου  $\sigma^2$ , δηλαδή

$$Var(\hat{Y}_{ij}) = \sigma^2 [I + X_f (X^T X)^{-1} X_f^T],$$

τα οποία και θα αντικαταστήσουμε στις σχέσεις (4.11) και (4.12) για να καταλήξουμε στις τιμές των πραγματικών εκτιμώμενων ατομικών ζημιών και τις τυπικές τους αποκλίσεις.

#### 4.4 Ανθεκτικό Λογαριθμοκανονικό Μοντέλο Παλινδρόμησης

##### Βήμα 1ο: Υπολογισμός ανθεκτικής εκτιμήτριας του $\beta$

Στο σημείο αυτό θα περιγράψουμε τον τρόπο υπολογισμού του ανθεκτικού M εκτιμητή παλινδρόμησης,  $\hat{\beta}^M$  εκτελώντας τον αλγόριθμο που πρότειναν οι Huber & Dutter το 1974 τον οποίο και περιγράφουμε αναλυτικά στην υποενότητα 3.1.4 της παρούσας εργασίας. Ουσιαστικά αποτελεί ένα πρόβλημα ελαχιστοποίησης, δηλαδή

$$\min_{\hat{\beta}} \sum_{i=1}^n \rho\left(\frac{r_i}{\sigma}\right) \quad (4.13),$$

όπου  $r_i = y_i - x_i^T \beta$  και  $\sigma$  μία παράμετρος κλίμακας την οποία πρακτικά θα πρέπει να εκτιμήσουμε παράλληλα με το  $\hat{\beta}^M$ . Το πρόβλημα ελαχιστοποίησης της (4.13) ισοδυναμεί με τη ταυτόχρονη λύση των παρακάτω δύο εξισώσεων:

$$\sum_{i=1}^n \psi\left(\frac{r_i}{\sigma}\right) x_i = 0,$$

$$\sum_{i=1}^n \chi\left(\frac{r_i}{\sigma}\right) = A,$$

όπου  $\psi(u) = \rho'(u)$  και  $\chi(u) = u\psi(u) - \rho(u)$ . Αν θέλουμε η παράμετρος κλίμακας  $\sigma$  να είναι ασυμπτωτικά αμερόληπτη για κανονικά σφάλματα, παίρνουμε

$$A = \left[\frac{n-p}{n}\right] E_{\Phi}(\chi),$$

με  $\Phi$  την κανονική κατανομή.

Ο Huber (1973, 1981), πρότεινε τη χρησιμοποίηση της παρακάτω  $\rho$  συνάρτησης

$$\rho(u) = \begin{cases} \frac{u^2}{2}, & \text{αν } |u| \leq c \\ c|u| - \frac{c^2}{2}, & \text{αν } |u| > c \end{cases}, \quad c > 0,$$

και  $\chi$  αποδεικνύεται ότι είναι ίση με  $\chi(u) = [\psi(u)]^2/2$ .

## Βήμα 2ο: Υπολογισμός ανθεκτικού πίνακα συνδιακύμανσης του $\hat{\beta}$

Ο πίνακας συνδιακύμανσης της ανθεκτικής εκτιμήτριας,  $\hat{\beta}^M$ , έχει την εξής μορφή

$$\text{Cov}(\hat{\beta}^M) = \widehat{\sigma}_M^2 (X'X)^{-1} = K^2 \frac{1}{n-p} \frac{\sum_{i=1}^n \psi\left(\frac{r_i}{\sigma}\right)^2 \sigma^2}{\left[\left(\frac{1}{n}\right) \sum_{i=1}^n \psi'\left(\frac{r_i}{\sigma}\right)\right]^2} (X'X)^{-1} \quad (4.14),$$

όπου  $K$  είναι ένας διορθωτικός παράγοντας για να πετύχουμε αμεροληψία, (βλέπε Huber (1981), σελ.173)



$$K = 1 + \frac{p \operatorname{Var} \left[ \psi' \left( \frac{r_n}{\sigma} \right) \right]}{n E \left[ \psi' \left( \frac{r_n}{\sigma} \right) \right]^2} \quad (4.15).$$

Οι ποσότητες  $E \left[ \psi' \left( \frac{r_n}{\sigma} \right) \right]$  και  $\operatorname{Var} \left[ \psi' \left( \frac{r_n}{\sigma} \right) \right]$  στην πράξη είναι άγνωστες και εκτιμούνται από τις σχέσεις

$$E \left[ \psi' \left( \frac{r_n}{\sigma} \right) \right] \approx \frac{1}{n} \sum \psi' \left( \frac{r_i}{\sigma} \right) = \mu,$$

$$\operatorname{Var} \left[ \psi' \left( \frac{r_n}{\sigma} \right) \right] \approx \frac{1}{n} \sum \left[ \psi' \left( \frac{r_i}{\sigma} \right) - \mu \right]^2,$$

όπου  $\psi'(u) = \frac{d}{du} \psi(u)$ .

Στην ιδιαίτερη περίπτωση που η συνάρτηση  $\psi$  έχει την εξής μορφή

$$\psi(x) = \min[c, \max(-c, x)],$$

η (4.15) απλοποιείται στη

$$K = 1 + \frac{p}{n} \frac{1 - \mu}{\mu},$$

όπου  $\mu$  είναι μία σχετική συχνότητα των καταλοίπων που ικανοποιεί τη  $-c < \frac{r_i}{\sigma} < c$ .

### Βήμα 3ο: Υπολογισμός ανθεκτικής M-εκτίμησης του $\tilde{\theta}_{ij}$

Λαμβάνοντας υπόψη τη σχέση (4.9) και αντικαθιστώντας με τις αντίστοιχες ανθεκτικές ποσότητες, παίρνουμε ασυμπτωτικά την αμερόληπτη ανθεκτική εκτίμηση για το  $\tilde{\theta}_{ij}$ ,  $\tilde{\theta}_{ij}^M$ , ως εξής

$$\tilde{\theta}_{ij}^M = \exp(x_{ij}^T \hat{\beta}^M) g_m \left[ \frac{1}{2} (1 - x_{ij}^T (X^T X)^{-1} x_{ij}) s_M^2 \right] \quad (4.16).$$

### Βήμα 4ο: Υπολογισμός ανθεκτικής M-εκτίμησης του $\tilde{\theta}_{ij}$

Λαμβάνοντας υπόψη τη σχέση (4.10) και αντικαθιστώντας με τις αντίστοιχες ανθεκτικές ποσότητες, παίρνουμε την ανθεκτική διακύμανση,  $\tilde{\tau}_{Mij}^2$ , ως εξής

$$\begin{aligned} \tilde{t}_{Mij}^2 = \exp(2x'_{ij}\hat{\beta}^M) & \left[ g_m \left( \frac{1}{2} (1 - x'_{ij}(X'X)^{-1}x_{ij})s_M^2 \right) \right. \\ & \left. - g_m \left( (1 - 2x'_{ij}(X'X)^{-1}x_{ij})s_M^2 \right) \right] \quad (4.17). \end{aligned}$$

#### 4.5 IBNR Ζημιές και Μοντέλο Ανάλυσης Διακύμανσης Δύο Παραγόντων (Kremer,1982)

Μία εναλλακτική μέθοδος για την εκτίμηση των αναμενόμενων τιμών των IBNR ζημιών προέρχεται από τον μετασχηματισμό του κλασσικού πολλαπλασιαστικού μοντέλου σε ένα μοντέλο ανάλυσης διακύμανσης δύο παραγόντων (two-way ANOVA), το οποίο μελετήθηκε εκτενώς από τον *E. Kremer* (“*IBNR-Claims and the Two-way Model of ANOVA*” (1982)) και παρουσιάζεται περιληπτικά παρακάτω.

Αν υποθέσουμε ότι  $C_{ij}, i, j = 1, 2, \dots, n$  είναι μη αρνητικές τυχαίες μεταβλητές, οι οποίες αντιπροσωπεύουν τις αθροιστικές ζημιές για όλα τα έτη εξέλιξης  $j$  σε κάθε έτος ατυχήματος  $i$ . Στην πράξη γνωρίζουμε μόνο τις ζημιές που βρίσκονται στο πάνω αριστερό μέρος του τριγώνου εξέλιξης ζημιών,  $C_{\Delta}$ , δηλαδή

$$C_{\Delta} = (C_{ij}, i = 1, 2, \dots, n, j = 1, 2, \dots, n - i + 1).$$

Ακολουθώς θεωρούμε το πρόβλημα εκτίμησης της αναμενόμενης τιμής της τελικής αθροιστικής ζημιάς  $C_{in}$ , δηλαδή

$$E_i = E(C_{in}).$$

Θεωρούμε ως  $P_{ij}$  τις ατομικές (μη σωρευτικές) ζημιές

$$P_{ij} = C_{ij} - C_{i,j-1}$$

$C_{i0} := 0$  και υποθέτουμε ότι:  $P_{ij} > 0, \forall j \leq n - i + 1, \forall i$ .

Οπότε, το βασικό μας μοντέλο είναι

$$P_{ij} = E_i S_j R_{ij} \quad (4.18),$$

όπου

- $S_j$ : ο λόγος (αναλογία) της αύξησης των αναμενόμενων ζημιών κατά το  $j$  έτος εξέλιξης και των αναμενόμενων τελικών ζημιών και θα ισχύει

$$\sum_{j=1}^n S_j = 1,$$

- $R_{ij}, i = 1, 2, \dots, n, j = 1, 2, \dots, n - i + 1$ : τυχαίες μεταβλητές με  $E(R_{ij}) = 1$ .

Στη συνέχεια θα μετασχηματίσουμε το βασικό μας μοντέλο (4.18) παίρνοντας λογαρίθμους

$$Y_{ij} = \mu + \alpha_i + b_j + e_{ij} \quad (4.19),$$

για  $\forall j \leq n - i + 1, \forall i$  και

$$Y_{ij} = \ln(P_{ij}),$$

$$e_{ij} = \ln(R_{ij}),$$

$$\alpha_i = \ln(E_i) - \frac{1}{n} \sum_{l=1}^n \ln(E_l),$$

$$b_j = \ln(S_j) - \frac{1}{n} \sum_{l=1}^n \ln(S_l),$$

$$\mu = \frac{1}{n} \sum_{l=1}^n (\ln(E_l) + \ln(S_l)).$$

Επίσης θα ισχύει

$$\sum_{i=1}^n \alpha_i = \sum_{j=1}^n b_j = 0 \quad (4.20).$$

Κάτω από τις πρόσθετες υποθέσεις για τα σφάλματα  $e_{ij}, j \leq n - i + 1, i \leq n$  είναι ασυσχέτιστα με

$$E(e_{ij}) = 0, \quad \text{Var}(e_{ij}) = \sigma^2 \in (0, \infty), \quad j \leq n - i + 1, \quad i \leq n.$$

Το μοντέλο που σχηματίζεται από τον συνδυασμό των σχέσεων (4.19) και (4.20) είναι σχεδόν ίδιο με το κλασσικό μοντέλο ανάλυσης διακύμανσης δύο παραγόντων (χωρίς αλληλεπίδραση), με τη μόνη διαφορά ότι οι ατομικές ζημιές  $P_{ij}$  είναι

άγνωστες για  $j > n - i + 1$ , δηλαδή τα δεδομένα μας έχουν την μορφή τριγώνου εξέλιξης ζημιών (βλέπε Πίνακα 4.2).

Παρά το γεγονός αυτό μπορούμε να υπολογίσουμε βέλτιστους εκτιμητές για τις παραμέτρους  $\mu, \alpha_i, b_j$ , (βλέπε *Kremer (1982)*, σελ. 49), ως εξής

**Θεώρημα 1.** Έστω  $\hat{\mu}, \hat{\alpha}_i, \hat{b}_j$  με  $i, j = 1, 2, \dots, n$  οι λύσεις των

$$\hat{\alpha}_i = \frac{1}{n-i+1} \sum_{j=1}^{n-i+1} \left( Y_{ij} - \frac{1}{n-j+1} \sum_{l=1}^{n-j+1} Y_{lj} - \frac{1}{n-j+1} \sum_{l=n-j+2}^n \hat{\alpha}_l \right),$$

$$\text{με } \hat{\alpha}_n = Y_{n1} - \frac{1}{n} \sum_{i=1}^n Y_{i1},$$

$$\hat{b}_j = \frac{1}{n-j+1} \sum_{i=1}^{n-j+1} \left( Y_{ij} - \frac{1}{n-i+1} \sum_{l=1}^{n-i+1} Y_{il} - \frac{1}{n-i+1} \sum_{l=n-i+2}^n \hat{b}_l \right),$$

$$\text{με } \hat{b}_n = Y_{1n} - \frac{1}{n} \sum_{j=1}^n Y_{1j},$$

$$\hat{\mu} = \frac{2}{n(n+1)} \sum_{i=1}^n \sum_{j=1}^{n-i+1} (Y_{ij} - \hat{\alpha}_i - \hat{b}_j).$$

Τότε οι  $\hat{\mu}, \hat{\alpha}_i, \hat{b}_j$  είναι οι καλύτερες γραμμικές αμερόληπτες εκτιμήτριες (**best linear unbiased estimators, BLUEs**) των  $\mu, \alpha_i, b_j$ .

## 4.6 Ανθεκτική Ανάλυση Διακύμανσης Δύο Παραγόντων

Στη στατιστική, συνήθως θεωρείται ότι τα σφάλματα είναι κανονικά κατανομημένα. Στην περίπτωση αυτή, η κλασική ανάλυση διακύμανσης (ANOVA) που βασίζεται στη μέθοδο ελαχίστων τετραγώνων μας δίνει ικανοποιητικά αποτελέσματα. Ωστόσο, σε αρκετές περιπτώσεις τα δεδομένα ενδέχεται να περιέχουν έκτοπες παρατηρήσεις, οι οποίες ασκούν σημαντική ανεπιθύμητη επιρροή στη κλασική προσαρμογή ελαχίστων τετραγώνων και επακόλουθη παραπλανητική ανάλυση στο πλαίσιο ANOVA.

Ανεξάρτητα από την αιτία εμφάνισης των ακραίων αυτών τιμών, αυτό που είναι αναγκαίο είναι να υπάρχει καλή ανθεκτική προσαρμογή του μοντέλου, που να

αντικατοπτρίζει καλά το μεγαλύτερο μέρος των δεδομένων. Τότε, οι ακραίες τιμές θα είναι σαφώς εκτεθειμένες στα κατάλοιπα για περαιτέρω μελέτη. Σε γενικές γραμμές, αποτελεί ένα πρόβλημα ελαχιστοποίησης, της μορφής

$$\min \sum_{i=1}^n \sum_{j=1}^n \rho(Y_{ij} - (\mu + \alpha_i + b_j + e_{ij})),$$

για μία αντικειμενική συνάρτηση  $\rho(\cdot)$ .

Η εφαρμογή του γενικού ανθεκτικού μοντέλου ανάλυσης διακύμανσης (ANOVA) πραγματοποιείται χρησιμοποιώντας διάφορες μεθόδους, όπως στην περίπτωση της ανθεκτικής γραμμικής παλινδρόμησης. Στην εφαρμογή θα χρησιμοποιήσουμε τον LAD εκτιμητή ο οποίος είναι αρκετά διαδεδομένος στη βιβλιογραφία και στις εφαρμογές ανθεκτικής εκτίμησης ANOVA.

#### 4.6.1 Εφαρμογή του LAD Εκτιμητή στο Γενικό Μοντέλο Ανθεκτικής ANOVA

Στη συνέχεια, θα παρουσιάσουμε περιληπτικά πως μπορούν να επιτευχθούν αποτελεσματικές LAD εκτιμήσεις για το γενικό μοντέλο ανθεκτικής ανάλυσης διακύμανσης δύο παραγόντων όπως παρουσιάζεται στους Armstrong & Frome (1979). Το γενικό μοντέλο ανάλυσης διακύμανσης δύο παραγόντων έχει τη μορφή

$$Y_{ijk} = \mu + \alpha_i + b_j + e_{ijk}, \quad \text{με } i = 1, \dots, r; j = 1, \dots, c; k = 1, \dots, n_{ij}.$$

Οι LAD εκτιμητές των παραμέτρων υπολογίζονται με την λύση του παρακάτω προβλήματος

$$\min Q = \sum_{i=1}^r \sum_{j=1}^c \sum_{k=1}^{n_{ij}} |Y_{ijk} - (\mu + \alpha_i + b_j)| \quad (4.21),$$

όπου,  $Y_{ijk}$  είναι η  $k$ -οστή παρατήρηση που βρίσκεται στο  $i$ -οστό επίπεδο του 1<sup>ου</sup> παράγοντα και στο  $j$ -οστό επίπεδο του 2<sup>ου</sup> παράγοντα,  $\alpha_i$  αντιπροσωπεύει την επίδραση της γραμμής,  $b_j$  αντιπροσωπεύει την επίδραση της  $j$  στήλης και  $\mu$  μία σταθερά. Υπάρχουν δύο βαθμοί ελευθερίας που ισοδυναμούν με το πλήθος των ανεξαρτήτων μεταβλητών, έτσι θα πρέπει να προστεθούν περιορισμοί στο παραπάνω

πρόβλημα ελαχιστοποίησης. Ομοίως με τη μέθοδο εκτίμησης ελαχίστων τετραγώνων, στη LAD ανάλυση, οι εκτιμήσεις που ελαχιστοποιούν τη σχέση

$$\sum_{i=1}^r |\alpha_i| + \sum_{j=1}^c |b_j| \quad (4.22),$$

παράγονται έτσι ώστε να επετεύχθη μία βέλτιστη τιμή για το Q της σχέσης (4.21). Αυτό το πρόσθετο κριτήριο δεν παράγει απαραίτητα μία μοναδική λύση. Περαιτέρω περιορισμοί ή διαφορετικό σύνολο κριτηρίων θα μπορούσε να αποδώσει μία και μοναδική λύση.

Θέτοντας  $\tau_i = \mu + \alpha_i$ , το πρόβλημα ελαχιστοποίησης της (4.22) γράφεται σαν ένα πρόβλημα γραμμικού προγραμματισμού ως εξής

$$\min \sum_{i=1}^r \sum_{j=1}^c \sum_{k=1}^{n_{ij}} (d_{ijk}^+ - d_{ijk}^-) \quad (4.23),$$

με τους παρακάτω περιορισμούς

$$\tau_i + b_j - Y_{ijk} + d_{ijk}^+ - d_{ijk}^- = 0,$$

$$d_{ijk}^+ \geq 0, \quad d_{ijk}^- \geq 0,$$

$$i = 1, 2, \dots, r; j = 1, 2, \dots, c; k = 1, 2, \dots, n_{ij},$$

όπου  $d_{ijk}^+$  και  $d_{ijk}^-$  είναι οι αντίστοιχες θετικές και αρνητικές αποκλίσεις της εξίσωσης παλινδρόμησης από την παρατήρηση (βλέπε *Armstrong & Frome, 1979*).

**Σημείωση:** Στη περίπτωση της ανθεκτικής εκτίμησης του τελικού αποθέματος, το εσωτερικό άθροισμα (αριθμός παρατηρήσεων σε κάθε κελί,  $k$  στη σχέση (4.21)) δεν απαιτείται, επειδή σε κάθε κελί έχουμε μία ζημιά ( $k = 1$ ) και ο αριθμός των επιπέδων για κάθε παράγοντα είναι ο ίδιος, δηλαδή  $r = c = n$ .



# Κεφάλαιο 5<sup>ο</sup>: Εφαρμογή

## 5.1 Εφαρμογή Λογαριθμικού Μοντέλου Παλινδρόμησης (0 outliers)

Το τρίγωνο εξέλιξης ζημιών (run-off triangle) που θα μελετήσουμε αποτελείται από ατομικές (μη σωρευτικές) ζημιές και χρησιμοποιήθηκε για πρώτη φορά από τους Taylor & Ashe, 1983), αλλά και από άλλους μεταγενέστερους ερευνητές (βλέπε Verall 1990, 1991, Mack 1993, Verdonck & Debruyne, 2010). Το τρίγωνο εμπεριέχει 10 έτη ατυχήματος ( $i$ ) και 10 έτη εξέλιξης ( $j$ ), όπως βλέπουμε στον Πίνακα 5.1.

Έτος Ατυχήματος ( $i$ )	Έτος Εξέλιξης ( $j$ )									
	1	2	3	4	5	6	7	8	9	10
1	357848	766940	610542	482940	527326	574398	146342	139950	227229	67948
2	352118	884021	933894	1183289	445745	320996	527804	266172	425046	
3	290507	1001799	926219	1016654	750816	146923	495992	280405		
4	310608	1108250	776189	1562400	272482	352053	206286			
5	443160	693190	991983	769488	504851	470639				
6	396132	937085	847498	805037	705960					
7	440832	847631	1131398	1063269						
8	359480	1061648	1443370							
9	376686	986608								
10	344014									

Πίνακας 5.1: Τρίγωνο Εξέλιξης Ζημιών Taylor and Ashe (1983)

Για να προβλέψουμε τις εκκρεμείς ζημιές και τα απαιτούμενα αποθεματικά θα εφαρμόσουμε το αναπροσαρμοσμένο λογαριθμοκανονικό μοντέλο παλινδρόμησης όπως ορίστηκε στην Ενότητα 4.2 και περιγράφηκε στην 4.3, δηλαδή

$$\ln(P_{ij}) = Y_{ij} = a_i + b_j + e_{ij} \quad i = 1, \dots, 10; j = 1, \dots, 10,$$

όπου  $P_{ij}$  οι ατομικές αποζημιώσεις,  $Y_{ij}$  οι λογάριθμοί τους,  $a_i$  η παράμετρος που προσδιορίζει το έτος ατυχήματος που συνέβη μία ζημιά (γραμμές) και  $b_j$  η παράμετρος που καθορίζει μετά από πόσα έτη αυτή πληρώθηκε εξ ολοκλήρου (στήλες). Στο μοντέλο μας οι λογάριθμοι των ατομικών ζημιών ( $Y_{ij}$ ) είναι οι

εξαρτημένες μεταβλητές και η θέση αυτών στο τρίγωνο εξέλιξης, που αντιπροσωπεύεται από τις στήλες του πίνακα σχεδιασμού  $X$ , είναι οι ανεξάρτητες μεταβλητές.

Συνεπώς, το τρίγωνο ζημιών που θα μελετήσουμε τελικά περιέχει τους λογαρίθμους των ατομικών ζημιών, όπως εμφανίζονται στον Πίνακα 5.2.

Έτος Ατυχήματος ( $i$ )	Έτος Εξέλιξης ( $j$ )									
	1	2	3	4	5	6	7	8	9	10
1	12.788	13.550	13.322	13.088	13.176	13.261	11.894	11.849	12.334	11.126
2	12.772	13.692	13.747	13.984	13.008	12.679	13.176	12.492	12.960	
3	12.579	13.817	13.739	13.832	13.529	11.898	13.114	12.544		
4	12.646	13.918	13.562	14.262	12.515	12.772	12.237			
5	13.002	13.449	13.807	13.553	13.132	13.062				
6	12.890	13.751	13.650	13.599	13.467					
7	12.996	13.650	13.939	13.877						
8	12.792	13.875	14.182							
9	12.839	13.802								
10	12.748									

Πίνακας 5.2: Τρίγωνο Εξέλιξης Λογαρίθμων Ζημιών Taylor and Ashe (1983)

Οι παράμετροι που θα εκτιμήσουμε είναι συνολικά 18 και είναι οι εξής

$$(\alpha_2, \alpha_3, \alpha_4, \alpha_5, \alpha_6, \alpha_7, \alpha_8, \alpha_9, \alpha_{10}, b_2, b_3, b_4, b_5, b_6, b_7, b_8, b_9, b_{10}).$$

Πιο γενικά το μοντέλο προς εκτίμηση είναι

$$Y = X\beta + e,$$

όπου  $Y$ :  $(55 \times 1)$  διάνυσμα λογαρίθμων ατομικών ζημιών,

$X$ :  $(55 \times 18)$  πίνακας σχεδιασμού που αποτελείται από 0 ή 1,

$\beta$ :  $(18 \times 1)$  διάνυσμα των παραμέτρων προς εκτίμηση,

$e$ :  $(55 \times 1)$  διάνυσμα σφαλμάτων,

$$\alpha_1 = 0, b_1 = 0.$$

Στον Πίνακα 5.3 αναπαριστώνται αναλυτικά οι ατομικές ζημιές  $P_{ij}$ , οι λογάριθμοι τους,  $Y_{ij} = \ln(P_{ij})$ , και ο πίνακας σχεδιασμού  $X$  που περιλαμβάνει τις παραμέτρους του μοντέλου μας και καθορίζει την θέση των ζημιών στο τρίγωνο εξέλιξης ανάλογα με την τιμή που αναγράφει (0 ή 1).

$i$	$j$	$P_{ij}$	$Y_{ij}$	Πίνακας Σχεδιασμού $X$																	
				$\mu$	$\alpha_2$	$\alpha_3$	$\alpha_4$	$\alpha_5$	$\alpha_6$	$\alpha_7$	$\alpha_8$	$\alpha_9$	$\alpha_{10}$	$b_2$	$b_3$	$b_4$	$b_5$	$b_6$	$b_7$	$b_8$	$b_9$
1	1	357848	12.79	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1	2	766940	13.55	1	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0
1	3	610542	13.32	1	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0
1	4	482940	13.09	1	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0
1	5	527326	13.18	1	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0
1	6	574398	13.26	1	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0
1	7	146342	11.89	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0
1	8	139950	11.85	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0
1	9	227229	12.33	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
1	10	67948	11.13	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
2	1	352118	12.77	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
...	...	...	...	.....																	
9	1	376686	12.84	1	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0
9	2	986608	13.80	1	0	0	0	0	0	0	0	1	0	1	0	0	0	0	0	0	0
10	1	344014	12.75	1	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0

Πίνακας 5.3: Ατομικές Ζημιές ( $P_{ij}$ ), λογάριθμοι ( $Y_{ij}$ ) και Πίνακας Σχεδιασμού  $X$

Συνεπώς κάθε γραμμή αντιστοιχεί σε μία ατομική ζημιά του τριγώνου εξέλιξης καθώς και στην θέση που κατέχει σε αυτό ανάλογα με την τιμή των παραμέτρων στον πίνακα σχεδιασμού. Για παράδειγμα, στον Πίνακα 5.3, η 10<sup>η</sup> γραμμή αντιπροσωπεύει την ζημιά που πραγματοποιήθηκε το 1<sup>ο</sup> έτος ατυχήματος και πληρώθηκε μετά από 10 έτη και υπολογίζεται σαν το άθροισμα των παραμέτρων της τεταγμένης  $\mu$  και των  $\alpha_1$  και  $b_{10}$  ( $Y_{1,10} = \mu + \alpha_1 + b_{10}$ ).

### 5.1.1 Εκτιμήσεις Παραμέτρων

Εφαρμόζοντας πολλαπλή παλινδρόμηση στα δεδομένα του Πίνακα 5.3, λαμβάνουμε τους συντελεστές παλινδρόμησης όπως εμφανίζονται στον Πίνακα 5.4.

<i>Παράμετροι</i>	<i>Εκτίμηση</i>	<i>Παράμετροι</i>	<i>Εκτίμηση</i>
<i>Intercept</i>	12.5198		
<i>a2<sup>LS</sup></i>	0.361	<i>b2<sup>LS</sup></i>	0.9112
<i>a3<sup>LS</sup></i>	0.2822	<i>b3<sup>LS</sup></i>	0.9387
<i>a4<sup>LS</sup></i>	0.1712	<i>b4<sup>LS</sup></i>	0.965
<i>a5<sup>LS</sup></i>	0.2822	<i>b5<sup>LS</sup></i>	0.3832
<i>a6<sup>LS</sup></i>	0.3117	<i>b6<sup>LS</sup></i>	-0.0049
<i>a7<sup>LS</sup></i>	0.392	<i>b7<sup>LS</sup></i>	-0.1181
<i>a8<sup>LS</sup></i>	0.4803	<i>b8<sup>LS</sup></i>	-0.4393
<i>a9<sup>LS</sup></i>	0.3452	<i>b9<sup>LS</sup></i>	-0.0535
<i>a10<sup>LS</sup></i>	0.2286	<i>b10<sup>LS</sup></i>	-1.3933

Πίνακας 5.4: Εκτιμήσεις Παραμέτρων με τη μέθοδο LS (0 outliers)

Στη συνέχεια, για να συμπληρώσουμε το τρίγωνο εξέλιξης ζημιών με τις μελλοντικές τιμές των λογαρίθμων των  $P_{ij}$  αρκεί να πολλαπλασιάσουμε τον πίνακα των εκτιμημένων παραμέτρων με τον μελλοντικό πίνακα σχεδιασμού προσθέτοντας πάντα τον όρο τεταγμένη (intercept). Για παράδειγμα, η προβλεπόμενη ζημιά που έγινε το 5<sup>ο</sup> έτος ατυχήματος και αποπληρώθηκε μετά από 8 χρόνια θα είναι ίση με 12.363 (δηλαδή,  $Y_{58} = \mu + \alpha_5 + b_8 = 12.5198 + 0.2822 - 0.4393 = 12.363$ ).

Κατά τον ίδιο τρόπο λαμβάνουμε το τετράγωνο εξέλιξης ζημιών (run-off square) όπως φαίνεται στον Πίνακα 5.5.

Έτος Ατυχήματος	Έτος Εξέλιξης									
	1	2	3	4	5	6	7	8	9	10
1	12.788	13.550	13.322	13.088	13.176	13.261	11.894	11.849	12.334	11.126
2	12.772	13.692	13.747	13.984	13.008	12.679	13.176	12.492	12.960	11.488
3	12.579	13.817	13.739	13.832	13.529	11.898	13.114	12.544	12.749	11.409
4	12.646	13.918	13.562	14.262	12.515	12.772	12.237	12.252	12.638	11.298
5	13.002	13.449	13.807	13.553	13.132	13.062	12.684	12.363	12.749	11.409
6	12.890	13.751	13.650	13.599	13.467	12.827	12.713	12.392	12.778	11.438
7	12.996	13.650	13.939	13.877	13.295	12.907	12.794	12.473	12.858	11.519
8	12.792	13.875	14.182	13.965	13.383	12.995	12.882	12.561	12.947	11.607
9	12.839	13.802	13.804	13.830	13.248	12.860	12.747	12.426	12.812	11.472
10	12.748	13.660	13.687	13.713	13.132	12.744	12.630	12.309	12.695	11.355

Πίνακας 5.5: Τετράγωνο Εξέλιξης Λογαριθμισμένων Ζημιών με τη μέθοδο LS

Οι εκτιμούμενες ατομικές ζημιές  $\hat{P}_{ij}$  υπολογίζονται με βάση τη διαδικασία της υποενότητας 4.3.2, δηλαδή

$$\hat{P}_{ij} = \exp(\hat{Y}_{ij} + 0.5 \text{Var}(\hat{Y}_{ij})).$$

Οι διακυμάνσεις των μελλοντικών τιμών του μοντέλου μας δίνονται από τη σχέση

$$\text{Var}(\hat{Y}_{ij}) = \sigma^2 [I + X_f(X'X)^{-1}X_f'],$$

και παρουσιάζονται αναλυτικά στο πίνακα Β<sub>1</sub> του παραρτήματος Β.

Στην περίπτωση της γραμμικής παλινδρόμησης (LS) η τιμή της διακύμανσης,  $\hat{\sigma}_{LS}^2$ , υπολογίζεται με βάση τη σχέση (4.8) και ισούται με

$$\hat{\sigma}_{LS}^2 = 0.11621.$$

Μετατρέποντας τις  $Y_{ij}$  ζημιές στις  $P_{ij}$  μέσω της σχέσης (4.11) υπολογίζουμε τις εκτιμούμενες ζημιές στον παρακάτω Πίνακα 5.6.

Έτος Ατυχήματος	Έτος Εξέλιξης									
	1	2	3	4	5	6	7	8	9	10
1	357848	766940	610542	482940	527326	574398	146342	139950	227229	67948
2	352118	884021	933894	1183289	445745	320996	527804	266172	425046	110586
3	290507	1001799	926219	1016654	750816	146923	495992	280405	379110	102290
4	310608	1108250	776189	1562400	272482	352053	206286	228588	339684	91637
5	443160	693190	991983	769488	504851	470639	350882	255827	380137	102533
6	396132	937085	847498	805037	705960	404285	362174	264050	392328	105802
7	440832	847631	1131398	1063269	646485	439495	393703	287022	426424	114970
8	359480	1061648	1443370	1268161	709884	482576	432274	315119	468110	126168
9	376686	986608	1089012	1119363	626553	425898	381471	278051	412955	111238
10	344014	970581	998455	1026113	574262	390281	349489	254655	377991	101663

Πίνακας 5.6: Τετράγωνο Εξέλιξης Ζημιών με τη μέθοδο LS (0 outliers)

## 5.2 Ανθεκτικές Μέθοδοι (0 outliers)

Στη συνέχεια, θα παραθέσουμε τα αποτελέσματα που παράχθηκαν χρησιμοποιώντας τις ανθεκτικές μεθόδους στις οποίες αναφερθήκαμε στο Κεφάλαιο 3. Για κάθε μέθοδο ξεχωριστά παρουσιάζουμε τους ανθεκτικούς συντελεστές παλινδρόμησης, την ανθεκτική εκτίμηση της διακύμανσης του μοντέλου και παρουσιάζουμε τα συμπληρωμένα τετράγωνα εξέλιξης ατομικών ζημιών ( $\hat{P}_{ij}$ ),

δηλαδή τις ζημιές μετά την μετατροπή τους από την λογαριθμική μορφή στην πραγματική μέσω της σχέσης (4.11).

Στο παράρτημα, που παραθέτεται στο τέλος της εργασίας έχουμε την ευκαιρία να δούμε και τα τετράγωνα εξέλιξης λογαριθμισμένων ζημιών (παράρτημα A<sub>1</sub>) και τις διακυμάνσεις των μελλοντικών τιμών (παράρτημα B<sub>1</sub>) όλων των χρησιμοποιούμενων μοντέλων.

### 5.2.1 Παλινδρόμηση με χρήση M-Εκτιμητή (0 outliers)

Ακολουθώντας τα βήματα της Ενότητας 4.4 και χρησιμοποιώντας τον ανθεκτικό M-Εκτιμητή του Huber με  $b = 2$ , του οποίου η συνάρτηση του οποίου ορίζεται ως εξής

$$\rho(u) = \begin{cases} 1, & \text{abs}(u) < b \\ \frac{b}{\text{abs}(u)}, & \text{abs}(u) \geq b \end{cases} \quad (5.1).$$

Οι εκτιμήσεις των παραμέτρων είναι:

Παράμετροι	Εκτίμηση	Παράμετροι	Εκτίμηση
<i>Intercept</i>	12.48534		
$a2^M$	0.387429	$b2^M$	0.91119
$a3^M$	0.38176	$b3^M$	0.93872
$a4^M$	0.182645	$b4^M$	0.936287
$a5^M$	0.30794	$b5^M$	0.400791
$a6^M$	0.348472	$b6^M$	0.058898
$a7^M$	0.433724	$b7^M$	-0.1119
$a8^M$	0.514772	$b8^M$	-0.44676
$a9^M$	0.379665	$b9^M$	-0.03222
$a10^M$	0.2631	$b10^M$	-1.35884

Πίνακας 5.7: Εκτιμήσεις Παραμέτρων με τη μέθοδο M εκτιμητή (0 outliers)

Η ανθεκτική M διακύμανση του μοντέλου δίνεται με βάση τη σχέση (4.8) είναι

$$\hat{\sigma}_M^2 = 0.1183 ,$$

και οι ανθεκτικές εκτιμούμενες ζημιές εμφανίζονται στον Πίνακα 5.8.



Έτος Ατυχήματος	Έτος Εξέλιξης									
	1	2	3	4	5	6	7	8	9	10
1	357848	766940	610542	482940	527326	574398	146342	139950	227229	67948
2	352118	884021	933894	1183289	445745	320996	527804	266172	425046	113806
3	290507	1001799	926219	1016654	750816	146923	495992	280405	414036	113256
4	310608	1108250	776189	1562400	272482	352053	206286	222079	339696	92905
5	443160	693190	991983	769488	504851	470639	350529	252128	385635	105452
6	396132	937085	847498	805037	705960	432563	365837	263128	402431	110024
7	440832	847631	1131398	1063269	663762	472597	399682	287457	439601	120158
8	359480	1061648	1443370	1234233	723657	515222	435708	313345	479131	130919
9	376686	986608	1090943	1089611	638822	454790	384569	276532	422748	115445
10	344014	972814	1000796	999407	585838	416991	352524	253405	387164	105562

Πίνακας 5.8: Τετράγωνο Εξέλιξης Ζημιών με τη μέθοδο του  $M$  εκτιμητή (0 outliers)

## 5.2.2 Παλινδρόμηση Ελαχίστων Απολύτων Αποκλίσεων - Least Absolute Deviation (0 outliers)

Εφαρμόζοντας την παλινδρόμηση ελαχίστων απολύτων αποκλίσεων (LAD) παίρνουμε τα εξής αποτελέσματα:

Παράμετροι	Εκτίμηση	Παράμετροι	Εκτίμηση
<i>Intercept</i>	12.40624		
$a2^{LAD}$	0.425015	$b2^{LAD}$	0.933961
$a3^{LAD}$	0.477108	$b3^{LAD}$	0.915864
$a4^{LAD}$	0.240048	$b4^{LAD}$	0.88044
$a5^{LAD}$	0.485358	$b5^{LAD}$	0.64557
$a6^{LAD}$	0.41033	$b6^{LAD}$	0.12525
$a7^{LAD}$	0.59018	$b7^{LAD}$	0.230969
$a8^{LAD}$	0.535133	$b8^{LAD}$	-0.33936
$a9^{LAD}$	0.461828	$b9^{LAD}$	0.128698
$a10^{LAD}$	0.342199	$b10^{LAD}$	-1.27974

Πίνακας 5.9: Εκτιμήσεις Παραμέτρων με τη μέθοδο LAD εκτιμητή (0 outliers)

Η ανθεκτική LAD διακύμανση του μοντέλου που δίνεται από τον τύπο (4.8) είναι

$$\hat{\sigma}_{LAD}^2 = 0.1503,$$

και οι ανθεκτικές εκτιμώμενες ζημιές δίνονται στον παρακάτω Πίνακα 5.10.

Έτος Ατυχήματος	Έτος Εξέλιξης									
	1	2	3	4	5	6	7	8	9	10
1	357848	766940	610542	482940	527326	574398	146342	139950	227229	67948
2	352118	884021	933894	1183289	445745	320996	527804	266172	425046	122337
3	290507	1001799	926219	1016654	750816	146923	495992	280405	507678	129015
4	310608	1108250	776189	1562400	272482	352053	206286	247884	401146	101920
5	443160	693190	991983	769488	504851	470639	557713	317447	513679	130484
6	396132	937085	847498	805037	705960	464873	518858	295316	477824	121347
7	440832	847631	1131398	1063269	937549	558783	623645	354935	574223	145785
8	359480	1061648	1443370	1127593	893381	532432	594197	338144	546972	138807
9	376686	986608	1098535	1061934	841293	501344	559440	318314	514751	130533
10	344014	1030363	1012950	978992	775418	461977	515358	293107	473635	119868

Πίνακας 5.10: Τετράγωνο Εξέλιξης Ζημιών με την μέθοδο LAD εκτιμητή (0 outliers)

### 5.2.3 Κλασσική & Ανθεκτική Ανάλυση Διακύμανσης (ANOVA) (0 outliers)

Θεωρούμε το μοντέλο του Kremer (1982)

$$Y_{ij} = \mu + \alpha_i + b_j + e_{ij}, \quad i = 1, 2, \dots, 10 \text{ και } j = 1, 2, \dots, 10$$

όπου  $Y_{ij}$ : η τιμή της μεταβλητής  $Y$  στη στάθμη  $i$  του παράγοντα  $A$  και στη στάθμη  $j$  του παράγοντα  $B$ ,

$\mu$ : ο μέσος όρος,

$\alpha_i$ : η επίδραση της στάθμης  $i$  του παράγοντα  $A$  στη διαμόρφωση των  $Y_{ij}$ ,

$b_j$ : επίδραση της στάθμης  $j$  του παράγοντα  $B$  στη διαμόρφωση των  $Y_{ij}$ ,

$e_{ij}$ : σφάλμα κατά την εκτίμηση των  $Y_{ij}$ .

Όπως αναμενόταν εφαρμόζοντας την κλασσική ANOVA δύο παραγόντων οδηγούμαστε στα ίδια αποτελέσματα με την κλασσική γραμμική παλινδρόμηση (Πίνακας 5.4).

Η εφαρμογή της ανθεκτικής ANOVA παράγει εκτιμήσεις παραμέτρων, οι οποίες όπως εμφανίζονται στον Πίνακα 5.11 και είναι διαφορετικές από τις αντίστοιχες της μη ανθεκτικής ANOVA.

Παράμετροι	Εκτίμηση	Παράμετροι	Εκτίμηση
<i>Intercept</i>	12.49292		
<i>a2<sup>R.Anova</sup></i>	0.361002	<i>b2<sup>R.Anova</sup></i>	0.91119
<i>a3<sup>R.Anova</sup></i>	0.443958	<i>b3<sup>R.Anova</sup></i>	0.93872
<i>a4<sup>R.Anova</sup></i>	0.155684	<i>b4<sup>R.Anova</sup></i>	0.957172
<i>a5<sup>R.Anova</sup></i>	0.267146	<i>b5<sup>R.Anova</sup></i>	0.383527
<i>a6<sup>R.Anova</sup></i>	0.340162	<i>b6<sup>R.Anova</sup></i>	0.25453
<i>a7<sup>R.Anova</sup></i>	0.420917	<i>b7<sup>R.Anova</sup></i>	-0.12771
<i>a8<sup>R.Anova</sup></i>	0.507186	<i>b8<sup>R.Anova</sup></i>	-0.46627
<i>a9<sup>R.Anova</sup></i>	0.37208	<i>b9<sup>R.Anova</sup></i>	-0.02659
<i>a10<sup>R.Anova</sup></i>	0.255514	<i>b10<sup>R.Anova</sup></i>	-1.36643

Πίνακας 5.11: Εκτιμήσεις Παραμέτρων με τη μέθοδο R. ANOVA (0 outliers)

Η ανθεκτική R.ANOVA διακύμανση του μοντέλου ισούται με

$$\hat{\sigma}_{R.Anova}^2 = 0.0988,$$

και οι ανθεκτικές εκτιμούμενες ζημιές δίνονται στον Πίνακα 5.12.

Έτος Ατυχήματος	Έτος Εξέλιξης									
	1	2	3	4	5	6	7	8	9	10
1	357848	766940	610542	482940	527326	574398	146342	139950	227229	67948
2	352118	884021	933894	1183289	445745	320996	527804	266172	425046	108522
3	290507	1001799	926219	1016654	750816	146923	495992	280405	439267	117991
4	310608	1108250	776189	1562400	272482	352053	206286	210485	329588	88518
5	443160	693190	991983	769488	504851	470639	329084	235620	368926	99069
6	396132	937085	847498	805037	705960	518369	354667	253928	397569	106744
7	440832	847631	1131398	1063269	639903	563501	385535	276017	432122	115998
8	359480	1061648	1443370	1241847	700681	617002	422121	302192	473051	126949
9	376686	986608	1073349	1094445	617481	543706	371948	266246	416704	111773
10	344014	952493	979758	998875	563482	496081	339301	242808	379836	101751

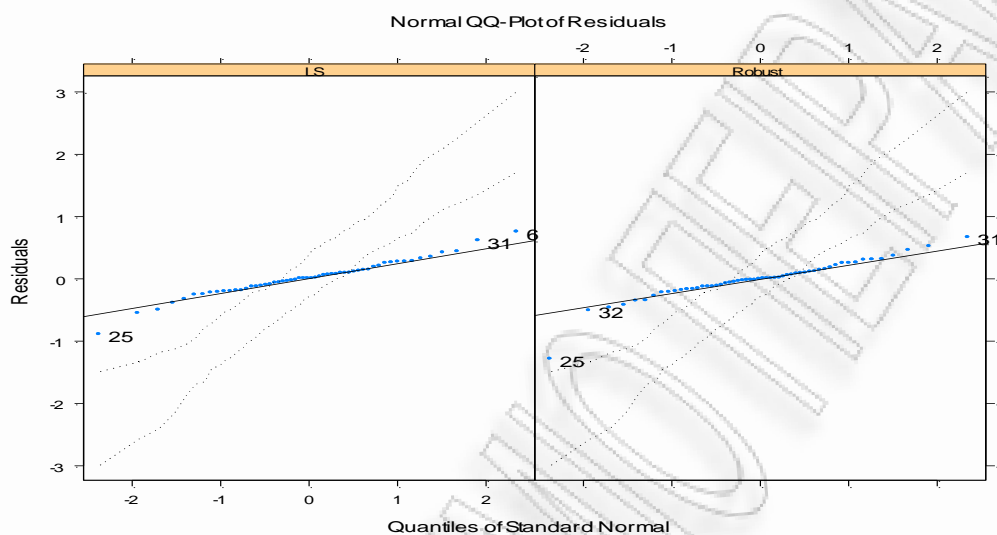
Πίνακας 5.12: Τετράγωνο Εξέλιξης Ζημιών με τη μέθοδο Robust ANOVA (0 outliers)

Τα διαφορετικά αυτά αποτελέσματα μεταξύ ανθεκτικής και κλασσικής προσαρμογής θα μπορούσαν να δικαιολογηθούν από την παρουσία μίας ή περισσοτέρων έκτοπων παρατηρήσεων στα δεδομένα μας. Με σκοπό, λοιπόν, να ερμηνευτούν οι διαφορές αυτές γίνονται τα παρακάτω διαγνωστικά γραφήματα.

## ΔΙΑΓΝΩΣΤΙΚΑ ΓΡΑΦΗΜΑΤΑ

### ➤ Γράφημα 1<sup>ο</sup>

Ο πιο διαδεδομένος τρόπος ελέγχου της κανονικότητας των καταλοίπων είναι η κατασκευή ενός “Normal Q-Q Plot of Residuals”, το οποίο βλέπουμε παρακάτω.

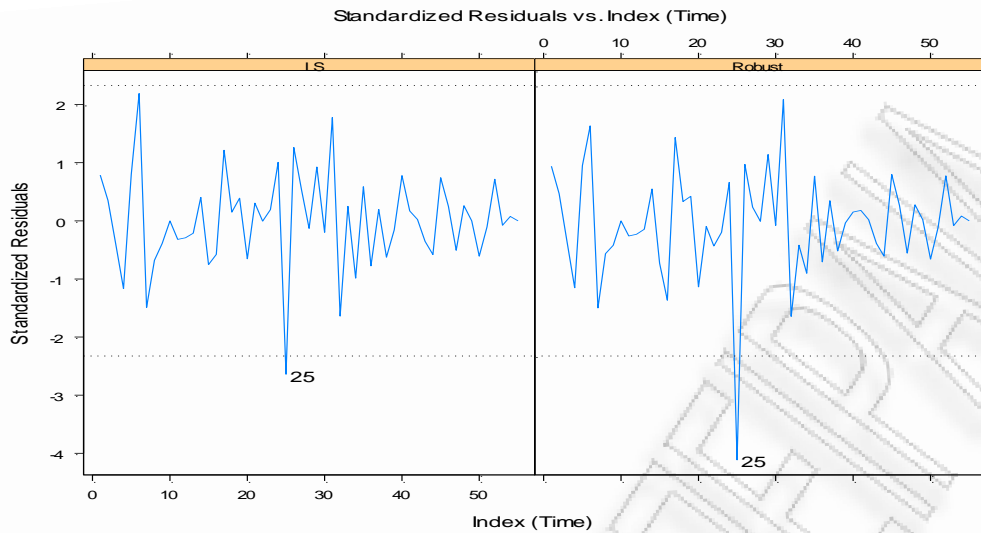


Γράφημα 5.1: Κανονικά Q-Q-Plots για Κλασσικά και Ανθεκτικά κατάλοιπα

Σύμφωνα με το Γράφημα 5.1, συμπεραίνουμε ότι για τη γραμμική παλινδρόμηση (αριστερά) τα κατάλοιπα είναι κατά προσέγγιση κανονικά κατανομημένα, με εξαίρεση ενδεχομένως μία ελαφρώς μεγαλύτερη έκτοπη παρατήρηση κάτω αριστερά. Το κανονικό Q-Q γράφημα των καταλοίπων της ανθεκτικής προσαρμογής επιβεβαιώνει την ύπαρξη αυτής της έκτοπης παρατήρησης, που αντιστοιχεί στη 25<sup>η</sup> παρατήρηση, ενώ τα υπόλοιπα κατάλοιπα προσαρμόζονται αρκετά καλά στην κανονική κατανομή.

### ➤ Γράφημα 2<sup>ο</sup>

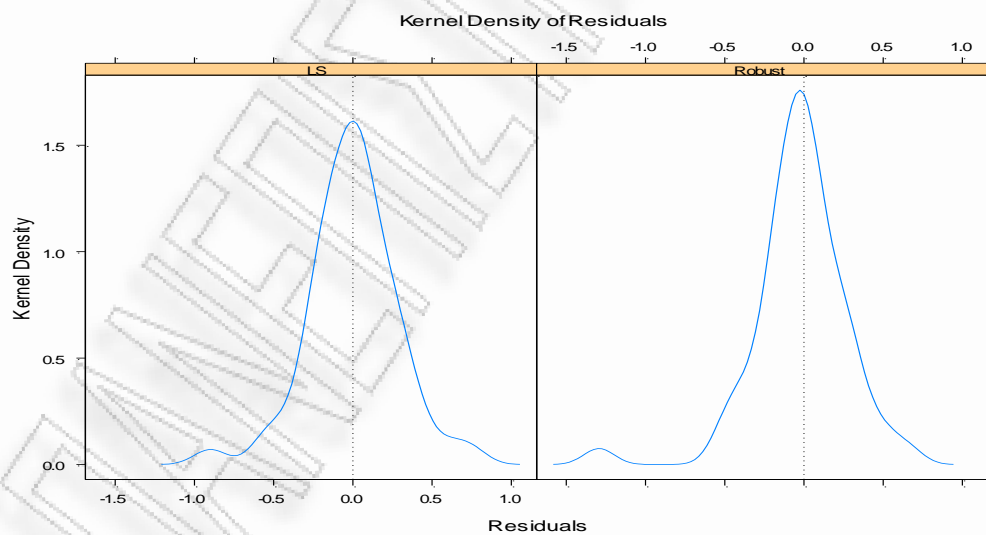
Τα γραφήματα των “Standardized Residuals versus Index (σε αυτή την περίπτωση ο αριθμός των παρατηρήσεων)” για την κλασσική και την ανθεκτική προσαρμογή φαίνονται παρακάτω. Οι οριζόντιες γραμμές αναφοράς στο  $\pm 2.5$  αντιστοιχούν σε ουρές πιθανοτήτων 0.006 μίας τυποποιημένης κανονικής τυχαίας μεταβλητής. Στο Γράφημα 5.2, παρατηρούμε ότι τα τυποποιημένα κατάλοιπα της κλασσικής προσαρμογής οριακά υπονοούν την παρουσία μίας έκτοπης παρατήρησης στην 25<sup>η</sup> θέση, ενώ η ανθεκτική προσαρμογή αναγνωρίζει ξεκάθαρα την 25<sup>η</sup> παρατήρηση σαν “outlier”.



Γράφημα 5.2: Standardized Residuals vs Index για Κλασσική & Ανθεκτική προσαρμογή

### ➤ Γράφημα 3<sup>ο</sup>

Στο Γράφημα 5.3 (“Kernel Density of residuals”), παρουσιάζονται οι εκτιμήσεις πυκνότητας πιθανότητας των καταλοίπων της κλασσικής και ανθεκτικής προσαρμογής.



Γράφημα 5.3: Kernel Density καταλοίπων της Κλασσικής & Ανθεκτικής προσαρμογής

Οι προσαρμογές αυτές φαίνονται σχεδόν ίδιες και για τα δύο μοντέλα και δίνουν μία σαφή εικόνα της κατανομής των όρων του σφάλματος στο μοντέλο. Η εκτίμηση της πυκνότητας των καταλοίπων και των δύο προσαρμογών είναι πολύ συμπαγής και η κεντρική της περιοχή φαίνεται σωστά κεντραρισμένη στο μηδέν.

## 5.2.4 Παλινδρόμηση Ελαχίστων Περικεκομμένων Τετραγώνων - Least Trimmed Squares- (0 outliers)

Η εφαρμογή του ανθεκτικού εκτιμητή παλινδρόμησης ελαχίστων περικεκομμένων τετραγώνων (LTS) στο λογαριθμικό μοντέλο (4.6) παρουσιάζει προβλήματα εκτίμησης λόγω του μεγάλου αριθμού επεξηγηματικών μεταβλητών στον πίνακα σχεδιασμού. Για το σκοπό αυτό υποθέτουμε ένα εναλλακτικό μοντέλο όπου ο πίνακας σχεδιασμού περιλαμβάνει μικρότερο αριθμό επεξηγηματικών μεταβλητών. Το μοντέλο αυτό ορίζεται ως (βλέπε *Christofides (1990)*, σελ. 19),

$$Y_{ij} = \alpha_i + d_j + e_{ij}, \quad i = 1, 2, \dots, 10 \text{ και } j = 1, 2, \dots, 10 \quad (5.2)$$

όπου  $Y_{ij}$ : η τιμή της μεταβλητής  $Y$  στη στάθμη  $i$  του παράγοντα  $A$  και στη στάθμη  $j$  του παράγοντα  $B$ ,

$\alpha_i$ : η επίδραση της στάθμης  $i$  του παράγοντα  $A$  στη διαμόρφωση των  $Y_{ij}$ ,

$d_j$ : η επίδραση της στάθμης  $j$  του παράγοντα  $B$  στη διαμόρφωση των  $Y_{ij}$  και υποθέτουμε ότι οι παράμετροι για τα έτη εξέλιξης 1 έως 10 ακολουθούν μία γραμμική σχέση με την ίδια κλίση ή παράμετρο  $s$ , δηλαδή,

$$d_j = s \times j, \quad \text{για } j > 1$$

$X$ : (55×12) πίνακας σχεδιασμού,

$e_{ij}$ : σφάλμα κατά την εκτίμηση των  $Y_{ij}$ .

Εφαρμόζοντας την παλινδρόμηση ελαχίστων περικεκομμένων τετραγώνων παίρνουμε τις εξής εκτιμήσεις:

Παράμετροι	Εκτίμηση	Παράμετροι	Εκτίμηση
$a1^{LTS}$	13.9713	$a7^{LTS}$	14.4546
$a2^{LTS}$	14.177	$a8^{LTS}$	14.2191
$a3^{LTS}$	14.1557	$a9^{LTS}$	14.1829
$a4^{LTS}$	13.6235	$a10^{LTS}$	14.0921
$a5^{LTS}$	14.3454	$d^{LTS}$	-1.1718
$a6^{LTS}$	14.2862	$s^{LTS}$	-0.1719

Πίνακας 5.13: Εκτιμήσεις Παραμέτρων με τη μέθοδο LTS εκτιμητή (0 outliers)



Η ανθεκτική LTS διακύμανση του συγκεκριμένου μοντέλου είναι

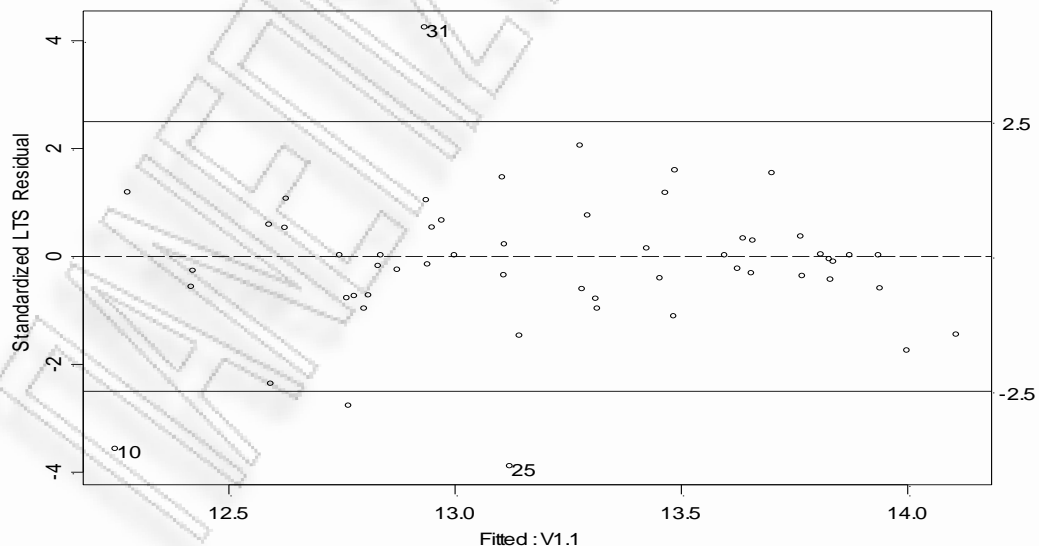
$$\hat{\sigma}_{LTS}^2 = 0.0985,$$

και οι ανθεκτικές εκτιμούμενες ζημιές δίνονται στον Πίνακα 5.14.

Έτος Ατυχήματος	Έτος Εξέλιξης									
	1	2	3	4	5	6	7	8	9	10
1	357848	766940	610542	482940	527326	574398	146342	139950	227229	67948
2	352118	884021	933894	1183289	445745	320996	527804	266172	425046	278009
3	290507	1001799	926219	1016654	750816	146923	495992	280405	322617	268210
4	310608	1108250	776189	1562400	272482	352053	206286	224673	186957	157875
5	443160	693190	991983	769488	504851	470639	548540	456930	385728	325804
6	396132	937085	847498	805037	705960	613535	511686	431806	364600	308026
7	440832	847631	1131398	1063269	862241	720188	607533	512788	433061	365934
8	359480	1061648	1443370	810205	678145	571822	482439	407255	343981	290701
9	376686	986608	932400	783223	660044	556550	469546	396365	334777	282918
10	344014	1032107	875274	736754	620504	522890	440879	371938	313954	265158

Πίνακας 5.14: Τετράγωνο Εξέλιξης Ζημιών με τη μέθοδο LTS εκτιμητή

Το Γράφημα 5.4 απεικονίζει τα τυποποιημένα κατάλοιπα τα οποία θεωρητικά θα πρέπει να κυμαίνονται στο διάστημα  $[-2.5, 2.5]$ . Όπως βλέπουμε παρακάτω έκτοπες παρατηρήσεις (outliers) εμφανίζονται να είναι οι 31<sup>η</sup>, 25<sup>η</sup> και 10<sup>η</sup>.



Γράφημα 5.4: Τυποποιημένα κατάλοιπα για τη μέθοδο LTS

### 5.2.5 Παλινδρόμηση Ελαχίστων Διαμέσων Τετραγώνων -Least Median Squares- (0 outliers)

Η εφαρμογή του ανθεκτικού εκτιμητή παλινδρόμησης ελαχίστων διαμέσων τετραγώνων (LMS) στο λογαριθμικό μοντέλο (4.6) καθώς και στο μοντέλο (5.2) της παλινδρόμησης του LTS εκτιμητή που περιγράψαμε στην προηγούμενη υποενότητα παρουσιάζει προβλήματα εκτίμησης λόγω του μεγάλου αριθμού επεξηγηματικών μεταβλητών στον πίνακα σχεδιασμού. Για το σκοπό αυτό υποθέτουμε ένα εναλλακτικό μοντέλο όπου ο πίνακας σχεδιασμού περιλαμβάνει μικρότερο αριθμό επεξηγηματικών μεταβλητών. Για το σκοπό αυτό, θεωρούμε το εξής εναλλακτικό μοντέλο που προτείνεται στη βιβλιογραφία

$$Y_{ij} = \alpha \cdot i + b \cdot j + e_{ij}, \quad i = 1, 2, \dots, 10 \text{ και } j = 1, 2, \dots, 10$$

όπου  $Y_{ij}$ : η τιμή της μεταβλητής  $Y$  στη στάθμη  $i$  του παράγοντα  $A$  και στη στάθμη  $j$  του παράγοντα  $B$ ,

$i$ : η στάθμη του έτους ατυχήματος,

$j$ : η στάθμη του έτους εξέλιξης,

$\alpha$ : η επίδραση του έτους ατυχήματος  $i$  στη διαμόρφωση των  $Y_{ij}$ ,

$b$ : η επίδραση του έτους εξέλιξης  $j$  στη διαμόρφωση των  $Y_{ij}$ ,

$X$ : (55×2) πίνακας σχεδιασμού,

$e_{ij}$ : σφάλμα κατά την εκτίμηση των  $Y_{ij}$ .

Εφαρμόζοντας την παλινδρόμηση ελαχίστων διαμέσων τετραγώνων παίρνουμε τα εξής αποτελέσματα:

Παράμετροι	Εκτίμηση
<b>Intercept</b>	13.78169
$\alpha^{LMS}$	0.052187
$\beta^{LMS}$	-0.13119

Πίνακας 5.15: Εκτιμήσεις Παραμέτρων με την μέθοδο LMS εκτιμητή (0 outliers)

Η ανθεκτική LMS διακύμανση του μοντέλου ισούται με

$$\hat{\sigma}_{LMS}^2 = 0.1681,$$

και οι ανθεκτικές εκτιμούμενες ζημιές δίνονται στον Πίνακα 5.16.

Έτος Ατυχήματος	Έτος Εξέλιξης									
	1	2	3	4	5	6	7	8	9	10
1	357848	766940	610542	482940	527326	574398	146342	139950	227229	67948
2	352118	884021	933894	1183289	445745	320996	527804	266172	425046	346781
3	290507	1001799	926219	1016654	750816	146923	495992	280405	399484	365790
4	310608	1108250	776189	1562400	272482	352053	206286	473611	421610	385958
5	443160	693190	991983	769488	504851	470639	565922	500125	445173	407431
6	396132	937085	847498	805037	705960	678813	598119	528548	470426	430428
7	440832	847631	1131398	1063269	816657	718504	633059	559385	497809	455331
8	359480	1061648	1443370	986252	866968	762724	671973	593709	528261	482954
9	376686	986608	1200461	1054674	927032	815484	718366	634585	564454	515613
10	344014	1503791	1320450	1159815	1019205	896327	789317	696929	619392	564540

Πίνακας 5.16: Τετράγωνο Εξέλιξης Ζημιών με τη μέθοδο LMS εκτιμητή

**Σημείωση:** Από τα αποτελέσματα της Ενότητας 5.2 συμπεραίνουμε ότι το αυθεντικό τρίγωνο εξέλιξης ζημιών των Taylor και Ashe (1983) εμπεριέχει τουλάχιστον μία έκτοπη παρατήρηση, γεγονός το οποίο έχει σαν αποτέλεσμα διαφορετικές εκτιμήσεις μεταξύ του λογαριθμοκανονικού μοντέλου παλινδρόμησης και του μοντέλου ανάλυσης διασποράς με δύο παράγοντες και των αντίστοιχων ανθεκτικών εκτιμήσεων των μοντέλων αυτών. Αυτό διαπιστώνεται αν παρατηρήσουμε τον Πίνακα 6.1 και το Γράφημα 6.1 στο Κεφάλαιο 6, όπου γίνεται σύγκριση της αποθεματοποίησης ζημιών των κλασσικών και ανθεκτικών μεθόδων. Αν το τρίγωνο εξέλιξης ζημιών των Taylor και Ashe (1983) δεν εμπεριείχε έκτοπες παρατηρήσεις, θα έπρεπε οι τιμές των τελικών αποθεμάτων με βάση τα ανθεκτικά και μη ανθεκτικά μοντέλα να είναι παραπλήσιες πράγμα που δεν συμβαίνει στην δική μας περίπτωση.

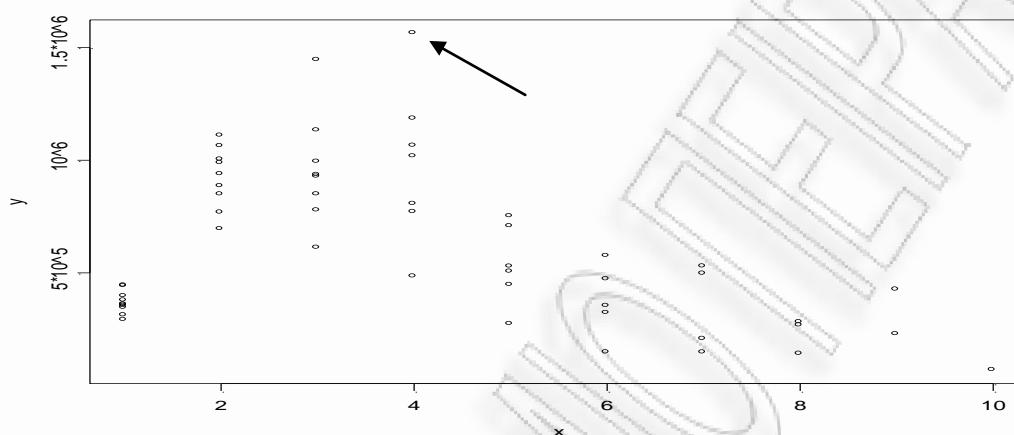
### 5.3 Εφαρμογή Λογαριθμικού μοντέλου παλινδρόμησης (1 outlier στη θέση (4,4) του τριγώνου εξέλιξης ζημιών)

Στο σημείο αυτό, επιθυμούμε να προσδιορίσουμε την επιρροή μίας έκτοπης παρατήρησης (outlier) στη πρόβλεψη των εκκρεμών ζημιών και του συνολικού απαιτούμενου αποθεματικού για κάθε μέθοδο ξεχωριστά.

Το Γράφημα 5.5 απεικονίζει τις ατομικές ζημιές σε σχέση με τη θέση τους στο τρίγωνο εξέλιξης ζημιών και η μεγαλύτερη ζημιά του αρχικού συνόλου δεδομένων

βρίσκεται στην 4<sup>η</sup> γραμμή και 4<sup>η</sup> στήλη. Θέλουμε να δούμε τι γίνεται στη περίπτωση που αυξήσουμε αυτή την ήδη μεγάλη ατομική ζημιά. Για το σκοπό αυτό, θα δεκαπλασιάσουμε την ατομική ζημιά που συνέβη κατά το 4<sup>ο</sup> έτος ατυχήματος και αποπληρώθηκε μετά από 4 χρόνια (4<sup>ο</sup> έτος εξέλιξης), δηλαδή

$$P'_{44} = 1562400 * 10 = 15624000.$$



Γράφημα 5.5: Ατομικές ζημιές σε σχέση με τη θέση τους στο τρίγωνο εξέλιξης ζημιών

Έτσι, το αναπροσαρμοσμένο τρίγωνο εξέλιξης ζημιών με τους λογαρίθμους των ατομικών ζημιών μετά την προσθήκη της έκτοπης παρατήρησης στη θέση (4,4) θα έχει την παρακάτω μορφή.

Έτος Ατυχήματος (i)	Έτος Εξέλιξης (j)									
	1	2	3	4	5	6	7	8	9	10
1	12.788	13.550	13.322	13.088	13.176	13.261	11.894	11.849	12.334	11.126
2	12.772	13.692	13.747	13.984	13.008	12.679	13.176	12.492	12.960	
3	12.579	13.817	13.739	13.832	13.529	11.898	13.114	12.544		
4	12.646	13.918	13.562	16.564	12.515	12.772	12.237			
5	13.002	13.449	13.807	13.553	13.132	13.062				
6	12.890	13.751	13.650	13.599	13.467					
7	12.996	13.650	13.939	13.877						
8	12.792	13.875	14.182							
9	12.839	13.802								
10	12.748									

Πίνακας 5.17: Τρίγωνο Εξέλιξης Ζημιών με 1 outlier (P<sub>44</sub>)

### 5.3.1 Εκτιμήσεις παραμέτρων

Εφαρμόζοντας γραμμική πολλαπλή παλινδρόμηση παίρνουμε τα εξής αποτελέσματα:

Παράμετροι	Εκτίμηση	Παράμετροι	Εκτίμηση
<i>Intercept</i>	12.4858		
$a2^{LS}$	0.361	$b2^{LS}$	0.9112
$a3^{LS}$	0.2822	$b3^{LS}$	0.9387
$a4^{LS}$	0.5001	$b4^{LS}$	1.2939
$a5^{LS}$	0.2685	$b5^{LS}$	0.3695
$a6^{LS}$	0.2827	$b6^{LS}$	-0.034
$a7^{LS}$	0.3438	$b7^{LS}$	-0.1663
$a8^{LS}$	0.5143	$b8^{LS}$	-0.4053
$a9^{LS}$	0.3792	$b9^{LS}$	-0.0195
$a10^{LS}$	0.2626	$b10^{LS}$	-1.3594

Πίνακας 5.18: Εκτιμήσεις Παραμέτρων με την LS μέθοδο (1 outlier)

Στη συνέχεια, για να συμπληρώσουμε το τρίγωνο εξέλιξης ζημιών με τις μελλοντικές τιμές των λογαρίθμων των αρκεί να πολλαπλασιάσουμε τον πίνακα των εκτιμημένων παραμέτρων με τον μελλοντικό πίνακα σχεδιασμού προσθέτοντας πάντα τον όρο τεταγμένη (intercept). Το συμπληρωμένο τετράγωνο εξέλιξης ζημιών (run-off square) δίνεται από τον παρακάτω πίνακα.

Έτος Ατυχήματος	Έτος Εξέλιξης									
	1	2	3	4	5	6	7	8	9	10
1	12.788	13.550	13.322	13.088	13.176	13.261	11.894	11.849	12.334	11.126
2	12.772	13.692	13.747	13.984	13.008	12.679	13.176	12.492	12.960	11.487
3	12.579	13.817	13.739	13.832	13.529	11.898	13.114	12.544	12.749	11.409
4	12.646	13.918	13.562	16.564	12.515	12.772	12.237	12.581	12.966	11.627
5	13.002	13.449	13.807	13.553	13.132	13.062	12.588	12.349	12.735	11.395
6	12.890	13.751	13.650	13.599	13.467	12.735	12.602	12.363	12.749	11.409
7	12.996	13.650	13.939	13.877	13.199	12.796	12.663	12.424	12.810	11.470
8	12.792	13.875	14.182	14.294	13.370	12.966	12.834	12.595	12.981	11.641
9	12.839	13.802	13.804	14.159	13.235	12.831	12.699	12.460	12.846	11.506
10	12.748	13.660	13.687	14.042	13.118	12.714	12.582	12.343	12.729	11.389

Πίνακας 5.19: Τρίγωνο Εξέλιξης Λογαριθμισμένων Ζημιών με την LS μέθοδο (1 outlier)



Στην περίπτωση της γραμμικής παλινδρόμησης η τιμή της διακύμανσης ισούται με

$$\hat{\sigma}_{LS(1outlier)}^2 = 0.3011,$$

και οι εκτιμούμενες ζημιές δίνονται στον Πίνακα 5.20.

Έτος Ατυχήματος	Έτος Εξέλιξης									
	1	2	3	4	5	6	7	8	9	10
1	357848	766940	610542	482940	527326	574398	146342	139950	227229	67948
2	352118	884021	933894	1183289	445745	320996	527804	266172	425046	135125
3	290507	1001799	926219	1016654	750816	146923	495992	280405	442417	125150
4	310608	1108250	776189	15624000	272482	352053	206286	365318	551827	156033
5	443160	693190	991983	769488	504851	470639	364547	290982	439468	124209
6	396132	937085	847498	805037	705960	420937	371850	296780	448146	126601
7	440832	847631	1131398	1063269	671662	451191	398539	318041	480142	135559
8	359480	1061648	1443370	2026710	807434	542340	478991	382173	576776	162703
9	376686	986608	1270865	1818440	724343	486442	429526	342597	516758	145555
10	344014	1186071	1221714	1747366	695735	467007	412118	328432	494653	138775

Πίνακας 5.20: Τετράγωνο Εξέλιξης Ζημιών με τη μέθοδο LS (1 outlier)

## 5.4 Ανθεκτικές Μέθοδοι (1 outlier)

Στην παρούσα ενότητα ακολουθούμε τα βήματα της Ενότητας 5.2 με τη διαφορά ότι έχουμε προσθέσει μία έκτοπη παρατήρηση στα δεδομένα, ως εξής  $P'_{44} = 15624000$ .

Στο παράρτημα, που παραθέτεται στο τέλος της εργασίας έχουμε την ευκαιρία να δούμε και τα τετράγωνα εξέλιξης λογαριασμένων ζημιών (παράρτημα A<sub>2</sub>) και τις διακυμάνσεις των μελλοντικών παρατηρήσεων (παράρτημα B<sub>2</sub>).

### 5.4.1 Παλινδρόμηση με χρήση M-Εκτιμητή (1 outlier)

Χρησιμοποιούμε τον ανθεκτικό M-Εκτιμητή του Huber με  $b = 2$ , τη συνάρτηση του οποίου είχαμε ορίσει με τη σχέση (5.1). Οι εκτιμήσεις των ανθεκτικών συντελεστών με τη μέθοδο αυτή εμφανίζονται στον Πίνακα 5.21.



Παράμετροι	Εκτίμηση	Παράμετροι	Εκτίμηση
<i>Intercept</i>	12.46846		
<i>a2<sup>M</sup></i>	0.392672	<i>b2<sup>M</sup></i>	0.8886648
<i>a3<sup>M</sup></i>	0.4884641	<i>b3<sup>M</sup></i>	0.9312818
<i>a4<sup>M</sup></i>	0.1104623	<i>b4<sup>M</sup></i>	0.8368361
<i>a5<sup>M</sup></i>	0.3254358	<i>b5<sup>M</sup></i>	0.4332284
<i>a6<sup>M</sup></i>	0.3827762	<i>b6<sup>M</sup></i>	0.1620504
<i>a7<sup>M</sup></i>	0.4877921	<i>b7<sup>M</sup></i>	-0.0970807
<i>a8<sup>M</sup></i>	0.5364851	<i>b8<sup>M</sup></i>	-0.4642581
<i>a9<sup>M</sup></i>	0.4078062	<i>b9<sup>M</sup></i>	-0.0179619
<i>a10<sup>M</sup></i>	0.2799786	<i>b10<sup>M</sup></i>	-1.341961

Πίνακας 5.21: Εκτιμήσεις Παραμέτρων με τη μέθοδο *M* εκτιμητή (1 outlier)

Η ανθεκτική *M* διακύμανση του μοντέλου (για τα δεδομένα με 1 outlier) ισούται με

$$\hat{\sigma}_{M(1outlier)}^2 = 0.3495,$$

και οι ανθεκτικές εκτιμούμενες ζημιές παρουσιάζονται στον Πίνακα 5.22.

Έτος Ατυχήματος	Έτος Εξέλιξης									
	1	2	3	4	5	6	7	8	9	10
1	357848	766940	610542	482940	527326	574398	146342	139950	227229	67948
2	352118	884021	933894	1183289	445745	320996	527804	266172	425046	146335
3	290507	1001799	926219	1016654	750816	146923	495992	280405	502216	145808
4	310608	1108250	776189	15624000	272482	352053	206286	264723	413290	119930
5	443160	693190	991983	769488	504851	470639	413892	301366	470411	136438
6	396132	937085	847498	805037	705960	510262	433805	315828	492882	142876
7	440832	847631	1131398	1063269	784647	561007	476897	347149	541620	156896
8	359480	1061648	1443370	1471010	864532	618050	525309	382308	596254	172551
9	376686	986608	1323412	1325514	778878	556700	473042	344141	536379	154955
10	344014	1250133	1288178	1289584	757390	541041	459417	333898	519513	149390

Πίνακας 5.22: Τετράγωνο Εξέλιξης Ζημιών με την μέθοδο *M* εκτιμητή (1 outlier)

#### 5.4.2 Παλινδρόμηση Ελαχίστων Απολύτων Αποκλίσεων - Least Absolute Deviation (1 outlier)

Εφαρμόζοντας την παλινδρόμηση ελαχίστων απολύτων αποκλίσεων (LAD) παίρνουμε τα εξής αποτελέσματα:

Παράμετροι	Εκτίμηση	Παράμετροι	Εκτίμηση
<i>Intercept</i>	12.40624		
$a2^{LAD}$	0.4250154	$b2^{LAD}$	0.9339609
$a3^{LAD}$	0.477108	$b3^{LAD}$	0.915864
$a4^{LAD}$	0.2400484	$b4^{LAD}$	0.8804398
$a5^{LAD}$	0.4853582	$b5^{LAD}$	0.6455698
$a6^{LAD}$	0.4103298	$b6^{LAD}$	0.1252499
$a7^{LAD}$	0.5901804	$b7^{LAD}$	0.2309685
$a8^{LAD}$	0.5351334	$b8^{LAD}$	-0.3393564
$a9^{LAD}$	0.4618282	$b9^{LAD}$	0.1286983
$a10^{LAD}$	0.3421993	$b10^{LAD}$	-1.27974

Πίνακας 5.23: Εκτιμήσεις Παραμέτρων με τη μέθοδο LAD εκτιμητή (1 outlier)

Η ανθεκτική LAD διακύμανση του μοντέλου (για τα δεδομένα με 1 outlier), δίνεται και ισούται με

$$\hat{\sigma}_{LAD(1outlier)}^2 = 0.3916,$$

και οι ανθεκτικές εκτιμούμενες ζημιές δίνονται στον Πίνακα 5.24.

Έτος Ατυχήματος	Έτος Εξέλιξης									
	1	2	3	4	5	6	7	8	9	10
1	357848	766940	610542	482940	527326	574398	146342	139950	227229	67948
2	352118	884021	933894	1183289	445745	320996	527804	266172	425046	158937
3	290507	1001799	926219	1016654	750816	146923	495992	280405	621057	167895
4	310608	1108250	776189	15624000	272482	352053	206286	297565	491947	132918
5	443160	693190	991983	769488	504851	470639	664406	382323	631939	170648
6	396132	937085	847498	805037	705960	552620	620915	357248	590357	159320
7	440832	847631	1131398	1063269	1116905	668693	751245	432163	713945	192523
8	359480	1061648	1443370	1353588	1075957	644092	723485	416095	687115	185082
9	376686	986608	1343869	1302305	1034977	619416	695565	399871	659842	177391
10	344014	1338613	1318218	1276739	1014096	606541	680578	390821	643657	172145

Πίνακας 5.24: Τετράγωνο Εξέλιξης Ζημιών με την μέθοδο LAD εκτιμητή (1 outlier)

### 5.4.3 Κλασσική & Ανθεκτική ANOVA (1 outlier)

Θεωρούμε το μοντέλο

$$Y_{ij} = \mu + \alpha_i + b_j + e_{ij}, \quad i = 1, 2, \dots, 10 \text{ και } j = 1, 2, \dots, 10$$

Εφαρμόζοντας την κλασσική ανάλυση διακύμανσης δύο παραγόντων για τα δεδομένα με ενσωματωμένη την μία έκτοπη παρατήρηση στη θέση (4,4) οδηγούμαστε στα ίδια αποτελέσματα με την κλασσική γραμμική παλινδρόμηση (Πίνακας 5.20).

Εφαρμόζοντας την ανθεκτική ανάλυση διακύμανσης παίρνουμε τους ανθεκτικούς συντελεστές του Πίνακα 5.25.

Παράμετροι	Εκτίμηση	Παράμετροι	Εκτίμηση
<i>Intercept</i>	12.50442		
<i>a2<sup>R.Anova</sup></i>	0.3610018	<i>b2<sup>R.Anova</sup></i>	0.9111896
<i>a3<sup>R.Anova</sup></i>	0.4480173	<i>b3<sup>R.Anova</sup></i>	0.9387197
<i>a4<sup>R.Anova</sup></i>	0.03749337	<i>b4<sup>R.Anova</sup></i>	0.8391747
<i>a5<sup>R.Anova</sup></i>	0.2716756	<i>b5<sup>R.Anova</sup></i>	0.3884434
<i>a6<sup>R.Anova</sup></i>	0.3512806	<i>b6<sup>R.Anova</sup></i>	0.27144808
<i>a7<sup>R.Anova</sup></i>	0.4389192	<i>b7<sup>R.Anova</sup></i>	-0.1106697
<i>a8<sup>R.Anova</sup></i>	0.495689	<i>b8<sup>R.Anova</sup></i>	-0.4791174
<i>a9<sup>R.Anova</sup></i>	0.3605823	<i>b9<sup>R.Anova</sup></i>	-0.0380883
<i>a10<sup>R.Anova</sup></i>	0.2440171	<i>b10<sup>R.Anova</sup></i>	-1.377923

Πίνακας 5.25: Εκτιμήσεις Παραμέτρων με μέθοδο R. ANOVA (1 outlier)

Οι ανθεκτικές εκτιμούμενες ζημιές δίνονται στον Πίνακα 5.26 και η ανθεκτική R.ANOVA διακύμανση του μοντέλου (για τα δεδομένα με 1 outlier), ισούται με

$$\hat{\sigma}_{RAnova(1outlier)}^2 = 0.09885,$$

Έτος Ατυχήματος	Έτος Εξέλιξης									
	1	2	3	4	5	6	7	8	9	10
1	357848	766940	610542	482940	527326	574398	146342	139950	227229	67948
2	352118	884021	933894	1183289	445745	320996	527804	266172	425046	108523
3	290507	1001799	926219	1016654	750816	146923	495992	280405	441055	118471
4	310608	1108250	776189	15624000	272482	352053	206286	186769	292849	78650
5	443160	693190	991983	769488	504851	470639	340147	236370	370602	99519
6	396132	937085	847498	805037	705960	539274	369014	256421	402016	107937
7	440832	847631	1131398	1063269	662311	590276	403901	280651	439972	118105
8	359480	1061648	1443370	1103630	704137	627531	429375	298334	467644	125498
9	376686	986608	1073352	972634	620526	552984	378340	262847	411942	110496
10	344014	952496	979761	887701	566261	504546	345132	239709	375495	100588

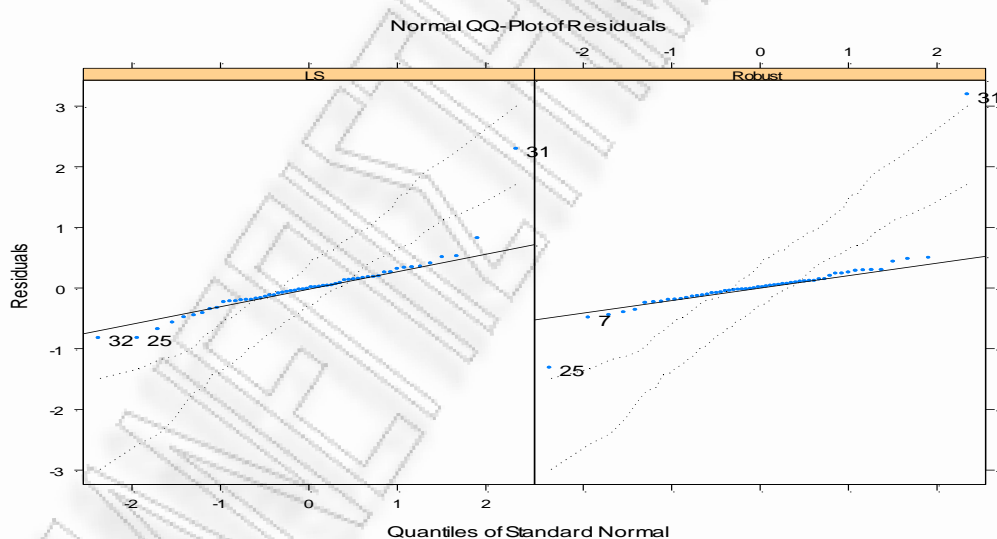
Πίνακας 5.26: Τετράγωνο Εξέλιξης Ζημιών με την μέθοδο Robust ANOVA (1 outlier)

Με σκοπό την επεξήγηση αυτής της διαφοράς μεταξύ κλασσικής και ανθεκτικής προσαρμογής, ακολουθούν τα διαγνωστικά γραφήματα μεταξύ κλασσικής και ανθεκτικής ανάλυσης διακύμανσης για την διάγνωση και την ερμηνεία των έκτοπων παρατηρήσεων.

### ΔΙΑΓΝΩΣΤΙΚΑ ΓΡΑΦΗΜΑΤΑ

#### ➤ Γράφημα 1<sup>ο</sup>

Σύμφωνα με το Γράφημα 5.6 συμπεραίνουμε ότι για την κλασσική γραμμική παλινδρόμηση (αριστερά) συμπεραίνουμε ότι τα κατάλοιπα είναι κατά προσέγγιση κανονικά κατανεμημένα, με εξαίρεση μία μετρίως μεγαλύτερη έκτοπη παρατήρηση πάνω δεξιά (την 31<sup>η</sup> παρατήρηση την οποία δεκαπλασιάσαμε). Το κανονικό Q-Q γράφημα των καταλοίπων της ανθεκτικής προσαρμογής επιβεβαιώνει την ύπαρξη αυτής της έκτοπης παρατήρησης, αλλά και ακόμα μίας κάτω αριστερά που αντιστοιχεί στη 25<sup>η</sup> παρατήρηση, ενώ τα υπόλοιπα κατάλοιπα προσαρμόζονται αρκετά καλά στην κανονική κατανομή.

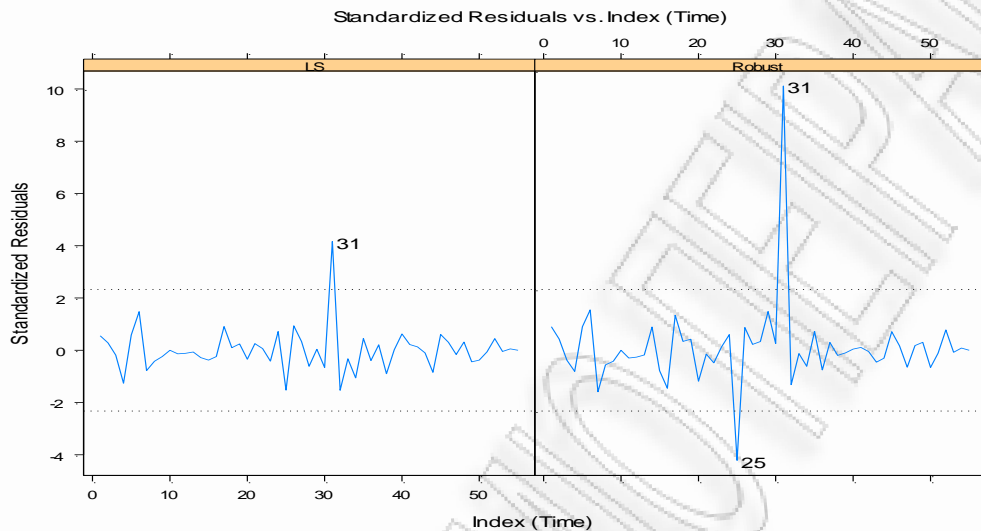


Γράφημα 5.6: Κανονικά Q-Q-Plots για Κλασσικά & Ανθεκτικά κατάλοιπα

#### ➤ Γράφημα 2<sup>ο</sup>

Τα γραφήματα των “Standardized Residuals versus Index (σε αυτή την περίπτωση ο αριθμός των παρατηρήσεων)” για την κλασσική και την ανθεκτική προσαρμογή φαίνονται παρακάτω. Οι οριζόντιες γραμμές αναφοράς στο  $\pm 2.5$  αντιστοιχούν σε ουρές πιθανοτήτων 0.006 μίας τυποποιημένης κανονικής τυχαίας μεταβλητής.

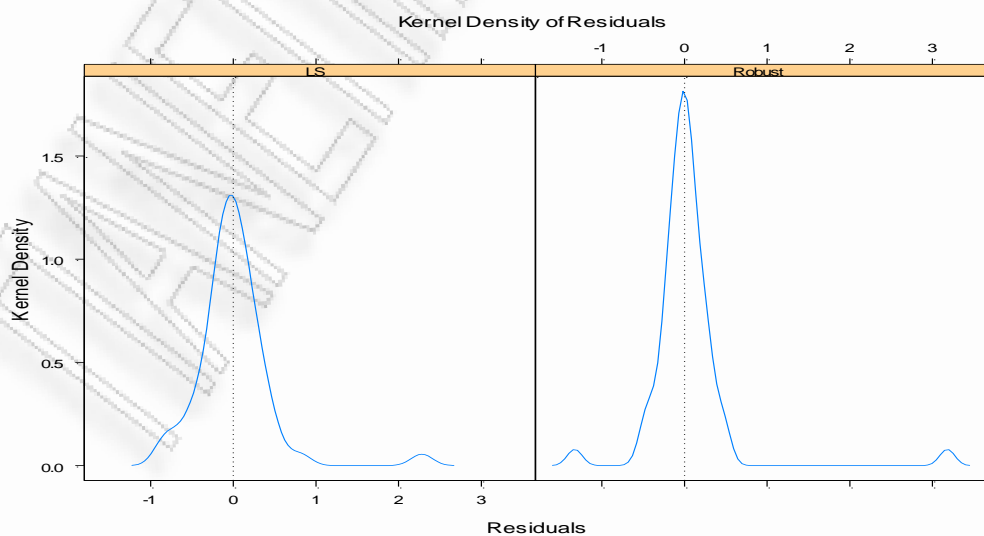
Σύμφωνα με το Γράφημα 5.7, τα τυποποιημένα κατάλοιπα της κλασσικής προσαρμογής οριακά υπονοούν την παρουσία μιας έκτοπης παρατήρησης στην 31<sup>η</sup> θέση, ενώ η ανθεκτική προσαρμογή όχι μόνο αναγνωρίζει ξεκάθαρα την 31<sup>η</sup> παρατήρηση σαν “outlier” αλλά και την 25<sup>η</sup>.



Γράφημα 5.7: Standardized Residuals vs Index για Κλασσική & Ανθεκτική προσαρμογή

### ➤ Γράφημα 3<sup>ο</sup>

Το Γράφημα 5.8 εμφανίζει διαφορετική προσαρμογή για τα δύο μοντέλα και δίνει μία ακριβή εικόνα της κατανομής των όρων του σφάλματος στο μοντέλο.



Γράφημα 5.8: Kernel Density καταλοίπων Κλασσικής & Ανθεκτικής προσαρμογής



Η εκτίμηση της πυκνότητας των καταλοίπων της ανθεκτικής προσαρμογής είναι πολύ συμπαγής και η κεντρική της περιοχή φαίνεται σωστά κεντραρισμένη στο μηδέν. Επίσης, εμφανίζει μία διακριτή «ανωμαλία» στο δεξί μέλος που υποδηλώνει την παρουσία μίας έκτοπης παρατήρησης.

#### 5.4.4 Παλινδρόμηση Ελαχίστων Περικεκομμένων Τετραγώνων - Least Trimmed Squares (1 outlier)

Εφαρμόζοντας στα δεδομένα με μία έκτοπη παρατήρηση την παλινδρόμηση ελαχίστων περικεκομμένων τετραγώνων (LTS) για το μοντέλο που περιγράψαμε στην υποενότητα 5.2.4, παίρνουμε τα εξής αποτελέσματα:

Παράμετροι	Εκτίμηση	Παράμετροι	Εκτίμηση
$a1^{LTS}$	14.2117	$a7^{LTS}$	14.3126
$a2^{LTS}$	14.3688	$a8^{LTS}$	14.1697
$a3^{LTS}$	14.4539	$a9^{LTS}$	14.2165
$a4^{LTS}$	13.9357	$a10^{LTS}$	14.1257
$a5^{LTS}$	14.3052	$d^{LTS}$	-1.1701
$a6^{LTS}$	14.3496	$s^{LTS}$	-0.2072

Πίνακας 5.27: Εκτιμήσεις Παραμέτρων με τη μέθοδο LTS εκτιμητή (1 outlier)

Οι ανθεκτικές εκτιμούμενες ζημιές δίνονται στον Πίνακα 5.28 και η ανθεκτική LTS διακύμανση του μοντέλου (για τα δεδομένα με 1 outlier) ισούται με

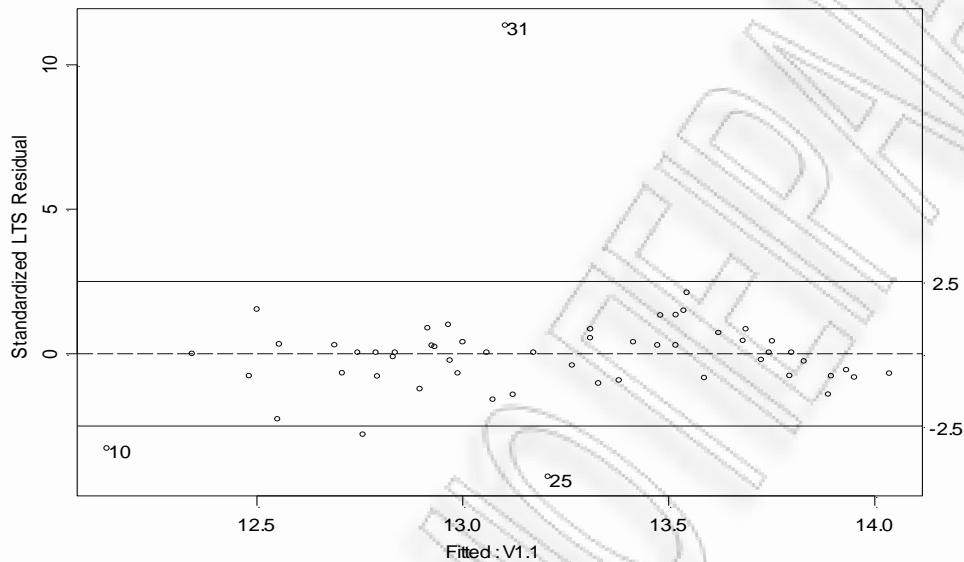
$$\hat{\sigma}_{LTS(1outlier)}^2 = 0.0933 .$$

Έτος Ατυχήματος	Έτος Εξέλιξης									
	1	2	3	4	5	6	7	8	9	10
1	357848	766940	610542	482940	527326	574398	146342	139950	227229	67948
2	352118	884021	933894	1183289	445745	320996	527804	266172	425046	235655
3	290507	1001799	926219	1016654	750816	146923	495992	280405	315126	253066
4	310608	1108250	776189	15624000	272482	352053	206286	230565	185321	151044
5	443160	693190	991983	769488	504851	470639	409980	329854	268761	219099
6	396132	937085	847498	805037	705960	526908	424414	345695	281727	229717
7	440832	847631	1131398	1063269	624670	503875	410276	334240	272441	222186
8	359480	1061648	1443370	667009	539097	438776	357314	291130	237331	193577
9	376686	986608	863788	700512	569842	463792	377679	307718	250850	204600
10	344014	989441	809685	657920	534885	435089	354101	288340	234917	191494

Πίνακας 5.28: Τετράγωνο Εξέλιξης Ζημιών με τη μέθοδο LTS εκτιμητή (1 outlier)



Το Γράφημα 5.9 απεικονίζει τα τυποποιημένα κατάλοιπα για το LTS μοντέλο και εμφανίζει τις παρατηρήσεις 31<sup>η</sup>, 25<sup>η</sup> και 10<sup>η</sup> ως έκτοπες (outliers).



Γράφημα 5.9: Γραφήματα τυποποιημένων καταλοίπων για το μοντέλο LTS

#### 5.4.5 Παλινδρόμηση Ελαχίστων Διαμέσων Τετραγώνων -Least Median Squares- (1 outlier)

Εφαρμόζοντας στα δεδομένα με μία έκτοπη παρατήρηση την παλινδρόμηση ελαχίστων διαμέσων τετραγώνων για το μοντέλο που περιγράψαμε στην υποενότητα 5.2.5, παίρνουμε τα εξής αποτελέσματα:

Παράμετροι	Εκτίμηση
<i>Intercept</i>	13.78169
$\alpha^{LMS}$	0.052187
$\beta^{LMS}$	-0.13119

Πίνακας 5.29: Εκτιμήσεις Παραμέτρων με τη μέθοδο LMS εκτιμητή (1 outlier)

Η ανθεκτική LMS διακύμανση του μοντέλου (για τα δεδομένα με 1 outlier) είναι

$$\hat{\sigma}_{LMS(1outlier)}^2 = 0.1565 ,$$

και οι ανθεκτικές εκτιμούμενες ζημιές παρουσιάζονται στον Πίνακα 5.30.

Έτος Ατυχήματος	Έτος Εξέλιξης									
	1	2	3	4	5	6	7	8	9	10
1	357848	766940	610542	482940	527326	574398	146342	139950	227229	67948
2	352118	884021	933894	1183289	445745	320996	527804	266172	425046	342433
3	290507	1001799	926219	1016654	750816	146923	495992	280405	395620	361173
4	310608	1108250	776189	1562400	272482	352053	206286	469458	417482	381048
5	443160	693190	991983	769488	504851	470639	561166	495661	440748	402193
6	396132	937085	847498	805037	705960	673177	592962	523718	465653	424815
7	440832	847631	1131398	1063269	809793	712309	627402	554100	492609	449266
8	359480	1061648	1443370	977603	859229	755754	665625	587800	522482	476300
9	376686	986608	1188850	1044348	917816	807211	710861	627644	557736	508043
10	344014	1484933	1303785	1145059	1006102	884637	778804	687328	610298	554773

Πίνακας 5.31: Τετράγωνο Εξέλιξης Ζημιών με τη μέθοδο LMS εκτιμητή (1 outlier)

## 5.5 Εφαρμογή Λογαριθμικού μοντέλου Παλινδρόμησης (2 outliers στις θέσεις (4,4) και (1,2))

Στο σημείο αυτό, επιθυμούμε να προσδιορίσουμε την επιρροή δύο έκτοπων παρατηρήσεων (outliers) στη πρόβλεψη των εκκρεμών ζημιών και του συνολικού απαιτούμενου αποθεματικού για κάθε μέθοδο ξεχωριστά.

Για το σκοπό αυτό, θα δεκαπλασιάσουμε την ατομική ζημιά που συνέβη κατά το 4<sup>ο</sup> έτος ατυχήματος και αποπληρώθηκε μετά από 4 χρόνια ( $P_{44}$ ) και την ατομική ζημιά που συνέβη κατά το 1<sup>ο</sup> έτος ατυχήματος και πληρώθηκε μετά από δύο έτη ( $P_{12}$ ). Οπότε

$$P'_{44} = 1562400 * 10 = 15624000,$$

$$P'_{12} = 766940 * 10 = 7669400.$$

Συνεπώς, το αναπροσαρμοσμένο τρίγωνο εξέλιξης ζημιών με τους λογαρίθμους των ατομικών ζημιών μετά την προσθήκη των δύο ακραίων τιμών στις θέσεις (4,4) και (1,2) εμφανίζεται στον Πίνακα 5.32.

Έτος Ατυχήματος (i)	Έτος Εξέλιξης (j)									
	1	2	3	4	5	6	7	8	9	10
1	12.788	15.853	13.322	13.088	13.176	13.261	11.894	11.849	12.334	11.126
2	12.772	13.692	13.747	13.984	13.008	12.679	13.176	12.492	12.960	
3	12.579	13.817	13.739	13.832	13.529	11.898	13.114	12.544		
4	12.646	13.918	13.562	16.564	12.515	12.772	12.237			
5	13.002	13.449	13.807	13.553	13.132	13.062				
6	12.890	13.751	13.650	13.599	13.467					
7	12.996	13.650	13.939	13.877						
8	12.792	13.875	14.182							
9	12.839	13.802								
10	12.748									

Πίνακας 5.32: Τρίγωνο Εξέλιξης Ζημιών με 2 outliers

### 5.5.1 Εκτιμήσεις παραμέτρων

Εκτελώντας γραμμική (κλασσική) πολλαπλή παλινδρόμηση παίρνουμε τα εξής αποτελέσματα:

Παράμετροι	Εκτίμηση	Παράμετροι	Εκτίμηση
<i>Intercept</i>	12.7672		
$a2^{LS}$	0.1051	$b2^{LS}$	1.1671
$a3^{LS}$	0.0104	$b3^{LS}$	0.9227
$a4^{LS}$	0.2139	$b4^{LS}$	1.2636
$a5^{LS}$	-0.0307	$b5^{LS}$	0.3261
$a6^{LS}$	-0.0319	$b6^{LS}$	-0.0927
$a7^{LS}$	0.0101	$b7^{LS}$	-0.2442
$a8^{LS}$	0.153	$b8^{LS}$	-0.5107
$a9^{LS}$	-0.0301	$b9^{LS}$	-0.1729
$a10^{LS}$	-0.0188	$b10^{LS}$	-1.6407

Πίνακας 5.33: Εκτιμήσεις Παραμέτρων με τη μέθοδο LS (2 outliers)

Στη συνέχεια, για να συμπληρώσουμε το τρίγωνο εξέλιξης ζημιών με τις μελλοντικές τιμές των λογαρίθμων των αρκεί να πολλαπλασιάσουμε τον πίνακα των

εκτιμημένων παραμέτρων με τον μελλοντικό πίνακα σχεδιασμού προσθέτοντας πάντα τον όρο τεταγμένη (intercept). Το συμπληρωμένο τετράγωνο εξέλιξης ζημιών (run-off square) δίνεται από τον παρακάτω πίνακα:

Έτος Ατυχήματος	Έτος Εξέλιξης									
	1	2	3	4	5	6	7	8	9	10
1	12.788	15.853	13.322	13.088	13.176	13.261	11.894	11.849	12.334	11.126
2	12.772	13.692	13.747	13.984	13.008	12.679	13.176	12.492	12.960	11.232
3	12.579	13.817	13.739	13.832	13.529	11.898	13.114	12.544	12.605	11.137
4	12.646	13.918	13.562	16.564	12.515	12.772	12.237	12.471	12.809	11.341
5	13.002	13.449	13.807	13.553	13.132	13.062	12.492	12.226	12.564	11.096
6	12.890	13.751	13.650	13.599	13.467	12.643	12.491	12.225	12.562	11.095
7	12.996	13.650	13.939	13.877	13.103	12.684	12.533	12.267	12.604	11.137
8	12.792	13.875	14.182	14.184	13.246	12.828	12.676	12.410	12.747	11.280
9	12.839	13.802	13.660	14.001	13.063	12.644	12.493	12.227	12.564	11.097
10	12.748	13.915	13.671	14.013	13.075	12.656	12.504	12.238	12.576	11.108

Πίνακας 5.34: Τετράγωνο Εξέλιξης Λογαριθμισμένων Ζημιών με τη μέθοδο LS (2 outliers)

Οι εκτιμούμενες ζημιές εμφανίζονται στον Πίνακα 5.35 και στην περίπτωση της Γραμμικής Παλινδρόμησης (LS) η τιμή της διακύμανσης,  $\hat{\sigma}_{LS}^2$ , ισούται με

$$\hat{\sigma}_{LS(2outliers)}^2 = 0.4336.$$

Έτος Ατυχήματος	Έτος Εξέλιξης									
	1	2	3	4	5	6	7	8	9	10
1	357848	7669400	610542	482940	527326	574398	146342	139950	227229	67948
2	352118	884021	933894	1183289	445745	320996	527804	266172	425046	120816
3	290507	1001799	926219	1016654	750816	146923	495992	280405	428025	110223
4	310608	1108250	776189	15624000	272482	352053	206286	361990	527327	135712
5	443160	693190	991983	769488	504851	470639	364641	284920	414960	106728
6	396132	937085	847498	805037	705960	422194	367171	286855	417670	107351
7	440832	847631	1131398	1063269	671836	445559	387441	302634	440504	113122
8	359480	1061648	1443370	2008043	790534	524201	455741	355891	517783	132804
9	376686	986608	1229401	1737531	683880	453362	394027	307556	447100	114428
10	344014	1768650	1389514	1962614	771996	511425	444107	346220	502225	127801

Πίνακας 5.35: Τετράγωνο Εξέλιξης Ζημιών με τη μέθοδο LS (2 outliers)

## 5.6 Ανθεκτικές Μέθοδοι (2 outliers)

Όπως και στα προηγούμενα εφαρμόζουμε τους ανθεκτικούς εκτιμητές με τη διαφορά ότι τώρα έχουμε δύο έκτοπες παρατηρήσεις στα δεδομένα μας.

### 5.6.1 Παλινδρόμηση με χρήση M-Εκτιμητή (2 outliers)

Χρησιμοποιώντας τον ανθεκτικό M-Εκτιμητή του Huber με  $b = 2$ , τη συνάρτηση του οποίου είχαμε ορίσει με τη σχέση (5.1), τα αποτελέσματα είναι τα εξής:

Παράμετροι	Εκτίμηση	Παράμετροι	Εκτίμηση
<i>Intercept</i>	12.55901		
$a2^M$	0.319107	$b2^M$	0.964327
$a3^M$	0.302637	$b3^M$	0.935399
$a4^M$	0.101274	$b4^M$	0.935917
$a5^M$	0.22936	$b5^M$	0.378437
$a6^M$	0.269383	$b6^M$	0.061266
$a7^M$	0.347693	$b7^M$	-0.125467
$a8^M$	0.424497	$b8^M$	-0.471279
$a9^M$	0.279427	$b9^M$	-0.071728
$a10^M$	0.18943	$b10^M$	-1.43251

Πίνακας 5.35: Εκτιμήσεις Παραμέτρων με τη μέθοδο M εκτιμητή (2 outliers)

Οι ανθεκτικές εκτιμούμενες ζημιές εμφανίζονται στον Πίνακα 5.36 και η ανθεκτική M διακύμανση του μοντέλου (για τα δεδομένα με 2 outliers) είναι

$$\hat{\sigma}_{M(2\text{ outliers})}^2 = 0.502253.$$

Έτος Ατυχήματος	Έτος Εξέλιξης									
	1	2	3	4	5	6	7	8	9	10
1	357848	7669400	610542	482940	527326	574398	146342	139950	227229	67948
2	352118	884021	933894	1183289	445745	320996	527804	266172	425046	161196
3	290507	1001799	926219	1016654	750816	146923	495992	280405	545517	159121
4	310608	1108250	776189	15624000	272482	352053	206286	287574	448322	130678
5	443160	693190	991983	769488	504851	470639	454641	329109	512939	149406
6	396132	937085	847498	805037	705960	567818	477674	345724	538673	156776
7	440832	847631	1131398	1063269	847036	622640	523714	378965	590245	171615
8	359480	1061648	1443370	1622780	935652	687662	578282	418323	651198	189068
9	376686	986608	1459210	1467490	845890	621505	522453	377735	587466	170139
10	344014	1555853	1516817	1524338	878032	644607	541333	390829	606317	174436

Πίνακας 5.36: Τετράγωνο Εξέλιξης Ζημιών με τη μέθοδο M εκτιμητή (2 outliers)

## 5.6.2 Παλινδρόμηση Ελαχίστων Απολύτων Αποκλίσεων -Least Absolute Deviation- (2 outliers)

Εφαρμόζοντας την παλινδρόμηση ελαχίστων απολύτων αποκλίσεων για τα δεδομένα με τις δύο ακραίες τιμές, παίρνουμε τα εξής αποτελέσματα:

Παράμετροι	Εκτίμηση	Παράμετροι	Εκτίμηση
<i>Intercept</i>	12.40624		
$a2^{LAD}$	0.425015	$b2^{LAD}$	0.933961
$a3^{LAD}$	0.477108	$b3^{LAD}$	0.915864
$a4^{LAD}$	0.240048	$b4^{LAD}$	0.88044
$a5^{LAD}$	0.266803	$b5^{LAD}$	0.64557
$a6^{LAD}$	0.41033	$b6^{LAD}$	0.12525
$a7^{LAD}$	0.59018	$b7^{LAD}$	0.230969
$a8^{LAD}$	0.535133	$b8^{LAD}$	-0.339356
$a9^{LAD}$	0.432929	$b9^{LAD}$	0.128698
$a10^{LAD}$	0.342199	$b10^{LAD}$	-1.27974

Πίνακας 5.37: Εκτιμήσεις Παραμέτρων με τη μέθοδο LAD εκτιμητή (2 outliers)

Η διακύμανση LAD του συγκεκριμένου μοντέλου είναι

$$\hat{\sigma}_{LAD(2\ outliers)}^2 = 0.56343$$

και οι ανθεκτικές εκτιμούμενες ζημιές δίνονται από τον Πίνακα 5.38.

Έτος Ατυχήματος	Έτος Εξέλιξης									
	1	2	3	4	5	6	7	8	9	10
1	357848	7669400	610542	482940	527326	574398	146342	139950	227229	67948
2	352118	884021	933894	1183289	445745	320996	527804	266172	425046	191499
3	290507	1001799	926219	1016654	750816	146923	495992	280405	716923	202537
4	310608	1108250	776189	15624000	272482	352053	206286	338905	568883	160586
5	443160	693190	991983	769488	504851	470639	604858	350770	588623	166026
6	396132	937085	847498	805037	705960	625032	705612	409120	686310	193406
7	440832	847631	1131398	1063269	1265184	759908	857732	497202	833720	234685
8	359480	1061648	1443370	1541636	1228300	737613	832368	482334	808305	227167
9	376686	986608	1507118	1463076	1165358	699580	789119	456998	765044	214409
10	344014	1612867	1590205	1542503	1227642	736311	829622	479688	800786	222759

Πίνακας 5.38: Τετράγωνο Εξέλιξης Ζημιών με τη μέθοδο LAD εκτιμητή (2 outliers)



### 5.6.3 Κλασσική & Ανθεκτική ANOVA (2 outliers)

Θεωρούμε το μοντέλο

$$Y_{ij} = \mu + \alpha_i + b_j + e_{ij}, \quad i = 1, 2, \dots, 10 \text{ και } j = 1, 2, \dots, 10$$

Εκτελώντας την Κλασσική Ανάλυση Διακύμανσης δύο παραγόντων στο S-Plus για τα δεδομένα με ενσωματωμένες τις δύο ακραίες τιμές στις θέσεις (4,4) και (1,2) οδηγούμαστε στα ίδια αποτελέσματα με την κλασσική γραμμική παλινδρόμηση (Πίνακας 5.35).

Με σκοπό την επεξήγηση αυτής της διαφοράς μεταξύ κλασσικής και ανθεκτικής προσαρμογής, στο τέλος αυτής της υποενότητας ακολουθούν τα διαγνωστικά γραφήματα μεταξύ κλασσικής και ανθεκτικής ανάλυσης διακύμανσης για την διάγνωση και την ερμηνεία των έκτοπων παρατηρήσεων.

Εφαρμόζοντας την ανθεκτική ανάλυση διακύμανσης παίρνουμε τους ανθεκτικούς συντελεστές του Πίνακα 5.39.

<b>Παράμετροι</b>	<b>Εκτίμηση</b>	<b>Παράμετροι</b>	<b>Εκτίμηση</b>
<b>Intercept</b>	12.4828		
<b>a2<sup>R.Anova</sup></b>	0.380524	<b>b2<sup>R.Anova</sup></b>	0.891668
<b>a3<sup>R.Anova</sup></b>	0.469419	<b>b3<sup>R.Anova</sup></b>	0.93994
<b>a4<sup>R.Anova</sup></b>	0.05973	<b>b4<sup>R.Anova</sup></b>	0.841921
<b>a5<sup>R.Anova</sup></b>	0.294427	<b>b5<sup>R.Anova</sup></b>	0.391735
<b>a6<sup>R.Anova</sup></b>	0.375356	<b>b6<sup>R.Anova</sup></b>	0.276944
<b>a7<sup>R.Anova</sup></b>	0.464431	<b>b7<sup>R.Anova</sup></b>	-0.104837
<b>a8<sup>R.Anova</sup></b>	0.523413	<b>b8<sup>R.Anova</sup></b>	-0.471136
<b>a9<sup>R.Anova</sup></b>	0.391966	<b>b9<sup>R.Anova</sup></b>	-0.026226
<b>a10<sup>R.Anova</sup></b>	0.26564	<b>b10<sup>R.Anova</sup></b>	-1.3563

Πίνακας 5.39: Εκτιμήσεις Παραμέτρων με τη μέθοδο R. ANOVA (2 outliers)

Η ανθεκτική R.ANOVA διακύμανση του μοντέλου (για τα δεδομένα με 2 outliers) ισούται με

$$\hat{\sigma}_{RAnova(2outliers)}^2 = 0.10699 ,$$

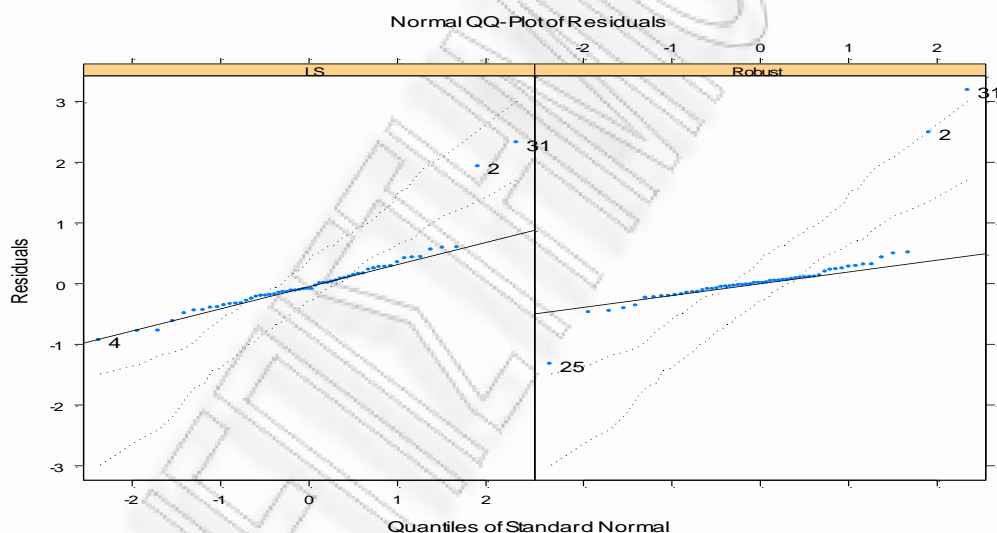
και οι ανθεκτικές εκτιμούμενες ζημιές παρουσιάζονται στον Πίνακα 5.40.

Έτος Ατυχήματος	Έτος Εξέλιξης									
	1	2	3	4	5	6	7	8	9	10
1	357848	7669400	610542	482940	527326	574398	146342	139950	227229	67948
2	352118	884021	933894	1183289	445745	320996	527804	266172	425046	111645
3	290507	1001799	926219	1016654	750816	146923	495992	280405	449267	122115
4	310608	1108250	776189	15624000	272482	352053	206286	189547	298576	81143
5	443160	693190	991983	769488	504851	470639	344555	240036	378085	102736
6	396132	937085	847498	805037	705960	546762	374348	260782	410736	111589
7	440832	847631	1131398	1063269	671041	599467	410420	285898	450257	122300
8	359480	1061648	1443370	1120329	715260	638946	437429	304693	479803	130286
9	376686	986608	1092617	991687	633094	565511	387124	269622	424491	115206
10	344014	942375	989723	898161	573300	512013	350427	243989	383932	104050

Πίνακας 5.40: Τετράγωνο Εξέλιξης Ζημιών με τη μέθοδο Robust ANOVA (2 outliers)

### ΔΙΑΓΝΩΣΤΙΚΑ ΓΡΑΦΗΜΑΤΑ

#### ➤ Γράφημα 1<sup>ο</sup>

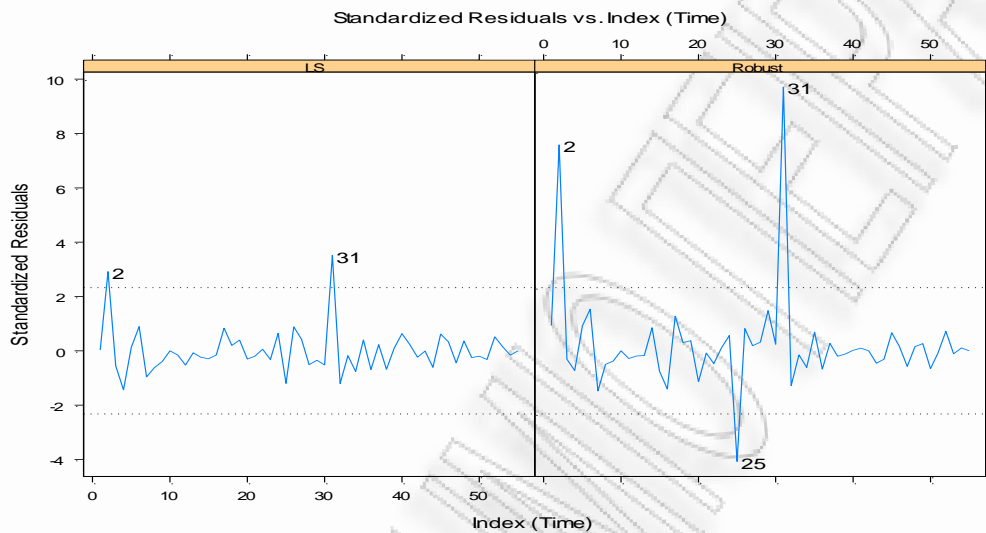


Γράφημα 5.10: Κανονικά QQ-Plots για Κλασσικά & Ανθεκτικά κατάλοιπα

Σύμφωνα με το Γράφημα 5.10, συμπεραίνουμε ότι τα κατάλοιπα για την κλασσική γραμμική παλινδρόμηση (αριστερά) είναι κατά προσέγγιση κανονικά κατανομημένα, με εξαίρεση δύο αρκετά μεγαλύτερες έκτοπες παρατηρήσεις πάνω δεξιά (τη 31<sup>η</sup> και τη 2<sup>η</sup> παρατήρηση τις οποίες δεκαπλασιάσαμε επιτούτου). Το κανονικό Q-Q γράφημα των καταλοίπων της ανθεκτικής προσαρμογής επιβεβαιώνει την ύπαρξη αυτών των έκτοπων παρατηρήσεων, αλλά και ακόμα μίας κάτω αριστερά που αντιστοιχεί στην 25<sup>η</sup> παρατήρηση, ενώ τα υπόλοιπα κατάλοιπα προσαρμόζονται αρκετά καλά στην κανονική κατανομή.

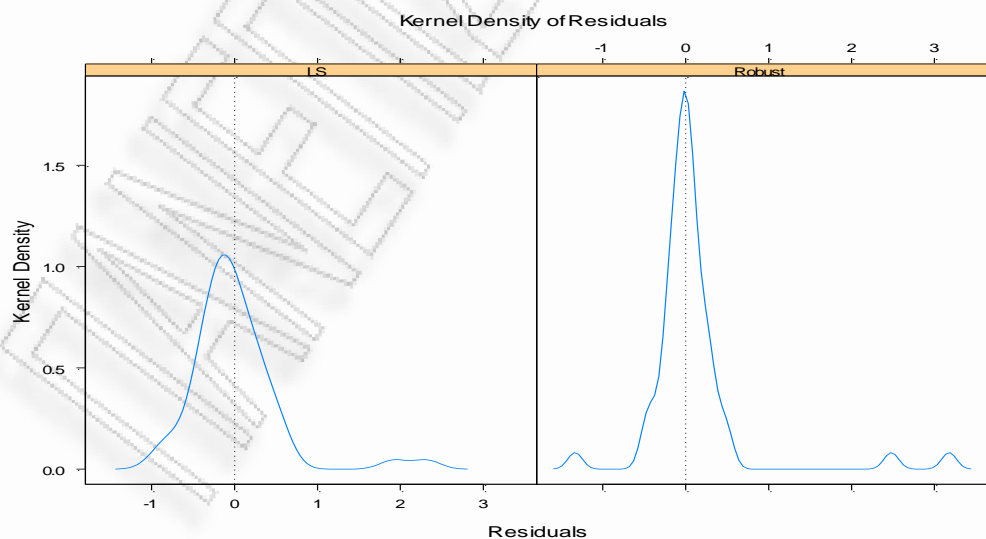
### ➤ Γράφημα 2<sup>ο</sup>

Στο Γράφημα 5.11, παρατηρούμε τα τυποποιημένα κατάλοιπα της κλασικής προσαρμογής οριακά υπονοούν την παρουσία δύο έκτοπων παρατηρήσεων στη 31<sup>η</sup> και 2<sup>η</sup> θέση, ενώ η ανθεκτική προσαρμογή όχι μόνο αναγνωρίζει ξεκάθαρα αυτές τις παρατηρήσεις σαν “outliers” αλλά και την 25<sup>η</sup>.



Γράφημα 5.11: Standardized Residuals vs Index για Κλασική & Ανθεκτική προσαρμογή

### ➤ Γράφημα 3<sup>ο</sup>



Γράφημα 5.12: Kernel Density των καταλοίπων Κλασικής & Ανθεκτικής προσαρμογής

Το Γράφημα 5.12, εμφανίζει διαφορετική προσαρμογή για τα δύο μοντέλα και δίνει μία ακριβή εικόνα της κατανομής των όρων του σφάλματος στο μοντέλο. Η εκτίμηση της πυκνότητας των καταλοίπων της ανθεκτικής προσαρμογής είναι πολύ συμπαγής και η κεντρική της περιοχή φαίνεται σωστά κεντραρισμένη στο μηδέν. Επίσης, εμφανίζει δύο διακριτές «ανωμαλίες» στο δεξί μέλος που υποδηλώνει την παρουσία δύο ακραίων τιμών.

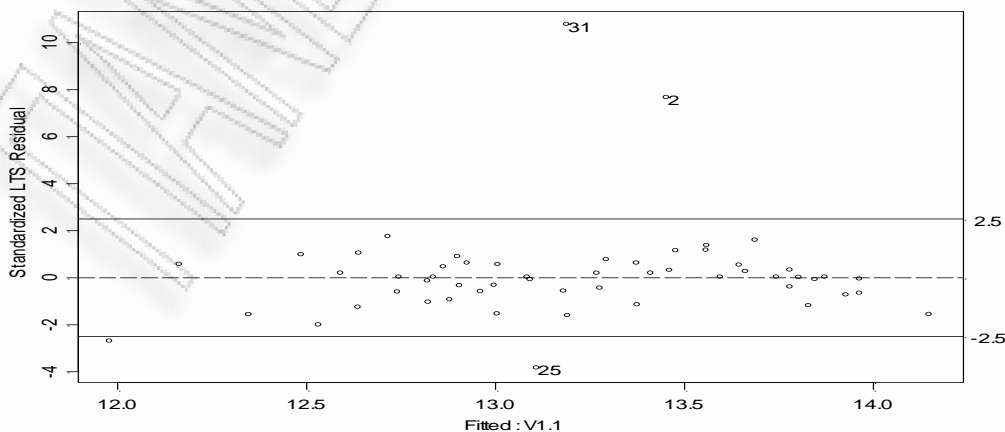
#### 5.6.4 Παλινδρόμηση Ελαχίστων Περικεκομμένων Τετραγώνων - Least Trimmed Squares- (2 outliers)

Εφαρμόζοντας για τα δεδομένα με δύο έκτοπες παρατηρήσεις την παλινδρόμηση ελαχίστων περικεκομμένων τετραγώνων για το μοντέλο που περιγράψαμε στην υποενότητα 5.2.4, παίρνουμε τα εξής αποτελέσματα:

Παράμετροι	Εκτίμηση	Παράμετροι	Εκτίμηση
$a1^{LTS}$	13.8245	$a7^{LTS}$	14.5197
$a2^{LTS}$	14.2998	$a8^{LTS}$	14.2438
$a3^{LTS}$	14.2179	$a9^{LTS}$	14.1744
$a4^{LTS}$	13.9292	$a10^{LTS}$	14.0837
$a5^{LTS}$	14.2012	$d^{LTS}$	-1.151
$a6^{LTS}$	14.3355	$s^{LTS}$	-0.1842

Πίνακας 5.41: Εκτιμήσεις Παραμέτρων με τη μέθοδο LTS εκτιμητή (2 outliers)

Το Γράφημα (5.13) απεικονίζει τα τυποποιημένα κατάλοιπα για το LTS μοντέλο και εμφανίζει τις παρατηρήσεις 31<sup>η</sup>, 25<sup>η</sup> και 2<sup>η</sup> ως έκτοπες (outliers).



Γράφημα 5.13: Γραφήματα τυποποιημένων καταλοίπων για το μοντέλο LTS

Η ανθεκτική LTS διακύμανση του μοντέλου (για τα δεδομένα με 2 outliers) ισούται με

$$\hat{\sigma}_{LTS(2\text{ outliers})}^2 = 0.305367,$$

και οι ανθεκτικές εκτιμούμενες ζημιές εμφανίζονται στον Πίνακα 5.42.

Έτος Ατυχήματος	Έτος Εξέλιξης									
	1	2	3	4	5	6	7	8	9	10
1	357848	7669400	610542	482940	527326	574398	146342	139950	227229	67948
2	352118	884021	933894	1183289	445745	320996	527804	266172	425046	277940
3	290507	1001799	926219	1016654	750816	146923	495992	280405	307328	252378
4	310608	1108250	776189	15624000	272482	352053	206286	276413	227202	189514
5	443160	693190	991983	769488	504851	470639	435682	358486	298925	249399
6	396132	937085	847498	805037	705960	598658	493179	411099	342872	286128
7	440832	847631	1131398	1063269	865311	713920	594883	495972	413738	345332
8	359480	1061648	1443370	790559	653615	544400	453688	378302	315620	263471
9	376686	986608	890972	739275	615391	512554	427141	356161	297142	248042
10	344014	998541	836452	695470	578573	481595	401096	334240	278683	232491

Πίνακας 5.42: Τετράγωνο Εξέλιξης Ζημιών με τη μέθοδο του LTS εκτιμητή (2 outliers)

### 5.6.5 Παλινδρόμηση Ελαχίστων Διαμέσων Τετραγώνων -Least Median Squares- (2 outliers)

Εφαρμόζοντας στα δεδομένα με μία έκτοπη παρατήρηση την παλινδρόμηση ελαχίστων διαμέσων τετραγώνων για το μοντέλο που περιγράψαμε στην υποενότητα 5.2.5, παίρνουμε τα εξής αποτελέσματα:

Παράμετροι	Εκτίμηση
<i>Intercept</i>	13.8712
$\alpha^{LMS}$	0.048789
$\beta^{LMS}$	-0.15079

Πίνακας 5.42: Εκτιμήσεις Παραμέτρων με τη μέθοδο LMS Εκτιμητή (2 outliers)

Η ανθεκτική LMS διακύμανση του μοντέλου (για τα δεδομένα με 2 outliers) είναι

$$\hat{\sigma}_{LMS(2\text{ outliers})}^2 = 0.17654,$$

και οι ανθεκτικές εκτιμούμενες ζημιές παρουσιάζονται στον Πίνακα 5.43.

Έτος Ατυχήματος	Έτος Εξέλιξης									
	1	2	3	4	5	6	7	8	9	10
1	357848	7669400	610542	482940	527326	574398	146342	139950	227229	67948
2	352118	884021	933894	1183289	445745	320996	527804	266172	425046	312512
3	290507	1001799	926219	1016654	750816	146923	495992	280405	365115	328543
4	310608	1108250	776189	15624000	272482	352053	206286	439641	384063	345507
5	443160	693190	991983	769488	504851	470639	533765	462731	404197	363529
6	396132	937085	847498	805037	705960	650646	562308	487446	425740	382797
7	440832	847631	1131398	1063269	795609	686513	593274	514251	449093	403653
8	359480	1061648	1443370	976782	842080	726568	627842	544157	475120	426833
9	376686	986608	1209145	1041782	898033	774764	669401	580067	506309	454453
10	344014	1542641	1328361	1144211	986082	850488	734570	636221	554835	496848

Πίνακας 5.43: Τετράγωνο Εξέλιξης Ζημιών με τη μέθοδο LMS εκτιμητή (2 outliers)

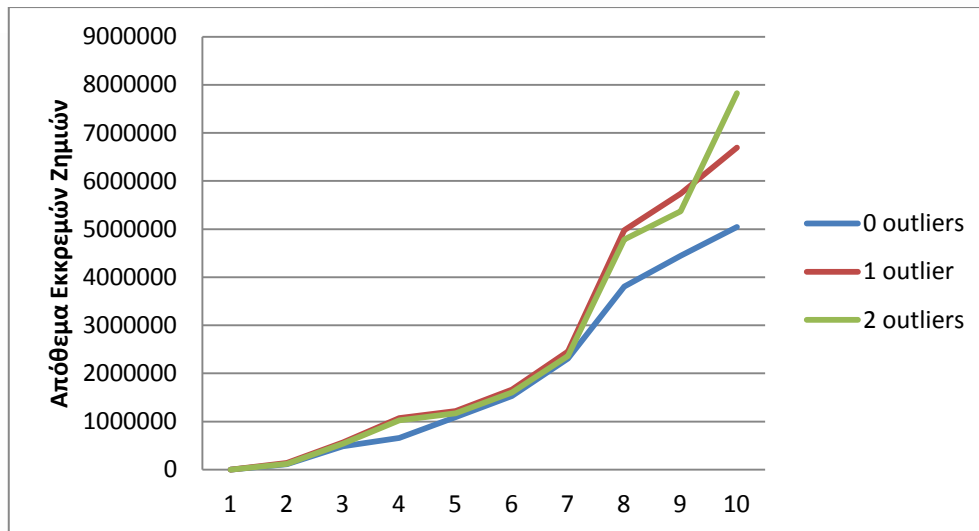
## 5.7 Σύγκριση εκκρεμών αποθεμάτων κάθε μεθόδου για 0, 1, 2 outliers

Στην παρακάτω Ενότητα συγκρίνουμε τα αποτελέσματα των αναμενόμενων εκκρεμών αποθεμάτων για κάθε μέθοδο ξεχωριστά, αρχικά για τα δεδομένα χωρίς έκτοπες παρατηρήσεις και στη συνέχεια με την προσθήκη μίας και δύο έκτοπων παρατηρήσεων. Για το σκοπό αυτό τα αποτελέσματα θα παρασταθούν υπό τη μορφή γραφημάτων, τα οποία θα απεικονίζουν τα εκτιμώμενα αποθέματα ζημιών για κάθε έτος ατυχήματος ανάλογα με τον αριθμό των έκτοπων παρατηρήσεων.

### 5.7.1 Λογαριθμικό Μοντέλο Παλινδρόμησης

Στο Γράφημα 5.14 παρατηρούμε ότι στην περίπτωση της γραμμικής παλινδρόμησης η προσθήκη ακραίων τιμών στα δεδομένα οδηγεί σε σαφής αύξηση του εκτιμώμενου αποθεματικού ζημιάς για τον 4<sup>ο</sup> χρόνο,  $R_4$ , έχουμε δηλαδή υπερεκτιμημένο  $R_4$  σε σχέση με το αντίστοιχο χωρίς ακραίες τιμές. Το  $R_4$  για την περίπτωση των δύο ακραίων τιμών είναι στα ίδια όμως επίπεδα με το αντίστοιχο για μία. Στη συνέχεια και για τα επόμενα τρία έτη τα αναμενόμενα αποθεματικά είναι παραπλήσια και για τις τρεις περιπτώσεις ύπαρξης ακραίων τιμών, ενώ από το 7<sup>ο</sup> έτος ατυχήματος και μετά παρατηρούμε απότομη αύξηση των προβλεπόμενων αποθεματικών για τα δεδομένα με ένα και δύο ακραίες τιμές.

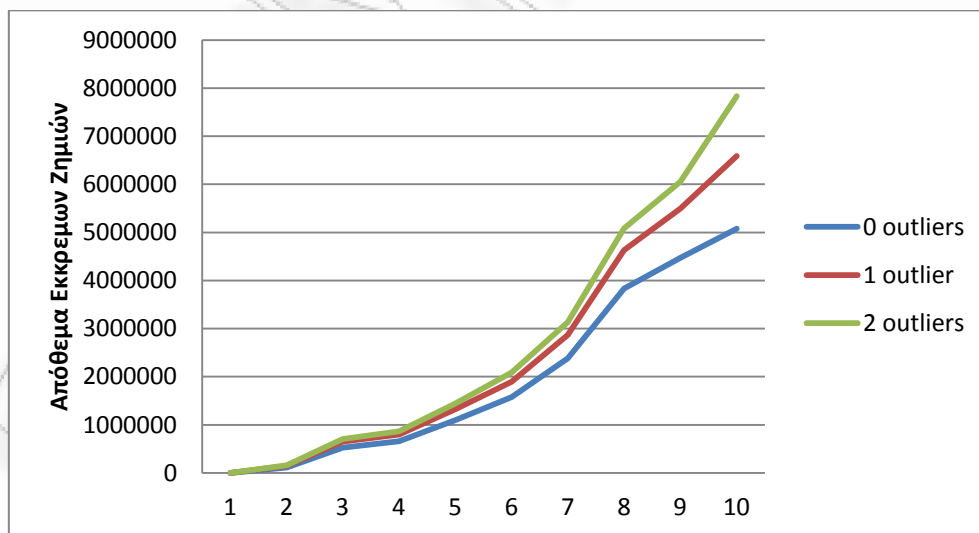




Γράφημα 5.14: Αποθέματα Εκκρεμών Ζημιών κάθε έτους ατυχήματος για το LS μοντέλο

### 5.7.2 Παλινδρόμηση με χρήση M-Εκτιμητή

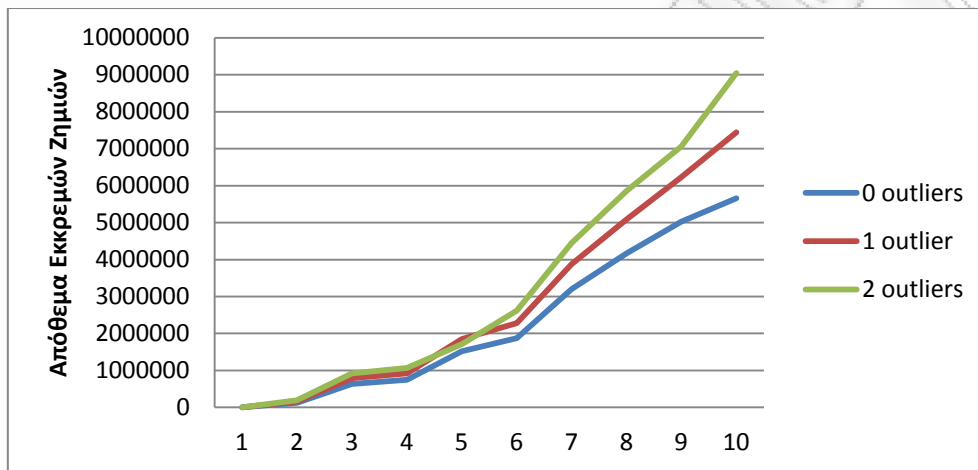
Στο μοντέλο του ανθεκτικού M-εκτιμητή τα αναμενόμενα αποθεματικά αυξάνονται με παρόμοιο τρόπο, για τα πρώτα 8 έτη, μετά την προσθήκη των δύο ακραίων τιμών στα δεδομένα. Όπως βλέπουμε στο Γράφημα 5.15, από το 8<sup>ο</sup> έτος ατυχήματος και μετά η διαφορά μεταξύ των τριών αυτών περιπτώσεων γίνεται πιο εμφανής, με το αναμενόμενο αποθεματικό για τα δύο τελευταία έτη αναφοράς, στη περίπτωση των δύο ακραίων τιμών, να φτάνει τη μέγιστη του τιμή.



Γράφημα 5.15: Αποθέματα Εκκρεμών Ζημιών κάθε έτους ατυχήματος για το μοντέλο M εκτιμητή

### 5.7.3 Παλινδρόμηση Ελαχίστων Απολύτων Αποκλίσεων (LAD)

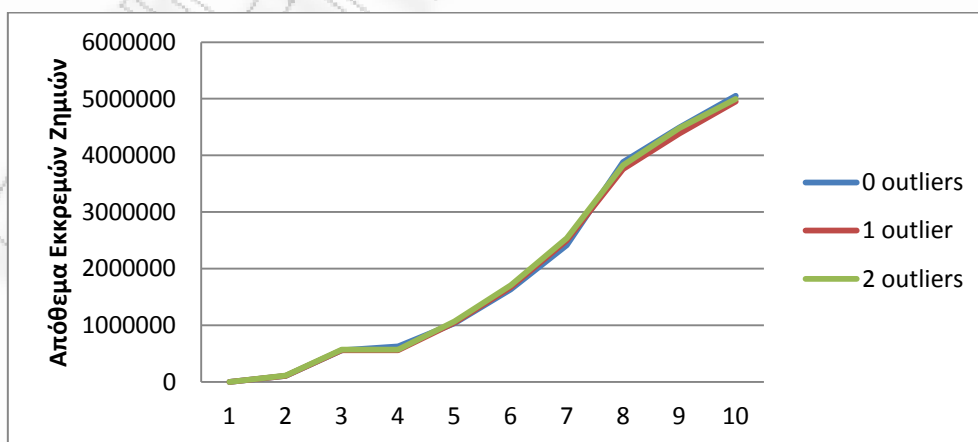
Στο μοντέλο του ανθεκτικού LAD εκτιμητή τα αναμενόμενα αποθεματικά αυξάνονται με παρόμοιο τρόπο, για τα πρώτα 6 έτη, μετά την προσθήκη των δύο ακραίων τιμών στα δεδομένα. Ωστόσο, όπως παρατηρούμε στο Γράφημα 5.16, από το 6<sup>ο</sup> έτος ατυχήματος και μετά η διαφορά μεταξύ των τριών αυτών περιπτώσεων γίνεται πιο εμφανής, με το αναμενόμενο αποθεματικό για τα δύο τελευταία έτη αναφοράς, στη περίπτωση των δύο ακραίων τιμών, να φτάνει τη μέγιστη του τιμή.



Γράφημα 5.16: Αποθέματα Εκκρεμών Ζημιών κάθε έτους ατυχήματος για το μοντέλο LAD εκτιμητή

### 5.7.4 Ανθεκτική Ανάλυση Διασποράς (Robust ANOVA)

Ακολουθεί το Γράφημα 5.17 για το μοντέλο της ανθεκτικής ανάλυσης διακύμανσης,

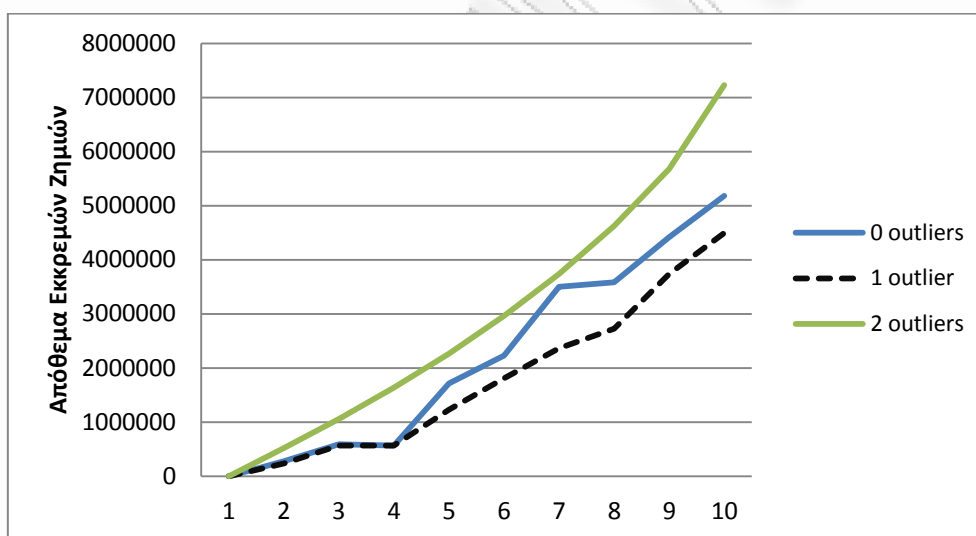


Γράφημα 5.17: Αποθέματα Εκκρεμών Ζημιών κάθε έτους ατυχήματος για το μοντέλο Robust Anova

Η ανθεκτική Ανάλυση Διακύμανσης (Robust ANOVA) δείχνει να μην επηρεάζεται από την προσθήκη των δύο ακραίων τιμών καθώς δίνει αναμενόμενες τελικές ζημιές και συνολικό αποθεματικό στα ίδια επίπεδα και για τις τρεις περιπτώσεις ύπαρξης ακραίων τιμών που μελετήσαμε.

### 5.7.5 Παλινδρόμηση Ελαχίστων Περικεκομμένων Τετραγώνων (LTS)

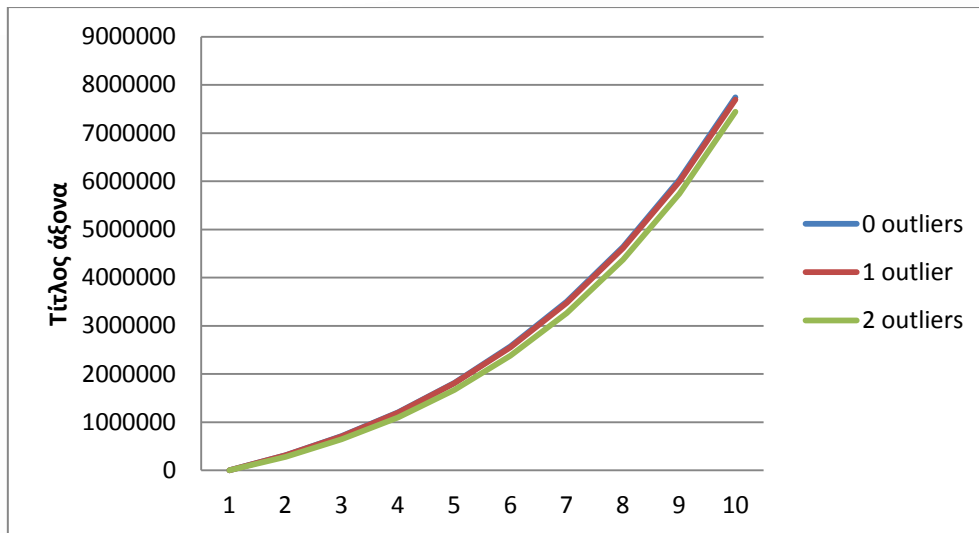
Στο Γράφημα 5.18 παρατηρούμε ότι η μέθοδος LTS παράγει ίδια αποτελέσματα για τα αναμενόμενα αποθεματικά πριν και μετά την προσθήκη της μίας έκτοπης παρατήρησης στα δεδομένα μέχρι και το 4<sup>ο</sup> έτος ατυχήματος. Στη συνέχεια οι εκτιμήσεις για τα αποθεματικά στη περίπτωση μίας έκτοπης παρατήρησης μειώνονται αισθητά. Στη περίπτωση των δύο ακραίων παρατηρήσεων στα δεδομένα είναι μεγαλύτερο από το αντίστοιχο για καμία και μία έκτοπη παρατήρηση.



Γράφημα 5.18: Αποθέματα Εκκρεμών Ζημιών κάθε έτους ατυχήματος για το μοντέλο LTS εκτιμητή

### 5.7.6 Παλινδρόμηση ελαχίστων Διαμέσων Τετραγώνων (LMS)

Η μέθοδος LMS παράγει σχεδόν ισοδύναμα αποτελέσματα για τα αναμενόμενα αποθεματικά πριν και μετά την προσθήκη ακραίων τιμών στα δεδομένα, όπως βλέπουμε στο Γράφημα 5.19.



Γράφημα 5.19: Αποθέματα Εκκρεμών Ζημιών κάθε έτους ατυχήματος για το μοντέλο LMS εκτιμητή

# Κεφάλαιο 6<sup>ο</sup>: Σύγκριση Αποθεματοποίησης Κλασσικών & Ανθεκτικών Μεθόδων

---

Στο παρόν κεφάλαιο γίνεται σύγκριση των αποτελεσμάτων αποθεματοποίησης όπως τα υπολογίσαμε στο Κεφάλαιο 5. Πιο συγκεκριμένα συγκρίνουμε το κλασσικό λογαριθμικό μοντέλο παλινδρόμησης και το μοντέλο ANOVA με τα αντίστοιχα ανθεκτικά μοντέλα παλινδρόμησης και ANOVA.

Παραθέτονται, λοιπόν, οι εκτιμήσεις των τελικών ζημιών, τα απαιτούμενα αποθεματικά για κάθε έτος ατυχήματος και τα συνολικά αποθεματικά, για καμία, μία και δύο έκτοπες παρατηρήσεις.

Για να υπολογιστεί το απόθεμα ζημιών για κάθε έτος ατυχήματος  $i$ , θα πρέπει να αφαιρεθούν οι πληρωθείσες ζημιές από τις εκτιμήσεις των τελικών ζημιών. Το συνολικό απόθεμα  $\hat{R}$ , που θα πρέπει να έχει μια ασφαλιστική εταιρεία προκειμένου να ανταποκριθεί στις μελλοντικές της υποχρεώσεις προς τους πελάτες της, είναι το άθροισμα των εκτιμώμενων αποθεματικών  $\hat{R}_i$ .

Όπως θα δούμε στους παρακάτω πίνακες, οι πληρωθείσες ζημιές είναι ίδιες για όλες τις μεθόδους καθώς πρόκειται για το άθροισμα των ήδη πραγματοποιηθεισών ζημιών (δηλαδή αυτές που βρίσκονται μέχρι και την κύρια διαγώνιο). Στην συνέχεια, ακολουθούν οι εκτιμήσεις για τις τελικές ζημιές (άθροισμα ζημιών κάθε έτους εξέλιξης για κάθε έτος ατυχήματος) και για τα αποθέματα ζημιών (αποτέλεσμα αφαίρεσης πληρωθεισών ζημιών από τις εκτιμήσεις των τελικών ζημιών).

## 6.1 Σύγκριση Αποθεματοποίησης Κλασικών & Ανθεκτικών Μεθόδων (0 outliers)

Πληρωθείσες Ζημιές	Παλινδρόμηση										Robust ANOVA	
	Γραμμική		M-Εκτιμητής		LAD		LMS		LTS		Τελικές Ζημιές	Απόθεμα Ζημιών
	Τελικές Ζημιές	Απόθεμα Ζημιών	Τελικές Ζημιές	Απόθεμα Ζημιών	Τελικές Ζημιές	Απόθεμα Ζημιών	Τελικές Ζημιές	Απόθεμα Ζημιών	Τελικές Ζημιές	Απόθεμα Ζημιών		
3901463	3901463	0	3901463	0	3901463	0	3901463	0	3901463	0	3901463	0
5339085	5449671	110586	5452891	113806	5461422	122337	5656114	317029	5617094	278009	5447607	108522
4909315	5390715	481400	5436608	527293	5546008	636693	5623019	713704	5500142	590827	5466572	557257
4588268	5248177	659909	5242948	654680	5339218	750950	5795500	1207232	5157773	569505	5216859	628591
3873311	4962691	1089380	4967054	1093743	5392635	1519324	5691787	1818476	5590313	1717002	4906010	1032699
3691712	5220351	1528639	5265696	1573984	5569929	1878217	6264449	2572737	5921365	2229653	5322989	1631277
3483130	5791230	2308100	5866387	2383257	6678050	3194920	6983841	3500711	6984874	3501744	5896206	2413076
2864498	6666790	3802292	6696712	3832214	7036025	4171527	7504150	4639652	6449046	3584548	6748341	3883843
1363294	5807834	4444540	5836754	4473460	6389437	5026143	7398090	6034796	5779118	4415824	5858946	4495652
344014	5387504	5043490	5418515	5074501	6005682	5661668	8085110	7741096	5523473	5179459	5398399	5054385
<b>34358090</b>	<b>53826426</b>	<b>19468336</b>	<b>54085027</b>	<b>19726937</b>	<b>57319869</b>	<b>22961779</b>	<b>62903524</b>	<b>28545434</b>	<b>56424661</b>	<b>22066571</b>	<b>54163392</b>	<b>19805302</b>

Πίνακας 6.1: Εκτιμώμενες Τελικές Ζημιές και Αποθεματικά για όλες τις μεθόδους (0 outliers)

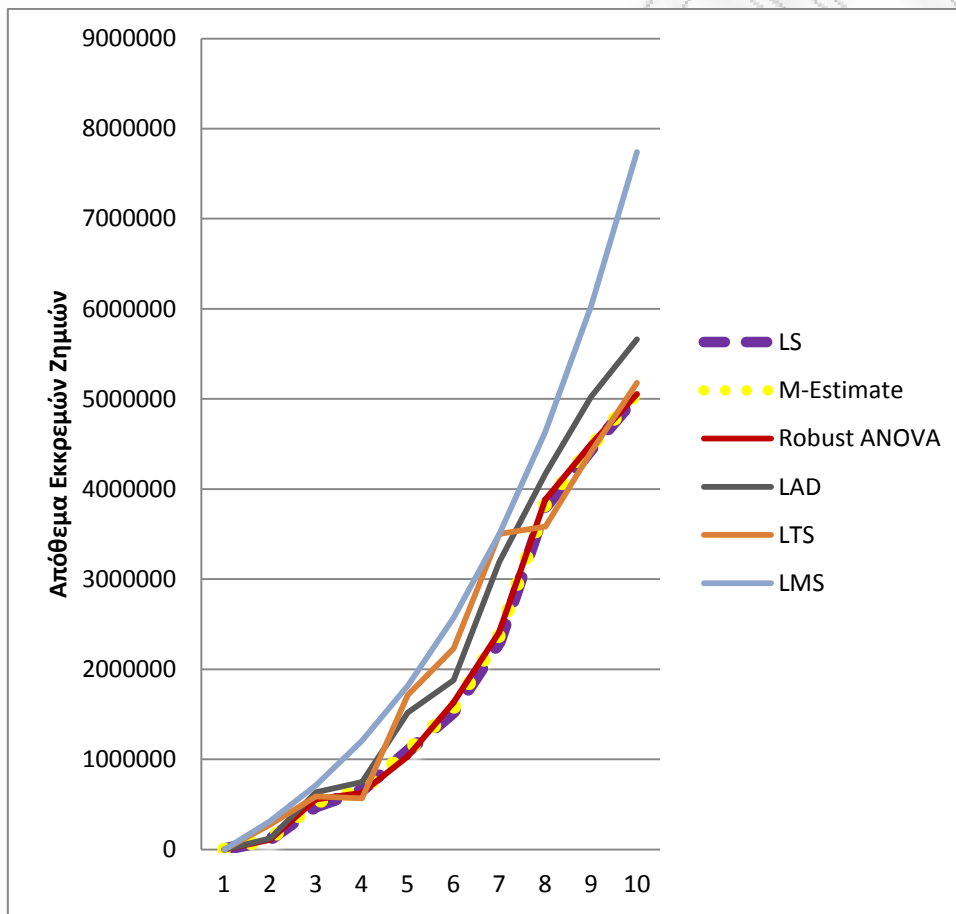
### Συμπεράσματα:

- Οι εκτιμήσεις των αποθεμάτων και των τελικών ζημιών που προέρχονται από τις μεθόδους της λογαριθμοκανονικής παλινδρόμησης, του ανθεκτικού M-Εκτιμητή και της ανθεκτικής ανάλυσης διακύμανσης (Robust ANOVA), είναι ισοδύναμες.



- Οι ανθεκτικοί LAD και LTS εκτιμητές παράγουν ελαφρώς υψηλότερες εκτιμήσεις για τις τελικές ζημιές, οπότε και για το απαιτούμενο συνολικό αποθεματικό.
- Ο ανθεκτικός LMS εκτιμητής παράγει εμφανώς μεγαλύτερες εκτιμήσεις εκκρεμών αποθεματικών σε σχέση με τις υπόλοιπες μεθόδους κλασσικές και ανθεκτικές.

Τα παραπάνω συμπεράσματα επιβεβαιώνονται από το Γράφημα 6.1,



Γράφημα 6.1: Απόθεμα Εκκρεμών Ζημιών κάθε έτους ατυχήματος για κάθε μέθοδο

## 6.2 Σύγκριση Αποθεματοποίησης Κλασσικών & Ανθεκτικών Μεθόδων (1outlier)

Πληρωθείσες Ζημιές	Παλινδρόμηση										Robust ANOVA	
	Γραμμική		M-Εκτιμητής		LAD		LMS		LTS		Τελικές Ζημιές	Απόθεμα Ζημιών
	Τελικές Ζημιές	Απόθεμα Ζημιών	Τελικές Ζημιές	Απόθεμα Ζημιών	Τελικές Ζημιές	Απόθεμα Ζημιών	Τελικές Ζημιές	Απόθεμα Ζημιών	Τελικές Ζημιές	Απόθεμα Ζημιών		
3901463	3901463	0	3901463	0	3901463	0	3901463	0	3901463	0	3901463	0
5339085	5474210	135125	5485420	146335	5498022	158937	5654088	315003	5574740	235655	5447608	108523
4909315	5476882	567567	5557339	648024	5698267	788952	5618537	709222	5477507	568192	5468840	559525
<b>18649868</b>	19723046	1073178	19447812	797944	19572298	922430	19849619	1199751	19216798	566930	19208136	558268
3873311	5092518	1219207	5195419	1322108	5722627	1849316	5680618	1807307	5101006	1227695	4919949	1046638
3691712	5356026	1664314	5587365	1895653	5972172	2280460	6248714	2557002	5500173	1808461	5366374	1674662
3483130	5938265	2455135	6351345	2868215	7358605	3875475	6962413	3479283	5850817	2367687	5978347	2495217
2864498	7841624	4977126	7494511	4630013	7949912	5085414	7475582	4611084	5588733	2724235	6620648	3756150
1363294	7097820	5734526	6856315	5493021	7596530	6233236	7360517	5997223	5102076	3738782	5746415	4383121
344014	7035884	6691870	6932558	6588544	7785422	7441408	8036121	7692107	4839886	4495872	5295702	4951688
<b>48419690</b>	<b>72937738</b>	<b>24518048</b>	<b>72809548</b>	<b>24389858</b>	<b>77055316</b>	<b>28635626</b>	<b>76787672</b>	<b>28367982</b>	<b>66153198</b>	<b>17733508</b>	<b>67953482</b>	<b>19533792</b>

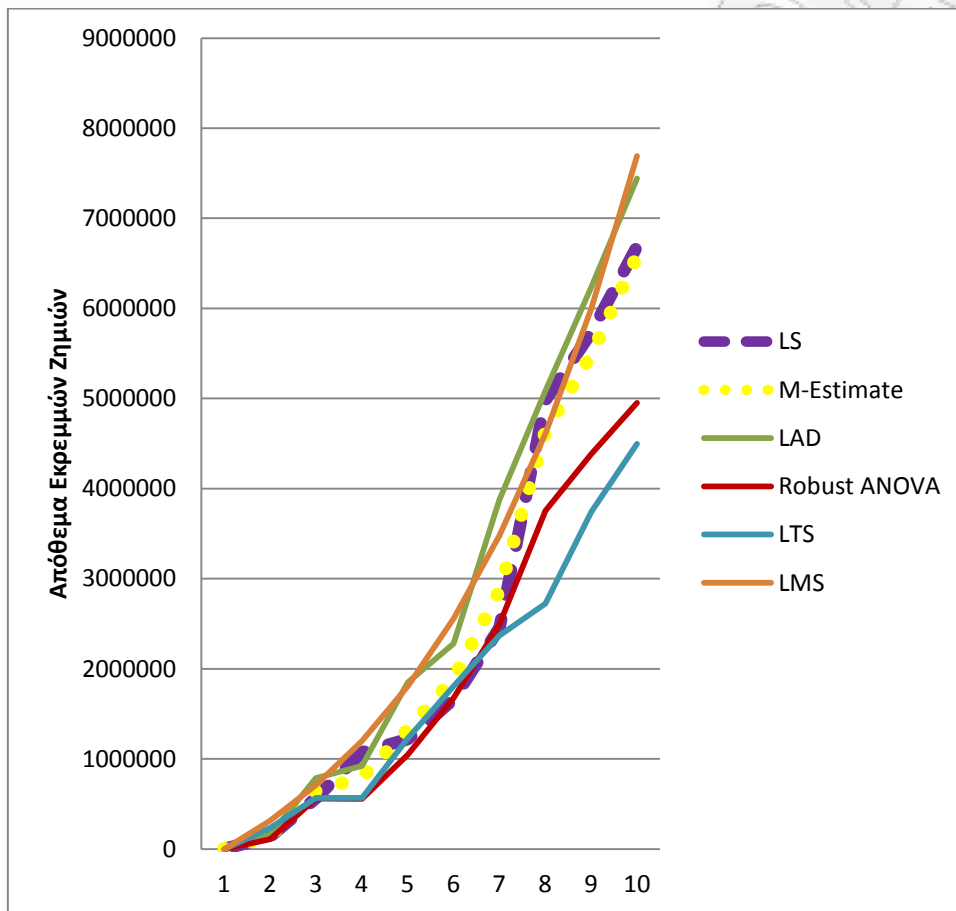
Πίνακας 6.2: Εκτιμώμενες Τελικές Ζημιές και Αποθεματικά για όλες τις μεθόδους (1 outlier)

### Συμπεράσματα:

- Οι εκτιμήσεις των αποθεμάτων και των τελικών ζημιών που προέρχονται από τις μεθόδους της λογαριθμοκανονικής παλινδρόμησης και του ανθεκτικού M-Εκτιμητή είναι παραπλήσιες.
- Η ανθεκτική ανάλυση διακύμανσης (Robust ANOVA) δείχνει να μην επηρεάζεται από την προσθήκη της μίας έκτοπης παρατήρησης καθώς δίνει αναμενόμενες τελικές ζημιές και συνολικό αποθεματικό στα ίδια επίπεδα με την περίπτωση χωρίς ακραίες τιμές.

- Οι ανθεκτικοί LAD και LMS εκτιμητές παράγουν ελαφρώς υψηλότερες εκτιμήσεις για τις τελικές ζημιές, οπότε και για το απαιτούμενο συνολικό αποθεματικό ενώ το εκτιμώμενό από τον ανθεκτικό LTS εκτιμητή συνολικό απόθεμα φαίνεται υποεκτιμημένο.

Τα παραπάνω συμπεράσματα επιβεβαιώνονται από το Γράφημα 6.2,



Γράφημα 6.2: Αποθέματα Εκκρεμιών Ζημιών κάθε έτους ατυχήματος για κάθε μέθοδο για 1 outlier

### 6.3 Σύγκριση Αποθεματοποίησης Κλασσικών & Ανθεκτικών Μεθόδων (2 outliers)

Πληρωθείσες Ζημιές	Παλινδρόμηση										Robust ANOVA	
	Γραμμική		M-Εκτιμητής		LAD		LMS		LTS		Τελικές Ζημιές	Απόθεμα Ζημιών
	Τελικές Ζημιές	Απόθεμα Ζημιών	Τελικές Ζημιές	Απόθεμα Ζημιών	Τελικές Ζημιές	Απόθεμα Ζημιών	Τελικές Ζημιές	Απόθεμα Ζημιών	Τελικές Ζημιές	Απόθεμα Ζημιών		
10803923	10803923	0	10803923	0	10803923	0	10803923	0	10803923	0	10803923	0
5339085	5459901	120816	5500281	161196	5530584	191499	5623500	284415	5861185	522100	5450730	111645
4909315	5447564	538249	5613953	704638	5828775	919460	5554095	644780	5966581	1057266	5480697	571382
18649868	19674896	1025028	19516442	866574	19718241	1068373	19748696	1098828	20284507	1634639	19219135	569267
3873311	5044560	1171249	5319406	1446095	5583588	1710277	5541702	1668391	6139709	2266398	4938722	1065411
3691712	5292953	1601241	5778378	2086666	6311192	2619480	6072061	2380349	6655185	2963473	5395928	1704216
3483130	5844225	2361095	6617344	3134214	7931562	4448432	6750954	3267824	7223468	3740338	6022512	2539382
2864498	7649496	4784998	7947463	5082965	8722221	5857723	7236139	4371641	7487456	4622958	6691243	3826745
1363294	6730579	5367285	7415182	6051888	8423997	7060703	7105443	5742149	7041762	5678468	5842645	4479351
344014	8168568	7824554	8176575	7832561	9386396	9042382	7785486	7441472	7570446	7226432	5341985	4997971
55322150	80116665	24794515	82688947	27366797	88240480	32918330	82222000	26899850	85034221	29712071	75187520	19865370

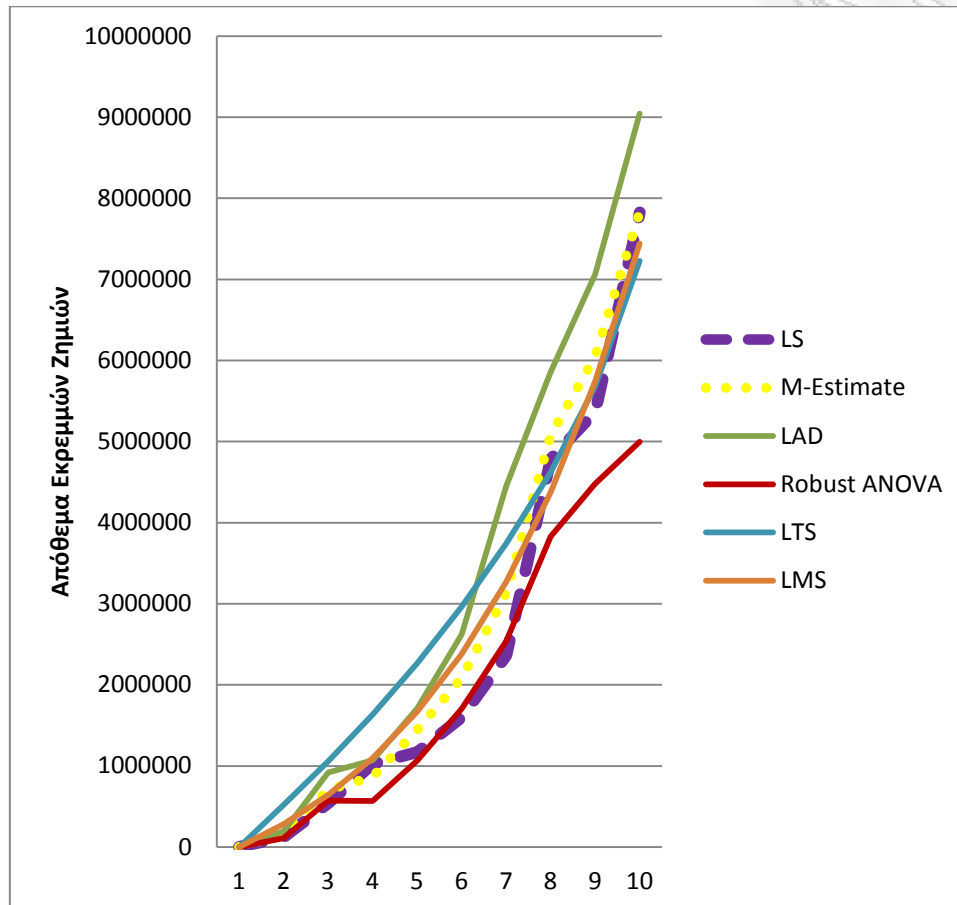
Πίνακας 6.3: Εκτιμώμενες Τελικές Ζημιές και Αποθεματικά για όλες τις μεθόδους (2 outliers)

#### Συμπεράσματα:

- Οι εκτιμήσεις των αποθεμάτων και των τελικών ζημιών που προέρχονται από τις μεθόδους της λογαριθμοκανονικής παλινδρόμησης κυμαίνονται στα ίδια επίπεδα με τη περίπτωση της μίας έκτοπης παρατήρησης και του ανθεκτικού M-Εκτιμητή είναι ελαφρώς αυξανόμενες.
- Η ανθεκτική ανάλυση διακύμανσης (Robust ANOVA) δείχνει να μην επηρεάζεται από την προσθήκη των δύο ακραίων τιμών καθώς δίνει αναμενόμενες τελικές ζημιές και συνολικό αποθεματικό στα ίδια επίπεδα με την περίπτωση χωρίς ακραίες τιμές.

- Ο ανθεκτικός LAD εκτιμητής δίνει υψηλότερες εκτιμήσεις για τις τελικές ζημιές, οπότε και για το απαιτούμενο συνολικό αποθεματικό ενώ οι ανθεκτικοί εκτιμητές LTS και LMS παράγουν ελαφρώς χαμηλότερες εκτιμήσεις.

Τα παραπάνω συμπεράσματα επιβεβαιώνονται από το Γράφημα 6.3,



Γράφημα 6.3: Αποθέματα Εκκρεμών Ζημιών κάθε έτους ατυχήματος για κάθε μέθοδο για 2 outliers

## 6.4 Σύγκριση Μεθόδων πριν και μετά την Προσθήκη Έκτοπων Παρατηρήσεων

Τα παραπάνω συμπεράσματα παρουσιάζονται στον Πίνακα 6.1 και έχουν σαν σκοπό την αξιολόγηση της αξιοπιστίας κάθε μοντέλου που εφαρμόστηκε ανάλογα με τον αριθμό των ακραίων τιμών που χρησιμοποιήθηκαν και το αντίστοιχο μέγεθος των εκτιμώμενων αποθεματικών. Ως  $R_i$  συμβολίζουμε το εκτιμώμενο αποθεματικό για το  $i$  έτος ατυχήματος (οπότε  $R_4$  το απαιτούμενο αποθεματικό για το 4<sup>ο</sup> έτος ατυχήματος) και  $R$  το εκτιμώμενο συνολικό απόθεμα.

Μοντέλο	# Ακραίων Τιμών	Αξιολόγηση	Επεξήγηση
Γραμμική Παλινδρόμηση	0	✓✓✓✓	
	1	✓✓✓	↑ $R_4, R$
	2	✓	↑ $R_i, R$
M-Εκτιμητής	0	✓✓✓✓	
	1	✓✓✓	↑ $R_4, R$
	2	✓✓✓	↑ $R_i, R$
LAD	0	✓✓✓	-
	1	✓✓	↑ $R_4, R$
	2	✓	↑↑ $R_i, R$
Robust ANOVA	0	✓✓✓✓	<u>παρόμοιο <math>R_i, R</math></u>
	1	✓✓✓✓	
	2	✓✓✓✓	
LTS	0	✓✓✓	
	1	✓✓	↓↓ $R_4, R$
	2	✓✓	↑ $R_i, R$
LMS	0	✓	<u>παρόμοιο <math>R_4, R</math></u>
	1	✓✓	
	2	✓✓✓	↓ $R_i, R$

Πίνακας 6.1: Σύγκριση μεθόδων για 0, 1 και 2 ακραίες τιμές

### Συμπεράσματα:

- ✓ Μετά τη προσθήκη μίας έκτοπης παρατήρησης στα δεδομένα μας στη θέση (4,4) του τριγώνου εξέλιξης ζημιών η 4<sup>η</sup> πληρωθείσα ζημιά αυξήθηκε, όπως βλέπουμε στον Πίνακα 6.1, γεγονός απόλυτα λογικό καθώς δεκαπλασιάσαμε την ήδη πληρωθείσα  $X_{44}$  ζημιά.



- ✓ Μετά τη προσθήκη δύο ακραίων τιμών στα δεδομένα μας στις θέσεις (4,4) και (1,2) του τριγώνου εξέλιξης ζημιών η 4<sup>η</sup> και η 1<sup>η</sup> πληρωθείσα ζημιά αυξήθηκαν όπως βλέπουμε στον Πίνακα 6.3, γεγονός απόλυτα λογικό καθώς δεκαπλασιάσαμε τις ήδη πληρωθείσες ζημιές,  $X_{44}$  και  $X_{12}$ .
- ✓ Επίσης παρατηρούμε μεγάλη αύξηση του εκτιμώμενου αποθεματικού ζημιάς για τον 4<sup>ο</sup> χρόνο, υπερεκτιμημένο  $R_4$ , στη περίπτωση της γραμμικής παλινδρόμησης ενώ στις περισσότερες ανθεκτικές μεθόδους η προσθήκη της έκτοπης παρατήρησης αύξησε μεν τα αναμενόμενα αποθεματικά σε μικρότερη δε κλίμακα, απορρόφησε δηλαδή την επιρροή της έκτοπης παρατήρησης.
- ✓ Το  $R_4$  για τις μεθόδους Robust Anova, LMS και LTS μειώθηκε ελαφρώς.
- ✓ Η μέθοδος LTS (ελαχίστων περικεκομμένων τετραγώνων) φαίνεται να παράγει υποεκτιμημένα αποτελέσματα για τα εκκρεμή αποθεματικά μετά την προσθήκη μίας έκτοπης παρατήρησης στα δεδομένα. Ωστόσο, η προσθήκη μίας δεύτερης έκτοπης παρατήρησης έχει ως συνέπεια τη διόγκωση των εκτιμώμενων αποθεματικών.
- ✓ Οι μέθοδοι LMS (ελαχίστων διαμέσων τετραγώνων) και Robust Anova δίνουν παραπλήσια αποτελέσματα για τα αναμενόμενα αποθεματικά και τις τελικές ζημιές πριν και μετά την προσθήκη μίας έκτοπης παρατήρησης στα δεδομένα.
- ✓ Για τις μεθόδους της γραμμικής παλινδρόμησης, του ανθεκτικού M-εκτιμητή και της LAD (ελαχίστων απολύτων αποκλίσεων), τα αναμενόμενα αποθεματικά και οι τελικές ζημιές μετά την προσθήκη της έκτοπης παρατήρησης στα δεδομένα αυξάνονται. Ιδιαίτερα στην περίπτωση του ανθεκτικού LAD εκτιμητή για δύο έκτοπες παρατηρήσεις τα αποθέματα φαίνονται υπερεκτιμημένα.

**Σημείωση:** Σε περίπτωση ύπαρξης έκτοπων παρατηρήσεων στο αρχικό τρίγωνο εξέλιξης ζημιών, που οφείλονται σε μεγάλες ζημιές, εμφανίζεται μεροληψία (bias). Αυτή οφείλεται στη διαφορά μεταξύ των προβλεπόμενων ανθεκτικών ζημιών,  $\hat{Y}_{ij}^R$ , και των τιμών  $Y_{ij}$  του αρχικού τριγώνου εξέλιξης ζημιών. Η διαφορά αυτή θα πρέπει να ληφθεί υπόψη στη τελική εκτίμηση των αποθεματικών και να προστεθεί σε αυτή με βάση τη στατιστική ανάλυση, τη φύση των δεδομένων και την κρίση του ερευνητή.

# Βιβλιογραφία

---

1. D.F Andrews, P.J. Bickel, F. R. Hampel, P.J. Huber, W.H. Rogers and J. W. Tukey (1972). *Robust Estimates of Location: Survey and Advances*, Princeton, N. J.: Princeton University Press.
2. R. D. Armstrong & E. L. Frome (1979). Least-absolute-value estimators for one-way and two-way tables.
3. D. Bradu & Y. Mundlak (1970). Estimation in Lognormal Linear Models, *JASA* Vol. 6, 198-211.
4. M. Busse, U. Müller & M. Dacorogna, 2010. Robust Estimation of Reserve Risk, *ASTIN Bulletin* 40(2), 453-489.
5. S. Christofides (1990). *Regression Models based on Log-incremental Payments Claims Reserving Manual*, Institute of Actuaries, London.
6. N. Draper & H. Smith (1966), Μετάφραση: Ε. Χατζηκωνσταντινίδης, Α. Καλαματιανού (1997). “Εφαρμοσμένη Ανάλυση Παλινδρόμησης”.
7. D. J. Finney (1941). On the Distribution of a Variate whose logarithm is Normally Distributed, *JBSS* , Suppl. 7, pp. 155-61.
8. F.R. Hampel, E.M. Ronchetti, P.J. Rousseeuw, W.A. Stahel (1986). *Robust Statistics: The Approach Based on Influence Functions*, Wiley, New York.
9. D. M. Hawkins (1980). *Identification of Outliers*, Chapman and Hall, London – New York.
10. P. J. Huber (1964). Robust Estimation of a Location Parameter, *Annals of Mathematical Statistics* 35, 73–101.
11. P. J. Huber & R. Dutter (1974). Numerical Solutions of Robust Regression. In: Bruckman, G. (Ed.), *Proceedings in Computational Statistics*. Physika Verlag, Vienna.
12. P. J. Huber (1981). *Robust Statistics*, Wiley, New York.
13. E. Kremer (1982). IBNR-Claims and the Two-Way Model of ANOVA. *Scand. Act. J.*, Vol. 1, 47-55.
14. T. Mack (1993). Distribution-free calculation of the standard error of chain ladder reserve estimates, *Astin Bulletin* 23/2, 213-225.
15. F. Mosteller & J. W. Tukey (1977). *Data Analysis and Regression*, Addison-Wesley, Reading, MA.

16. G. Pitselis (2008). Robust Regression Credibility: The influence Function approach, *Insurance: Mathematics & Economics* 42, 288-300.
17. G. Pitselis & V. Grigoriadou, 2011. Robust Regression Techniques Applied to Loss Reserving, presented on the 15<sup>th</sup> IME Conference, June 14-16, Trieste, Italy.
18. A. E. Renshaw (1989). Chain Ladder and Interactive Modeling & J.I.A. Vol. 116, 559-58.
19. P. J. Rousseeuw (1984). Least Median of Squares Regression, *Journal of the American Statistical Association* 79, 871-881.
20. P. J. Rousseeuw, B. Daniels, A. Leroy (1987). Applying Robust Regression to Insurance, *Insurance: Mathematics and Economics* 3, 67-72.
21. P. J. Rousseeuw & A. Leroy (1987). Robust Regression and Outlier Detection, Wiley, New York.
22. P. J. Rousseeuw & C. Croux (1993). Alternatives to the Median Absolute Deviation, *Journal of the American Statistical Association*, 88, 1273-1283.
23. G. C. Taylor & F. R. Ashe (1983). Second Moments of Estimates of Outstanding Claims. *J. of Econometrics*, Vol. 23, 37-61.
24. G. Taylor (1986). Regression Models in Claims Analysis I: Theory, Casualty Loss Reserve Seminar Transcript, Arlington, Virginia, 354-382.
25. R. J. Verall (1990). Statistical Aspects of Outstanding Claims Reserving.
26. R. J. Verall (1991a). On the Unbiased Estimation of Reserves from loglinear Models, *Insurance: Mathematics and Economics*, Vol. 10, 75-80.
27. R. J. Verall (1991b). Chain Ladder and Maximum Likelihood J.I.A. Vol.118, 489-499.
28. T. Verdonck & M. Debruyne (2010). A robustification of the Chain-Ladder Method. *Insurance mathematics & Economics* 48, 85-98.
29. T. Verdonck, M. Van Wouwe & J. Dhaene (2009). A Robustification of the Chain-Ladder Method. *NAAJ*, 13(2), 280-298.
30. M. Wüthrich & M. Merz, 2008. Stochastic Claims Reserving Methods in Insurance, Wiley Finance.
31. H. Zaharopoulou & G. Pitselis (2004). Diagnostic of outlier events and robust estimation Ανίχνευση Έκτοπων Μετρήσεων, Διαγνωστική και Ανθεκτική Εκτίμηση. Proceedings of the 17th Congress of the Greek Statistical Institute, Leykada, April 14-18, 2004, Greece.
32. B. Zehnwirth (1985). Interactive Claims Reserving Forecasting System, Benhar Nominees Pty Ltd., Tunawarra, NSW, Australia.

# Παράρτημα

## Α1. Συμπληρωμένα Τρίγωνα Εξέλιξης Ζημιών με Λογαρίθμους ατομικών ζημιών (0 outliers)

Έτος Ατυχήματος	Έτος Εξέλιξης									
	1	2	3	4	5	6	7	8	9	10
1	12.788	13.550	13.322	13.088	13.176	13.261	11.894	11.849	12.334	11.126
2	12.772	13.692	13.747	13.984	13.008	12.679	13.176	12.492	12.960	11.488
3	12.579	13.817	13.739	13.832	13.529	11.898	13.114	12.544	12.749	11.409
4	12.646	13.918	13.562	14.262	12.515	12.772	12.237	12.252	12.638	11.298
5	13.002	13.449	13.807	13.553	13.132	13.062	12.684	12.363	12.749	11.409
6	12.890	13.751	13.650	13.599	13.467	12.827	12.713	12.392	12.778	11.438
7	12.996	13.650	13.939	13.877	13.295	12.907	12.794	12.473	12.858	11.519
8	12.792	13.875	14.182	13.965	13.383	12.995	12.882	12.561	12.947	11.607
9	12.839	13.802	13.804	13.830	13.248	12.860	12.747	12.426	12.812	11.472
10	12.748	13.660	13.687	13.713	13.132	12.744	12.630	12.309	12.695	11.355

Πίνακας (1): Γραμμική Παλινδρόμηση

Έτος Ατυχήματος	Έτος Εξέλιξης									
	1	2	3	4	5	6	7	8	9	10
1	12.788	13.550	13.322	13.088	13.176	13.261	11.894	11.849	12.334	11.126
2	12.772	13.692	13.747	13.984	13.008	12.679	13.176	12.492	12.960	11.514
3	12.579	13.817	13.739	13.832	13.529	11.898	13.114	12.544	12.835	11.508
4	12.646	13.918	13.562	14.262	12.515	12.772	12.237	12.221	12.636	11.309
5	13.002	13.449	13.807	13.553	13.132	13.062	12.681	12.347	12.761	11.434
6	12.890	13.751	13.650	13.599	13.467	12.893	12.722	12.387	12.802	11.475
7	12.996	13.650	13.939	13.877	13.320	12.978	12.807	12.472	12.887	11.560
8	12.792	13.875	14.182	13.936	13.401	13.059	12.888	12.553	12.968	11.641
9	12.839	13.802	13.804	13.801	13.266	12.924	12.753	12.418	12.833	11.506
10	12.748	13.660	13.687	13.685	13.149	12.807	12.637	12.302	12.716	11.390

Πίνακας (2): Μ Εκτιμητής

Έτος Ατυχήματος	Έτος Εξέλιξης									
	1	2	3	4	5	6	7	8	9	10
1	12.788	13.550	13.322	13.088	13.176	13.261	11.894	11.849	12.334	11.126
2	12.772	13.692	13.747	13.984	13.008	12.679	13.176	12.492	12.960	11.552
3	12.579	13.817	13.739	13.832	13.529	11.898	13.114	12.544	13.012	11.604
4	12.646	13.918	13.562	14.262	12.515	12.772	12.237	12.307	12.775	11.367
5	13.002	13.449	13.807	13.553	13.132	13.062	13.123	12.552	13.020	11.612
6	12.890	13.751	13.650	13.599	13.467	12.942	13.048	12.477	12.945	11.537
7	12.996	13.650	13.939	13.877	13.642	13.122	13.227	12.657	13.125	11.717

8	12.792	13.875	14.182	13.822	13.587	13.067	13.172	12.602	13.070	11.662
9	12.839	13.802	13.784	13.749	13.514	12.993	13.099	12.529	12.997	11.588
10	12.748	13.682	13.664	13.629	13.394	12.874	12.979	12.409	12.877	11.469

Πίνακας (3): LAD Εκτιμητής

Έτος Ατυχήματος	Έτος Εξέλιξης									
	1	2	3	4	5	6	7	8	9	10
1	12.788	13.550	13.322	13.088	13.176	13.261	11.894	11.849	12.334	11.126
2	12.772	13.692	13.747	13.984	13.008	12.679	13.176	12.492	12.960	11.487
3	12.579	13.817	13.739	13.832	13.529	11.898	13.114	12.544	12.910	11.570
4	12.646	13.918	13.562	14.262	12.515	12.772	12.237	12.182	12.622	11.282
5	13.002	13.449	13.807	13.553	13.132	13.062	12.632	12.294	12.733	11.394
6	12.890	13.751	13.650	13.599	13.467	13.088	12.705	12.367	12.806	11.467
7	12.996	13.650	13.939	13.877	13.297	13.168	12.786	12.448	12.887	11.547
8	12.792	13.875	14.182	13.957	13.384	13.255	12.872	12.534	12.974	11.634
9	12.839	13.802	13.804	13.822	13.249	13.120	12.737	12.399	12.838	11.499
10	12.748	13.660	13.687	13.706	13.132	13.003	12.621	12.282	12.722	11.382

Πίνακας (4): Robust Anova

Έτος Ατυχήματος	Έτος Εξέλιξης									
	1	2	3	4	5	6	7	8	9	10
1	12.788	13.550	13.322	13.088	13.176	13.261	11.894	11.849	12.334	11.126
2	12.772	13.692	13.747	13.984	13.008	12.679	13.176	12.492	12.960	12.774
3	12.579	13.817	13.739	13.832	13.529	11.898	13.114	12.544	12.876	12.808
4	12.646	13.918	13.562	14.262	12.515	12.772	12.237	12.978	12.910	12.843
5	13.002	13.449	13.807	13.553	13.132	13.062	13.080	13.012	12.945	12.877
6	12.890	13.751	13.650	13.599	13.467	13.181	13.114	13.046	12.979	12.911
7	12.996	13.650	13.939	13.877	13.283	13.216	13.148	13.081	13.013	12.946
8	12.792	13.875	14.182	13.385	13.317	13.250	13.182	13.115	13.047	12.980
9	12.839	13.802	13.487	13.419	13.352	13.284	13.217	13.149	13.082	13.014
10	12.748	13.589	13.521	13.454	13.386	13.319	13.251	13.184	13.116	13.049

Πίνακας (5): LTS Εκτιμητής

Έτος Ατυχήματος	Έτος Εξέλιξης									
	1	2	3	4	5	6	7	8	9	10
1	12.788	13.550	13.322	13.088	13.176	13.261	11.894	11.849	12.334	11.126
2	12.772	13.692	13.747	13.984	13.008	12.679	13.176	12.492	12.960	12.574
3	12.579	13.817	13.739	13.832	13.529	11.898	13.114	12.544	12.758	12.626
4	12.646	13.918	13.562	14.262	12.515	12.772	12.237	12.941	12.810	12.678
5	13.002	13.449	13.807	13.553	13.132	13.062	13.124	12.993	12.862	12.731
6	12.890	13.751	13.650	13.599	13.467	13.308	13.176	13.045	12.914	12.783
7	12.996	13.650	13.939	13.877	13.491	13.360	13.229	13.097	12.966	12.835
8	12.792	13.875	14.182	13.674	13.543	13.412	13.281	13.150	13.018	12.887
9	12.839	13.802	13.858	13.727	13.595	13.464	13.333	13.202	13.071	12.939



10	12.748	14.041	13.910	13.779	13.648	13.516	13.385	13.254	13.123	12.992
----	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------

Πίνακας (6): LMS Εκτιμητής

## A2. Συμπληρωμένα Τρίγωνα Εξέλιξης Ζημιών με Λογαρίθμους ατομικών ζημιών (1 outlier)

Έτος Ατυχήματος	Έτος Εξέλιξης									
	1	2	3	4	5	6	7	8	9	10
1	12.788	13.550	13.322	13.088	13.176	13.261	11.894	11.849	12.334	11.126
2	12.772	13.692	13.747	13.984	13.008	12.679	13.176	12.492	12.960	11.487
3	12.579	13.817	13.739	13.832	13.529	11.898	13.114	12.544	12.749	11.409
4	12.646	13.918	13.562	16.564	12.515	12.772	12.237	12.581	12.966	11.627
5	13.002	13.449	13.807	13.553	13.132	13.062	12.588	12.349	12.735	11.395
6	12.890	13.751	13.650	13.599	13.467	12.735	12.602	12.363	12.749	11.409
7	12.996	13.650	13.939	13.877	13.199	12.796	12.663	12.424	12.810	11.470
8	12.792	13.875	14.182	14.294	13.370	12.966	12.834	12.595	12.981	11.641
9	12.839	13.802	13.804	14.159	13.235	12.831	12.699	12.460	12.846	11.506
10	12.748	13.660	13.687	14.042	13.118	12.714	12.582	12.343	12.729	11.389

Πίνακας (7): Γραμμική Παλινδρόμηση

Έτος Ατυχήματος	Έτος Εξέλιξης									
	1	2	3	4	5	6	7	8	9	10
1	12.788	13.550	13.322	13.088	13.176	13.261	11.894	11.849	12.334	11.126
2	12.772	13.692	13.747	13.984	13.008	12.679	13.176	12.492	12.960	11.515
3	12.579	13.817	13.739	13.832	13.529	11.898	13.114	12.544	12.835	11.508
4	12.646	13.918	13.562	16.564	12.515	12.772	12.237	12.222	12.636	11.310
5	13.002	13.449	13.807	13.553	13.132	13.062	12.680	12.347	12.761	11.435
6	12.890	13.751	13.650	13.599	13.467	12.892	12.720	12.387	12.802	11.475
7	12.996	13.650	13.939	13.877	13.319	12.977	12.805	12.472	12.887	11.561
8	12.792	13.875	14.182	13.937	13.401	13.059	12.887	12.553	12.968	11.642
9	12.839	13.802	13.804	13.802	13.265	12.923	12.751	12.418	12.833	11.507
10	12.748	13.660	13.687	13.685	13.149	12.807	12.635	12.302	12.716	11.390

Πίνακας (8): M Εκτιμητής

Έτος Ατυχήματος	Έτος Εξέλιξης									
	1	2	3	4	5	6	7	8	9	10
1	12.788	13.550	13.322	13.088	13.176	13.261	11.894	11.849	12.334	11.126
2	12.772	13.692	13.747	13.984	13.008	12.679	13.176	12.492	12.960	11.552
3	12.579	13.817	13.739	13.832	13.529	11.898	13.114	12.544	13.012	11.604
4	12.646	13.918	13.562	16.564	12.515	12.772	12.237	12.307	12.775	11.367
5	13.002	13.449	13.807	13.553	13.132	13.062	13.123	12.552	13.020	11.612
6	12.890	13.751	13.650	13.599	13.467	12.942	13.048	12.477	12.945	11.537
7	12.996	13.650	13.939	13.877	13.642	13.122	13.227	12.657	13.125	11.717



8	12.792	13.875	14.182	13.822	13.587	13.067	13.172	12.602	13.070	11.662
9	12.839	13.802	13.784	13.749	13.514	12.993	13.099	12.529	12.997	11.588
10	12.748	13.682	13.664	13.629	13.394	12.874	12.979	12.409	12.877	11.469

Πίνακας (9): LAD Εκτιμητής

Έτος Ατυχήματος	Έτος Εξέλιξης									
	1	2	3	4	5	6	7	8	9	10
1	12.788	13.550	13.322	13.088	13.176	13.261	11.894	11.849	12.334	11.126
2	12.772	13.692	13.747	13.984	13.008	12.679	13.176	12.492	12.960	11.487
3	12.579	13.817	13.739	13.832	13.529	11.898	13.114	12.544	12.914	11.575
4	12.646	13.918	13.562	16.564	12.515	12.772	12.237	12.063	12.504	11.164
5	13.002	13.449	13.807	13.553	13.132	13.062	12.665	12.297	12.738	11.398
6	12.890	13.751	13.650	13.599	13.467	13.127	12.745	12.377	12.818	11.478
7	12.996	13.650	13.939	13.877	13.332	13.215	12.833	12.464	12.905	11.565
8	12.792	13.875	14.182	13.839	13.389	13.272	12.889	12.521	12.962	11.622
9	12.839	13.802	13.804	13.704	13.253	13.136	12.754	12.386	12.827	11.487
10	12.748	13.660	13.687	13.588	13.137	13.020	12.638	12.269	12.710	11.371

Πίνακας (10): Robust Anova

Έτος Ατυχήματος	Έτος Εξέλιξης									
	1	2	3	4	5	6	7	8	9	10
1	12.788	13.550	13.322	13.088	13.176	13.261	11.894	11.849	12.334	11.126
2	12.772	13.692	13.747	13.984	13.008	12.679	13.176	12.492	12.960	12.903
3	12.579	13.817	13.739	13.832	13.529	11.898	13.114	12.544	12.984	12.935
4	12.646	13.918	13.562	16.564	12.515	12.772	12.237	13.064	13.015	12.966
5	13.002	13.449	13.807	13.553	13.132	13.062	13.144	13.095	13.046	12.997
6	12.890	13.751	13.650	13.599	13.467	13.225	13.175	13.126	13.077	13.028
7	12.996	13.650	13.939	13.877	13.305	13.256	13.207	13.158	13.108	13.059
8	12.792	13.875	14.182	13.385	13.336	13.287	13.238	13.189	13.140	13.091
9	12.839	13.802	13.465	13.416	13.367	13.318	13.269	13.220	13.171	13.122
10	12.748	13.546	13.497	13.448	13.398	13.349	13.300	13.251	13.202	13.153

Πίνακας (11): LTS Εκτιμητής

Έτος Ατυχήματος	Έτος Εξέλιξης									
	1	2	3	4	5	6	7	8	9	10
1	12.788	13.550	13.322	13.088	13.176	13.261	11.894	11.849	12.334	11.126
2	12.772	13.692	13.747	13.984	13.008	12.679	13.176	12.492	12.960	12.574
3	12.579	13.817	13.739	13.832	13.529	11.898	13.114	12.544	12.758	12.626
4	12.646	13.918	13.562	16.564	12.515	12.772	12.237	12.941	12.810	12.678
5	13.002	13.449	13.807	13.553	13.132	13.062	13.124	12.993	12.862	12.731
6	12.890	13.751	13.650	13.599	13.467	13.308	13.176	13.045	12.914	12.783
7	12.996	13.650	13.939	13.877	13.491	13.360	13.229	13.097	12.966	12.835

8	12.792	13.875	14.182	13.674	13.543	13.412	13.281	13.150	13.018	12.887
9	12.839	13.802	13.858	13.727	13.595	13.464	13.333	13.202	13.071	12.939
10	12.748	14.041	13.910	13.779	13.648	13.516	13.385	13.254	13.123	12.992

Πίνακας (12): LMS Εκτιμητής

### Α3. Συμπληρωμένα Τρίγωνα Εξέλιξης Ζημιών με Λογαρίθμους ατομικών ζημιών (2 outliers)

Έτος Ατυχήματος	Έτος Εξέλιξης									
	1	2	3	4	5	6	7	8	9	10
1	12.788	15.853	13.322	13.088	13.176	13.261	11.894	11.849	12.334	11.126
2	12.772	13.692	13.747	13.984	13.008	12.679	13.176	12.492	12.960	11.232
3	12.579	13.817	13.739	13.832	13.529	11.898	13.114	12.544	12.605	11.137
4	12.646	13.918	13.562	16.564	12.515	12.772	12.237	12.471	12.809	11.341
5	13.002	13.449	13.807	13.553	13.132	13.062	12.492	12.226	12.564	11.096
6	12.890	13.751	13.650	13.599	13.467	12.643	12.491	12.225	12.562	11.095
7	12.996	13.650	13.939	13.877	13.103	12.684	12.533	12.267	12.604	11.137
8	12.792	13.875	14.182	14.184	13.246	12.828	12.676	12.410	12.747	11.280
9	12.839	13.802	13.660	14.001	13.063	12.644	12.493	12.227	12.564	11.097
10	12.748	13.915	13.671	14.013	13.075	12.656	12.504	12.238	12.576	11.108

Πίνακας (13): Γραμμική Παλινδρόμηση

Έτος Ατυχήματος	Έτος Εξέλιξης									
	1	2	3	4	5	6	7	8	9	10
1	12.788	15.853	13.322	13.088	13.176	13.261	11.894	11.849	12.334	11.126
2	12.772	13.692	13.747	13.984	13.008	12.679	13.176	12.492	12.960	11.446
3	12.579	13.817	13.739	13.832	13.529	11.898	13.114	12.544	12.790	11.429
4	12.646	13.918	13.562	16.564	12.515	12.772	12.237	12.189	12.589	11.228
5	13.002	13.449	13.807	13.553	13.132	13.062	12.663	12.317	12.717	11.356
6	12.890	13.751	13.650	13.599	13.467	12.890	12.703	12.357	12.757	11.396
7	12.996	13.650	13.939	13.877	13.285	12.968	12.781	12.435	12.835	11.474
8	12.792	13.875	14.182	13.919	13.362	13.045	12.858	12.512	12.912	11.551
9	12.839	13.802	13.774	13.774	13.217	12.900	12.713	12.367	12.767	11.406
10	12.748	13.713	13.684	13.684	13.127	12.810	12.623	12.277	12.677	11.316

Πίνακας (14): M Εκτιμητής

Έτος Ατυχήματος	Έτος Εξέλιξης									
	1	2	3	4	5	6	7	8	9	10
1	12.788	15.853	13.322	13.088	13.176	13.261	11.894	11.849	12.334	11.126
2	12.772	13.692	13.747	13.984	13.008	12.679	13.176	12.492	12.960	11.552
3	12.579	13.817	13.739	13.832	13.529	11.898	13.114	12.544	13.012	11.604
4	12.646	13.918	13.562	16.564	12.515	12.772	12.237	12.307	12.775	11.367
5	13.002	13.449	13.807	13.553	13.132	13.062	12.904	12.334	12.802	11.393

6	12.890	13.751	13.650	13.599	13.467	12.942	13.048	12.477	12.945	11.537
7	12.996	13.650	13.939	13.877	13.642	13.122	13.227	12.657	13.125	11.717
8	12.792	13.875	14.182	13.822	13.587	13.067	13.172	12.602	13.070	11.662
9	12.839	13.802	13.755	13.720	13.485	12.964	13.070	12.500	12.968	11.559
10	12.748	13.682	13.664	13.629	13.394	12.874	12.979	12.409	12.877	11.469

Πίνακας (15): LAD Εκτιμητής

Έτος Ατυχήματος	Έτος Εξέλιξης									
	1	2	3	4	5	6	7	8	9	10
1	12.788	15.853	13.322	13.088	13.176	13.261	11.894	11.849	12.334	11.126
2	12.772	13.692	13.747	13.984	13.008	12.679	13.176	12.492	12.960	11.507
3	12.579	13.817	13.739	13.832	13.529	11.898	13.114	12.544	12.926	11.596
4	12.646	13.918	13.562	16.564	12.515	12.772	12.237	12.071	12.516	11.186
5	13.002	13.449	13.807	13.553	13.132	13.062	12.672	12.306	12.751	11.421
6	12.890	13.751	13.650	13.599	13.467	13.135	12.753	12.387	12.832	11.502
7	12.996	13.650	13.939	13.877	13.339	13.224	12.842	12.476	12.921	11.591
8	12.792	13.875	14.182	13.848	13.398	13.283	12.901	12.535	12.980	11.650
9	12.839	13.802	13.815	13.717	13.267	13.152	12.770	12.404	12.849	11.518
10	12.748	13.640	13.688	13.590	13.140	13.025	12.644	12.277	12.722	11.392

Πίνακας (16): Robust Anova

Έτος Ατυχήματος	Έτος Εξέλιξης									
	1	2	3	4	5	6	7	8	9	10
1	12.788	15.853	13.322	13.088	13.176	13.261	11.894	11.849	12.334	11.126
2	12.772	13.692	13.747	13.984	13.008	12.679	13.176	12.492	12.960	12.834
3	12.579	13.817	13.739	13.832	13.529	11.898	13.114	12.544	12.911	12.857
4	12.646	13.918	13.562	16.564	12.515	12.772	12.237	12.987	12.933	12.879
5	13.002	13.449	13.807	13.553	13.132	13.062	13.063	13.009	12.956	12.902
6	12.890	13.751	13.650	13.599	13.467	13.140	13.086	13.032	12.978	12.924
7	12.996	13.650	13.939	13.877	13.216	13.162	13.108	13.054	13.000	12.946
8	12.792	13.875	14.182	13.292	13.238	13.184	13.131	13.077	13.023	12.969
9	12.839	13.802	13.369	13.315	13.261	13.207	13.153	13.099	13.045	12.991
10	12.748	13.445	13.391	13.337	13.283	13.229	13.175	13.121	13.068	13.014

Πίνακας (17): LTS Εκτιμητής

Έτος Ατυχήματος	Έτος Εξέλιξης									
	1	2	3	4	5	6	7	8	9	10
1	12.788	15.853	13.322	13.088	13.176	13.261	11.894	11.849	12.334	11.126
2	12.772	13.692	13.747	13.984	13.008	12.679	13.176	12.492	12.960	12.461
3	12.579	13.817	13.739	13.832	13.529	11.898	13.114	12.544	12.660	12.510
4	12.646	13.918	13.562	16.564	12.515	12.772	12.237	12.860	12.709	12.558
5	13.002	13.449	13.807	13.553	13.132	13.062	13.060	12.909	12.758	12.607

6	12.890	13.751	13.650	13.599	13.467	13.259	13.108	12.958	12.807	12.656
7	12.996	13.650	13.939	13.877	13.459	13.308	13.157	13.006	12.856	12.705
8	12.792	13.875	14.182	13.658	13.508	13.357	13.206	13.055	12.904	12.754
9	12.839	13.802	13.858	13.707	13.556	13.406	13.255	13.104	12.953	12.802
10	12.748	14.058	13.907	13.756	13.605	13.454	13.304	13.153	13.002	12.851

Πίνακας (18): LMS Εκτιμητής

## B. Διακυμάνσεις των $\hat{Y}_{ij}$

Οι διακυμάνσεις των μελλοντικών τιμών του μοντέλου μας είναι το άθροισμα των διαγωνίων του πίνακα Διακυμάνσεων- Συνδιακυμάνσεων συν την διακύμανση του μοντέλου  $\sigma^2$ , δηλαδή:

$$Var(\hat{Y}_{ij}) = \sigma^2 [I + X_f(X'X)^{-1}X_f']$$

### B1. Διακυμάνσεις των $\hat{Y}_{ij}$ (0 outliers)

i	j	Γραμμική Παλινδρόμηση	M-Εκτιμητής	LAD	Robust Anova	LTS	LMS
2	10	0.2521	0.2566359	0.3260	0.2144	0.15482	0.185254
3	9	0.1942	0.19765657	0.2511	0.1651	0.15124	0.1805
3	10	0.2537	0.2582934	0.3281	0.2158	0.12565	0.183799
4	8	0.1760	0.17912247	0.2276	0.1497	0.148199	0.177104
4	9	0.1965	0.20007799	0.2542	0.1672	0.124467	0.179724
4	10	0.2558	0.26037938	0.3308	0.2176	0.130119	0.182771
5	7	0.1686	0.17164594	0.2181	0.1434	0.14583	0.175067
5	8	0.1791	0.18233794	0.2317	0.1524	0.124169	0.177007
5	9	0.1996	0.20316765	0.2581	0.1698	0.129177	0.179375
5	10	0.2585	0.2631336	0.3343	0.2199	0.135305	0.182169
6	6	0.1665	0.16954607	0.2154	0.1417	0.144385	0.174387
6	7	0.1730	0.17607378	0.2237	0.1471	0.125135	0.175649
6	8	0.1834	0.18668527	0.2372	0.1560	0.129463	0.177337
6	9	0.2037	0.20737411	0.2635	0.1733	0.134912	0.179452
6	10	0.2622	0.26696435	0.3392	0.2231	0.141481	0.181994
7	5	0.1686	0.17164594	0.2181	0.1434	0.14438	0.175067
7	6	0.1730	0.17607378	0.2237	0.1471	0.128135	0.175649
7	7	0.1793	0.18253106	0.2319	0.1525	0.131725	0.176658
7	8	0.1896	0.19304191	0.2452	0.1613	0.136435	0.178094
7	9	0.2098	0.21355464	0.2713	0.1784	0.142265	0.179957
7	10	0.2679	0.27267526	0.3464	0.2278	0.149216	0.182247
8	4	0.1760	0.17912247	0.2276	0.1497	0.147085	0.177104
8	5	0.1791	0.18233794	0.2317	0.1524	0.135033	0.177007
8	6	0.1834	0.18668527	0.2372	0.1560	0.137765	0.177337

8	7	0.1896	0.19304191	0.2452	0.1613	0.141619	0.178094
8	8	0.1998	0.20340899	0.2584	0.1700	0.146592	0.179278
8	9	0.2197	0.22367014	0.2842	0.1869	0.152686	0.180888
8	10	0.2771	0.28211988	0.3584	0.2357	0.1599	0.182926
9	3	0.1942	0.19765657	0.2511	0.1651	0.156635	0.1805
9	4	0.1965	0.20007799	0.2542	0.1672	0.151746	0.179724
9	5	0.1996	0.20316765	0.2581	0.1698	0.153325	0.179375
9	6	0.2037	0.20737411	0.2635	0.1733	0.156024	0.179452
9	7	0.2098	0.21355464	0.2713	0.1784	0.159843	0.179957
9	8	0.2197	0.22367014	0.2842	0.1869	0.164782	0.180888
9	9	0.2392	0.24349103	0.3093	0.2034	0.170842	0.182247
10	10	0.2955	0.30076672	0.3821	0.2513	0.178022	0.184032
10	2	0.2521	0.2566359	0.3260	0.2144	0.197627	0.185254
10	3	0.2537	0.2582934	0.3281	0.2158	0.211784	0.183799
10	4	0.2558	0.26037938	0.3308	0.2176	0.21102	0.182771
10	5	0.2585	0.2631336	0.3343	0.2199	0.211376	0.182169
10	6	0.2622	0.26696435	0.3392	0.2231	0.212853	0.181994
10	7	0.2679	0.27267526	0.3464	0.2278	0.215449	0.182247
10	8	0.2771	0.28211988	0.3584	0.2357	0.219167	0.182926
10	9	0.2955	0.30076672	0.3821	0.2513	0.224004	0.184032
10	10	0.3486	0.35491162	0.4509	0.2965	0.229962	0.185565

## B2. Διακυμάνσεις των $\hat{Y}_{ij}$ (1 outlier)

I	j	Γραμμική Παλινδρόμηση	M-Εκτιμητής	LAD	Robust Anova	LTS	LMS
2	10	0.6531	0.758268999	0.849489	0.214429	0.146645	0.172432
3	9	0.5030	0.584005782	0.654262	0.16515	0.143254	0.168007
3	10	0.6573	0.763166319	0.854976	0.215814	0.119015	0.171077
4	8	0.4558	0.529244024	0.592912	0.149664	0.140373	0.164846
4	9	0.5092	0.591160228	0.662277	0.167173	0.117895	0.167284
4	10	0.6626	0.769329661	0.861881	0.217557	0.123249	0.17012
5	7	0.4368	0.507153495	0.568164	0.143417	0.13813	0.162949
5	8	0.4640	0.53874459	0.603556	0.15235	0.117613	0.164756
5	9	0.5170	0.600289087	0.672504	0.169754	0.122356	0.166959
5	10	0.6696	0.777467416	0.870997	0.219858	0.12816	0.16956
6	6	0.4315	0.500949116	0.561214	0.141662	0.136761	0.162317
6	7	0.4481	0.520236213	0.582821	0.147116	0.118527	0.163491
6	8	0.4751	0.551589442	0.617946	0.155983	0.122627	0.165063
6	9	0.5277	0.612717699	0.686428	0.173269	0.127788	0.167031
6	10	0.6794	0.788785938	0.883677	0.223059	0.13401	0.169398
7	5	0.4368	0.507153495	0.568164	0.143417	0.136756	0.162949
7	6	0.4481	0.520236213	0.582821	0.147116	0.121369	0.163491
7	7	0.4645	0.53931519	0.604195	0.152512	0.124769	0.164431



7	8	0.4913	0.57037106	0.638987	0.161294	0.129231	0.165767
7	9	0.5435	0.630978982	0.706886	0.178433	0.134753	0.167501
7	10	0.6939	0.805659662	0.902581	0.22783	0.141337	0.169632
8	4	0.4558	0.529244024	0.592912	0.149664	0.139318	0.164846
8	5	0.4640	0.53874459	0.603556	0.15235	0.127902	0.164756
8	6	0.4751	0.551589442	0.617946	0.155983	0.130491	0.165063
8	7	0.4913	0.57037106	0.638987	0.161294	0.13414	0.165767
8	8	0.5177	0.601002162	0.673303	0.169956	0.138851	0.166869
8	9	0.5692	0.660866738	0.74037	0.186885	0.144623	0.168368
8	10	0.7180	0.83356521	0.933844	0.235722	0.151456	0.170265
9	3	0.5030	0.584005782	0.654262	0.16515	0.148364	0.168007
9	4	0.5092	0.591160228	0.662277	0.167173	0.143734	0.167284
9	5	0.5170	0.600289087	0.672504	0.169754	0.145229	0.166959
9	6	0.5277	0.612717699	0.686428	0.173269	0.147785	0.167031
9	7	0.5435	0.630978982	0.706886	0.178433	0.151402	0.167501
9	8	0.5692	0.660866738	0.74037	0.186885	0.156081	0.168368
9	9	0.6197	0.719430512	0.805979	0.203446	0.161821	0.169632
10	10	0.7654	0.888660051	0.995567	0.251302	0.168622	0.171294
10	2	0.6531	0.758268999	0.849489	0.214429	0.187191	0.172432
10	3	0.6573	0.763166319	0.854976	0.215814	0.200601	0.171077
10	4	0.6626	0.769329661	0.861881	0.217557	0.199877	0.17012
10	5	0.6696	0.777467416	0.870997	0.219858	0.200214	0.16956
10	6	0.6794	0.788785938	0.883677	0.223059	0.201613	0.169398
10	7	0.6939	0.805659662	0.902581	0.22783	0.204073	0.169632
10	8	0.7180	0.83356521	0.933844	0.235722	0.207594	0.170265
10	9	0.7654	0.888660051	0.995567	0.251302	0.212176	0.171294
10	10	0.9032	1.048639228	1.174791	0.296542	0.217819	0.172721

### Β3. Διακυμάνσεις των $\hat{Y}_{ij}$ (2 outliers)

i	j	Γραμμική Παλινδρόμηση	M-Εκτιμητής	LAD	Robust Anova	LTS	LMS
2	10	0.941	1.090	1.222	0.232	0.154721	0.194558
3	9	0.724	0.839	0.941	0.179	0.151144	0.189565
3	10	0.947	1.097	1.230	0.234	0.12557	0.19303
4	8	0.657	0.760	0.853	0.162	0.148104	0.185998
4	9	0.733	0.849	0.953	0.181	0.124388	0.18875
4	10	0.954	1.105	1.240	0.235	0.130036	0.191949
5	7	0.629	0.729	0.817	0.155	0.145737	0.183859
5	8	0.668	0.774	0.868	0.165	0.12409	0.185897
5	9	0.745	0.863	0.968	0.184	0.129095	0.188383
5	10	0.964	1.117	1.253	0.238	0.135218	0.191318
6	6	0.621	0.720	0.807	0.153	0.144293	0.183145
6	7	0.645	0.748	0.839	0.159	0.125055	0.18447



6	8	0.684	0.793	0.889	0.169	0.129381	0.186243
6	9	0.760	0.880	0.988	0.188	0.134826	0.188464
6	10	0.979	1.133	1.271	0.241	0.141391	0.191134
7	5	0.629	0.729	0.817	0.155	0.144288	0.183859
7	6	0.645	0.748	0.839	0.159	0.128053	0.18447
7	7	0.669	0.775	0.869	0.165	0.131641	0.18553
7	8	0.708	0.820	0.919	0.175	0.136348	0.187038
7	9	0.783	0.907	1.017	0.193	0.142175	0.188994
7	10	0.999	1.158	1.299	0.247	0.149121	0.191399
8	4	0.657	0.760	0.853	0.162	0.146992	0.185998
8	5	0.668	0.774	0.868	0.165	0.134947	0.185897
8	6	0.684	0.793	0.889	0.169	0.137678	0.186243
8	7	0.708	0.820	0.919	0.175	0.141528	0.187038
8	8	0.746	0.864	0.969	0.184	0.146499	0.188281
8	9	0.820	0.950	1.065	0.202	0.152588	0.189973
8	10	1.034	1.198	1.344	0.255	0.159798	0.192113
9	3	0.724	0.839	0.941	0.179	0.156536	0.189565
9	4	0.733	0.849	0.953	0.181	0.151165	0.18875
9	5	0.745	0.863	0.968	0.184	0.153227	0.188383
9	6	0.760	0.880	0.988	0.188	0.155924	0.188464
9	7	0.783	0.907	1.017	0.193	0.159741	0.188994
9	8	0.820	0.950	1.065	0.202	0.164677	0.189973
9	9	0.892	1.034	1.160	0.220	0.170733	0.191399
10	10	1.102	1.277	1.432	0.272	0.177908	0.193274
10	2	0.941	1.090	1.222	0.232	0.197501	0.194558
10	3	0.947	1.097	1.230	0.234	0.211649	0.19303
10	4	0.954	1.105	1.240	0.235	0.210885	0.191949
10	5	0.964	1.117	1.253	0.238	0.211241	0.191318
10	6	0.979	1.133	1.271	0.241	0.212717	0.191134
10	7	0.999	1.158	1.299	0.247	0.215312	0.191399
10	8	1.034	1.198	1.344	0.255	0.219027	0.192113
10	9	1.102	1.277	1.432	0.272	0.223861	0.193274
10	10	1.301	1.507	1.690	0.321	0.229815	0.194884