ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΕΙΡΑΙΩΣ

ΤΜΗΜΑ ΣΤΑΤΙΣΤΙΚΗΣ ΚΑΙ ΑΣΦΑΛΙΣΤΙΚΗΣ ΕΠΙΣΤΗΜΗΣ

Π.Μ.Σ. ΑΝΑΛΟΓΙΣΤΙΚΗΣ ΕΠΙΣΤΗΜΗΣ ΚΑΙ ΔΙΟΙΚΗΤΙΚΗΣ ΚΙΝΔΥΝΟΥ

ΑΝΘΕΚΤΙΚΕΣ ΜΕΘΟΔΟΙ ΑΠΟΘΕΜΑΤΟΠΟΙΗΣΗΣ

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

ΛΥΜΠΕΡΗΣ ΓΕΩΡΓΙΟΣ

ΕΠΙΒΛΕΠΩΝ ΚΑΘΗΓΗΤΗΣ: ΠΙΤΣΕΛΗΣ ΓΕΩΡΓΙΟΣ

ΠΕΙΡΑΙΑΣ 2019

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

Τα μέλη της Επιτροπής ήταν:

- Πιστέλης Γεώργιος
- Τζαβελάς Γεώργιος
- Ψαρράκος Γεώργιος

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

UNIVERSITY OF PIRAEUS

DEPARTMENT OF STATISTICS AND INSURANCE SCIENCE

MSc. IN ACTUARIAL SCIENCE AND RISK MANAGEMENT

ROBUST LOSS RESERVING

MASTER THESIS

LYMPERIS GEORGIOS

ADVISOR PROFESSOR: PITSELIS GEORGIOS

PIRAEUS 2019

ΠΕΡΙΛΗΨΗ

Στόχος της παρούσας διπλωματικής είναι η εκτίμηση του αποθέματος με μεθόδους που δεν επηρεάζονται από ακραίες τιμές (outliers) ή μεγάλα άλματα (large jumps), μεταξύ των ετών εξέλιξης των ζημιών. Η μελέτη επικεντρώνεται στην μέθοδο αποθεματοποίησης Chain Ladder, που είναι μια μέθοδος «ευαίσθητη» σε ακραίες τιμές. Παρουσιάζονται τρόποι και μεθοδολογίες ώστε η Chain Ladder να γίνει πιο ανθεκτική (Robust Chain Ladder) και τα αποτελέσματά της να μην επηρεάζονται από την εμφάνιση ακραίων τιμών. Παράλληλα, προτείνονται τρόποι ανίχνευσης των έκτοπων τιμών, σε ένα τρίγωνο εξέλιξης ζημιών αλλά και η επεξεργασία (ή αφαίρεση) των ακραίων αυτών τιμών, χωρίς να επηρεάζεται ο γενικότερος υπολογισμός του αποθέματος ζημιών. Τέλος, μέσω δύο ακόμη μεθόδων της Mack Method και της Bootstrap Method, χρησιμοποιούνται για την εκτίμηση του τυπικού σφάλματος, για το αντίστοιχο αποτέλεσμα της Chain Ladder. Λόγω της «ευαισθησίας» των μεθόδων αυτών σε ακραίες τιμές, παρουσιάζονται παρεμβάσεις ώστε οι μέθοδοι Mack και Bootstrap να γίνουν πιο ανθεκτικές (Robust Mack Method και Robust Bootstrap Method), για να μην επηρεάζονται από την εμφάνιση έκτοπων τιμών (outliers) ή μεγάλων αλμάτων (large jumps). Με τον τρόπο αυτό, γίνεται πιο εμφανές αν το απόθεμα που υπολογίζεται είναι ακριβές ή αποκλίνει σημαντικά από την πραγματική του τιμή. Για να γίνουν πιο κατανοητές οι μέθοδοι και οι εφαρμογές τους, δίνονται αναλυτικά τόσο η θεωρία τους, όσο και ένα αναλυτικό παράδειγμα για την κάθε μία.

ABSTRACT

The purpose of this thesis is the robust estimation of reserve, in run off triangles with outliers or large jumps. The main reserving method is Chain Ladder, which is sensitive to outliers or large claims. In this thesis, we propose some changes in the way Chain Ladder is applied, in order to give a robust estimation of reserve and not be affected by outliers (Robust Chain Ladder). However, a method which detects outliers in a run off triangle, is very important. Therefore, is present not only a way to detect outliers but also an idea to deal with those outliers or large jumps. Last but not least, it is essential for the standard error of the method to be measured, in order to be sure that the reserve estimated was the real one. Two well-known methods, Mack method and Bootstrap Method, are presented. However, these methods are also sensitive to outliers or large jumps, therefore a robust transform of these methods is also presented. Finally, for each method the theory and an analytical example are given, so that the methods be more understood.

ΠΕΡΙΕΧΟΜΕΝΑ

Εισαγωγή

Για μια ασφαλιστική εταιρία, η πρόβλεψη των μελλοντικών (αναμενόμενων) ζημιών καθώς και η διατήρηση του κατάλληλου αποθέματος για αυτές, είναι ένα από τα σημαντικότερα κομμάτια στην λειτουργία της. Πράγματι, τα κέρδη μιας ασφαλιστικής εξαρτώνται όχι μόνο από τις απαιτήσεις που έχει ήδη πληρώσει αλλά και από της μελλοντικές απαιτήσεις που θα κληθεί να καταβάλει. Τα τελευταία χρόνια, έχει δημιουργηθεί μια νέα νομοθεσία για τις ασφαλιστικές εταιρίες (Solvency II – Φερεγγυότητα II) η οποία αποτελεί το νέο κανονιστικό πλαίσιο που οι ασφαλιστικές επιχειρήσεις πρέπει να ακολουθούν. Η νέα νομοθεσία ρυθμίζει την χρηματοοικονομική λειτουργία των ασφαλιστικών και αντασφαλιστικών εταιριών, οι οποίες δραστηριοποιούνται στην Ε.Ε. (ΕΑΕΕ, 2016) . Στόχος της νέας νομοθεσίας είναι η προστασία των ασφαλισμένων, καθώς θέτει ένα ευρύτατο πλέγμα ποιοτικών και ποσοτικών απαιτήσεων, με κύριο μέλημα την προστασία των ασφαλισμένων. Επίσης, ενισχύεται η κεφαλαιακή επάρκεια των ασφαλιστικών επιχειρήσεων και οι τελευταίες θωρακίζονται έναντι κάθε είδους κινδύνων, στους οποίους εκτίθενται πέραν του ασφαλιστικού στων οποίων ήταν προσηλωμένο το προηγούμενο εποπτικό πλαίσιο (Solvency I – Φερεγγυότητα I). Τα προηγούμενα σε συνδυασμό με τα νέα Λογιστικά Πρότυπα (IFRS 17) έρχονται να αυστηροποιήσουν το πλαίσιο λειτουργίας των ασφαλιστικών εταιριών, με σκοπό την προστασία των ασφαλισμένων και την εύρυθμή λειτουργία του ασφαλιστικού κλάδου.

Εντός των νέων κανονιστικών αυτών πλαισίων κάθε ασφαλιστική επιχείρηση οφείλει να διακρατά ένα απόθεμα ώστε να μπορεί να ανταπεξέλθει στις υποχρεώσεις που έχει αναλάβει. Η διαδικασία που πρέπει να ακολουθείται από κάθε ασφαλιστική αλλά και οι μέθοδοι, προκειμένου να υπολογίζεται το απόθεμα, ορίζονται από το Solvency II. Ωστόσο, κατά τη διάρκεια της συλλογής των δεδομένων, που χρειάζονται για την αποθεματοποίηση, υπάρχουν κίνδυνοι που πρέπει να αντιμετωπίσει η ασφαλιστική εταιρία. Τυχόν λάθος ή ελλιπή δεδομένα μπορούν να αυξήσουν τον λεγόμενο Κίνδυνο Αποθέματος (Reserve Risk) και έτσι να οδηγήσουν σε μια λάθος εκτίμηση. Αυτό μπορεί να προκαλέσει τόσο ένα κενό στον ισολογισμό μιας ασφαλιστικής όσο και την ανάκληση της άδειας λειτουργίας της (σε πιο ακραίες συνθήκες). Κατά συνέπεια του προηγούμενου, προκύπτει η ανάγκη αξιόπιστων μεθόδων για την εκτίμηση των αποθεμάτων της ασφαλιστικής. Έτσι, είναι ζωτικής σημασίας μία αξιόπιστη μέθοδος πρόβλεψης των μελλοντικών απαιτήσεων (ή αποζημιώσεων), ώστε η ασφαλιστική εταιρία να διασφαλίσει την οικονομική σταθερότητα.

Ένας από τους κινδύνους που αντιμετωπίζει η ασφαλιστική κατά τη συλλογή δεδομένων είναι το γεγονός ότι τα δεδομένα μπορεί να είναι *ελλιπή,* δηλαδή να έχουν πληρωθεί ζημιές οι οποίες δεν φαίνονται στα δεδομένα που ο αναλογιστής θα χρησιμοποιήσει. Ένας άλλος είναι οι έκτοπες παρατηρήσεις. Έκτοπες χαρακτηρίζονται οι ζημιές που αποκλίνουν πολύ από ό,τι συνήθως πληρώνει η ασφαλιστική βάσει των κινδύνων που έχει αναλάβει ή βάσει της ιστορικότητας των πληρωμών. Και οι δύο κίνδυνοι είναι σημαντικοί καθώς αν δεν υπάρχουν σωστά δεδομένα μπορεί η ασφαλιστική να υπολογίσει απόθεμα κατώτερο από αυτό που χρειάζεται ενώ στη δεύτερη περίπτωση κινδύνου να υπολογίσει απόθεμα περισσότερο από ότι χρειάζεται και άρα να κρατά κεφάλαια που θα μπορούσε να τα χρησιμοποιήσει αλλιώς. Ακόμη, ένας άλλο κίνδυνος έχει να κάνει με τα μεγάλα άλματα. Αυτά προκύπτουν όταν η ασφαλιστική εμφανίζει μεγάλη διακύμανση στα ποσά που πληρώνει, πχ μπορεί να πληρώνει εύλογα ποσά σε ζημιές για κάποια χρόνια, στη συνέχεια να πληρώνει πολύ μικρά ποσά (πχ. Δικαστικά έξοδα, έξοδα εμπειρογνωμόνων κ.α.) για 2 χρόνια και στη συνέχεια να επανέρχεται σε εύλογα ποσά, τα οποία όμως διαφέρουν αισθητά από τα 2 χρόνια που πλήρωνε μικροποσά. Η διαφορά των μεγάλων αλμάτων σε σχέση με τις έκτοπες παρατηρήσεις έγκειται στο ότι οι έκτοπες παρατηρήσεις είναι συνήθως μεγάλα ποσά, πολύ ανώτερα των όσων έχει συνηθίσει να πληρώνει η ασφαλιστική.

Από τα παραπάνω, είναι απαραίτητη η χρήση μιας μεθόδου αποθεματοποίησης, η οποία θα υπολογίζει το απόθεμα χωρίς να λαμβάνει υπόψιν της μεγάλα άλματα, ελλιπή δεδομένα ή έκτοπες παρατηρήσεις. Δυστυχώς, σχεδόν όλες οι μέθοδοι αποθεματοποίησης επηρεάζονται από τα φαινόμενα που αναφέρθηκαν παραπάνω. Για το λόγο αυτό, οι μέθοδοι μετασχηματίστηκαν ώστε να μπορούν αν αντιμετωπίσουν τέτοιες περιπτώσεις.

Η παρούσα διπλωματική εκτείνεται σε δύο σκέλη. Πρώτα, παρουσιάζει και αναλύει δύο γνωστές μεθόδους αποθεματοποίησης, την Chain Ladder και την Mack Method. Όπως θα φανεί και στη συνέχεια, η Mack Method χρησιμοποιεί την Chain Ladder και εύλογα αναρωτιέται κανείς, πώς χρησιμοποιείται. Η Mack Method αποσκοπεί στην εύρεση του μέσου τετραγωνικού σφάλματος. Το τελευταίο, μας δίνει μία εκτίμηση για το πόσο πιο έγκυρη είναι η εκτίμηση του αποθέματος ενώ σε περίπτωση που γίνει σύγκριση με άλλες μεθόδους αποθεματοποίησης – οι οποίες δίνουν διαφορετικά αποτελέσματα- μπορεί να δείξει αν η διαφορά αυτή είναι σημαντική ή όχι (Mack 1993). Σε σύγκριση της μεθόδου Mack έρχεται η μέθοδος Bootstrap. Η τελευταία δίνει τόσο ένα απόθεμα αλλά και μία εκτίμηση σφάλματος, όπως η Mack, τα οποία και συγκρίνονται. Επιπλέον, η μέθοδος Chain Ladder παρουσιάζεται και με τη χρήση της Ανάλυσης Παλινδρόμησης. Τέλος παρουσιάζεται μια ανθεκτική μορφή της Chain Ladder με τη χρήση της Ανάλυσης Παλινδρόμησης αλλά και με κατάλληλες μετατροπές στην μεθοδολογία της Chain Ladder. Θα παρουσιαστεί το απόθεμα, επεμβαίνοντας στα δεδομένα και δημιουργώντας έκτοπες τιμές, ώστε να δούμε πόσο αποκλίνει το απόθεμα με τις έκτοπες τιμές, από το απόθεμα που προκύπτει χωρίς να υπάρχουν έκτοπες τιμές.

1 Αποθεματοποίηση

1.1 Ορισμοί και έννοιες

Πριν προχωρήσουμε στην παρουσίαση των μεθόδων αποθεματοποίησης, καλό είναι να δώσουμε κάποιους ορισμούς, που θα βοηθήσουν - τον αναγνώστη – να καταλάβει καλύτερα τις διαδικασίες αποθεματοποίησης (Pavlovic, 2012).

Πληρωθείσες Ζημιές (Paid Losses)

Είναι οι απαιτήσεις, εκ μέρους των ασφαλισμένων, που η ασφαλιστική πλήρωσε κατά την χρήση ενός οικονομικού έτους.

Εκκρεμείς Ζημιές (Outstanding Losses)

Είναι οι απαιτήσεις, που η ασφαλιστική έχει κληθεί να πληρώσει αλλά δεν τις έχει διευθετήσει ακόμα, τη στιγμή που κλείνει η χρήση του οικονομικού έτους.

Επισυμβείς Ζημιές (Incurred Losses)

Πρόκειται για τις ζημιές που έχουν πληρωθεί (σε μία οικονομική περίοδο), στις οποίες προστίθεται και η μεταβολή του αποθέματος του προηγούμενου έτους. Δεν περιέχουν, πάντα, Ζημιές που έγιναν αλλά δεν έχουν ακόμα αναφερθεί στην ασφαλιστική (IBNR- Incurred But Not Reported)

Όπως γίνεται αντιληπτό, οι επισυμβείς ζημιές είναι το άθροισμα των ζημιών που έχουν ήδη πληρωθεί και των εκκρεμών ζημιών, δηλαδή:

Επισυμβείς Ζημιές = Πληρωθείσες Ζημιές + Εκκρεμείς Ζημιές

IBNR (Incurred But Not Reported)

Είναι οι ζημιές που έχουν γίνει αλλά ακόμα δεν έχουν αναφερθεί στην ασφαλιστική.

Συνολικές Ζημιές (Ultimate Losses)

Πρόκειται για το άθροισμα των ποσών που μία ασφαλιστική πληρώνει για μία ζημιά (ή για όλες τις ζημιές όταν μιλάμε σε εύρος χαρτοφυλακίου). Το άθροισμα αυτό προκύπτει ως εξής:

Συνολικές Ζημιές = Πληρωθείσες Ζημιές + Εκκρεμείς Ζημιές + IBNR

Δεν είναι πάντα εύκολο να υπολογιστούν οι συνολικές ζημιές, με ακρίβεια. Συνήθως, ο υπολογισμός γίνεται με χρονικό ορίζοντα τη διάρκεια του συμβολαίου.

Τεχνικό Απόθεμα (Technical Reserve)

Πρόκειται για το ποσό που προκύπτει, όταν από την παρούσα αξία μελλοντικών ζημιών αφαιρέσουμε την παρούσα αξία μελλοντικών ασφαλίστρων. Το ποσό κρατείται από την ασφαλιστική για εκκρεμείς και μελλοντικές απαιτήσεις ή καταστροφικούς κινδύνους.

Είναι τα κεφάλαια που πρέπει η ασφαλιστική εταιρία να διαθέτει, ώστε να πληρώσει τις ζημιές που θα προκύψουν από κάποιο ασφαλιστικό γεγονός (ατύχημα, περίθαλψη, σύνταξη κλπ.), όταν αυτό συμβεί και εφόσον η ασφαλιστική έχει αναλάβει την υποχρέωση να πληρώσει. Το ύψος του αποθέματος βασίζεται σε μία πρόβλεψη που γίνεται από την ασφαλιστική, με τη χρήση στατιστικών μεθόδων. Η ασφαλιστική υπολογίζει τις ζημιές, στις οποίες πιθανόν να εκτεθεί (βάσει των κινδύνων που έχει αναλάβει), για ένα χρονικό διάστημα. Εξαιτίας της τυχαιότητας που υπάρχει μέσα σε κάθε πρόβλεψη, το απόθεμα μπορεί να είναι αλλά μπορεί και να μην είναι αρκετό, για τους κινδύνους που η ασφαλιστική έχει αναλάβει. Το τεχνικό απόθεμα αναγράφεται στο Παθητικό του ισολογισμού.

Με βάση τη νομοθεσία το Τεχνικό απόθεμα επιμερίζεται σε 4 επιμέρους αποθέματα:

- Απόθεμα μη Δεδουλευμένων Ασφαλίστρων.
- Απόθεμα Κινδύνων εν Ισχύ.
- Απόθεμα Εκκρεμών Ζημιών.
- Απόθεμα εξόδων Διακανονισμού.

Απόθεμα Μη Δεδουλευμένων Ασφαλίστρων - ΑΜΔΑ (Unearned Premium Reserve)

Είναι το απόθεμα που προορίζεται να καλύψει απαιτήσεις, οι οποίες μπορεί να προκύψουν στο μέλλον από ασφαλίσεις που δεν έχουν ακόμα λήξει. Περιλαμβάνει το ποσό των Ασφαλίστρων που πρέπει να καταλογιστούν στο χρονικό διάστημα από την ημερομηνία υπολογισμού μέχρι την ημερομηνία λήξης της περιόδου, για την οποία έχουν καταχωρηθεί τα ασφάλιστρα.

Απόθεμα Κινδύνων Εν Ισχύ – ΑΚΕΙ (Unexpired Risk Reserve)

Το απόθεμα Κινδύνων εν Ισχύ αναφέρεται σε ασφαλιστικές αποζημιώσεις (και Διοικητικά έξοδα) που πιθανόν να προκύψουν μετά την ημερομηνία υπολογισμού του Τεχνικού Αποθέματος και να επιβαρύνουν τις υφιστάμενες ασφαλιστικές συμβάσεις. Ουσιαστικά, πρόκειται για μία επιπλέον πρόβλεψη του ΑΜΔΑ, σε περίπτωση που το ΑΜΔΑ δεν αρκεί για να καλύψει όλες τις ζημιές που βρίσκονται εν ισχύ, κατά το κλείσιμο του προϋπολογισμού.

Απόθεμα Εκκρεμών Ζημιών – ΑΕΖ (Outstanding Claims Reserve)

Είναι απόθεμα, που η ασφαλιστική τοποθετεί κατά μέρος ώστε να αντιμετωπίσει ζημιές, οι οποίες κατά το κλείσιμο της οικονομικής χρήσεως παραμένουν σε εκκρεμότητα, περιμένοντας τον οριστικό διακανονισμό τους. Στο απόθεμα αυτό υπολογίζονται και τα άμεσα και έμμεσα έξοδα διακανονισμού των ζημιών αυτών, που αναμένεται να προκύψουν.

Επιπρόσθετα, για τις ζημιές που έχουν συμβεί αλλά δεν έχουν αναφερθεί στην ασφαλιστική εταιρία (IBNR), πρέπει να κρατηθεί ένα απόθεμα. Η πρόβλεψη για το απόθεμα αυτό περιλαμβάνει αξιώσεις που θα προκύψουν από ασφαλιστικούς κινδύνους που έχουν επέλθει αλλά τη στιγμή που γινόταν ο υπολογισμός του συνολικού αποθέματος, δεν είχαν γίνει γνωστοί στην ασφαλιστική. Μέσα στο απόθεμα για της IBNR ζημιές περιλαμβάνονται Άμεσα και Έμμεσα έξοδα Διακανονισμού που μπορεί να προκύψουν από αυτές τις ζημιές. Το απόθεμα για τις IBNR ζημιές προστίθεται στο Απόθεμα των Εκκρεμών Ζημιών.

Επομένως, το συνολικό απόθεμα, που οφείλει να διαθέτει μια ασφαλιστική προκύπτει ως:

Συνολικό Απόθεμα = ΑΕΖ + IBNR

Διαχωρισμός Αποθέματος Ζημιών

Το απόθεμα ζημιών χωρίζεται σε 5 επιμέρους πεδία (Πιτσέλης, 2018).

1. Απόθεμα για συγκεκριμένες απαιτήσεις (ζημιές).
2. Εκτίμηση μελλοντικής εξέλιξης απαιτήσεων που είναι γνωστές στην ασφαλιστική.
3. Εκτίμηση για ζημιές των οποίων ο φάκελος έχει κλείσει αλλά είναι πιθανόν να ανοίξει πάλι λόγων νεών δεδομένων.
4. Πρόβλεψη για απαιτήσεις που έχουν συμβεί ως γεγονός αλλά μέχρι την ημερομηνία όπου γίνεται η αποθεματοποίηση, η ασφαλιστική δεν είναι ενήμερη. Είναι οι γνωστές IBNR ζημιές.
5. Πρόβλεψη για τις απαιτήσεις που έχουν αναγγελθεί -άρα η ασφαλιστική είναι ενήμερη - αλλά δεν έχει ακόμα υπολογίσει το ύψος της απαίτησης.

Συγκεντρώνοντας τα παραπάνω 5 πεδία σε δύο υποκατηγορίες, μπορούμε να πούμε ότι, το απόθεμα ζημιών επιμερίζεται σε 2 μεγάλες κατηγορίες:

A. Το απόθεμα γνωστών απαιτήσεων. Το εν λόγω αφορά το ύψος των ζημιών που έχουν αναγγελθεί στην ασφαλιστική.
B. Το απόθεμα άγνωστων απαιτήσεων. Πρόκειται για το ύψος ζημιών που προορίζεται για ασφαλιστικά γεγονότα που είτε έχουν γίνει αλλά δεν έχουν αναγγελθεί στην εταιρεία είτε για ασφαλιστικά γεγονότα που έχουν αναγγελθεί αλλά ακόμα η ασφαλιστική δεν τα έχει υπολογίσει.

Πριν προχωρήσουμε σε πιο τεχνικά θέματα, όσον αφορά την εκτίμηση αποθέματος, καλό είναι να αναφέρουμε μία ευρύτερη στρατηγική που οφείλει να ακολουθήσει ένας αναλογιστής. Η στρατηγική αυτή χωρίζεται σε 4 στάδια:

i. Ο αναλογιστής οφείλει να κάνει μία διασταύρωση των στοιχείων που έχει στη διάθεσή του, με έγκυρες πηγές. Τέτοιες είναι το σύστημα καταγραφής της ασφαλιστικής και η συχνή επικοινωνία του με τον αντίστοιχο κλάδο. Έτσι θα μπορέσει να μειώσει τον κίνδυνο που ελλοχεύει από μεγάλες αποκλίσεις στις ζημιές.
ii. Στη συνέχεια, ο αναλογιστής πρέπει αν εφαρμόσει την κατάλληλη τεχνική εκτίμησης του αποθέματος.
iii. Επιπλέον, είναι ζωτικής σημασίας να κάνει επανέλεγχο των αποτελεσμάτων, συγκρίνοντας διαφορετικέ μεθόδους αποθεματοποίησης. Σε περίπτωση μεγάλων αποκλίσεων των μεθόδων, ο αναλυτής πρέπει να είναι σε θέση να εξηγήσει τις αποκλίσεις αυτές ή να τις συμβιβάσει.
iv. Τέλος, οφείλει να παρουσιάσει και να προετοιμάσει την εξέλιξη του αποθέματος, όπως αυτή διαγράφεται σε μελλοντικά ημερολογιακά έτη.

Πρέπει να επισημάνουμε ότι ο υπολογισμός των παραπάνω αποθεμάτων, γίνεται ξεχωριστά για κάθε ζημιά (η διαδικασία είναι γνωστή ως «Μέθοδος φάκελο προς φάκελο»). Επομένως, σε κάθε ξεχωριστό κλάδο ασφάλισης υπολογίζονται τα προαναφερθέντα αποθέματα φάκελο προς φάκελο, το οποία όταν τα προσθέσουμε μας δίνουν το τεχνικό απόθεμα για τον κλάδο που μελετάμε. Το άθροισμα των τεχνικών αποθεμάτων από όλους τους κλάδους, δίνει το συνολικό τεχνικό απόθεμα που η ασφαλιστική εταιρία πρέπει να διαθέτει.

Επιπλέον, είναι ήσσονος σημασίας η ομοιογένεια του χαρτοφυλακίου. Για το λόγο αυτό, το συνολικό χαρτοφυλάκιο ζημιών, επιμερίζεται σε μικρότερα υποχαρτοφυλάκια τα οποία περιέχουν ζημιές του ίδιου κλάδου (πχ. Αστικής Ευθύνης, Πυρός, Μεταφορών κλπ.). Στη συνέχεια, εφαρμόζεται η επιλεγμένη – από την εταιρία- μέθοδος εκτίμησης αποθέματος σε κάθε υποχαρτοφυλάκιο. Τέλος, το συνολικό απόθεμα που πρέπει να διαθέτει η εταιρία, για μελλοντικές ζημιές είναι το άθροισμα των επιμέρους αποθεμάτων.

Ο διαχωρισμός του ευρύτερου χαρτοφυλακίου σε επιμέρους χαρτοφυλάκια απαιτεί και επιπλέον ενέργειες που χρήζουν προσοχής. Για παράδειγμα, σε ένα κλάδο, πρέπει οι ζημιές που εμφανίζονται να έχουν κοινά χαρακτηριστικά. Τέτοια είναι, λόγου χάρη, η ημερομηνία, ο τόπος ατυχήματος, το είδος ζημιάς, το είδος κάλυψης κ.α. Παραδείγματος χάριν, πρέπει όλες οι ζημιές να είναι καταγεγραμμένες ως προς το ίδιο επίπεδο ημερομηνίας, πχ την ημερομηνία αγγελίας ή την ημερομηνία πληρωμής. Αν κάποιες ζημιές στο χαρτοφυλάκιο έχουν καταγραφεί με ημερομηνία αναγγελίας του ατυχήματος και κάποιες άλλες με την ημερομηνία πληρωμής, το αποτέλεσμα της εκτίμησης θα είναι εσφαλμένο.

Γίνεται προφανές, ότι η διαδικασία της αποθεματοποίησης απαιτεί λεπτομέρεια σε κάθε της βήμα. Ο αναλογιστής πρέπει να έχει ένα επαρκές δείγμα, τόσο σε πλήθος όσο και σε ποσό ζημιών. Από τη μία, ένα μεγάλο δείγμα ζημιών οδηγεί σε πιο αξιόπιστα αποτελέσματα εκτίμησης Από την άλλη, ένα ακριβές δείγμα (χωρίς έκτοπες παρατηρήσεις ή ελλιπή στοιχεία) θα προσδώσει μία εκτίμηση, πολύ κοντά στο πραγματικό αποτέλεσμα.

2 Μέθοδοι Αποθεματοποίησης

2.1 Μέθοδος Αναμενόμενου Δείκτη Ζημιών

Μία από τις μεθόδους που χρησιμοποιείται για την εκτίμηση αποθέματος είναι η Μέθοδος Αναμενόμενου Δείκτη Ζημιών (Πιτσέλης 2018). Όπως γίνεται αντιληπτό, η μέθοδος αυτή βασίζεται στον τελικό (αναμενόμενο) δείκτη ζημιών που θα έχει η ασφαλιστική, σε ένα χρόνο. Ο δείκτης ζημιών (Loss Ratio) είναι ο λόγος των συνολικών αποζημιώσεων προς τα συνολικά δεδουλευμένα ασφάλιστρα (Earned Premiums).

$$\text{Δείκτης Εκτιμώμενων Ζημιών} = \frac{\text{Συνολικές Ζημιές}}{\text{Δεδουλευμένα Ασφάλιστρα}}.$$

Ο παραπάνω λόγος ισχύει για κάθε κλάδο ξεχωριστά και αναφέρεται σε μία περίοδο ασφάλισης.

Αν τώρα πολλαπλασιάσουμε το Δείκτη Εκτιμώμενων Ζημιών με τα Δεδουλευμένα Ασφάλιστρα, θα μας δώσει τις εκτιμώμενες Ζημιές, για τη χρονική περίοδο και τον κλάδο που μελετάμε.

$$(\text{Εκτιμώμενες Τελικές Ζημιές})_{ij} = (\text{Δείκτης Εκτιμώμενων Ζημιών})_{ij} * (\text{Δεδουλευμένα Ασφάλιστρα})_{ij}.$$

Όπου i δηλώνει τον κλάδο ασφάλισης και j δηλώνει την περίοδο ασφάλισης.

Έτσι το Εκτιμώμενο Απόθεμα Ζημιών για τον κλάδο ασφάλισης i και την περίοδο ασφάλισης j είναι:

$$(\text{Εκτιμώμενο Αποθέματα Ζημιών})_{ij} = (\text{Εκτιμώμενες Τελικές Ζημιές})_{ij} - (\text{Ζημιές που ήδη πληρώθηκαν ως σήμερα})_{ij}.$$

Τέλος το Συνολικό Εκτιμώμενο Απόθεμα Ζημιών για όλους τους κλάδους, σε μία περίοδο ασφάλισης είναι:

$$\text{Συνολικό Εκτιμώμενο Απόθεμα Ζημιών} = \sum_{i} (\text{Εκτιμώμενο Απόθεμα Ζημιών})_{i}.$$

Αξίζει να αναφέρουμε ότι στο Δείκτης Ζημιών διακρίνεται μία ιστορικότητα. Μπορεί κανείς να δει την αλλαγή στο ύψος των ζημιών ή την αλλαγή στην τιμολογιακή πολιτική μιας ασφαλιστικής. Λόγου χάρη, ο δείκτης μπορεί να μεγαλώσει, αν αυξηθούν οι συνολικές ζημιές. Ωστόσο μπορεί να αυξηθεί αν η Συνολικές Ζημιές μείνουν (σχεδόν) σταθερές και τα Δεδουλευμένα Ασφάλιστρα μειωθούν, δηλαδή η εταιρία αλλάξει τιμολογιακή πολιτική (ίσως λόγω ανταγωνισμού). Αυτό θα έχει επίδραση στο Εκτιμώμενο Τελικό Απόθεμα του κλάδου και κατ’ επέκταση στο συνολικό απόθεμα της εταιρίας.

Ωστόσο, υπάρχουν περιπτώσεις όπου η μέθοδος Αναμενόμενου Δείκτη Ζημιών είναι και η μοναδική που μπορεί να εμπιστευτεί μία ασφαλιστική. Για παράδειγμα, τα πρώτα έτη λειτουργίας τής, η ασφαλιστική δεν έχει αρκετή εμπειρία ζημιών για το συγκεκριμένο κλάδο, οπότε η μέθοδος αυτή να αποτελεί μονόδρομο. Τέλος, η εποπτική αρχή μπορεί να επιβάλει κατώτατο επίπεδο στο δείκτη Ζημιών ενός κλάδου, σε μία ασφαλιστική, προκειμένου η τελευταία να ικανοποιεί τα αποθέματα ζημιών.

2.2 Η Μέθοδος Chain Ladder – Μέθοδος Τριγώνου Εξέλιξης Ζημιών

Σκοπός της μεθόδου (Werner G., Modlin C., 2016) είναι να εκτιμήσει τις μελλοντικές απαιτήσεις, που μια ασφαλιστική μπορεί να κληθεί να πληρώσει. Δηλαδή ζημιές που έγιναν αλλά για τις οποίες ακόμα η ασφαλιστική δεν έχει ενημερωθεί (Incurred But Not Reported Claims, IBNR Claims). Η μέθοδος Chain Ladder βασίζεται σε ένα μοντέλο το οποίο σχετίζει τις συνολικές ζημιές κάθε έτους, με τα έτη στα οποία οι ζημιές αυτές εξελίσσονται. Θεωρούμε ότι τα δεδομένα μιας ασφαλιστικής εταιρίας εμφανίζονται στην μορφή του παρακάτω πίνακα (Πίνακας 1), δηλαδή στη μορφή ενός τριγώνου. Κάθε γραμμή του πίνακα, αντιστοιχεί σε ένα έτος ατυχήματος (δηλαδή σε ζημιές που αναγγέλθηκαν στο κάθε έτος) και κάθε στήλη του πίνακα αναφέρεται στο πώς οι ζημιές του αντίστοιχου έτους εξελίσσονται στη διάρκεια των επόμενων ετών. Συνήθως οι ζημιές παρουσιάζονται σε αθροιστικά ποσά (cumulative claims), αν και είναι εξίσου συχνό να παρουσιάζονται σε σταδιακά ποσά (incremental claims).

Οι λόγοι, που κάποιες ζημιές μπορεί να χρειάζονται πολλά χρόνια μέχρι να διευθετηθούν συνολικά είναι πολλοί. Κυρίως έχουν να κάνουν με τους χρόνους που χρειάζονται οι ασφαλιστικές ώστε να διερευνήσουν πλήρως τις υποθέσεις, με νομικές διαδικασίες που τελεσιδικούν έπειτα από μεγάλα χρονικά διαστήματα, με διαδικασίες διαπραγμάτευσης μεταξύ των ασφαλιστικών εταιριών, για το ποσό που θα καταβληθεί

Στον Πίνακα 1 φαίνεται πώς παρουσιάζονται τα δεδομένα και στη συνέχεια ακολουθεί ο τρόπος υπολογισμού των μελλοντικών απαιτήσεων.

Πίνακας 1 Τρίγωνο Εξέλιξης Ζημιών

		Έτος Εξέλιξης Ατυχήματος						
		1	**2**	**...**	**i**	**...**	**n-1**	**n**
	1	C_{11}	C_{12}	...	C_{1i}	...	$C_{1,n-1}$	C_{1n}
	2	C_{21}	C_{22}	...	C_{2i}	...	$C_{2,n-1}$	
	...	...	...	...	...	...		
Έτος Αναγγελίας Ατυχήματος	**i**	C_{i1}	C_{i2}	...	C_{ii}			
	...	...	...	...				
	n-1	$C_{n-1,1}$	$C_{n-1,2}$					
	n	C_{n1}						

Με C_{ij} θεωρούμε τις συνολικές (αθροιστικές) ζημιές που έγιναν στο έτος i και πώς εξελίσσονται μέχρι το έτος j. Προφανώς $C_{ij} < C_{i,j+1}$ για κάθε $i+j<= n+1$, και θεωρούνται γνωστές. Οι απαιτήσεις που φαίνονται στη διαγώνιο αφορούν το έτος n. Η Chain Ladder προκειμένου να εκτιμήσει τις μελλοντικές απαιτήσεις, κάνει χρήση των συντελεστών εξέλιξης.

Ως συντελεστής εξέλιξης θεωρείται ο λόγος των αθροισμάτων των αθροιστικών ζημιών, δύο διαδοχικών ετών. Οι ζημιές αυτές αναφέρονται στο έτος αναγγελίας ατυχήματος i και στα έτη εξέλιξης j, j+1. Ο συντελεστής εξέλιξης συμβολίζεται με f_k και ορίζεται ως:

$$f_k = \frac{\sum_{i=1}^{n-k} C_{i,k+1}}{\sum_{i=1}^{n-k} C_{i,k}} \quad \text{για} \quad 1 \le k \le n-1. \qquad (2.2.1)$$

Αφού έχουμε υπολογίσει τους συντελεστές εξέλιξης ζημιών, οι εκτιμώμενες μελλοντικές αθροιστικές ζημιές προκύπτουν από την ακόλουθη σχέση, ως:

$$C_{i,n} = C_{i,n+1-i}\, \widehat{f}_{n+1-i} \dots \widehat{f}_{n-1}. \qquad (2.2.2)$$

Το τρίγωνο στον Πίνακα 1 συμπληρώνεται με την παραπάνω μεθοδολογία ως εξής:

Πίνακας 2:Εύρεση αποθέματος με την μέθοδο Chain Ladder

		Έτος Εξέλιξης Ατυχήματος								
		1	2	...	i	...	n-1	n	Πληρωθείσες Ζημιές	Απόθεμα
Έτος Αναγγελίας Ατυχήματος	1	C_{11}	C_{12}	...	C_{1i}	...	$C_{1,n-1}$	C_{1n}	C_{1n}	R_1
	2	C_{21}	C_{22}	...	C_{2i}		$C_{2,n-1}$	U_{2n}	$C_{2,n-1}$	R_2
	...	...	...	...	...	...	...	...	...	...
	i	C_{i1}	C_{i2}	...	C_{ii}	...	$U_{i,n-1}$	U_{in}	C_{ii}	R_i
	...	...	...	...	...	...	...	...	...	...
	n-1	$C_{n-1,1}$	$C_{n-1,2}$	...	$U_{n-1,i}$	...	$U_{n-1,n-1}$	$U_{n-1,n}$	$C_{n-1,2}$	R_{n-1}
	n	C_{n1}	U_{n2}	...	U_{ni}	...	$U_{n,n-1}$	$U_{n,n}$	C_{n1}	R_n

Όπως βλέπουμε, οι πληρωμένες ζημιές είναι οι ζημιές που βρίσκονται στην κύρια διαγώνιο. Αυτές έχουν πληρωθεί μέχρι το έτος i. Τα αποθέματα R_i υπολογίζονται με το να αφαιρέσουμε από τις εκτιμώμενες ζημιές του έτους εξέλιξής n (και για κάθε έτος αναγγελίας ατυχήματος), τις ζημιές της κυρίας διαγώνιου. Αυτό που μένει είναι το απόθεμα για κάθε έτος. Σε μαθηματική μορφή είναι:

$$R = U_{i,n} - C_{i,k} \text{ για } i=1,...,n \text{ και } k = n, n\text{-}1,...,1. \qquad (2.2.3)$$

Προφανώς, για i=1 και k=n,

$$U_{1n} = C_{1n}$$

Η σχέση (2.2.3) μας δίνει το απόθεμα που πρέπει η ασφαλιστική να κρατήσει για κάθε έτος αναγγελίας ατυχήματος i. Επομένως, για να βρούμε το συνολικό απόθεμα, δεν έχουμε παρά να προσθέσουμε όλα τα επιμέρους αποθέματα των ετών αναγγελίας. Έτσι, για τη δεδομένη στιγμή που γίνεται η μελέτη, προκύπτει το συνολικό απόθεμα, που πρέπει να διαθέτει η ασφαλιστική.

2.3 Μέθοδος Chain Ladder με τη Χρήση Ανάλυσης Παλινδρόμησης

Η κλασσική μέθοδος Chain Ladder μπορεί να εφαρμοστεί, χρησιμοποιώντας την ανάλυση παλινδρόμησης. Η μεθοδολογία αυτή προτάθηκε πρώτη φορά από την Kremer (1982) και έκτοτε συναντάται σε πολλές δημοσιεύσεις, όπως Christofides (1990), Verall (1991) και Pitselis et al. (2015). Μάλιστα ο Verall (1991) εκτίμησε το απόθεμα και τις εκκρεμείς ζημιές εφαρμόζοντας τη Λογαριθμοκανονική Παλινδρόμηση. Βέβαια, για αυτό, στηρίχθηκε στη γενικότερη θεωρία για την εκτίμηση γραμμικών μοντέλων όπως αυτή παρουσιάστηκε από τους Bradu και Mundlak (1970). Επιπρόσθετα, οι Zehnwirth (1985) και Renshaw (1989) έδειξαν πώς από ένα τρίγωνο εξέλιξης ζημιών με λογαριθμημένες ατομικές ζημιές μπορεί – με τη χρήση πολλαπλής παλινδρόμησης - να εκτιμηθεί το απόθεμα πράγμα που χρησιμοποίησε και ο Christofides (1990).

Έχοντας παρουσιάσει την Chain Ladder στην προηγούμενη ενότητα, θα παρουσιάσουμε τον τρόπο με τον οποίων η ανάλυση παλινδρόμησης χρησιμοποιείται στα τρίγωνα εξέλιξης ζημιών. Επιπλέον, στα πλαίσια της ανθεκτικής εκτίμησης του αποθέματος, θα αναφέρουμε κάποιους ανθεκτικούς εκτιμητές και πώς αυτοί χρησιμοποιούνται ώστε να δώσουν πιο αξιόπιστα αποτελέσματα.

Σύμφωνα με τον Christofides (1990) οι βασικές υποθέσεις της Chain Ladder μπορούν να περιγραφούν ξανά ως εξής:

a. Κάθε έτος ατυχήματος έχει το δικό του μοναδικό επίπεδο (has a unique level), δηλαδή έχει ένα συγκεκριμένο πλήθος ζημιών που κατανέμεται στα έτη εξέλιξης και δεν επηρεάζεται από ζημιές σε άλλα έτη.
b. Οι συντελεστές εξέλιξης ζημίων για κάθε περίοδο είναι ανεξάρτητοι από τα έτη ατυχήματος.

Υποθέτουμε, λοιπόν, ότι:

A_i : είναι οι αθροιστικές ζημιές για κάθε έτος ατυχήματος i.

B_j : είναι το ποσοστό των πληρωμένων ζημιών κατά τη διάρκεια του έτους εξέλιξης j.

P_{ij}: είναι οι ατομικές ζημιές για το έτος ατυχήματος i κατά τη διάρκεια του έτους εξέλιξης j.

Επομένως, το μοντέλο της Chain Ladder μπορεί να περιγραφεί από τις παρακάτω σχέσεις:

$$P_{ij} = A_i \times B_j, \quad \gamma\iota\alpha \;\; 1 \le i, j \le n, \qquad (2.3.1)$$

με τη συνθήκη ότι

$$\sum B_j = 1, \quad \gamma\iota\alpha \;\; 1 \le j \le n. \qquad (2.3.2)$$

Καθώς η σχέση (2.3.1) είναι σε μορφή γινομένου, η συνήθης τακτική για να μπορέσουμε να προχωρήσουμε στην επίλυση της είναι να τη φέρουμε σε γραμμική μορφή. Αυτό επιτυγχάνεται λογαριθμώντας και τα δύο μέρη της σχέσης (2.3.1). Στη συνέχεια, εφαρμόζουμε πολλαπλή παλινδρόμηση για να μπορέσουμε να υπολογίσουμε τους εκτιμητές των παραμέτρων. Τέλος, αντιστρέφουμε το μετασχηματισμό που κάναμε με τη χρήση του λογάριθμου, ώστε να πάρουμε τις κανονικές τιμές.

Έτσι παίρνοντας τους λογαρίθμους της σχέσης (2.3.1) έχουμε ότι:

$$\ln\left(P_{ij}\right) = \ln\left(A_i\right) + \ln\left(B_j\right). \qquad (2.3.3)$$

Ωστόσο, παίρνοντας τους λογαρίθμους της σχέσης (2.3.2) δεν έχουμε μια γραμμική σχέση αφού:

$$\ln\left(\sum B_j\right) \neq \sum \ln\left(\mathrm{B}_j\right). \qquad (2.3.4)$$

Έτσι δε θα ληφθεί υπόψιν η σχέση (2.3.4) και θα στραφεί την προσοχή περισσότερο στην σχέση (2.3.3). Ωστόσο, αυτή η επιλογή έχει σαν αποτέλεσμα να πρέπει να μετασχηματιστεί κατάλληλα η σχέση (2.3.3), στην οποία εισάγεται ένας παράγοντας σφάλματος. Με τον τρόπο αυτό, η σχέση (2.3.3) ξαναγράφεται ως:

$$\ln\left(P_{ij}\right) = \mathrm{Y}_{ij} = a_i + b_j + e_{ij}, \qquad (2.3.5)$$

όπου $\ln\left(A_i\right) = a_i$, $\ln\left(B_j\right) = b_j$ $\kappa\alpha\iota$ e_{ij} είναι ο όρος σφάλματος. Μάλιστα, θεωρείται ότι ο όρος σφάλματος ακολουθεί κανονική κατανομή δηλαδή $e_{ij} \sim \mathrm{N}\left(0, \sigma^2\right)$.

Σε μια πιο γενική μορφή η σχέση (2.2.8) γράφεται ως:

$$\mathrm{Y} = \mathrm{X}\beta + e,$$

με

$\mathrm{Y} : (n \times 1)$ να είναι το διάνυσμα των ζημιών σε λογαριθμική μορφή,

$\mathrm{X} : (n \times p)$ να ορίζεται ο πίνακας σχεδιασμού,

$\beta : (p \times 1)$ το διάνυσμα των παραμέτρων που θα εκτιμηθούν,

και $e_{ij} \sim \mathrm{N}\left(0, \sigma^2\right)$ ένα $(n \times 1)$ διάνυσμα με άγνωστη την παράμετρο της διακύμανσης.

Από τα παραπάνω προκύπτει ότι οι $\mathrm{Y}_{ij}, 1 \leq i \leq n, 1 \leq j \leq n - i + 1$ είναι τ.μ. ανεξάρτητες και κανονικά κατανεμημένες. Παράλληλα ισχύει ότι:

$$\mathrm{E}\left(\mathrm{Y}_{ij}\right) = x_{ij}^T \beta \;\; \kappa\alpha\iota \;\; Var\left(\mathrm{Y}_{ij}\right) = \sigma^2, \qquad (2.3.6)$$

στην οποία x_{ij}^T είναι ένα διάνυσμα με p το πλήθος μεταβλητές και β ένα διάνυσμα διάστασης $(p \times 1)$ που αποτελεί τους άγνωστους συντελεστές παλινδρόμησης (Verall, 1991). Τέλος για να μπορέσουμε να πάρουμε την αναμενόμενη τιμή των μελλοντικών ζημιών $\widehat{P}_{ij}$ αντιστρέφουμε τις λογαριθμημένες τιμές και παίρνουμε ότι:

$$\mathrm{E}\left(\widehat{P}_{ij}\right) = \exp\left(x_{ij}^T \widehat{\beta} + \frac{1}{2}\widehat{\sigma}^2\right). \qquad (2.3.7)$$

Το $\widehat{\beta}$ είναι μια εκτίμηση του διανύσματος β και για $\widehat{\sigma}^2$ να είναι μια εκτίμηση για το σ^2 έχουμε τον εκτιμητή μέγιστης πιθανοφάνειας του $\mathrm{E}\left(\widehat{P}_{ij}\right)$. Ακόμη, το διάνυσμα $\widehat{\beta}$ υπολογίζεται ως:

$$\widehat{\beta} = \left(\mathrm{X}^{\mathrm{T}}\mathrm{X}\right)^{-1}\mathrm{X}^{\mathrm{T}}\mathrm{Y}, \qquad (2.3.8)$$

με X να είναι ο πίνακας σχεδιασμού, X^{T} ο ανάστροφος του και Y είναι το διάνυσμα δεδομένων. Όσον αφορά τη συνδιακύμανση έχουμε ότι:

$$Cov\left(\widehat{\beta}\right) = \sigma^2\left(\mathrm{X}^{\mathrm{T}}\mathrm{X}\right)^{-1} \text{ και } Cov\left(x_{ij}^T\widehat{\beta}\right) = \sigma^2 x_{ij}^T\left(\mathrm{X}^{\mathrm{T}}\mathrm{X}\right)^{-1}x_{ij}. \qquad (2.3.9)$$

Μια αμερόληπτη εκτίμηση του $\widehat{\sigma}^2$ είναι η $\widehat{s}^2$ που δίνεται από τη σχέση:

$$\widehat{s}^2 = \frac{1}{(n-p)}\left(Y - X\widehat{\beta}\right)^{\mathrm{T}}\left(Y - X\widehat{\beta}\right). \qquad (2.3.10)$$

Έτσι, σύμφωνα με το Verall (1991) μία αμερόληπτη εκτιμήτρια του $\mathrm{E}\left(\widehat{P}_{ij}\right)$ δίνεται από τη σχέση:

$$\mathrm{E}\left(\widehat{P}_{ij}\right) = \exp\left(x_{ij}^T\widehat{\beta}\right)g_m\left[\frac{1}{2}\left(1 - x_{ij}^T\left(X^TX\right)^{-1}x_{ij}\right)s^2\right], \qquad (2.3.11)$$

με

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

μία συνάρτηση που πρότεινε ο Finney (1941) ώστε να μην υπάρχει μεροληψία και m να ορίζονται οι βαθμοί ελευθερίας του $\widehat{\sigma}^2$.

Τέλος, η διακύμανση της αμερόληπτης εκτίμησης $\mathrm{E}\left(\widehat{P}_{ij}\right)$ είναι τ_{ij}^2. Μία αμερόληπτη εκτιμήτρια του τ_{ij}^2 είναι (βλ. Verall, 1991, p.79):

$$\widehat{\tau}_{ij}^2 = \exp\left(2x_{ij}^T\widehat{\beta}\right)\left[\left(g_m\left(\frac{1}{2}\left(1 - x_{ij}^T\left(X^TX\right)^{-1}x_{ij}\right)s^2\right)\right)^2 - g_m\left(\left(1 - 2x_{ij}^T\left(X^TX\right)^{-1}x_{ij}\right)s^2\right)\right]. \qquad (2.3.12)$$

3 Αποθεματοποίηση και Μέτρα Αβεβαιότητας

Η μέθοδος Chain Ladder είναι η πιο διαδεδομένη, για την εκτίμηση του αποθέματος των ζημιών που έχουν γίνει αλλά δεν έχουν αναφερθεί στην ασφαλιστική (για τις ονομαζόμενες ως IBNR). Ο κύριος λόγος είναι η απλότητα της μεθόδου αλλά και το γεγονός ότι είναι μία μέθοδος που δεν απαιτεί κάποια κατανομή – άρα και κάποιες υποθέσεις που πρέπει να γίνουν. Ωστόσο, είναι ευρέως γνωστό ότι οι εκτιμήσεις των αποθεμάτων μέσω της Chain Ladder είναι πολύ ευαίσθητες σε διακυμάνσεις των δεδομένων, κάτι που θα δούμε στο κεφάλαιο αυτό.

3.1 Η Μέθοδος του Mack Μοντέλου (Mack, 1993)

Στη σύγχρονη βιβλιογραφία υπάρχουν πολλές μέθοδοι εκτιμήσεις των αποθεμάτων IBNR, των οποίων τα αποτελέσματα διαφέρουν (άλλοτε πολύ άλλοτε λίγο) από αυτά της Chain Ladder. Προέκυψε, λοιπόν η ανάγκη, να εκτιμηθεί το τυπικό σφάλμα (standard error) της Chain Ladder ως ένας τρόπος για να αξιολογηθούν οι διαφορές αυτές. Με άλλα λόγια, να είναι σε θέση ο μελετητής να κρίνει, αν οι διαφορές αυτές είναι σημαντικές ή όχι.

Μια αναλυτική εκτίμηση δόθηκε από τη μέθοδο του Mack (Mack, 1993). Ο Mack κινήθηκε πάνω στη λογική κάποιων δημοσιεύσεων στο δεύτερο μισό της δεκαετίας του ’80, η οποίες πρότειναν τρόπους να εκτιμηθεί το τυπικό σφάλμα στην εκτίμηση του αποθέματος (Taylor and Ashe (1983), Zehnwirth (1985) Renshaw (1989), Christofides (1990) κ.α.). Ωστόσο, για την εποχή εκείνη απαιτούνταν σύνθετες μέθοδοι προγραμματισμού, καθότι έπρεπε να υπολογιστούν πολλές διακυμάνσεις και συνδιακυμάνσεις μεταξύ των παραμέτρων εκτίμησης.

Ο Mack στάθηκε περισσότερο στο paper του Schnieper (1991), στο οποίο ο τελευταίος προσπαθούσε να υπολογίσει την παραπάνω έκθεση που μπορούσε να έχει μία ασφαλιστική, στις καλύψεις τραυματισμού που είχε εγγυηθεί. Στο paper αυτό ο Schnieper χρησιμοποιούσε ένα μοντέλο που αποτελούσε μίξη των μεθόδων Chain Ladder και Bornhuetter-Ferguson (B-F method). Μέσα στο μοντέλο, ο Schnieper ανέπτυξε έναν υπολογισμό του τυπικού σφάλματος, για το εκτιμώμενο ασφάλιστρο, χρησιμοποιώντας σειρές Taylor.

Έτσι, ο Mack έλαβε υπόψιν τις παραδοχές του Schnieper προσθέτοντας τα ακόλουθα στοιχεία:

- Η μέθοδος αναφέρεται καθαρά στην περίπτωση εφαρμογής της Chain Ladder. Αυτό κάνει πιο εύκολα τα πράγματα, καθότι οι σειρές Taylor μπορούν να αντικατασταθούν από πιο ευκρινή διαδικασία.
- Μέσα στην εκτίμηση του τυπικού σφάλματος για το απόθεμα, εμπεριέχεται και μία εκτίμηση της διακύμανσης διαδικασίας (process variance). Η διακύμανση αυτή είναι, στην ουσία, η εκτιμώμενη διαφορά μεταξύ των παρατηρήσεων και των πραγματικών τιμών τούς.
- Ο Schnieper στο paper του ισχυρίστηκε ότι οι συντελεστές εξέλιξης της Chain Ladder είναι «ελαφρώς συσχετισμένοι». Ωστόσο, ο Mack απέδειξε, όπως θα δούμε παρακάτω ότι οι συντελεστές αυτοί είναι ασυσχέτιστοι. Κατ’ επέκταση, το απόθεμα που προκύπτει είναι αμερόληπτο.
- Τέλος, πέραν από το τυπικό σφάλμα αποθέματος για κάθε χρόνο ατυχήματος, δίνει και μία φόρμουλα για το τυπικό σφάλμα του συνολικού αποθέματος, η οποία λαμβάνει υπόψιν της, τις συσχετίσεις μεταξύ των εκτιμήσεων, που έχουν τα έτη ατυχήματος.

Θεωρούμε ότι με L_{jk} ορίζονται οι αθροιστικές ζημιές για το έτος j, με 1≤j≤N, οι οποίες είτε έχουν πληρωθεί είτε συνέβησαν μέχρι το έτος εξέλιξης k, με 1≤k≤N. Ακόμη οι τιμές L_{jk} θεωρούνται τυχαίες μεταβλητές , από τις οποίες έχουμε μία παρατήρηση αν j+k≤N+1. Έτσι σχηματίζεται και το τρίγωνο ζημιών. Σκοπός είναι να εκτιμηθούν το ποσό των συνολικών ζημιών και το ποσό των εκκρεμών ζημιών U_j:

$$U_j = L_{jN} - L_{j,N+1-j}, \qquad (3.1.1)$$

για τα έτη ατυχήματος j=2,…N.

Όπως αναφέρθηκε και παραπάνω για την Chain Ladder, θεωρείται ότι υπάρχουν συντελεστές εξέλιξης τέτοιοι ώστε:

$$\mathrm{E}(L_{j,k+1} \mid L_{j,1},...L_{j,k}) = L_{j,k} f_k,\ 1 \le j \le N, 1 \le k \le N-1. \qquad (I)$$

Επίσης, η μέθοδος Chain Ladder απαιτεί τον υπολογισμό των f_k ως εξής:

$$f_k = \frac{\sum_{j=1}^{N-k} L_{j,k+1}}{\sum_{j=1}^{N-k} L_{j,k}},\ 1 \le k \le N-1. \qquad (3.1.2)$$

αλλά και το ποσό των συνολικών ζημιών L_{jN} ως εξής:

$$L_{j\mathrm{N}} = L_{j,N+1-j} \cdot \widehat{f}_{N+1-j} \dots \cdot \widehat{f}_{N-1}, \qquad (3.1.3)$$

ή ισοδύναμα το απόθεμα U_j

$$\widehat{U}_j = L_{j,N+1-j}(\widehat{f}_{N+1-j}\cdots\widehat{f}_{N-1}-1). \qquad (3.1.4)$$

Ο αλγόριθμος της Chain Ladder δεν λαμβάνει υπόψιν του τυχόν εξαρτήσεις μεταξύ των ετών ατυχήματος. Έτσι μπορεί να υποτεθεί ότι οι μεταβλητές L_{jk} από διαφορετικά έτη ατυχήματος είναι ανεξάρτητες:

$$\left\{L_{i,1},\dots L_{i,N}\right\},\ \left\{L_{j,1},\dots L_{j,N}\right\},\ i\neq j \text{ είναι ανεξάρτητες} \qquad \text{(II)}$$

Το τελευταίο πρέπει να θεωρηθεί ως μία επιπρόσθετη – έμμεση- υπόθεση της Chain Ladder. Στην πράξη, οι ανεξαρτησία των ετών ατυχήματος μπορεί να διαταραχθεί από μεγάλες αλλαγές στις ζημιές που διαχειρίζεται η ασφαλιστική.

Το ακόλουθο θεώρημα (Mack, 1993) επιβεβαιώνει ότι οι σχέσεις (I) και (II) είναι πράγματι υποθέσεις που προκύπτουν από τη μέθοδο Chain Ladder.

Θεώρημα 1

Έστω ότι $D=\left\{L_{jk}\,\middle|\, j+k\leq N+1\right\}$ είναι το σύνολο των δεδομένων που έχουμε. Κάτω από τις υποθέσεις (I) και (II) ισχύει ότι:

$$\mathrm{E}\left(L_{jk}\middle|D\right)=L_{j,\mathrm{N}+1-j}\cdot f_{N+1-j}\cdot\ldots\cdot f_{N-1}. \qquad (3.1.5)$$

Απόδειξη: Θα χρησιμοποιήσουμε τη σχέση, $\mathrm{E}_j\left(X\right)=E\left(X\middle|L_{jN},\dots,L_{j,N+1-j}\right)$.

Τότε η (II) και η επαναλαμβανόμενη εφαρμογή της (I) δίνει:

$$\begin{aligned}
\mathrm{E}\left(L_{j\mathrm{N}}\middle|D\right)&=\mathrm{E}_j\left(L_{j\mathrm{N}}\right)\\
&=\mathrm{E}_j\left(\mathrm{E}\left(L_{j\mathrm{N}}\middle|L_{j1},\dots,L_{j,\mathrm{N}-1}\right)\right)\\
&=\mathrm{E}_j\left(L_{j,\mathrm{N}-1}f_{N-1}\right)\\
&=\mathrm{E}_j\left(L_{j,\mathrm{N}-1}\right)f_{N-1}\\
&=\mathrm{E}_j\left(L_{j,\mathrm{N}+1-j}\right)f_{N+1-j}\cdot\ldots\cdot f_{N-1}\\
&=L_{j,\mathrm{N}+1-j}f_{N+1-j}\cdot\ldots\cdot f_{N-1}.
\end{aligned} \qquad (3.1.6)$$

Από το Θεώρημα 1 προκύπτει ότι η εκτίμηση του $\widehat{L}_{j\mathrm{N}}$ έχει την ίδια μορφή με την $\mathrm{E}\left(L_{j\mathrm{N}}\middle|D\right)$, η οποία αποτελεί την καλύτερη πρόβλεψη της $L_{j\mathrm{N}}$, σύμφωνα με τα δεδομένα του δείγματος D. Το επόμενο θεώρημα αποδεικνύει ότι εκτιμώντας το πολλαπλασιαστή της σχέσης (3.1.6) $f_{N+1-j}\cdot\ldots\cdot f_{N-1}$ μέσω του $\widehat{f}_{N+1-j}\cdot\ldots\cdot\widehat{f}_{N-1}$ είναι λογικό.

Θεώρημα 2

Κάτω από τις υποθέσεις (I) και (II) οι εκτιμητές $\widehat{f}_k,\ 1 \le k \le N-1$, είναι αμερόληπτοι και ασυσχέτιστοι.

Απόδειξη: Έστω ότι, $\mathrm{B}_k = \left\{ L_{ij} \,\middle|\, j \le k,\ i + j \le N+1 \right\},\ 1 \le k \le N.$

Τότε οι υποθέσεις (II) και (I) δίνουν ότι,

$$\mathrm{E}\left(L_{i,k+1} \middle| B_k \right) = \mathrm{E}\left(L_{i,k+1} \middle| L_{iN}, \ldots, L_{i,k} \right) = L_{i,k} \cdot f_k. \qquad (3.1.7)$$

Από το τελευταίο έχουμε

$$\mathrm{E}\left(\widehat{f}_k \middle| \mathrm{B}_k \right) = \frac{\sum_{j=1}^{N-k} \mathrm{E}\left(L_{j,k+1} \middle| B_k \right)}{\sum_{j=1}^{N-k} L_{jk}} = f_k, \qquad \gamma\iota\alpha\ k = N \qquad (3.1.8)$$

το οποίο δίνει την αμεροληψία για τους εκτιμητές των παραμέτρων. Επιπρόσθετα, οι συντελεστές $\widehat{f}_k$ είναι ασυσχέτιστοι διότι για j<k ισχύει:

$$\begin{aligned} \mathrm{E}\left(\widehat{f}_j \widehat{f}_k \right) &= \mathrm{E}\left(\mathrm{E}\left(\widehat{f}_j \widehat{f}_k \middle| B_k \right) \right) \\ &= \mathrm{E}\left(\widehat{f}_j \mathrm{E}\left(\widehat{f}_k \middle| B_k \right) \right) \\ &= \mathrm{E}\left(\widehat{f}_j \right) \widehat{f}_k \\ &= \mathrm{E}\left(\widehat{f}_j \right) \mathrm{E}\left(\widehat{f}_k \right). \end{aligned} \qquad (3.1.9)$$

Το γεγονός ότι οι συντελεστές $\widehat{f}_k$ είναι ασυσχέτιστοι, είναι θετικά παράξενο, αν σκεφτούμε ότι για να υπολογιστούν, λόγου χάρη οι $\widehat{f}_k$ και $\widehat{f}_{k-1}$, χρησιμοποιήθηκαν τα ίδια δεδομένα. Επίσης, το αποτέλεσμα του προηγούμενου θεωρήματος, μπορεί να επεκταθεί στις εκτιμήσεις των ζημιών αλλά και των αποθεμάτων. Προκύπτει δηλαδή ότι:

- $\widehat{L}_{j,\mathrm{N}} = L_{j,\mathrm{N}+1-j} \widehat{f}_{N+1-j} \cdot \ldots \cdot \widehat{f}_{N-1}$ είναι ένας αμερόληπτος εκτιμητής του $\mathrm{E}\left(L_{j,\mathrm{N}} \middle| D \right) = L_{j,\mathrm{N}+1-j} f_{N+1-j} \cdot \ldots \cdot f_{N-1}$.
- $\widehat{U}_j = \widehat{L}_{j,\mathrm{N}} - L_{j,I+1-j}$ είναι ένας αμερόληπτος εκτιμητής του αποθέματος $U_j = L_{j,\mathrm{N}} - L_{j,I+1-j}$.

Στη συνέχεια, ο Mack (1993) παρουσιάζει τον υπολογισμό του μέσου τετραγωνικού σφάλματος, το οποίο αποδεικνύεται ότι χωρίζεται σε δύο μέρη, 1) την process variance δηλαδή το διασπορά διαδικασίας και 2) την parameter variance που είναι το σφάλμα εκτίμησης, όπως θα δούμε παρακάτω.

Ξεκινώντας από τον ορισμό, το μέσο τετραγωνικό σφάλμα του εκτιμητή $\widehat{L}_{jk}$, ο οποίος εκτιμά το L_{jk} ορίζεται ως:

$$mse(\widehat{L}_{jk}) = E\left(\left(\widehat{L}_{jk} - L_{jk}\right)^2 \middle| D\right), \qquad (3.1.10)$$

όπου $D = \left\{L_{jk} \middle| i+k \leq N+1\right\}$ είναι το σύνολο των δεδομένων του τριγώνου. Εδώ πρέπει να σημειωθεί ότι ο Mack δε χρησιμοποιεί το μέσο τετραγωνικό σφάλμα χωρίς δέσμευση, δηλαδή να ισχύει ότι $E\left(\left(\widehat{L}_{jk} - L_{jk}\right)^2\right) - E\left(\left(\widehat{L}_{jk} - L_{jk}\right)^2 \middle| D\right)$. Ο λόγος είναι ότι, το τελευταίο θα ίσχυε για όλα τα πιθανά σύνολα που θα μπορούσαν να σχηματιστούν. Αντιθέτως, το κύριο ενδιαφέρον εστιάζεται στο υπό συνθήκη μέσο τετραγωνικό σφάλμα των συγκεκριμένων τιμών $\widehat{L}_{jk}$, οι οποίες βασίζονται σε ένα συγκεκριμένο δείγμα D και γι' αυτό πρέπει να χρησιμοποιήσουμε την έκφραση $E\left(\left(\widehat{L}_{jk} - L_{jk}\right)^2 \middle| D\right)$, η οποία μας δίνει την μέση απόκλιση μεταξύ του $\widehat{L}_{jk}$ και του L_{jk}, λόγω της μελλοντικής τυχαιότητας και μόνο.

Βλέπουμε, επίσης, ότι:

$$mse(\widehat{R}_i) = E\left(\left(\widehat{R}_i - R_i\right)^2 \middle| D\right) = E\left(\left(\widehat{L}_{iN} - L_{iN}\right)^2 \middle| D\right) = mse\left(\widehat{L}_{iN}\right). \qquad (3.1.11)$$

Χρησιμοποιώντας τη σχέση $\mathrm{E}\left(\mathrm{X}\text{-}\alpha\right)^2 = \mathrm{Var}\left(\mathrm{X}\right) + \left(\mathrm{E}\left(\mathrm{X}\right)\text{-}\alpha\right)^2$ έχουμε ότι:

$$mse(\widehat{L}_{jk}) = Var\left(L_{jN} \middle| D\right) + \left(E\left(L_{jN} \middle| D\right) - \widehat{L}_{jk}\right)^2, \qquad (3.1.12)$$

το οποίο δείχνει ότι το μέσο τετραγωνικό σφάλμα είναι το άθροισμα της διασποράς διαδικασίας (process variance) και του σφάλματος εκτίμησης (parameter error).

Τώρα, προκειμένου να μελετήσουμε περαιτέρω το μέσο τετραγωνικό σφάλμα (mean square error-mse) χρειαζόμαστε μία φόρμουλα για τη διακύμανση του L_{jk}. Από το γεγονός ότι κάθε συντελεστής εξέλιξης $\widehat{f}_k$ είναι ένας L_{ik}-σταθμισμένος μέσος των ανεξάρτητων συντελεστών εξέλιξης $L_{j,k+1}/L_{jk}$, $1 \leq j \leq N-k$, μπορούμε να συμπεράνουμε ότι η διακύμανση $Var\left(L_{j,k+1}/L_{jk} \middle| L_{j1}, \ldots, L_{jN}\right)$ πρέπει να είναι αντιστρόφως ανάλογη του L_{jk} ή ισοδύναμα:

$$Var\left(L_{j,k+1} \middle| L_{j1}, \ldots, L_{jN}\right) = L_{jk}\sigma_k^2, \; 1 \leq j \leq N, \; 1 \leq k \leq N-1, \qquad (\mathrm{III})$$

με τις παραμέτρους σ_k^2, $1 \le k \le N-1$, να είναι άγνωστες. Αυτή είναι η διακύμανση που προκύπτει από τη μέθοδο Chain Ladder.

Όπως γίνεται αντιληπτό, χρειαζόμαστε έναν εκτιμητή για τη σ_k^2. Με το ίδιο σκεπτικό όπως στην $\widehat{f}_k$, αποδεικνύεται ότι

$$\widehat{\sigma}_k^2 = \frac{1}{\mathrm{N}-k-1} \sum_{j=1}^{N-k} L_{jk} \left(\frac{L_{j,k+1}}{L_{jk}} - f_k \right)^2, \ 1 \le k \le N-2, \qquad (3.1.13)$$

είναι ένας αμερόληπτος εκτιμητής της σ_k^2, $1 \le k \le N-2$. Ωστόσο, λείπει ένας εκτιμητής για την $\sigma_{\mathrm{N}-1}$. Αν θεωρήσουμε ότι $\widehat{f}_{\mathrm{N}-1} = 1$και ότι οι η εξέλιξη των ζημιών σταματάει μετά από N-1 χρόνια, μπορούμε να πούμε ότι $\sigma_{\mathrm{N}-1} = 0$. Αν όχι, τότε επεκτείνουμε την εκθετικά μειούμενη σειρά $\widehat{\sigma}_{\mathrm{N}}, \ldots \widehat{\sigma}_{\mathrm{N}-3}, \widehat{\sigma}_{\mathrm{N}-2}$ κατά έναν επιπλέον παράγοντα, απαιτώντας να ισχύει ότι:

$\widehat{\sigma}_{\mathrm{N}-3} / \widehat{\sigma}_{\mathrm{N}-2} = \widehat{\sigma}_{\mathrm{N}-2} / \widehat{\sigma}_{\mathrm{N}-1}$, το οποίο ισχύει τουλάχιστον όσο $\widehat{\sigma}_{\mathrm{N}-3} > \widehat{\sigma}_{\mathrm{N}-2}$. Το τελευταίο οδηγεί στο ακόλουθο:

$$\widehat{\sigma}_{\mathrm{N}-1}^2 = \min\left[\widehat{\sigma}_{\mathrm{N}-2}^4 / \widehat{\sigma}_{\mathrm{N}-3}^2 , \ \min\left(\widehat{\sigma}_{\mathrm{N}-3}^2, \widehat{\sigma}_{\mathrm{N}-2}^2 \right) \right]. \qquad (3.1.14)$$

Από τα παραπάνω, προκύπτει το ακόλουθο θεώρημα, που αναφέρεται στον υπολογισμό του μέσου τετραγωνικού σφάλματος (mse).

Θεώρημα 3

Κάτω από τις υποθέσεις, (Ι), (ΙΙ), (ΙΙΙ) το μέσο τετραγωνικό σφάλμα (mean square error-mse) μπορεί να υπολογιστεί από τον κάτωθι τύπο:

$$mse(R_j) = \widehat{U}_j^2 \sum_{k=N+1-j}^{N-1} \frac{\widehat{\sigma}_k^2}{\widehat{f}_k} \left(\frac{1}{\widehat{L}_{jk}} + \frac{1}{\sum_{l=1}^{N-k} L_{lk}} \right), \qquad (3.1.15)$$

όπου $\widehat{L}_{jk} = L_{j,N+1-j} \widehat{f}_{N+1-j} \cdot \ldots \cdot \widehat{f}_{k-1}, \ k > N+1-j,$ είναι οι εκτιμώμενες τιμές των μελλοντικών απαιτήσεων L_{jk} και $\widehat{L}_{j,N+1-j} = L_{j,N+1-j}$.

Απόδειξη: Θα χρειαστούμε τις σχέσεις:

$$\begin{aligned} &\mathrm{E}_j(\mathrm{X}) = \mathrm{E}\left(\mathrm{X}\,\middle|\,L_{N1},\dots,L_{j,N+1-j}\right), \\ &Var_j(X) = Var\left(X\,\middle|\,L_{N1},\dots,L_{j,N+1-j}\right). \end{aligned} \tag{3.1.16}$$

Χρησιμοποιώντας τη σχέση (3.1.12) έχουμε ότι:

$$mse\left(\widehat{R}_j\right) = Var\left(L_{jN}\,\middle|\,D\right) + E\left(\left(L_{jN}\,\middle|\,D\right) - \widehat{L}_{jN}\right)^2. \tag{3.1.17}$$

Εφαρμόζοντας, επαναληπτικά, την υπόθεση (I) της Chain Ladder και τις υπόθεσης (III) για την διακύμανση, τότε για τον πρώτο όρο του $mse\left(\widehat{R}_j\right)$προκύπτει:

$$\begin{aligned} Var\left(L_{jN}\,\middle|\,D\right) &= Var_j\left(L_{jN}\right) \\ &= \mathrm{E}_j\left(Var\left(L_{jN}\,\middle|\,L_{jN},\dots,L_{j,N-1}\right)\right) + \\ &\quad + Var_j\left(\mathrm{E}\left(L_{jN}\,\middle|\,L_{jN},\dots,L_{j,N-1}\right)\right) \\ &= \mathrm{E}_j\left(L_{j,N-1}\right)\sigma_{\mathrm{N}-1}^2 + Var_j\left(L_{j,N-1}\right) f_{N-1}^2 \\ &= \mathrm{E}_j\left(L_{j,N-2}\right) f_{N-2}^2\sigma_{\mathrm{N}-1}^2 + \mathrm{E}_j\left(L_{j,N-2}\right) f_{N-1}^2\sigma_{\mathrm{N}-2}^2 + \\ &\quad + Var_j\left(L_{j,N-2}\right) f_{N-2}^2 f_{N-1}^2 \\ &= L_{j,N+1-j}\sum_{k=N+1-j}^{N-1} f_{N+1-j}\cdot\ldots\cdot f_{k-1}\sigma_k^2 f_{k+1}^2\cdot\ldots\cdot f_{N-1}^2, \end{aligned} \tag{3.1.18}$$

επειδή $Var_j\left(L_{j,\mathrm{N}+1-j}\right) = 0.$

Τώρα από το Θεώρημα 1, για το δεύτερο όρο του $mse\left(\widehat{R}_j\right)$, παίρνουμε ότι:

$$E\left(\left(L_{jN}\,\middle|\,D\right) - \widehat{L}_{jN}\right)^2 = L_{j,N+1-j}^2\left(f_{N+1-j}\cdot\ldots\cdot f_{k-1} - \widehat{f}_{N+1-j}\cdot\ldots\cdot\widehat{f}_{k-1}\right)^2. \tag{*}$$

Στην πραγματικότητα, πρέπει να βρούμε εκτιμητές για τους δύο όρους του $mse\left(\widehat{R}_j\right)$. Για τον πρώτο όρο, αυτό επιτυγχάνεται με το να αντικαταστήσουμε τις άγνωστες παραμέτρους f_k και σ_k^2 με τους εκτιμητές $\widehat{f}_k$ και $\widehat{\sigma}_k^2$, που σημαίνει ότι εκτιμούμε το $Var_j\left(L_{jN}\,\middle|\,D\right)$ από τη σχέση:

$$\begin{aligned} &L_{j,N+1-j}\left(\sum_{k=N+1-j}^{N-1}\widehat{f}_{N+1-j}\cdot\ldots\cdot\widehat{f}_{N-1}\cdot\sigma_k^2\cdot\widehat{f}_{k+1}^2\cdot\ldots\cdot\widehat{f}_{N-1}^2\right) \\ &= \widehat{L}_{jN}^2\sum_{k=N+1-j}^{N-1}\frac{\frac{\sigma_k^2}{\widehat{f}_k^2}}{L_{jk}}. \end{aligned} \tag{3.1.20}$$

Παρόλα αυτά, στο δεύτερο όρο της σχέσης (*) του $mse\left(\widehat{R}_j\right)$ δεν μπορούμε να αντικατασταθεί το f_k με $\widehat{f}_k$ διότι κάτι τέτοιο θα έδινε 0. Ο Mack χρησιμοποιεί μια διαφορετική προσέγγιση. Θεωρεί ότι:

$$\begin{aligned} F &= f_{N+1-j}\cdot\ldots\cdot f_{N-1} - \widehat{f}_{N+1-j}\cdot\ldots\cdot\widehat{f}_{N-1} \\ &= S_{N+1-j}+\ldots+S_{N-1}, \end{aligned} \tag{3.1.21}$$

με

$$S_k = \widehat{f}_{N+1-j}\cdot\ldots\cdot\widehat{f}_{k-1}\left(f_k-\widehat{f}_k\right)\cdot f_{k+1}\cdot\ldots\cdot f_{N-1}. \tag{3.1.22}$$

Έτσι προκύπτει ότι:

$$\begin{aligned} F^2 &= \left(S_{N+1-j}+\ldots+S_{N-1}\right)^2 \\ &= \sum_{N+1-j}^{N-1} S_k^2 + 2\sum_{i<k} S_i S_k. \end{aligned} \tag{3.1.23}$$

Στη συνέχεια, αντικαθιστά το S_k^2 με $\mathrm{E}\left(S_k^2\left|\mathrm{B}_k\right.\right)$ και $S_iS_k,\ i<k,$ με $\mathrm{E}\left(S_iS_k\left|\mathrm{B}_k\right.\right)$. Αυτό σημαίνει ότι προσεγγίζει τα S_k^2 και τα $S_iS_k,\ i<k,$ χρησιμοποιώντας όσο το δυνατόν λιγότερα δεδομένα από το δείγμα. Εξαιτίας του ότι $\mathrm{E}\left(f_k-\widehat{f}\left|\mathrm{B}_k\right.\right)=0$ (όπως φαίνεται και στη απόδειξη του Θεωρήματος 2) τότε $\mathrm{E}\left(S_iS_k\left|\mathrm{B}_k\right.\right)=0$, $i<k$. Επίσης, εξαιτίας του ότι:

$$\begin{aligned} Var\left(\widehat{f}_k\left|B_k\right.\right) &= \mathrm{E}\left({}_k\left(f_k-\widehat{f}\right)^2\left|\mathrm{B}_k\right.\right) \\ &= \frac{\sum_{i=1}^{N-k} Var\left(L_{i,N+1-i}\left|B_k\right.\right)}{\left(\sum_{i=1}^{N-k}\left(L_{ik}\right)\right)^2} \\ &= \frac{\sigma_k^2}{\left(\sum_{i=1}^{N-k} L_{ik}\right)}, \end{aligned} \tag{3.1.24}$$

παίρνουμε ότι

$$\mathrm{E}\left(S_k^2\left|B_k\right.\right) = \frac{\widehat{f}_{N+1-j}^2\cdot\ldots\cdot\widehat{f}_{N-1}^2\sigma_k^2 f_{k+1}^2\cdot\ldots\cdot\widehat{f}_{N-1}^2}{\sum_{j=1}^{N-1} L_{jk}}. \tag{3.1.25}$$

Τώρα, αντικαθιστώντας $F^2 = \left(\sum S_k\right)^2$ με $\sum_k \mathrm{E}\left(S_k^2 \middle| B_k\right)$ και επειδή όλοι οι όροι του αθροίσματος αυτού είναι θετικοί, γίνεται να αντικατασταθούν όλες οι άγνωστοι παράμετροι f_k και σ_k^2 με τους αμερόληπτους εκτιμητές τους $\widehat{f}_k, \widehat{\sigma}_k^2$. Τέλος , η σχέση $F^2 = \left(f_{N+1-j} \cdot\ldots\cdot f_{N-1} - \widehat{f}_{N+1-j} \cdot\ldots\cdot \widehat{f}_{N-1}\right)^2$ εκτιμάται από :

$$\sum_{k=N+1-j}^{N-1} \left(\frac{\widehat{f}_{N+1-j}^2 \cdot\ldots\cdot \widehat{f}_{N-1}^2 \sigma_k^2 f_{k+1}^2 \cdot\ldots\cdot \widehat{f}_{N-1}^2}{\sum_{j=1}^{N-1} L_{jk}} \right) = \widehat{f}_{N+1-j}^2 \cdot\ldots\cdot \widehat{f}_{N-1}^2 \sum_{k=N+1-j}^{N-1} \frac{\frac{\widehat{\sigma}_k^2}{\widehat{f}_k^2}}{\sum_{j=1}^{N-1} L_{jk}}. \qquad (3.1.26)$$

Το τελευταίο οδηγεί στον εκτιμητή που αναφέρθηκε παραπάνω στο θεώρημα.

Η τετραγωνική ρίζα $s.e.\left(\widehat{R}_j\right)$ ενός εκτιμητή του μέσου τετραγωνικού σφάλματος, ορίζεται να είναι η τυπική απόκλιση του $\widehat{R}_j$.

Συχνά η τυπική απόκλιση της συνολικής εκτίμησης του αποθέματος είναι αντικείμενο διερεύνησης. Στην περίπτωση αυτή, δεν γίνεται να απλά να προστεθούν όλες οι τιμές του τυπικού σφάλματος $\left(s.e.\left(\widehat{R}_j\right)\right)^2$, $2 \le j \le \mathrm{N}$, διότι θα υπήρχε συσχέτιση μεταξύ των εκτιμητών $\widehat{f}_k, \widehat{\sigma}_k$. Έτσι για να εκτιμηθεί το συνολικό μέσο τετραγωνικό σφάλμα του συνολικού αποθέματος, ο Mack χρησιμοποιεί το παρακάτω πόρισμα:

Πόρισμα 1

Με τις υποθέσεις του Θεωρήματος 3 το μέσο τετραγωνικό σφάλμα του συνολικού αποθέματος $\widehat{R} = \widehat{R}_2 + \ldots \widehat{R}_{\mathrm{N}}$ μπορεί να υπολογιστεί ως:

$$\widehat{mse}(R) = \sum_{j=2}^{N} \left\{ \widehat{mse}(R_j) + \widehat{L}_{jN} \left(\sum_{l=j+1}^{N} \widehat{U}_{jl} \right) \sum_{k=N+1-j}^{N-1} \frac{\frac{2\widehat{\sigma}_k^2}{\widehat{f}_k^2}}{\sum_{n=1}^{N-k} L_{nk}} \right\}. \qquad (3.1.27)$$

Απόδειξη:

Έχουμε ότι:

$$mse\left(\sum_{j=2}^{N}\widehat{R}_j\right)=E\left(\left(\sum_{j=2}^{N}\widehat{R}_j-\sum_{j=2}^{N}R_j\right)^2\middle|D\right)$$

$$=E\left(\left(\sum_{j=2}^{N}\widehat{L}_{jN}-\sum_{j=2}^{N}L_{jN}\right)^2\middle|D\right) \quad (3.1.28)$$

$$=Var\left(\sum_{j=2}^{N}L_{jN}\,|D\right)+\left(E\left(\sum_{j=2}^{N}L_{jN}\,|D\right)-\sum_{j=2}^{N}\widehat{L}_{jN}\right)^2.$$

Η ανεξαρτησία των ετών ατυχήματος δίνει ότι:

$$Var\left(\sum_{j=2}^{N}L_{jN}\,|D\right)=\sum_{j=2}^{N}Var\left(L_{jN}\,|D\right), \quad (3.1.29)$$

από το οποίο τα υποαθροίσματα έχουν ήδη υπολογιστεί στην απόδειξη του Θεωρήματος 3.

Επιπλέον,

$$\left(E\left(\sum_{j=2}^{N}L_{jN}\,|D\right)-\sum_{j=2}^{N}\widehat{L}_{jN}\right)^2=\left(\sum_{j=2}^{N}\left(\mathrm{E}\left(L_{jN}\,|D\right)-\widehat{L}_{jN}\right)\right)^2$$

$$=\sum_{jl}\left(\mathrm{E}\left(L_{jN}\,|D\right)-\widehat{L}_{jN}\right)\cdot\left(\mathrm{E}\left(L_{lN}\,|D\right)-\widehat{L}_{lN}\right) \quad (3.1.30)$$

$$=\sum_{jl}L_{j,N+1-j}L_{j,N+1-l}F_jF_l,$$

με $F=f_{N+1-j}\cdot\ldots\cdot f_{N-1}-\widehat{f}_{N+1-j}\cdot\ldots\cdot\widehat{f}_{N-1}$.

Παρατηρώντας ότι,

$$mse\left(\widehat{R}_j\right)=Var\left(L_{jN}\,|D\right)+\left(L_{j,N+1-j}F_j\right)^2, \quad (3.1.31)$$

(από απόδειξη Θεωρήματος 3) βλέπουμε ότι,

$$mse\left(\sum_{j=2}^{N}\widehat{R}_j\right)=\sum_{j=2}^{N}mse\left(\widehat{R}_j\right)+\sum_{2\le l\le j\le N}2\cdot L_{j,N+1-j}L_{j,N+1-l}F_jF_l. \quad (3.1.32)$$

Μία ίδια διαδικασία ακολουθείται για το F^2, στην παραπάνω απόδειξη, δίνοντας για F_jF_l, $j<l$ τον εκτιμητή:

$$\sum_{k=N+1-j}^{N-1}\frac{\widehat{f}_{N+1-j}\cdot\ldots\cdot\widehat{f}_{N-1}\cdot\sigma_k^2\cdot\widehat{f}_{k+1}^2\cdot\ldots\cdot\widehat{f}_{N-1}^2}{\sum_{j=1}^{N-1}L_{jk}}, \quad (3.1.33)$$

ολοκληρώνοντας την απόδειξη.

3.2 Η Μέθοδος Bootstrap (Bootstrap Method) στο Mack Model

Η μέθοδος Bootstrap προτάθηκε από τον Efron (1979) και είναι μια γενική προσέγγιση του στατιστικού συμπεράσματος. Η πρώτη της εφαρμογή στον στοχαστικό υπολογισμό του αποθέματος έγινε από τους England και Verall (1999). Η λογική της μεθόδου στηρίζεται στο μοντέλο της Over-Dispersed Poisson. Η επέκταση της μεθόδου στα γενικευμένα γραμμικά μοντέλα (McCullag και Nelder (1989)), συμπεριλαμβάνοντας και το μοντέλο Mack, παρουσιάστηκε πάλι από τους England και Verall (2006). Θα παρουσιάσουμε τη μέθοδο ενώ σε παρακάτω κεφάλαιο, θα δούμε και την εφαρμογή της, στον κίνδυνο αποθέματος.

Η κύρια ιδέα πίσω από τη μέθοδο Bootstrap είναι η εξής. Ας υποθέτουμε ότι θέλουμε να προσεγγίσουμε την τιμή ενός στατιστικού μεγέθους, όπως πχ η μέση τιμή ενός πληθυσμού με συγκεκριμένα χαρακτηριστικά. Προκειμένου να υπολογίσουμε αυτό το μέγεθος, δεν μπορούμε να μετρήσουμε όλο τον πληθυσμό με τα χαρακτηριστικά που μας ενδιαφέρουν. Έτσι, εφαρμόζοντας κάποιες μετρήσεις, παίρνουμε ένα δείγμα του πληθυσμού που μας ενδιαφέρει, μεγέθους n. Οφείλουμε να επισημάνουμε ότι το δείγμα αυτό πρέπει να είναι ανεξάρτητο και ομοιόμορφα κατανεμημένο. Από αυτό το δείγμα, μπορούμε να πάρουμε μόνο μία εκτίμηση του μεγέθους που μελετάμε, πχ μία μόνο εκτίμηση της μέσης τιμής.

Προκειμένου, να σιγουρευτούμε, αν η μέση τιμή του δείγματος είναι τέτοια που χαρακτηρίζει όλο τον πληθυσμό, χρειάζεται να δούμε τη διακύμανση της μέσης τιμής. Εδώ έρχεται η χρησιμότητα της μεθόδου Bootstrap. Παίρνοντας το δείγμα (μεγέθους n) που ήδη έχουμε και επιλέγοντας με τυχαίο τρόπο και με επανάθεση τιμές, από το δείγμα αυτό, φτιάχνουμε άλλα δείγματα μεγέθους n το καθένα. Τα δείγματα αυτά ονομάζονται Bootstrap δείγματα. Η διαδικασία αυτή πρέπει να επαναληφθεί πολλές φορές, πχ 1.000, και για το κάθε νέο δείγμα μπορούμε να υπολογίσουμε το μέγεθος που μας ενδιαφέρει. Στη συνέχεια με τη χρήση ιστογράμματος, μπορούμε να δούμε την κατανομή του εκτιμώμενου μεγέθους και έτσι μπορούμε να απαντήσουμε στο αν η διακύμανσή του είναι μεγάλη ή μικρή.

Με ένα πιο επίσημο τρόπο, μπορούμε να πούμε ότι με τη μέθοδο Bootstrap παίρνουμε μία εκτίμηση για την συνάρτηση κατανομής, του πληθυσμού, μέσα από την εμπειρική κατανομή, ακριβώς επειδή χρησιμοποιούμε δείγματα που προέκυψαν με επανάθεση από το αρχικό. Τα δείγματα, που προέκυψαν από επιλογή με επανάθεση του αρχικού, ονομάζονται ψευδο-δεδομένα (pseudo data).

Παρουσιάζοντας τις παραπάνω παραγράφους με μαθηματικές εκφράσεις, θα λέγαμε ότι: Δεδομένου ενός ανεξάρτητου και ομοιόμορφα κατανεμημένου (iid) δείγματος μεγέθους n, $x=\left(x_1,\ldots x_n\right)$, μιας άγνωστης κατανομής F και ενός στατιστικού $\widehat{\theta}(x)$, πχ έναν εκτιμητή, να μπορέσουμε να βρούμε τη κατανομή του $\widehat{\theta}$.

Σκοπός της μεθόδου είναι να προσεγγίσουμε την κατανομή F μέσω της εμπειρικής κατανομής $\widehat{F}=\sum_{i=1}^{n}\frac{\delta\left(x-x_i\right)}{n}$. Ένα δείγμα (iid) -μεγέθους n- που επιλέγεται από την κατανομή $\widehat{F}$, ονομάζεται Bootstrap δείγμα $x^*=\left(x_1^*,\ldots x_n^*\right)$. Αυτό το δείγμα προκύπτει μέσω επιλογής με επανάθεση από το δείγμα $x=\left(x_1,\ldots x_n\right)$. Η επιλογή αυτή επαναλαμβάνεται B φορές δίνοντας ένα σύνολο $\left\{x_1^*,\ldots x_B^*\right\}$ και το στατιστικό που μας ενδιαφέρει υπολογίζεται για κάθε δείγμα. Τέλος, η κατανομή $P\left(\widehat{\theta}\right)$ προσεγγίζεται από την Bootstrap κατανομή $\widehat{P}=\sum_{i=1}^{B}\frac{\delta\left(\widehat{\theta}-\widehat{\theta}\left(x_i^*\right)\right)}{B}$.

Όταν εφαρμόζεται η μέθοδος Bootstrap στα γενικευμένα γραμμικά μοντέλα, πρώτα πρέπει να ορίσουμε το στατιστικό μοντέλο. Στη συνέχεια, πρέπει να δημιουργήσουμε νέα σύνολα δεδομένων, τα οποία ονομάζονται ψευδο-δεδομένα (pseudo-data) και τα οποία τα χρησιμοποιούμε στο τρίγωνο εξέλιξης ζημιών.

Στα γενικευμένα γραμμικά μοντέλα, οι υποθέσεις που γίνονται – για κάθε τιμή u μιας τ.μ. X είναι οι ακόλουθες:

$$\mathrm{E}\left[\mathrm{X}_u\right]=m_u \qquad \kappa\alpha\iota \qquad Var\left[\mathrm{X}_u\right]=\frac{\varphi V\left(m_u\right)}{w_m}, \tag{3.2.1}$$

όπου με φ δηλώνεται η παράμετρος κλίμακας, $V\left(m_u\right)$ είναι η συνάρτηση διακύμανσης (πρόκειται για μία συνάρτηση του μέσου) και τα w_u είναι βάρη(που συνήθως ορίζονται με τη μονάδα, για όλες τις παρατηρήσεις). Η επιλογή της κατανομής είναι αυτή που καθορίζει τις τιμές των φ και $V\left(m_u\right)$ (McCullagh and Nelder, 1989).

Σύμφωνα, λοιπόν, με τους England and Verall (2006) η Bootstrap μέθοδος μπορεί να εφαρμόζεται στο Over-Dispersed Poisson μοντέλο όσο και στο Over-Dispersed Negative Binomial μοντέλο, με τα οποία δε θα ασχοληθούμε αλλά οφείλουμε να τα αναφέρουμε για ακαδημαϊκούς λόγους. Θα εστιάσουμε στην εφαρμογή της Bootstrap μεθόδου πάνω στο μοντέλο του Mack. Η διαδικασία της Bootstrap πάνω στο μοντέλο του Mack ορίζει μία αναδρομική διαδικασία. Η διαφορά της με τις δύο άλλες εφαρμογές (Over-Dispersed Poisson και Over-Dispersed Binomial) είναι στις υποθέσεις κατανομής. Χρησιμοποιεί διαφορετικό ορισμό για τα υπόλοιπα και επιπλέον δε χρειάζεται τον υπολογισμό της παραμέτρου κλίμακας, που οι προαναφερθείσες εφαρμογές απαιτούν.

Στο πλαίσιο αυτό, η Bootstrap μέθοδος δεν μπορεί να χαρακτηριστεί «ελεύθερη κατανομής» (distribution free), καθώς απαιτούνται υποθέσεις μιας κατανομής, όταν ορίζουμε το στατιστικό μοντέλο και θέλουμε να πάρουμε εκτίμηση για κάποιες σημαντικές παραμέτρους.

Πιο αναλυτικά, από τη σχέση με την οποία ορίσαμε το μοντέλο του Mack (1993), χρησιμοποιώντας τους συντελεστές εξέλιξης ως μεταβλητές που θέλουμε να εκτιμήσουμε και να εξηγήσουμε (response variable), έχουμε ότι:

$$E\left[f_{jk}\left|L_{j,k-1}\right.\right]=\lambda_k \quad \kappa\alpha\iota \quad Var\left[f_{jk}\left|L_{j,k-1}\right.\right]=\frac{\sigma_j^2}{L_{j,k-1}} \quad \gamma\iota\alpha \;\; k\geq 2. \qquad (3.2.2)$$

Για το λόγο αυτό, και σύμφωνα με τη σχέση (3.2.1) έχουμε ότι $\mathrm{X}_u = f_{jk}, \quad m_u=\lambda_k, \quad w_u = L_{j,k-1} \;\; \kappa\alpha\iota \;\; V\left(m_u\right)=1$. Επίσης, το μοντέλο, ορίζεται χρησιμοποιώντας μη-σταθερές παραμέτρους κλίμακας $\varphi_k=\sigma_k^2$. Έτσι, τα μετασχηματισμένα υπόλοιπα Pearson ορίζονται ως εξής:

$$r_{jk}=r_{PS}\left(f_{jk},\widehat{\lambda}_k,L_{jk},\widehat{\sigma}_k\right)=\frac{\sqrt{L_{jk}}\left(f_{jk}-\widehat{\lambda}_k\right)}{\widehat{\sigma}_k}, \quad j\leq N-k, \;\; k\leq N-1. \qquad (3.2.3)$$

3.3 Εφαρμογή της Μεθόδου Bootstrap στον Κίνδυνο αποθέματος

Με τη χρήση της Bootstrap μεθόδου σκοπός είναι να εκτιμηθεί η κατανομή του αποθέματος R, όπως αυτό υπολογίζεται από την Chain Ladder. Για να γίνει αυτό, πρέπει αρχικά να επιλεγεί ένα δείγμα x. Σύμφωνα με τους England and Verall (2006) ως δείγμα προτείνονται τα υπόλοιπα του Pearson (Pearson residuals), όπως αυτά παρουσιάστηκαν στην παραπάνω παράγραφο.

Από το δείγμα των υπολοίπων, που προκύπτει, εφαρμόζεται η μέθοδος Bootstrap. Αυτό σημαίνει ότι για κάθε θέση του τριγώνου επιλέγουμε τυχαία- με επανάθεση - ένα από τα υπόλοιπα και τα τοποθετούμε στην κάθε θέση. Έτσι δημιουργείται ένα Bootstrapped τρίγωνο. Αυτό επαναλαμβάνεται πολλές φορές και το πλήθος των επαναλήψεων το συμβολίζουμε με B. Στη συνέχεια, η εξίσωση (3.2.3) αντιστρέφεται και λύνεται ως προς f_{jk}. Έτσι έχουμε ένα νέο τρίγωνο με τους year-to-year συντελεστές εξέλιξης, οι οποίοι δίνονται από την ακόλουθη σχέση:

$$f_{jk}^{*} = r_{jk}^{*} \frac{\widehat{\sigma}_k}{\sqrt{L_{jk}}} + \widehat{f}_k. \qquad (3.3.1)$$

Τώρα, με τη χρήση της (3.2.4) προκύπτουν οι συντελεστές εξέλιξης της Chain Ladder ως εξής:

$$\widehat{f}_k^{*} = \sum_{j=1}^{N-k} \frac{L_{jk}}{\sum_{l=1}^{N-k} L_{lk}} f_{jk}^{*}. \qquad (3.3.2)$$

Οπότε, οι μελλοντικές εκτιμώμενες ζημιές υπολογίζονται με αναδρομικό τρόπο σύμφωνα με τους Verall και England (2002,2006). Αυτό γίνεται ως εξής: Θεωρείται ότι οι μελλοντικές ζημιές ακολουθούν μία κατανομή, π.χ. Lognormal. Έτσι για κάθε έτος αναγγελίας και κάθε έτος ανάπτυξης επιλέγεται τυχαία, μία τιμή της παραπάνω κατανομής. Οι παράμετροι της κατανομής είναι αυτοί που ορίζουν την τυχαία τιμή που θα μπει σε κάθε θέση του τριγώνου. Παραδείγματος χάριν, για την Lognormal έχουμε:

$$L_{j,k+1}^{*} \sim \text{Lognormal}\left(\widehat{f}_k^{*} L_{jk}^{*}, \widehat{\sigma}^2 L_{jk}^{*} \right), \;\; k \geq N-j+2. \qquad (3.3.3)$$

Εφόσον συμπληρωθεί το τρίγωνο, τότε βρίσκουμε το απόθεμα. Σημειωτέων, ότι αυτή η διαδικασία επαναλαμβάνεται B φορές, άρα θα βρούμε B το πλήθος αποθέματα. Σκοπός, όπως αναφέραμε, είναι να εκτιμήσουμε την κατανομή του εκτιμητή αποθέματος. Αυτό γίνεται μέσω της εμπειρικής κατανομής $\hat{P}$, η οποία υπολογίζεται ως εξής:

$$\widehat{P} = \frac{\sum_{i=1}^{B} \delta\left(\widehat{R} - \widehat{R}\left(r_i^{*}\right)\right)}{B} \;\; . \qquad (3.3.4)$$

Εν κατακλείδι μπορούμε να υπολογίσουμε το μέσο τετραγωνικό σφάλμα (mean square error- mse), το οποίο μας δίνει ένα μέτρο κινδύνου. Το mse υπολογίζεται από τη σχέση:

$$m\widehat{s}e\left(\widehat{R}\right) = \frac{1}{B-1} \sum_{i=1}^{B} \left(\widehat{R}\left(r_i^{*}\right) - \overline{R}^{*} \right)^2, \quad \mu\varepsilon \quad \overline{R} = \frac{1}{B} \sum_{i=1}^{B} \widehat{R}\left(r_i^{*}\right). \qquad (3.3.5)$$

3.3.1 Αλγόριθμος της Μεθόδου Bootstrap

Στο σημείο αυτό είναι καλό να δείξουμε τον τρόπο που δουλεύει η μέθοδο Bootstrap ώστε να γίνει πιο κατανοητή. Θα παρουσιάσουμε σε απλά βήματα τον τρόπο με τον οποίο λειτουργεί και πώς οδηγούμαστε στο να υπολογίσουμε το απόθεμα. Τα βήματα υπολογισμού είναι αυτά που παρουσίασαν οι Verall and England (1999) αλλά και οι Leong et al. (2014).

Βήματα αλγορίθμου:

1. Αρχικά έχουμε ένα τρίγωνο εξέλιξης ζημιών nxn, με τις αθροιστικές ζημιές (cumulative losses) και όχι με τις ατομικές ζημιές (incremental losses)
2. Για το παραπάνω τρίγωνο, βρίσκουμε τους year-to-year development factors. Οι εν λόγω συντελεστές υπολογίζονται διαιρώντας σε κάθε γραμμή, το κελί της στήλης j+1 με το κελί της στήλης j.
3. Στη συνέχεια υπολογίζουμε τα υπόλοιπα (residuals) με τον τύπο που δίνει η σχέση (3.2.3). τα υπόλοιπα αυτά είναι τα μη μετασχηματισμένα υπόλοιπα του Pearson (unscaled Pearson's Residuals). Είναι συχνό να πολλαπλασιάζονται με μία παράμετρο κλίμακας ώστε να μετασχηματιστούν ανάλογα με τους βαθμούς ελευθερίας και τις ποσότητες που θέλουμε να υπολογίσουμε. Η παράμετρος αυτή συμβολίζεται ως ϕ_m.
4. Εφόσον έχουμε υπολογίσει τα υπόλοιπα (τα οποία εμφανίζονται σε έναν τρίγωνο (n-1)x(n-1)), τότε ακολουθούμε την εξής διαδικασία:
    a. Δημιουργούμε ένα νέο τρίγωνο υπολοίπων, διαλέγοντας τυχαία αλλά με επανάθεση υπόλοιπα από το πρωταρχικό τρίγωνο υπολοίπων.
    b. Για κάθε ένα νέο τρίγωνο υπολοίπων λύνουμε τη σχέση (3.2.3) ως προς f_{jk}^{*} όπως δείχνει η σχέση (3.3.1). Έτσι υπολογίσουμε πάλι τους year-to year συντελεστές εξέλιξης, όπως κάναμε στο βήμα 2.
    c. Βρίσκουμε τους νέους συντελεστές εξέλιξης $\widehat{f}_k^{*}$ σύμφωνα με τη σχέση (3.3.2).
    d. Τώρα, επιστρέφοντας στο αρχικό τρίγωνο εξέλιξης αθροιστικών ζημιών, εφαρμόζουμε τη σχέση (3.3.3), ώστε να γεμίσουμε το τρίγωνο εξέλιξης ζημιών, όπως θα γέμιζε και με την Chain Ladder.
    e. Το κάτω δεξιό μέρος του τριγώνου – το οποίο γεμίσαμε με το βήμα d – αποτελείται από ζημιές σε ατομικό επίπεδο (incremental losses). Για να μπορέσουμε, τώρα, να υπολογίσουμε το απόθεμα, αθροίζουμε όλες τις τιμές του κάτω δεξιού τριγώνου, όπως αυτές υπολογίστηκαν από τη σχέση (3.3.3).
    f. Αποθηκεύουμε το άθροισμα αυτό, καθώς αποτελεί το απόθεμα για μία επανάληψη της μεθόδου Bootstrap.
5. Η διαδικασία αυτή επαναλαμβάνεται πολλές φορές (πχ 1000, 2500). Η μέση τιμή των αποθηκευμένων αποθεμάτων κάθε επαναληπτικής διαδικασίας μας δίνει το απόθεμα που

πρέπει να κρατήσει η ασφαλιστική. Η τυπική απόκλιση των αποθηκευμένων αποθεμάτων μας δίνει το σφάλμα διαδικασίας (process error).

3.4 Ακραίες Τιμές

3.4.1 Λάθη και Προβλήματα Δεδομένων

Στα προηγούμενα κεφάλαια είδαμε πώς εκτιμάται το απόθεμα με τη χρήση της Μεθόδου Chain Ladder και πώς υπολογίζεται το σφάλμα της εκτίμησης αυτής με την Mack Method και την Bootstrap. Ωστόσο, η εκτίμηση του αποτελέσματος μπορεί να μην είναι σωστή, όχι λόγω της μεθόδου που χρησιμοποιείται αλλά λόγω της μη αρτιότητας των δεδομένων. Αυτό συμβαίνει όταν τα δεδομένα επηρεάζονται από διάφορους παράγοντες ενδογενείς (π.χ. λάθος καταχώριση στο σύστημα) ή εξωγενείς (π.χ. πληθωρισμός, μη επιθυμητή δικαστική έκβαση κλπ.), με αποτέλεσμα να εμφανίζονται μεγάλες ζημιές ή μεγάλα άλματα μεταξύ των ζημιών. Με αυτές τις επιρροές και χωρίς κανένα μέτρο πρόληψης, η κάθε ασφαλιστική εταιρία μπορεί να εκτιμήσει ένα απόθεμα το οποίο πιθανόν να μην αρκεί για τις απαιτήσεις που θα κληθεί να πληρώσει ή να είναι υπερτιμημένο και άρα να επηρεάζει τα κέρδη της εταιρίας ή να την αναγκάζει να κρατάει δεσμευμένα μεγαλύτερα ποσά από αυτά που χρειάζεται.

Οι μέθοδοι εκτίμησης του αποθέματος, τόσο αυτές που αναφέρθηκαν όσο και οι υπόλοιπες της βιβλιογραφίας, δεν έχουν σαν σκοπό να παρατηρήσουν τυχόν έκτοπες τιμές στα δεδομένα ή να ερμηνεύσουν μεγάλες ζημιές που εμφανίζονται. Σκοπός τους είναι να δώσουν ένα αποτέλεσμα, ένα νούμερο που να δείχνει τι ποσό πρέπει να έχει κάθε φορά διαθέσιμο η ασφαλιστική. Άρα προκύπτει η ανάγκη ώστε να δημιουργηθούν τρόποι, με τους οποίους θα ανιχνεύονται τέτοιες έκτοπες παρατηρήσεις. Επίσης, εξίσου σημαντικό είναι να βρεθεί ένας κατάλληλος τρόπος αντιμετώπισης αυτών των έκτοπων ή ακραίων τιμών, ούτως ώστε να μην επηρεάζουν την εκτίμηση του αποθέματος. Με άλλα λόγια, να επηρεάζουν όσο το δυνατόν λιγότερο το αποτέλεσμα της Chain Ladder ή όποιας άλλης μεθόδου αποθεματοποίησης χρησιμοποιείται.

Όπως αναφέραμε, τα πραγματικά δεδομένα πολλές φορές έχουν λάθη διαφορετικής φύσεως. Εφαρμόζοντας τις παραπάνω μεθόδους αποθεματοποίησης θα οδηγηθούμε σίγουρα σε λανθασμένη εκτίμηση αποθέματος. Παρακάτω θα παρουσιάσουμε κάποιες μεθόδους με τις οποίες θα εντοπίζονται οι έκτοπες παρατηρήσεις και αφού εντοπιστούν, πώς μπορεί κανείς να τις διαχειριστεί.

Συνήθως, τα προβλήματα που αντιμετωπίζουμε στα δεδομένα είναι τα ακόλουθα:

1. **Ελλιπή δεδομένα**: Να λείπουν δεδομένα από το τρίγωνο εξέλιξης ζημιών είτε στα έτη εξέλιξης είτε στα έτη ατυχήματος.
2. **Λάθος καταχώρηση**: να έχει καταχωρηθεί μία ζημιά λάθος (π.χ. αντί για 10.000€ να καταχωρηθεί 100.000€) και έτσι να διαφέρει δραστικά από το προηγούμενο έτος. Συνήθως αυτά είναι λάθη που διορθώνονται στην επόμενη εκτίμηση του αποθέματος αλλά μέχρι να βρεθούν σίγουρα θα υπάρξει κάποια λάθος μέτρηση του αποθέματος.
3. **Μικρά μεγέθη ζημιών:** αναφέρονται κυρίως σε ζημιές, οι οποίες παίρνουν πολλά χρόνια για να διευθετηθούν. Έτσι, μπορεί σε κάποιο έτος να εμφανίζονται μικρά ποσά ή και μηδενικά.

Πέρα από αυτές τις κατηγορίες λαθών ή έκτοπων τιμών, μία ασφαλιστική έρχεται αντιμέτωπη και με άλλα προβλήματα στα δεδομένα, που επηρεάζουν το αποτέλεσμα του εκτιμώμενου αποθέματος. Τέτοια προβλήματα είναι οι διακυμάνσεις που προκύπτουν μεταξύ των ασφαλιστικών κλάδων μιας ασφαλιστικής, π.χ. κλάδος αυτοκινήτου, κλάδος πυρός, κλάδος υγείας κλπ. Κάθε κλάδος έχει ιδιαιτερότητες ως προς τους χρόνους αποπληρωμής, το ύψος των αποπληρωμών αλλά και τη συχνότητα εκτίμησης του αποθέματος. Η εν λόγω διπλωματική δεν θα ασχοληθεί περαιτέρω με αυτό το κομμάτι, ωστόσο ήταν καίριο να αναφερθεί.

3.5 Έλεγχος Λαθών ή Ακραίων Τιμών

Για τις περιπτώσεις των ελλιπών δεδομένων ή των λάθος καταχωρήσεων είναι απαραίτητη η ανίχνευσή τους αλλά και η σωστή διαχείρισή τους.. Παρόλα αυτά, η ανίχνευση των λαθών δεν είναι τετριμμένη. Ο μελετητής πρέπει να διαχωρίσει τις εύλογα μεγάλες ζημιές (π.χ. μιας φυσικής καταστροφής, μιας επιδημίας κλπ.) από τις έκτοπες τιμές λόγω των παραπάνω σφαλμάτων. Οι Busse et al. (2010) προτείνουν ένα τρόπο αυτόματης ανίχνευσης των έκτοπων παρατηρήσεων, τον οποίον θα παρουσιάσουμε παρακάτω. Ωστόσο, όπως προτείνουν τόσο οι παραπάνω ερευνητές όσο και άνθρωποι της αγοράς, ο αναλογιστής θα κληθεί να αποφασίσει εάν μία μεγάλη ή μικρή ζημιά είναι δικαιολογημένη – και άρα τη λαμβάνει υπόψιν στην εκτίμηση αποθέματος- ή είναι λάθος και άρα διερευνά περαιτέρω το θέμα – είτε αποκλείοντάς την, είτε προσπαθώντας να πάρει ό,τι πληροφορία μπορεί να του δώσει.

Όπως αναφέρθηκε, άνωθεν, μία μεγάλη κατηγορία λαθών σε ένα τρίγωνο εξέλιξης ζημιών είναι οι έκτοπες παρατηρήσεις (ή ακραίες τιμές). Ως τέτοιες ορίζονται μεμονωμένες τιμές, οι οποίες διαφέρουν πάρα πολύ από εκείνες του προηγούμενου ή του επόμενου έτους. Οι τιμές αυτές μπορεί να προκύπτουν είτε από κάποιο λάθος είτε από συγκυριακές περιπτώσεις, πχ μία μεγάλη ζημιά (σεισμός, θεομηνία κλπ) που εμφανίστηκε και η ασφαλιστική πρέπει να την καλύψει.

Ως ένα παράδειγμα, δίνεται το παρακάτω τρίγωνο εξέλιξης ζημιών, για να γίνει πιο κατανοητή η έννοια της ακραίας τιμής.

Τρίγωνο εξέλιξης αθροιστικών ζημιών (σε χιλ.€) χωρίς ακραία τιμή

Πίνακας 3 Τρίγωνο Εξέλιξης Ζημιών χωρίς έκτοπη παρατήρηση

		Έτος Εξέλιξης Ατυχήματος				
		1	**2**	**3**	**4**	**5**
Έτος Αναγγελίας Ατυχήματος	**1**	100	250	420	710	890
	2	120	256	448	687	
	3	145	296	500		
	4	182	323			
	5	158				

Έστω ότι παρατηρείται μία ακραία τιμή στο κελί (3,3). Η τιμή στο κελί, κατά λάθος, καταχωρήθηκε 5.000. Έτσι έχουμε το παρακάτω τρίγωνο:

Πίνακας 4 Τρίγωνο Εξέλιξης Ζημιών με έκτοπη παρατήρηση

		Έτος Εξέλιξης Ατυχήματος				
		1	**2**	**3**	**4**	**5**
Έτος Αναγγελίας Ατυχήματος	**1**	100	250	420	710	890
	2	120	256	448	687	
	3	145	296	5.000		
	4	182	323			
	5	158				

Αν εφαρμόσουμε τη μέθοδο Chain Ladder τόσο στο τρίγωνο του Πίνακα 3 όσο και στου Πίνακα 4 θα δούμε ότι, το απόθεμα υπολογίζεται σε:

Πίνακας 5 Απόθεμα με / χωρίς έκτοπη παρατήρηση

Πίνακας 3	**Πίνακας 4**
2.431 χιλ €	**14.345 χιλ €**

Όπως παρατηρείται, το απόθεμα που πρέπει να κρατήσει μία ασφαλιστική, εξαιτίας μιας έκτοπης παρατήρησης είναι πολύ μεγαλύτερο από εκείνο που θα έπρεπε να κρατήσει, αν τα δεδομένα της ήταν σωστά. Επομένως, είναι ιδιαίτερα σημαντικό να μπορούμε να εντοπίσουμε τις έκτοπες παρατηρήσεις.

Συνήθως, όπως έγινε κατανοητό, τέτοιες ακραίες τιμές προκύπτουν από λάθος πληκτρολόγηση ή λάθος ερμηνεία των ζημιών. Ακολουθώντας το paper των Busse et al. (2010), παρουσιάζουμε την πρότασή τους για την ανίχνευση τέτοιων ακραίων τιμών. Οι εν λόγω ερευνητές πρότειναν να συγκρίνεται η δήθεν έκτοπη ζημιά με τη συνολική ζημιά $\hat{U}_j$ και στη συνέχεια να ονομάζεται ως μίας *προς τα πάνω έκτοπη παρατήρηση* ή ως μία *προς τα κάτω έκτοπη ζημιά* , ανάλογα με τις συνθήκες που πληροί.

- Μία παρατήρηση χαρακτηρίζεται ως *προς τα πάνω έκτοπη παρατήρηση* αν ικανοποιεί και τις δύο ακόλουθες ανισότητες:

$$L_{jk} - L_{j,k-1} \geq a\hat{U}_j \quad and \quad L_{j,k+1} - L_{jk} \leq -a\hat{U}_j. \tag{3.5.1}$$

Μία παρατήρηση χαρακτηρίζεται ως *προς τα κάτω έκτοπη παρατήρηση* αν ικανοποιεί και τις δύο ακόλουθες συνθήκες

$$L_{jk} - L_{j,k-1} \leq -a\hat{U}_j \quad and \quad L_{j,k+1} - L_{jk} \geq a\hat{U}_j. \tag{3.5.2}$$

Η παράμετρος *α* χαρακτηρίζεται ως *κατώφλι* (threshold) για την ανίχνευση της έκτοπης τιμής και προτείνεται να επιλέγεται μεταξύ 10% και 30%. Με άλλα λόγια – και όπως φαίνεται και από τις ανισότητες παραπάνω – δηλώνει ότι για να χαρακτηριστεί μία τιμή ως έκτοπη πρέπει να περνά (ή να μην ξεπερνά) ένα συγκεκριμένο ποσοστό των συνολικών ζημιών για το έτος εξέλιξης j.

Μία άλλη ανωμαλία που μπορεί να υπάρξει στα δεδομένα είναι αυτή των *μεγάλων αλμάτων* . Πρόκειται για μεγάλη απόκλιση της ζημιάς ενός έτος από το προηγούμενο. Η διαφορά με τις άλλες μορφές σφαλμάτων (ελλιπή δεδομένα, λάθη καταχώρισης) είναι ότι τα μεγάλα άλματα συνήθως εμφανίζονται σε ζημιές οι οποίες αργούν να διευθετηθούν. Εδώ αντί για το ύψος της ζημιάς, προτείνεται να χρησιμοποιηθεί ο συντελεστής εξέλιξης f_{jk} (year to year) συγκρινόμενος με τον συντελεστής εξέλιξης της Chain Ladder f_k. Έτσι προκύπτει η ακόλουθη ανισότητα, που πρέπει να πληρείται, ώστε να χαρακτηριστεί μία παρατήρηση ως *μεγάλο άλμα*.

$$f_{jk} \geq b\hat{f}_k. \tag{3.5.3}$$

Πάλι εδώ εμφανίζεται ένα *κατώφλι* συμβολιζόμενο ως b. Εδώ προτείνεται το ποσοστό b να κυμαίνεται μεταξύ 10% και 30%. Πρέπει να επισημάνουμε ότι η ελευθερία στην επιλογή του συντελεστή b (όσο και αυτή στην επιλογή του συντελεστή α) δηλώνει και τα όρια που έχει το φιλτράρισμα των δεδομένων, όπως αυτό συμβαίνει με τις παραπάνω ανισότητες. Πιο συγκεκριμένα, στην περίπτωση των *μεγάλων αλμάτων* είναι αρκετά δύσκολο να αποφασίσει κανείς, αν όντως είναι λάθος μία ζημιά.

Στη συνέχεια και αφού οι ακραίες τιμές εντοπιστούν, προκύπτει η ανάγκη ώστε να διαχειριστεί ο μελετητής (ή ο αναλογιστής) τις έκτοπες παρατηρήσεις. Αυτό επιτυγχάνεται μέσω της ανθεκτικοποίησης της μεθόδου. Σκοπός είναι ο μελετητής να μπορέσει να έχει μία όσο το δυνατόν καλύτερη εκτίμηση για το απόθεμα, δίχως το τελευταίο να επηρεάζεται από τυχόν τιμές του δείγματος που είναι ή φαίνονται ακραίες.

4 Ανθεκτική Εκτίμηση Αποθέματος

Όπως αναφέρθηκε στην προηγούμενη παράγραφο, είναι ιδιαίτερο σημαντικό ο μελετητής να μπορεί να ξεχωρίσει τις έκτοπες παρατηρήσεις (1) σε σχέση με το υπόλοιπο δείγμα και (2) από το αν είναι πραγματικές (μεγάλες ή μικρές ζημιές) ή προέκυψαν από λάθος (πληκτρολόγηση, καταγραφή κλπ). Εφόσον ο μελετητής ξεχωρίσει τις ακραίες τιμές, πρέπει να δει πώς θα τις διαχειριστεί. Οι ασφαλιστικές εταιρίες ακολουθούν συνήθως ακολουθούν την κάτωθι μεθοδολογία:

1. Βάσει στατιστικής μελέτης που ποικίλει από εταιρία σε εταιρία, θέτουμε ένα όριο ως προς τι χαρακτηρίζεται μεγάλη ζημιά (συνήθως οι μεγάλες ζημιές είναι που απασχολούν περισσότερο και όχι οι μικρές). Πχ. Ατομικές ζημιές άνω των 100.000€ ή αθροιστικές ζημιές άνω των 250.000€ θεωρούνται μεγάλες. Βέβαια, αυτό έχει να κάνει και με τον κλάδο ασφάλισης στον οποίον προκύπτει η ζημιά. Για παράδειγμα, μία ζημιά 50.000 για ένα τροχαίο (χωρίς απώλεια ζωής) θεωρείται μεγάλη ενώ αντίστοιχου ύψους ζημιά για ένα σεισμό, δεν θεωρείται μεγάλη.
2. Τις ζημιές που βρίσκονται πάνω από το όριο τις εξαιρούμε.
3. Με τις εναπομείναντες ζημιές φτιάχνει ένα τρίγωνο εξέλιξης ζημιών και εφαρμόζει, όποια μέθοδο αποθεματοποίησης θεωρεί καλύτερη ή επιβάλλεται από τη ρυθμιστική αρχή, από όπου προκύπτει το απόθεμα που θέλει να κρατήσει.
4. Για τις μεγάλες –και μόνο- ζημιές φτιάχνει ένα νέο τρίγωνο εξέλιξης ζημιών και εφαρμόζει την αντίστοιχη μέθοδο αποθεματοποίησης προκειμένου να βγάλει ένα απόθεμα για μεγάλες ζημιές. Από την άλλη είναι συχνό, η ασφαλιστική να διαχειρίζεται με έναν πιο εμπειρικό τρόπο τέτοιες μεγάλες ζημιές, χωρίς να αποσταθεροποιεί τη θέση της εταιρίας.

Κάποιοι ερευνητές (βλ. Busse et al. 2010), προτείνουν την εξάλειψη των ζημιών οι οποίες είναι ακραίες, όταν ψάχνουν να εκτιμήσουν το mse. Χρησιμοποιώντας μία *συνάρτηση φιλτραρίσματος (*filter function) βγάζουν εκτός τις ακραίες τιμές και υπολογίζουν ξανά τους συντελεστές εξέλιξης της Chain Ladder ως σταθμισμένους μέσους των year-to year- συντελεστών. Επίσης, υπολογίζουν τη διακύμανση για κάθε συντελεστή εξέλιξης ως το σταθμισμένο μέσο της απόκλισης του συντελεστή f_{jk} (δηλαδή των year-to – year) από τη μέση τιμή των συντελεστών εξέλιξης. Από την άλλη, οι Verdonck et al. (2009) προτείνουν την ανθεκτικοποίηση της Chain Ladder, εστιάζοντας στους συντελεστές εξέλιξης. Πιο συγκεκριμένα, αντί να δέχονται ως συντελεστή εξέλιξης τη μέση τιμή των year-to-year συντελεστών f_{jk}, προτείνουν τη διάμεσό τους για κάθε έτος ανάπτυξης, η οποία διάμεσος είναι πιο ανθεκτική σε ακραίες παρατηρήσεις - σε σχέση με τη μέση τιμή. Επίσης, οι Gisler and Wuthrich (2008) πρότειναν τον υπολογισμό των συντελεστών εξέλιξης, κάνοντας χρήση της Θεωρίας Αξιοπιστίας. Τέλος οι Pitselis et al. (2015) εκτιμούν το απόθεμα, μέσω λογαριθμοκανονικής παλινδρόμησης, χρησιμοποιώντας ανθεκτικούς εκτιμητές, όπως οι LAD-εκτιμητές, M-εκτιμητές, MLS-εκτιμητές κ.α. (για περισσότερα βλ. Huber (1973) και Hampel et al. (1986)).

4.1 Ανθεκτική Chain Ladder

Όπως αναφέρθηκε και παραπάνω, ένα από τα σημαντικότερα βήματα, για να γίνει πιο ανθεκτική η Chain Ladder, είναι να διευκρινιστεί τι είναι αυτό που κάνει την Chain Ladder τόσο ευάλωτη στις έκτοπες τιμές. Ο ορισμός των συντελεστών εξέλιξης ζημιών, όπως αυτός περιγράφεται στη σχέση (2.2.1), στηρίζεται στις αθροιστικές ζημιές. Έτσι, μία έκτοπη ζημιά πχ. Στην 1η στήλη, σίγουρα δε θα έδινε μία καλή εκτίμηση για το τελικό απόθεμα. Από την άλλη, αν κάποιος δούλευε με τις ατομικές ζημιές, τότε μία έκτοπη τιμή θα επηρέαζε το πολύ δύο συντελεστές εξέλιξης ζημιών.

Μία πρόταση για την ανθεκτικοποίηση της Chain Ladder ήρθε από τους Verdonck *et al.* (2009). Πιο συγκεκριμένα, πρότειναν μία ανθεκτικοποίηση των συντελεστών εξέλιξης ζημιών. Η ιδέα έγκειται στον τρόπο υπολογισμού των συντελεστών. Αντί κάποιος να διαιρεί το άθροισμα της στήλης i+1 με το άθροισα της στήλης i, οι Verdonck *et al.* (2009) πρότειναν να εστιάσει κανείς στους λόγους των στηλών, για κάθε γραμμή, όπως φαίνεται παρακάτω:

$$\left\{ \frac{L_{ij}}{L_{i,j-1}} \mid i = 1,\dots,n-j+1,\ j = 2,\dots n \right\}. \tag{4.1.1}$$

Στη συνέχεια, λαμβάνοντας κανείς τη μέση τιμή των λόγων αυτών για κάθε στήλη, θα μπορούσε να έχει -προσεγγιστικά- τον ίδιο συντελεστή εξέλιξης ζημιών για κάθε έτος, με αυτόν που δίνει η σχέση (2.1.1). Οι Hampel *et al.* (1986) απέδειξαν -με τη χρήση της συνάρτησης επιρροής- ότι η μέση τιμή ως στατιστικό μέγεθος είναι αρκετά ευαίσθητη απέναντι σε έκτοπες τιμές. Παρατηρώντας τη σχέση (2.1.1) βλέπει κανείς ότι οι συντελεστές εξέλιξης ζημιών μπορούν να θεωρηθούν ως μία μέση τιμή, πράγμα που εξηγεί και την ευαισθησία της Chain Ladder σε μία έκτοπη τιμή. Για να λυθεί αυτό το πρόβλημα, οι Verdonck *et al.* (2009) πρότειναν να αντικατασταθεί η μέση τιμή από μία πιο ανθεκτική εκτίμηση. Τη λύση εδώ έδωσαν στις δημοσιεύσεις τους οι Huber (1981), Hampel *et al.* (1986), Rousseeuw και Leroy (1987), αντικαθιστώντας την μέση τιμή με τη διάμεσο, η οποία δίνει μια πιο ανθεκτική εκτίμηση, σε σχέση με τη μέση τιμή. Εν αντιθέσει με τη μέση τιμή, η διάμεσος δεν επηρεάζεται από έκτοπες τιμές.

Με το σκεπτικό αυτό, οι Verdonck *et al.* (2009) πρότειναν να χρησιμοποιηθεί η διάμεσος (εφαρμοσμένη πάνω στα ατομικά δεδομένα) η οποία δίνει την ακόλουθη έκφραση για τους συντελεστές εξέλιξης ζημιών:

$$\widehat{f}_j = median\left\{ \frac{L_{ij}}{L_{i,j-1}} \mid i=1,\ldots,n-j+1 \right\}, \quad 2 \le j \le n. \qquad (4.1.2)$$

Εφαρμόζοντας τη σχέση (4.1.2) στο τρίγωνο εξέλιξης ζημιών φαίνεται να δουλεύει αρκετά καλά, καθώς το απόθεμα που προκύπτει είναι πολύ κοντά, σχεδόν το ίδιο, με το απόθεμα που προκύπτει αν εφαρμοστεί η κλασσική Chain Ladder, σε δεδομένα χωρίς έκτοπη παρατήρηση, όπως θα φανείς στο επόμενο κεφάλαιο όπου παρουσιάζονται οι εφαρμογές των μεθόδων.

Παρόλα αυτά, κάποια προβλήματα παραμένουν. Για παράδειγμα, μία πολύ μεγάλη ζημιά στην προτελευταία στήλη, ακόμη και με χρήση της σχέσης (4.1.2) θα επηρέαζε σημαντικά το τελικό απόθεμα. Συγκεκριμένα, στο σημείο αυτό, θα είχαμε δύο συντελεστές εξέλιξης ζημιών, όπου η διάμεσος και η μέση τιμή ταυτίζονται. Επίσης, στη στήλη αυτή έχουμε δύο ζημιές, πράγμα που καθιστά δύσκολο να επιλέξει κανείς ποια είναι η έκτοπη και ποια η λογική.

Ο λόγος, που η πιο ανθεκτική μορφή των συντελεστών εξέλιξης αδυνατεί να δώσει ένα καλό αποτέλεσμα στην περίπτωση αυτή, είναι ότι η έκτοπη ζημιά πάλι χρησιμοποιείται για να εκτιμηθεί ο συντελεστής εξέλιξης του τελευταίου έτους. Όπως προκύπτει, η ανθεκτικοποίηση στον υπολογισμό των συντελεστών εξέλιξης δεν είναι αρκετή. Για το λόγο αυτό, οι Verdonck et al. (2009) πρότειναν έναν αλγόριθμο, για να προσαρμόσουν τις έκτοπες τιμές.

Ο αλγόριθμος που πρότειναν οι Verdonck et al. (2009), προκειμένου να βρουν -ανάμεσα στα δεδομένα ενός τριγώνου εξέλιξης ζημιών- την /τις έκτοπη/-ες τιμή/-ες στηρίχθηκε στις υποθέσεις που έκαναν οι England και Verall (1999) και οι οποίες φαίνονται στη σχέση (3.2.1). Ουσιαστικά, πρόκειται για τη βασική εκδοχή της Bootstrap μεθόδου (όπου τα υπόλοιπα Pearson είναι τώρα ως προς τις ατομικές ζημιές C_{ij}) και είναι αυτή πάνω στην οποία στηρίχθηκαν και οι Busse et al. (2010) όπως παρουσιάστηκε σε προηγούμενη παράγραφο.

4.2 Περιγραφή Αλγορίθμου

Βήμα 1*: Ο*ι συντελεστές εξέλιξης ζημιών υπολογίζονται από την σχέση, στηριζόμενοι στις αθροιστικές ζημιές, (4.1.2). Βέβαια, θα μπορούσε κανείς να επιλέξει αν επιλέξει τις ατομικές ζημιές (από όπου οι συντελεστές εξέλιξης μπορεί να προκύψουν και πιο ανθεκτικοί) ωστόσο η προσέγγιση που ακολουθείται δουλεύει μόνο στα αθροιστικά δεδομένα. Οι ζημιές που φαίνονται πάνω στην κύρια διαγώνιο αποτελούν τις πραγματικές πληρωθείσες ζημιές, όπως προαναφέρθηκε. Έτσι δομείται ένα νέο τρίγωνο εξέλιξης ζημιών (fitted cumulative triangle), όπου έχει την ίδια κύρια διαγώνιο με το τρίγωνο των αρχικών μας δεδομένων. Οι ζημιές που αναφέρονται στις θέσεις πάνω και αριστερά από την κύρια διαγώνιο (πάντα σε αθροιστική μορφή) υπολογίζονται με αναδρομικό τρόπο, διαιρώντας τις ζημιές στο έτος t με τον συντελεστή εξέλιξης t-1. Προφανώς, οι ατομικές ζημιές m προκύπτουν, αφαιρώντας τις ζημιές στο έτος εξέλιξης t-1 από τις ζημιές στο έτος εξέλιξης t.

Στη συνέχεια, τα υπόλοιπα Pearson προκύπτουν ως εξής:

$$r_{ij} = \frac{C_{ij} - m_{ij}}{\sqrt{\phi m_{ij}}}, \tag{4.2.1}$$

όπου τα C_{ij} είναι οι ατομικές ζημιές, φ είναι μία παράμετρος κλίμακας και m είναι οι ατομικές ζημιές του νέου τριγώνου (fitted triangle). Στο τρίγωνο που, πλέον, αποτελείται από τα υπόλοιπα Pearson η πάνω δεξιά γωνία όσο και η κάτω αριστερά γωνία έχουν τιμή 0.

Βήμα 2*:* Προκειμένου να βρεθεί η έκτοπη τιμή χρησιμοποιείται το εργαλείο θηκογράμματος (boxplot) όπως αυτό παρουσιάστηκε από τον Tukey (1977), πάνω στο τρίγωνο των υπολοίπων. Τα υπόλοιπα των έκτοπων παρατηρήσεων θα βρίσκονται εκτός του του διαστήματος:

$$\left[Q_1 - 3IQR, Q_3 + 3IQR\right],$$

όπου τα Q_1, Q_3 είναι το πρώτο και το τρίτο τεταρτημόριο. Ο κανόνας αυτός προϋποθέτει κανονικότητα των υπολοίπων, κάτι το οποίο μπορεί να αποδειχθεί εφαρμόζοντας το τεστ Shapiro-Wilk.

Βήμα 3*:* Όταν, στο τρίγωνο των υπολοίπων παρατηρηθεί μία έκτοπη τιμή στην πρώτη στήλη τότε η αντίστοιχη ζημιά θεωρείται ως έκτοπη παρατήρηση. Αν τώρα, η ζημιά στην αμέσως επόμενη στήλη (και στην ίδια γραμμή) χαρακτηριστεί επίσης ως έκτοπη, τότε η ζημιά στην πρώτη στήλη αντικαθίσταται από τη διάμεσο των ζημιών της πρώτης στήλης (σε κάποιες περιπτώσεις μπορεί να αντικατασταθεί και με τη μέση τιμή). Σε κάθε άλλη περίπτωση, διαιρείται η ζημιά στην αμέσως επόμενη στήλη με έναν ανθεκτικό συντελεστή εξέλιξης και αυτό που προκύπτει, παίρνει την θέση της έκτοπης τιμής στη στήλη αυτή. Προφανώς, θα δημιουργηθεί μία άλλη τιμή για το αντίστοιχο υπόλοιπο Pearson αλλά δεν λαμβάνεται υπόψιν.

Βήμα 4*:* Για να εντοπιστούν έκτοπες τιμές στις άλλες στήλες, η προσοχή στρέφεται ξανά στα υπόλοιπα Pearson αλλά τώρα τα δεδομένα από νέο τρίγωνο (fitted triangle) υπολογίζονται με διαφορετικό τρόπο. Από το προηγούμενο βήμα είναι γνωστό ότι η πρώτη στήλη είναι καθαρή από ακραίες τιμές. Έτσι, υπολογίζονται ξανά οι συντελεστές εξέλιξης ζημιών αλλά χρησιμοποιώντας τις ατομικές ζημιές της πρώτης στήλης, όπως δείχνει η παρακάτω σχέση:

$$\widehat{f}_j^1 = median\left\{ \frac{C_{ij}}{C_{i1}} \middle| i = 1,2,...,n-j+1 \right\}, \; 2 \le j \le n. \qquad (4.2.2)$$

Κατά συνέπεια, οι μελλοντικές ζημιές εκτιμώνται ως:

$$\widehat{C}_{ij}^1 = C_{i1}\widehat{f}_j^1, \quad 2 \le i \le n, \quad n-i+2 \le j \le n. \qquad (4.2.3)$$

Επίσης, το πάνω μέρος του τριγώνου αναπροσαρμόζεται, πολλαπλασιάζοντας τις ζημιές στην πρώτη στήλη με τους νέους συντελεστές εξέλιξης ζημιών (4.1.4), έτσι ώστε να δημιουργηθεί ένα νέο σύνολο ατομικών δεδομένων. Υπολογίζονται ξανά τα υπόλοιπα Pearson σύμφωνα με τη σχέση (4.1.3), όπου τη θέση του m_{ij} παίρνει η εκτίμηση $\widehat{C}_{ij}^1$. Με τον τρόπο αυτό, μία έκτοπη ζημιά επηρεάζει μόνο το αντίστοιχο υπόλοιπο Pearson. Είναι προφανές, ότι στο νέο τρίγωνο υπολοίπων Pearson, η πρώτη στήλη είναι 0. Τα εναπομείναντα υπόλοιπα μπορούν να εξεταστούν, για τυχόν έκτοπες παρατηρήσεις, χρησιμοποιώντας πάλι το θηκόγραμμα (boxplot). Σε περίπτωση, που ένα υπόλοιπο βρίσκεται εκτός του διαστήματος που αναφέρθηκε στο Βήμα 2, τότε το υπόλοιπο αυτό αντικαθίσταται από τη διάμεσο των υπολοίπων. Αντιστρέφοντας τη σχέση των υπολοίπων Pearson (4.1.3) προκύπτουν οι τελικές παρατηρήσεις του τριγώνου εξέλιξης ζημιών, οι οποίες αναπαρίστανται ως X_{ij}^r.

Βήμα 5: Αφότου έχει αντιστραφεί η σχέση των υπολοίπων Pearson, όπως αναφέρθηκε στο τέλος του Βήματος 4, δημιουργείται ένα τρίγωνο εξέλιξης ζημιών των X_{ij}^r. Σε αυτό το τρίγωνο, μπορεί τώρα να εφαρμοστεί η μέθοδος Chain Ladder.

Αξίζει να αναφερθεί ότι, με τη μεθοδολογία που παρουσιάστηκε πιο πάνω, δεν είναι δυνατόν να εντοπιστούν ακραίες τιμές, αν οι τελευταίες βρίσκονται στις γωνίες του τριγώνου ζημιών, δηλαδή στις θέσεις X_{1n} και X_{n1}. Για τη δεύτερη περίπτωση, μπορεί κανείς να υπολογίσει τη διάμεσο των ζημιών της πρώτης στήλης και να δει αν η ζημιά στη θέση X_{n1} διαφέρει σημαντικά από τη διάμεσο. Για την πρώτη περίπτωση, έχει προταθεί από τους Verdonck et al. (2009) η χρήση μίας καμπύλης εκτίμησης, με χρήση των συντελεστών εξέλιξης.

4.3 Ανθεκτικό Λογαριθμοκανονικό Μοντέλο

Στην ενότητα αυτή θα παρουσιαστούν τα βήματα που απαιτούνται για να πάρουμε μία ανθεκτική εκτίμηση του αποθέματος, χρησιμοποιώντας την Ανάλυση Παλινδρόμησης για τους Μ-εκτιμητές. Θα παρουσιαστεί ο αλγόριθμος, με τον οποίον παίρνουμε ανθεκτικό εκτιμητή για το διάνυσμα β, έτσι όπως τον εισήγαγαν πρώτοι οι Huber and Dutter (1974) και χρησιμοποιήθηκε και από τους Pitselis et al. (2015). Ο αλγόριθμος χωρίζεται σε 4 βήματα:

Βήμα 1: Σκοπός είναι να προκύψει ένας ανθεκτικός εκτιμητής $\widehat{\beta}^{\mathrm{M}}$ για το διάνυσμα β της σχέσης (2.3.8). Αυτό επιτυγχάνεται με την ελαχιστοποίηση της ακόλουθης σχέσης:

$$\min_{\widehat{\beta}} \sum_{t=1}^{T} \rho\left(\frac{r_t}{S}\right). \qquad (4.3.1)$$

Η σχέση (4.3.1) χρησιμοποιείται για να βρεθεί το $\widehat{\beta}^{\mathrm{M}}$, ένας ανθεκτικός εκτιμητής των β, S,με τη διτετραγωνική συνάρτηση ρ του Tukey που ορίζεται ως:

$$\rho(t)=\begin{cases}\left(\frac{t}{c}\right)^6-3\left(\frac{t}{c}\right)^4+3\left(\frac{t}{c}\right)^2 & \gamma\iota\alpha\ \ |t|\le c\\ 1 & \gamma\iota\alpha\ \ |t|>c.\end{cases} \qquad (4.3.2)$$

Η λύση της ελαχιστοποιούμενης συνάρτησης είναι αντίστοιχη της λύσης των παρακάτω εξισώσεων:

$$\begin{gathered}\sum_{t=1}^{T}\psi\left(\frac{r_t}{S}\right)\sum_{t=1}^{T}x_{kt}=0, \quad k=1,2,\ldots,p\\ \kappa\alpha\iota\\ \sum_{t=1}^{T}\chi\left(\frac{r_t}{S}\right)=a,\end{gathered} \qquad (4.3.3)$$

όπου $\psi(u)=\rho'(u)$ και $\chi(u)=u\psi(u)-\rho(u)$. Αν χρειάζεται το S να είναι ασυμπτωτικά αμερόληπτο για κανονικά κατανεμημένα σφάλματα, τότε το $a=\left[(n-p)/n\right]E_{\varphi}(\chi)$ με Φ να είναι η κανονική κατανομή.

Βήμα 2: Υπολογίζεται ο ανθεκτικός πίνακας διακύμανσης του $\widehat{\beta}$ ως εξής:

$$Cov\left(\beta^{\mathrm{M}}\right)=\widehat{\sigma}_{\mathrm{M}}^2\left(\mathrm{X}^{\mathrm{T}}\mathrm{X}\right)^{-1}, \qquad (4.3.4)$$

με

$$\widehat{\sigma}_{\mathrm{M}}^2=L^2\frac{1}{(t-p)}\frac{\sum_{i=1}^{t}\psi\left(\frac{r_i}{S}\right)^2 S^2}{\left[(1/t)\sum_{i=1}^{t}\psi'\left(\frac{r_i}{S}\right)\right]^2}. \qquad (4.3.5)$$

Να σημειωθεί ότι το L είναι ένας παράγοντας διόρθωσης για την μεροληψία (βλ. Huber, 1973) και ορίζεται ως εξής:

$$L = 1 + \frac{p}{T} \frac{Var\left[\psi'\left(\frac{r_i}{S}\right)\right]}{\mathrm{E}\left[\psi'\left(\frac{r_i}{S}\right)\right]^2}. \qquad (4.3.6)$$

Στην ουσία, οι ποσότητες $\mathrm{E}\left[\psi'\left(\frac{r_i}{S}\right)\right]$ και $Var\left[\psi'\left(\frac{r_i}{S}\right)\right]$ είναι άγνωστες και θα εκτιμηθούν από τις ακόλουθες σχέσεις:

$$\mathrm{E}\left[\psi'\left(\frac{r_i}{S}\right)\right] \simeq \frac{1}{t}\sum_{i=1}^{T} \psi'\left(\frac{r_i}{S}\right) = m, \qquad (4.3.7)$$

$$Var\left[\psi'\left(\frac{r_i}{S}\right)\right] \simeq \frac{1}{t}\sum_{i=1}^{T} \left[\psi'\left(\frac{r_i}{S}\right) - m\right]^2. \qquad (4.3.8)$$

Επίσης, στην ειδική περίπτωση όπου $\psi(u) = \min\left[c, \max(-c, u)\right]$ τότε η σχέση (4.3.6) γίνεται:

$$L = 1 + \frac{1}{t}\frac{(1-m)}{m}. \qquad (4.3.9)$$

Το m αποτελεί τη σχετική συχνότητα των υπολοίπων που ικανοποιούν την ανισότητα $-c < \frac{r_i}{S} < c.$

Βήμα 3: Ο ανθεκτικός Μ-εκτιμητής του $\hat{\theta}_{ij}$ (ή αλλιώς της αναμενόμενης τιμής $\mathrm{E}\left(\hat{P}_{ij}\right)$) υπολογίζεται με το αντικαταστήσουμε τις παραμέτρους της σχέσης (2.3.11) με τις ανθεκτικές τους μορφές, δηλαδή:

$$\mathrm{E}\left(\hat{P}_{ij}\right) = \exp\left(x_{ij}^T \beta^{\mathrm{M}}\right) g_m\left[\frac{1}{2}\left(1 - x_{ij}^T \left(X^T X\right)^{-1} x_{ij}\right)\sigma_{\mathrm{M}}^2\right]. \qquad (4.3.10)$$

Επιπρόσθετα, παίρνοντας υπόψιν τη σχέση (2.2.14), η ανθεκτική διακύμανση του ασυμπτωτικά αμερόληπτου εκτιμητή τ_{ij}^2 υπολογίζεται ως εξής:

$$\tilde{\tau}_{\mathrm{M}_{ij}}^2 = \exp\left(2x_{ij}^T \beta^{\mathrm{M}}\right)\left[\begin{array}{r}\left(g_m\left(\frac{1}{2}\left(1 - x_{ij}^T\left(X^T X\right)^{-1} x_{ij}\right)s_{\mathrm{M}}^2\right)\right)^2 \\ -g_m\left(\left(1 - 2x_{ij}^T\left(X^T X\right)^{-1} x_{ij}\right)s_{\mathrm{M}}^2\right)\end{array}\right]. \qquad (4.3.11)$$

4.4 Ανθεκτική Μέθοδος Mαck (Robust Mack Method)

Θα επικεντρώσουμε την προσοχή μας στο πρόβλημα των *μεγάλων αλμάτων* και θα περιγράψουμε έναν μετασχηματισμό των μεθόδων Mack και Bootstrap ώστε η εκτίμηση του μέσου τετραγωνικού σφάλματος να γίνει ανθεκτική και να μην επηρεάζεται από ακραίες τιμές.

Τα *μεγάλα άλματα* οδηγούν σε πολύ μεγάλες τιμές όταν υπολογίζεται η διακύμανση. Ένας τρόπος για να αντιμετωπιστούν είναι να δημιουργηθεί ένας ανθεκτικός εκτιμητής.

Ο εκτιμητής της διακύμανσης της σχέσης (3.1.13) μπορεί να ξαναγραφτεί ως:

$$\widehat{\sigma}_k^2 = \frac{1}{N-k-1}\sum_{j=1}^{N-k}\frac{1}{L_{jk}}\left(L_{j,k+1} - \widehat{f}_k L_{jk}\right)^2. \qquad (4.4.1)$$

Όπως φαίνεται για $L_{jk}=0$ ο εκτιμητής απειρίζεται. Από τη στιγμή που η πραγματοποιθείσα ζημιά βρίσκεται στον παρονομαστή, η εκτίμηση του Mack είναι πολύ ευαίσθητη σε μικρές τιμές του L_{jk} αλλά και σε σφάλματα αυτών. Για να γίνει ο εκτιμητής πιο ανθεκτικός, προτείνεται (βλ. Busse et al. (2010)) να αντικατασταθεί ο παρονομαστής από την αναμενόμενη τιμή του. Έτσι, η (4.4.1) γίνεται:

$$(\widehat{\sigma}_k^2)^r = \frac{1}{N-k-1}\sum_{j=1}^{N-k}\frac{1}{E(L_{jk})}\left(L_{j,k+1} - \widehat{f}_k L_{jk}\right)^2. \qquad (4.4.2)$$

Η αλλαγή αυτή, είναι λογικό να επηρεάζει την αμεροληψία του εκτιμητή. Ωστόσο αυτό θα αναλυθεί παρακάτω. Στην πράξη, η θεωρητική τιμή $E(L_{jk})$ είναι αδύνατον να υπολογιστεί. Αυτό σημαίνει ότι πρέπει να εισάγουμε μία κατάλληλη εκτίμηση του $E(L_{jk})$. Σύμφωνα με τους England and Verall (2002), η $E(L_{jk})$ μπορεί να υπολογιστεί με αντίστροφη αναδρομή, ξεκινώντας από την τελευταία πραγματοποιθείσα ζημιά, στην τελευταία διαγώνιο. Δηλαδή:

$$E(L_{j,N-j+1}) = L_{j,N-j+1}. \qquad (4.4.3)$$

Ενώ από εκεί και πέρα ακολουθείται η αντίστροφη αναδρομή.

$$\widehat{E}(L_{j,N-j+1}) = \widehat{E}(L_{j,k-1})\,\widehat{f}_{k-1}^{-1},\ k \le N-j+1. \qquad (4.4.4)$$

Έχει αποδειχθεί ότι αυτός ο εκτιμητής είναι ίδιος με τον $E(L_{jk}|L_f)$:

$$E(L_{jk}\left|L_f\right.) = L_{j,N+1} f_{N-1}^{-1}\cdot\ldots\cdot f_k^{-1},\ k \le N-j+1. \qquad (4.4.5)$$

ο οποίος αποτελεί την καλύτερη πρόβλεψη για το L_{jk} δοσμένης της μελλοντικής τιμής

$$L_f = \left\{L_{jk}\left|j+k \ge N+1\right.\right\}. \qquad (4.4.6)$$

Οι ζημιές τις τελευταίας διαγώνιου καθώς και οι συντελεστές εξέλιξης της Chain Ladder, θεωρητικά μιλώντας, επηρεάζονται λιγότερο από τυχόν λάθη στα δεδομένα σε σχέση με τις ατομικές ζημιές L_{jk} των πρώτων ετών. Επιπλέον, η αναμενόμενη τιμή $E(L_{jk})$, του παρονομαστή, είναι πιο ανθεκτική από ότι οι ατομικές ζημιές L_{jk}. Επομένως και ο εκτιμητής $(\sigma_k^2)^r$ είναι ανθεκτικός – υπό την έννοια ότι δεν επηρεάζεται από κάποια έκτοπη παρατήρηση – και έτσι μπορεί να αντικαταστήσει τον εκτιμητή της σχέσης (3.1.13).

Όπως προαναφέραμε, οφείλουμε να εξετάσουμε την μεροληψία του εκτιμητή προκειμένου να δούμε, αν η ανθεκτικοποίησή του έπαιξε κάποιο ρόλο ώστε να πάψει να είναι αμερόληπτος.

Η αναμενόμενη τιμή δίδεται ως:

$$\mathrm{E}((\sigma_k^2)^r) = \frac{1}{N-k-1}\sum_{j=1}^{N-k}\frac{E(L_{j,k+1}^2 - 2\widehat{f}_k L_{jk}L_{j,k+1} + \widehat{f}_k^{\,2} L_{jk}^2)}{E(L_{jk})}. \qquad (4.4.7)$$

Ο πρώτος όρος της (4.4.7) στον αριθμητή, γίνεται:

$$\begin{aligned} E(L_{j,k+1}^2) &= E(E(L_{j,k+1}^2 \mid L_{j1},...,L_{jk})) \\ &= E(Var(L_{j,k+1} \mid L_{j1},...,L_{jk}) + E(L_{j,k+1} \mid L_{j1},...,L_{jk})^2) \qquad (4.4.8) \\ &= \sigma_k^2 E(L_{jk}) + f_k^2 E(L_{jk}^2), \end{aligned}$$

όπου χρησιμοποιείται η συνάρτηση (3.2.2) από τον ορισμό της μεθόδου Mack.

Για τον δεύτερο όρο στον αριθμητή χρησιμοποιούμε την ακόλουθη υπόθεση, $L_k = \{L_{jk} \mid j \le k,\ j \le N+1-i\},\ 1 \le k \le N$

και προκύπτει ότι:

$$\mathrm{E}(\widehat{f}_k L_{jk} L_{j,k+1}) = E\left(L_{jk} E\left[L_{j,k+1} \frac{\sum_{l=1}^{N-k} L_{l,k+1}}{\sum_{l=1}^{N-k} L_{l,k}} \middle| L_k \right] \right)$$

$$= E\left(\frac{L_{jk}}{\sum_{l=1}^{N-k} L_{lk}} E\left[L_{j,k+1} \sum_{l=1,\, l\neq j}^{N-k} L_{l,k+1} + L_{j,k+1}^2 \middle| L_{jk} \right] \right) \qquad (4.4.9)$$

$$= E\left(\frac{L_{jk}}{\sum_{l=1}^{N-k} L_{lk}} f_k L_{jk} \sum_{l=1,\, l\neq j}^{N-k} f_k L_{lk} \right) + E\left(\frac{L_{jk}}{\sum_{l=1}^{N-k} L_{lk}} \left[\sigma_k^2 L_{jk} + f_k^2 L_{jk}^2 \right] \right)$$

$$= f_k^2 E(L_{jk}^2) + \sigma_k^2 E\left(\frac{L_{jk}^2}{\sum_{l=1}^{N-k} L_{lk}} \right).$$

Ενώ ο τρίτος όρος γίνεται:

$$\mathrm{E}(\widehat{f}_k^2 L_{jk}^2) = E\left(L_{jk}^2 \left[Var(\widehat{f}_k \mid L_k) + E^2(\widehat{f}_k \mid L_k) \right] \right)$$

$$= E\left(L_{jk}^2 \left[\frac{\sum_{j=1}^{N-k} Var(L_{j,k+1} \mid L_k)}{\sum_{j=1}^{N-k} (L_{jk})^2} + f_k^2 \right] \right) \qquad (4.4.10)$$

$$= \sigma_k^2 E\left(\frac{L_{jk}^2}{\sum_{l=1}^{N-k} L_{lk}} \right) + f_k^2 E(L_{jk}^2).$$

Τώρα η αν αντικαταστήσουμε τις (4.4.8) – (4.4.10) στην (4.4.7) παίρνουμε ότι:

$$\mathrm{E}\left((\widehat{\sigma}_k^2)^r \right) = \frac{1}{\mathrm{N}-k-1} \sum_{j=1}^{N-k} \frac{1}{E(L_{jk})} \left[\sigma_k^2 E(L_{jk}) - \sigma_k^2 E\left(\frac{L_{jk}^2}{\sum_{l=1}^{N-k} L_{lk}} \right) \right] \qquad (4.4.11)$$

$$= \sigma_k^2 + \frac{\sigma_k^2}{\mathrm{N}-k-1} \left[1 - \sum_{j=1}^{N-k} \frac{1}{E(L_{jk})} E\left(\frac{L_{jk}^2}{\sum_{l=1}^{N-k} L_{jk}} \right) \right].$$

Βλέπουμε, λοιπόν, ότι ο εκτιμητής $(\widehat{\sigma}_k^2)^r$ έχει μία μεροληψία που δεν μπορούμε να εξαλείψουμε:

$$\mathrm{B}_{\left(\hat{\sigma}_k^2\right)^r} = E\left(\left(\hat{\sigma}_k^2\right)^r\right) - \sigma_k^2$$

$$= \frac{\sigma_k^2}{N-k-1}\left[1 - \sum_{j=1}^{N-k}\frac{1}{E(L_{jk})}E\left(\frac{L_{jk}^2}{\sum_{l=1}^{N-k}L_{lk}}\right)\right]. \qquad (4.4.12)$$

Για να απλοποιήσουμε την σχέση (4.4.12) χρησιμοποιούμε την παρακάτω παραδοχή

$$E\left(\frac{L_{jk}^2}{\sum_{l=1}^{N-k}L_{lk}}\right) \simeq \frac{E\left(L_{jk}^2\right)}{E\left(\sum_{l=1}^{N-k}L_{lk}\right)}. \qquad (4.4.13)$$

Η σχέση (4.4.13) ισχύει στην περίπτωση που οι ζημιές L_{jk} είναι ανεξάρτητες και ισόνομες ακολουθώντας την κανονική κατανομή με μέση τιμή μ_k και διακύμανση σ_k^2. Η υπόθεση αυτή υποδηλώνει ότι το σφάλμα που προκύπτει από την παραδοχή στη σχέση (4.4.13) υπολογίζεται ως:

$$\text{Error}= \frac{2}{N-k}\frac{\sigma_k^2}{\mu_k^2 + \sigma_k^2}. \qquad (4.4.14)$$

Λόγου χάρη το σφάλμα της σχέσης (4.4.14) είναι μικρότερο από 4% για N-k ≥ 3 και σ_k/μ_k=0.25 ή είναι μικρότερο του 6% για N-k ≥ 5 και σ_k/μ_k=0.4. Ακόμη, από την (4.4.12) παίρνουμε το ακόλουθο:

$$\mathrm{B}_{\left(\hat{\sigma}_k^2\right)^r} \approx \frac{\sigma_k^2}{N-k-1}\left[1 - \sum_{j=1}^{N-k}\frac{Var(L_{jk}) + E^2(L_{jk})}{E(L_{jk})E\left(\sum_{l=1}^{N-k}L_{lk}\right)}\right] \qquad (4.4.15)$$

$$= \frac{-\sigma_k^2}{N-k-1}\sum_{j=1}^{N-k}\frac{Var(L_{jk})}{E(L_{jk})\sum_{l=1}^{N-k}E\left(L_{lk}\right)}.$$

Η τάξη μεγέθους της σχέσης μπορεί αν εκτιμηθεί, θεωρώντας ότι οι διακυμάνσεις της ζημιάς L_{jk} φράσσονται από ένα ποσοστό της αναμενόμενης τιμής $E(L_{jk})$,

$$\left[Var(L_{jk})\right]^{1/2} \leq \mu \mathrm{E}(L_{jk}). \qquad (4.4.16)$$

Αντικαθιστώντας τη σχέση (4.4.16) στη σχέση (4.4.15) προκύπτει ότι:

$$\frac{|\mathrm{B}|}{\sigma_k^2} \leq \frac{\mu^2}{\mathrm{N} - \kappa - 1}. \qquad (4.4.17)$$

Από τα παραπάνω προκύπτει ότι η μεροληψία (4.4.17), για τα πρώτα έτη εξέλιξης k (k<<N) είναι αμελητέα. Για τα μετέπειτα έτη εξέλιξης k, όπου k>>1 τότε η μεροληψία μπορεί να αξίζει προσοχής ωστόσο η συμβολή της στο τελικό λάθος της μεθόδου Mack είναι μικρή και άρα μπορούμε να την θεωρήσουμε αποδεκτή. Βέβαια, μπορεί να υποστηριχθεί, ότι αυτό που εμφανίζεται ως μεροληψία, στην πραγματικότητα είναι μία «διόρθωση» περνώντας από την κλασσική στην ανθεκτική εκτίμηση (Busse et al. 2010).

4.5 Ανθεκτική Μέθοδος Bootstrap (Robust Bootstrapping)

Η μέθοδος Bootstrap, εφαρμοσμένη στα πλαίσια της μεθόδου Mack, έχει μία παρόμοια ευαισθησία στις μικρές τιμές ζημιών. Αυτό, φυσικά, δεν αποτελεί έκπληξη από τη στιγμή που ο εκτιμητής της διακύμανσης στη μέθοδο Mack είναι μέρος του αλγορίθμου της μεθόδου Bootstrap. Όπως και στην προηγούμενη παράγραφο, προκειμένου να γίνει ανθεκτική η μέθοδος, γίνεται χρήση του ανθεκτικού εκτιμητή της σχέσης (4.5.2), αντί για τον εκτιμητή του μοντέλου. Με άλλα λόγια, αντικαθίσταται το σ_k^2 με το $(\hat{\sigma}_k^2)^r$ στις σχέσεις (4.4.2), (4.4.3), (4.4.4). Η σχέση (4.4.2) που αφορά τα υπόλοιπα, μπορεί τώρα να γραφτεί ως:

$$r_{jk} = \frac{L_{j,k+1} - \hat{f}_k L_{jk}}{\hat{\sigma}_k \sqrt{L_{jk}}}. \qquad (4.5.1)$$

Όπως φαίνεται, για L_{jk}=0 μπορεί να απειρίζεται. Για να αποκλειστεί ένα τέτοιο σενάριο, προτείνεται ο μετασχηματισμός της σχέσης (4.4.2), πριν εφαρμοστεί η μέθοδος Bootstrap. Έτσι η (4.4.2) γίνεται:

$$r_{jk}^{'} = r_{jk} \sqrt{\frac{L_{jk}}{\hat{E}(L_{jk})}} \qquad (4.5.2)$$

Η αναμενόμενη τιμή της (4.4.2), $\hat{E}(L_{jk})$, προκύπτει από τις σχέσεις (4.4.5). Αφού γίνει η επαναδειγματοληψία (resampling), τότε τα Bootstrapped υπόλοιπα μετασχηματίζονται σε :

$$r_{jk}^{*} = r_{j'k'}^{'} \sqrt{\frac{\hat{E}(L_{jk})}{L_{jk}}} \qquad (4.5.3)$$

Όπως θα δείξουμε και στην αριθμητικό εφαρμογή, η ανθεκτική Bootstrap μέθοδος οδηγεί σε παρόμοια εκτίμηση για τον κίνδυνο του αποθέματος με τη μέθοδο Mack.

5 Εφαρμογή των Μεθόδων

Στο κεφάλαιο αυτό θα παρουσιάσουμε την εφαρμογή των μεθόδων, όπως αυτές παρουσιάστηκαν στα προηγούμενα κεφάλαια. Ως δεδομένα, θα χρησιμοποιήσουμε ένα τρίγωνο εξέλιξης ζημιών, που χρησιμοποιείται ευρέως στην επιστημονική κοινότητα. Είναι τα δεδομένα που χρησιμοποίησαν πρώτοι οι Taylor and Ashe (1983) και τα οποία παραθέτουμε στον παρακάτω πίνακα (Πίνακας 4) σε αθροιστικές ζημιές. Στη συνέχεια, θα παρουσιάσουμε τα αποτελέσματα με την ακόλουθη σειρά:

- Chain Ladder χωρίς έκτοπη παρατήρηση
- Chain Ladder με έκτοπη παρατήρηση
- Robust Chain Ladder με έκτοπη παρατήρηση
- Log-Linear Chain Ladder χωρίς έκτοπη παρατήρηση
- Log-Linear Chain Ladder με έκτοπη παρατήρηση
- Mack Method χωρίς έκτοπη παρατήρηση
- Robust Mack Method με έκτοπες παρατηρήσεις
- Bootstrap Method χωρίς έκτοπη παρατήρηση
- Robust Bootstrap Method με έκτοπη παρατήρηση

Για τις ανάγκες των υπολογισμών χρησιμοποιήθηκε το στατιστικό πακέτο R (για περισσότερα βλ.: *https://cran.r-project.org/)*. Ο κώδικας που γράφτηκε για την παραγωγή αποτελεσμάτων είναι διαθέσιμος στο Παράρτημα Β. Ο κώδικας που χρησιμοποιήθηκε για την εξαγωγή των αποτελεσμάτων στηρίχθηκε στην θεωρία, όπως αυτή παρουσιάστηκε στα προηγούμενα κεφάλαια. Ο κώδικα που αναφέρεται στη μέθοδο Chain Ladder με τη χρήση Ανάλυσης Παλινδρόμησης αλλά και την ανθεκτική τού εκδοχή προέρχεται από έτοιμες εντολές του στατιστικού πακέτου R. Ο υπόλοιπος κώδικας, που δίνει τα αποτελέσματα για την Μέθοδο Mack, την Μέθοδο Bootstrap και τις ανθεκτικές εκδοχές τους, γράφτηκε από την αρχή, καθώς το πακέτο της R δεν διέθετε κώδικα για την μελέτη που θέλαμε να πραγματοποιήσουμε.

Πίνακας 6: Taylor and Ashe (1983)

Έτος Αναγγελίας Ατυχήματος	Έτος Εξέλιξης Ατυχήματος: 1	2	3	4	5	6	7	8	9	10
1	357.848	1.124.788	1.735.330	2.218.270	2.745.596	3.319.994	3.466.336	3.606.286	3.833.515	3.901.463
2	352.118	1.236.139	2.170.033	3.353.322	3.799.067	4.120.063	4.647.867	4.914.039	5.339.085	
3	290.507	1.292.306	2.218.525	3.235.179	3.985.995	4.132.918	4.628.910	4.909.315		
4	310.608	1.418.858	2.195.047	3.757.447	4.029.929	4.381.982	4.588.268			
5	443.160	1.136.350	2.128.333	2.897.821	3.402.672	3.873.311				
6	396.132	1.333.217	2.180.715	2.985.752	3.691.712					
7	440.832	1.288.463	2.419.861	3.483.130						
8	359.480	1.421.128	2.864.498							
9	376.686	1.363.294								
10	344.014									

Προχωρούμε στην παρουσίαση των αποτελεσμάτων, σύμφωνα με τη σειρά που αναφέρθηκε παραπάνω.

5.1 Chain Ladder χωρίς έκτοπη παρατήρηση

Λαμβάνοντας τα δεδομένα του *Πίνακα 5* και εφαρμόζοντας την Chain Ladder, όπως αυτή παρουσιάστηκε στο Κεφάλαιο 2, παίρνουμε το παρακάτω πίνακα (*Πίνακα 7*), ο οποίος δείχνει την πρόβλεψη των ζημιών για την εν λόγω μέθοδο.

Οι συντελεστές εξέλιξης, όπως προέκυψαν από την εφαρμογή της Chain Ladder, παρουσιάζονται στον Πίνακας 6, στρογγυλοποιημένοι στο τρίτο δεκαδικό.

Πίνακας 7: Συντελεστές Εξέλιξης Ζημιών

Development Factors

1	2	3	4	5	6	7	8	9
3,491	1,753	1,456	1,173	1,103	1,086	1,054	1,076	1,018

Χρησιμοποιώντας τους συντελεστές και εφαρμόζοντας τη σχέση (2.2.1) προκύπτει ο Πίνακας 7.

Πίνακας 8: Τρίγωνο ολοκληρωμένο με Chain Ladder και το Απόθεμα

	Development Year										
Accident Year	1	2	3	4	5	6	7	8	9	10	R_j
2008	357.856	1.124.804	1.735.354	2.218.302	2.745.636	3.320.042	3.466.392	3.606.350	3.833.587	3.901.543	0
2009	352.126	1.236.155	2.230.057	3.413.354	3.859.107	4.180.111	4.707.923	4.974.103	5.399.157	5.494.865	95.708
2010	290.515	1.292.322	2.218.549	3.235.211	3.986.035	4.132.966	4.628.966	4.909.379	5.282.593	5.376.235	466.856
2011	310.616	1.418.874	2.195.071	3.757.479	4.029.969	4.382.030	4.588.324	4.834.364	5.201.875	5.294.086	705.762
2012	443.168	1.136.366	2.128.357	2.897.853	3.402.712	3.873.359	4.206.263	4.431.816	4.768.725	4.853.258	979.899
2013	396.140	1.333.233	2.180.739	2.985.784	3.691.752	4.073.771	4.423.900	4.661.123	5.015.464	5.104.370	1.412.618
2014	440.840	1.288.479	2.419.885	3.483.162	4.086.755	4.509.649	4.897.240	5.159.845	5.552.099	5.650.518	2.167.356
2015	359.488	1.421.144	2.864.522	4.169.584	4.892.126	5.398.358	5.862.332	6.176.688	6.646.243	6.764.058	3.899.536
2016	376.694	1.363.310	2.390.131	3.479.063	4.081.946	4.504.341	4.891.476	5.153.772	5.545.565	5.643.868	4.280.558
2017	344.022	1.200.834	2.105.282	3.064.437	3.595.470	3.967.526	4.308.523	4.539.559	4.884.659	4.971.247	4.627.225
										Συνολικό Απόθεμα	18.635.518

Με κόκκινο είναι τα κενά του τριγώνου που συμπληρώθηκαν σύμφωνα με την Chain Ladder. Με πράσινο γέμισμα είναι το απόθεμα που προκύπτει για κάθε έτος ατυχήματος (accident year). Επομένως, το απόθεμα – R_j- για κάθε έτος προκύπτει αφαιρώντας από τα στοιχεία της στήλης 10, τα στοιχεία της διαγώνιου. Τέλος, προσθέτουμε τα επιμέρους αποθέματα για να προκύψει το συνολικό απόθεμα R.

5.2 Chain Ladder με δύο έκτοπες παρατηρήσεις

Εδώ θα δούμε την ευαισθησία της Chain Ladder, στις έκτοπες παρατηρήσεις. Στον *Πίνακα 6* θεωρούμε ότι έγινε λάθος καταχώριση, με αποτέλεσμα η τιμή στο κελί [2011,7] αντί για 445,753 να γραφτεί 445.753. Αυτό έχει σαν αποτέλεσμα να αλλάξουν τόσο οι συντελεστές εξέλιξης όσο και τα νούμερα, που προκύπτουν από την εφαρμογή της Chain Ladder. Ο *Πίνακας 8* δείχνει τη διαφορά που προκύπτει στους συντελεστές εξέλιξης.

Πίνακας 9: Συντελεστές Εξέλιξης με 2 έκτοπες παρατηρήσεις

Development Factors με δύο outliers								
1	2	3	4	5	6	7	8	9
304,2636	1,0076	1,0068	5,3758	1,0003	1,0003	1,0001	1,0001	1,0177

Με κίτρινο φαίνεται ποιος συντελεστής εξέλιξης επηρεάστηκε πρώτος από την αλλαγή. Από εκεί και πέρα , βλέπουμε ότι και οι υπόλοιποι συντελεστές έχουν αλλάξει σε μεγάλο βαθμό. Ως απόρροια αυτής της αλλαγής είναι ο ακόλουθος *Πίνακας 9* που δείχνει πώς αλλάζει το απόθεμα. Με έντονο κίτρινο φαίνεται το σημείο όπου καταχωρήθηκε λάθος η ζημιά.

Πίνακας 10: Υπολογισμός Αποθέματος σε Τρίγωνο με 2 έκτοπες παρατηρήσεις

Απόθεμα με δύο έκτοπες παρατηρήσεις

Development Year

Accident Year	1	2	3	4	5	6	7	8	9	10	R_j
2008	357.856	1.124.804	1.735.354	2.218.302	2.745.636	3.320.042	3.466.392	3.606.350	3.833.587	3.901.543	0
2009	352.126	1.236.155	2.230.057	3.413.354	4.460.943.354	4.461.264.358	4.461.792.170	4.462.058.350	4.462.483.404	4.541.587.523	79.104.119
2010	290.515	1.002.097.515	1.003.023.742	1.004.040.404	1.004.791.228	1.004.938.159	1.005.434.159	1.005.714.572	1.005.861.475	1.023.691.857	17.977.285
2011	310.616	1.418.874	2.195.071	3.757.479	4.029.969	4.382.030	4.588.324	4.588.900	4.589.570	4.670.927	82.603
2012	443.168	1.136.366	2.128.357	2.897.853	3.402.712	3.873.359	3.874.333	3.874.819	3.875.385	3.944.082	70.723
2013	396.140	1.333.233	2.180.739	2.985.784	3.691.752	3.693.009	3.693.938	3.694.402	3.694.941	3.760.440	68.688
2014	440.840	1.288.479	2.419.885	3.483.162	18.724.717	18.731.094	18.735.804	18.738.156	18.740.893	19.073.103	15.589.941
2015	359.488	1.421.144	2.864.522	2.883.930	15.503.377	15.508.657	15.512.557	15.514.504	15.516.770	15.791.828	12.927.306
2016	376.694	1.363.310	1.373.721	1.383.029	7.434.859	7.437.392	7.439.262	7.440.195	7.441.282	7.573.190	6.209.880
2017	344.022	104.673.382	105.472.743	106.187.355	570.839.988	571.034.412	571.178.003	571.249.683	571.333.125	581.460.849	581.116.827
										Συνολικό Απόθεμα	713.147.372

Παρατηρούμε ότι το απόθεμα εκτοξεύεται κατά πολύ και αποκλίνει πολύ από την εκτίμηση που δίνει η Chain Ladder χωρίς κάποια έκτοπη τιμή. Βλέπουμε, επίσης, ότι λόγω της έκτοπης τιμής δημιουργούνται μεγάλα άλματα δυσανάλογα.

5.3 Robust Chain Ladder με δύο έκτοπες παρατηρήσεις

Είδαμε στην προηγούμενη ενότητα, πώς ένα μικρό πλήθος έκτοπων παρατηρήσεων μπορούν να επηρεάσουν το αποτέλεσμα της Chain Ladder. Τώρα, θα δώσουμε τα αποτελέσματα που προκύπτουν αν εφαρμοστεί η Ανθεκτική Chain Ladder, όπως αυτή περιεγράφηκε στο 4ο Κεφάλαιο. Θα χρησιμοποιήσουμε το τρίγωνο που χρησιμοποιήσαμε στην προηγούμενη ενότητα (τρίγωνο του Πίνακα 10). Εφαρμόζεται στο ακέραιο, ο αλγόριθμος που προτάθηκε από τους Verdonck *et al.* (2009) και ο οποίος παρατέθηκε στο Κεφάλαιο 4, στις σελίδες 37-39. Για πιο αναλυτικά, ο αναγνώστης μπορεί να δει στον κώδικα της R στο Παράρτημα Β.

Οι συντελεστές εξέλιξης, μετά την ανθεκτικοποίηση της Chain Ladder σύμφωνα με τους Verdonck et al (2009) προκύπτουν ως εξής:

Development Factors Verdonck

1	2	3	4	5	6	7	8	9
3,511	1,688	1,369	1,204	1,109	1,089	1,057	1,074	1,018

Το αποτέλεσμα που προκύπτει, εφαρμόζοντας τον αλγόριθμο των Verdonck *et al.* (2009) φαίνεται στον Πίνακα 11.

Πίνακας 11: Robust Chain Ladder - Verdonck et al. (2009)

Robust Chain Ladder
Development Year

Accident Year	1	2	3	4	5	6	7	8	9	10	Rj
2008	357.856	1.124.804	1.735.354	2.218.302	2.745.636	3.320.042	3.466.392	3.606.350	3.833.587	3.901.543	0
2009	352.126	1.236.155	2.230.057	3.413.354	3.859.107	4.180.111	4.707.923	4.974.103	5.399.157	5.494.865	95.708
2010	359.080	1.260.567	2.186.794	3.203.456	3.954.280	4.101.211	4.597.211	4.877.624	5.239.699	5.332.581	454.956
2011	310.616	1.418.874	2.195.071	2.944.256	3.216.746	3.568.807	3.775.101	3.988.541	4.284.617	4.360.569	585.467
2012	443.168	1.136.366	2.128.357	2.897.853	3.402.712	3.873.359	4.219.530	4.458.097	4.789.030	4.873.922	1.000.563
2013	396.140	1.333.233	2.180.739	2.985.784	3.691.752	4.095.801	4.461.852	4.714.119	5.064.057	5.153.825	1.462.073
2014	440.840	1.288.479	2.419.885	3.483.162	4.194.768	4.653.870	5.069.797	5.356.437	5.754.055	5.856.054	2.372.892
2015	359.488	1.421.144	2.331.614	3.192.357	3.844.552	4.265.324	4.646.525	4.909.234	5.273.655	5.367.139	3.035.524
2016	376.694	1.363.310	2.300.879	3.150.275	3.793.872	4.209.098	4.585.274	4.844.520	5.204.138	5.296.389	3.933.079
2017	344.022	1.207.706	2.038.263	2.790.711	3.360.850	3.728.683	4.061.924	4.291.580	4.610.152	4.691.873	4.347.851
										Συνολικό Απόθεμα	**17.288.114**

5.4 Log-Linear Chain Ladder χωρίς έκτοπη παρατήρηση

Στο σημείο αυτό θα παρουσιάσουμε την εκτίμηση αποθέματος με τη χρήση της Ανάλυσης παλινδρόμησης. Το τρίγωνο του Πίνακα 6, παρατίθενται στην κάτωθι μορφή (Πίνακας 12), όπου acc είναι το έτος ατυχήματος, dev είναι το έτος εξέλιξης ζημιών, claims είναι οι ζημιές, logclaims οι λογάριθμοι των ζημιών. Επίσης, έχουν προστεθεί δύο ακόμη στήλες (accf και devf) που αναφέρονται στις μελλοντικές ζημιές, οι οποίες πρόκειται αν εκτιμηθούν.

Πίνακας 12:Δεδομένα Taylor and Ashe (1983) για Ανάλυση Παλινδρόμησης

acc	dev	claims	logclaims	accf	devf
0	0	357.856	12,79	0	0
0	1	766.948	13,55	0	1
0	2	610.550	13,32	0	2
0	3	482.948	13,09	0	3
0	4	527.334	13,18	0	4
0	5	574.406	13,26	0	5
0	6	146.350	11,89	0	6
0	7	139.958	11,85	0	7
0	8	227.237	12,33	0	8
0	9	67.956	11,13	0	9
1	0	352.126	12,77	1	0
1	1	884.029	13,69	1	1
1	2	993.902	13,81	1	2
1	3	1.183.297	13,98	1	3
1	4	445.753	13,01	1	4
1	5	321.004	12,68	1	5
1	6	527.812	13,18	1	6
1	7	266.180	12,49	1	7
1	8	425.054	12,96	1	8
1	9	#N/A	#N/A	1	9
2	0	290.515	12,58	2	0
2	1	1.001.807	13,82	2	1
…	…	…	…	…	…
9	5	#N/A	#N/A	9	5
9	6	#N/A	#N/A	9	6
9	7	#N/A	#N/A	9	7
9	8	#N/A	#N/A	9	8
9	9	#N/A	#N/A	9	9

Το μοντέλο γραμμικής παλινδρόμησης που εφαρμόζεται είναι αυτό που εμφανίζεται στους τύπους (2.2.8)-(2.2.14), με τις αντίστοιχες συνθήκες. Το αποτέλεσμα που παίρνουμε-για τις μελλοντικές ζημιές-από την εφαρμογή αυτή παρουσιάζεται στον Πίνακα 13.

Πίνακας 13: Μελλοντικές Ζημιές με χρήση Ανάλυσης Παλινδρόμησης

i	j	Z	VarZ	Y	VarY	seY
1	9	11,49	0,26	111.713	3.670.572.776	60.585
2	8	12,74	0,20	377.998	30.901.370.604	175.788
2	9	11,41	0,26	102.617	3.130.144.352	55.948
3	7	12,25	0,18	228.014	10.054.367.307	100.271
3	8	12,63	0,20	338.616	25.183.545.065	158.693
3	9	11,30	0,26	91.926	2.542.206.193	50.420
4	6	12,68	0,17	350.028	22.587.412.037	150.291
4	7	12,36	0,18	255.118	12.855.496.975	113.382
4	8	12,74	0,20	378.868	32.132.253.319	179.255
4	9	11,41	0,27	102.853	3.230.129.189	56.834
5	5	12,82	0,17	403.567	29.615.796.245	172.092
5	6	12,71	0,17	361.318	24.787.337.872	157.440
5	7	12,39	0,18	263.347	14.084.544.495	118.678
5	8	12,77	0,21	391.089	35.109.383.536	187.375
5	9	11,44	0,27	106.171	3.510.321.438	59.248
6	4	13,29	0,17	645.034	76.705.705.661	276.958
6	5	12,90	0,17	438.438	36.497.713.682	191.044
6	6	12,79	0,18	392.539	30.501.066.386	174.646
6	7	12,47	0,19	286.102	17.292.466.027	131.501
6	8	12,85	0,21	424.882	42.946.239.937	207.235
6	9	11,52	0,28	115.345	4.261.647.604	65.281
7	3	13,96	0,18	1.264.647	309.294.164.368	556.142
7	4	13,38	0,18	707.724	98.931.358.573	314.534
7	5	12,99	0,18	481.049	46.996.172.223	216.786
7	6	12,88	0,19	430.689	39.186.893.659	197.957
7	7	12,56	0,20	313.908	22.143.241.135	148.806
7	8	12,94	0,22	466.175	54.688.771.943	233.856
7	9	11,60	0,29	126.555	5.365.230.538	73.248
8	2	13,81	0,20	1.098.227	260.844.785.544	510.730
8	3	13,83	0,20	1.120.126	275.572.712.207	524.950
8	4	13,25	0,20	626.847	87.960.694.729	296.582
8	5	12,86	0,21	426.075	41.672.086.958	204.137
8	6	12,75	0,21	381.471	34.618.752.951	186.061
8	7	12,42	0,22	278.036	19.453.624.887	139.476
8	8	12,81	0,25	412.902	47.596.610.713	218.166
8	9	11,47	0,31	112.093	4.577.946.455	67.661
9	1	13,66	0,26	972.890	278.387.599.777	527.624
9	2	13,69	0,26	1.009.021	302.638.025.371	550.125
9	3	13,71	0,26	1.029.142	318.628.654.347	564.472
9	4	13,13	0,27	575.930	101.279.622.205	318.245
9	5	12,74	0,27	391.467	47.722.550.345	218.455
9	6	12,63	0,28	350.485	39.347.624.673	198.362
9	7	12,31	0,29	255.452	21.859.763.837	147.850
9	8	12,69	0,31	379.363	52.435.500.823	228.988
9	9	11,35	0,38	102.988	4.827.928.746	69.483
			Συνολικό Απόθεμα	19.478.845		

Όπου Z θεωρούνται οι λογάριθμοι των μελλοντικών ζημιών, VarZ η διακύμανση σύμφωνα με τη σχέση (2.2.9), Y οι πραγματικές μελλοντικές ζημιές με χρήση του τύπου (2.2.10), VarY η διακύμανση των μελλοντικών πραγματικών ζημιών και seY είναι το τυπικό σφάλμα της εκτίμησης. Το συνολικό απόθεμα προκύπτει ως άθροισμα της στήλης Y. Πιο αναλυτικά ο κώδικας και η επεξήγησή του στο Παράρτημα Α.

5.5 Robust Log-Linear Chain Ladder

Στην περίπτωση αυτή, θα πάρουμε το τρίγωνο εξέλιξης ζημιών με δύο έκτοπες παρατηρήσεις (Πίνακας 10). Αντίστοιχα με την προηγούμενη παράγραφο, εφαρμόζονται οι αντίστοιχες σχέσεις για την Robust

Log-Linear Chain Ladder (4.2.1)-(4.2.11) και με χρήση του Μ-εκτιμητή του Huber. Ο πλήρης κώδικας παρουσιάζεται στο Παράρτημα Α.

Πίνακας 14:Robust Log Linear Chain Ladder

i	j	Z	VarZ	Y	VarY	seY
1	9	11,50	0,19	108.782	2.412.484.056	49.117
2	8	12,94	0,12	441.901	25.229.499.471	158.838
2	9	11,61	0,19	120.664	3.012.094.257	54.883
3	7	12,18	0,10	204.900	4.504.068.175	67.112
3	8	12,63	0,12	324.660	13.957.911.858	118.144
3	9	11,30	0,19	88.651	1.652.913.571	40.656
4	6	12,65	0,09	326.600	10.520.931.788	102.572
4	7	12,32	0,11	235.107	6.148.492.531	78.412
4	8	12,76	0,13	372.522	18.937.794.502	137.615
4	9	11,43	0,19	101.720	2.220.906.490	47.126
5	5	12,90	0,09	418.088	16.819.614.157	129.690
5	6	12,69	0,10	340.219	12.027.224.836	109.669
5	7	12,35	0,11	244.911	6.991.985.175	83.618
5	8	12,80	0,13	388.056	21.371.798.915	146.191
5	9	11,47	0,20	105.962	2.475.468.455	49.754
6	4	13,34	0,09	654.444	42.244.065.303	205.534
6	5	12,99	0,10	459.016	21.892.934.230	147.963
6	6	12,78	0,11	373.525	15.576.909.144	124.807
6	7	12,44	0,12	268.886	8.993.825.494	94.836
6	8	12,89	0,14	426.045	27.213.915.447	164.966
6	9	11,56	0,21	116.335	3.099.645.324	55.674
7	3	13,88	0,10	1.124.793	135.727.068.172	368.412
7	4	13,41	0,11	703.999	55.129.133.086	234.796
7	5	13,05	0,11	493.774	28.421.045.801	168.585
7	6	12,85	0,12	401.809	20.083.780.664	141.717
7	7	12,51	0,13	289.247	11.486.420.653	107.175
7	8	12,96	0,15	458.305	34.261.659.695	185.099
7	9	11,63	0,22	125.144	3.807.605.057	61.706
8	2	13,79	0,12	1.030.830	137.288.387.579	370.524
8	3	13,74	0,12	987.939	129.248.223.106	359.511
8	4	13,27	0,13	618.343	52.177.651.793	228.424
8	5	12,91	0,13	433.696	26.694.589.826	163.385
8	6	12,70	0,14	352.920	18.673.884.196	136.652
8	7	12,37	0,15	254.054	10.528.124.554	102.607
8	8	12,82	0,17	402.543	30.711.178.635	175.246
8	9	11,49	0,24	109.918	3.278.463.962	57.258
9	1	13,68	0,19	962.179	188.739.194.851	434.441
9	2	13,68	0,19	960.041	190.673.017.225	436.661
9	3	13,64	0,19	920.095	178.053.201.247	421.964
9	4	13,17	0,19	575.880	71.183.758.500	266.803
9	5	12,81	0,20	403.913	35.969.556.283	189.656
9	6	12,60	0,21	328.685	24.742.930.999	157.299
9	7	12,27	0,22	236.608	13.610.999.666	116.666
9	8	12,71	0,24	374.899	38.138.703.494	195.291
9	9	11,38	0,31	102.369	3.760.132.132	61.320
			Συνολικό Απόθεμα	18.772.978		

Όπως αναφέρθηκε και σε προηγούμενο κεφάλαιο, θα εξετάσουμε την περίπτωση όπου τα δεδομένα μας είναι ελλιπή ή περιλαμβάνουν έξοδα πολύ μικρά σε σχέση με το επίπεδο των υπολοίπων ετών. Αυτό σημαίνει ότι στο τρίγωνο εξέλιξης ζημιών παρουσιάζονται σημεία όπου τα έξοδα είναι μηδενικά είτε πολύ μικρά (αν μιλάμε για χιλ. € τότε αυτά τα ποσά μπορεί να είναι μερικές εκατοντάδες €). Έτσι προκύπτουν τα λεγόμενα *μεγάλα άλματα*. Πριν από αυτό θα παρουσιάσουμε το μοντέλο του Mack (Mack model) με το οποίο ο τελευταίος εκτιμά το μέσο τετραγωνικό σφάλμα της Chain Ladder. Η περιγραφή θα γίνει τόσο με κανονικά δεδομένα όσο και με δεδομένα που έχουν κάποια έκτοπη τιμή. Στη συνέχεια, θα παρουσιάσουμε τα αποτελέσματα σύμφωνα με το ανθεκτικό μοντέλο του Mack (Robust Mack model). Τέλος η ίδια διαδικασία θα γίνει με τη μέθοδο Bootstrap και θα συγκριθούν τα αποτελέσματα.

5.6 Mack Method χωρίς έκτοπη παρατήρηση

Κάνοντας χρήση της θεωρίας, όπως παρουσιάστηκε αναλυτικά στο Κεφάλαιο 3 προχωρήσαμε στην εφαρμογή του μοντέλου χρησιμοποιώντας τα δεδομένα του Πίνακα 9.

Ξεκινάμε από την εκτίμηση αποθέματος. Η διαδικασία είναι ίδια με την κλασσική μέθοδο Chain Ladder, μιας και η τελευταία αποτελεί κομμάτι του μοντέλου. Επομένως, παίρνουμε ως δεδομένο τον Πίνακα 7 από την προηγούμενη ενότητα. Στη συνέχεια, υπολογίζουμε την διακύμανση σ^2 από τη σχέση (3.13) Το διάνυσμα της διακύμανσης είναι:

Πίνακας 15: Διάνυσμα Διακύμανσης Mack Method

Διάνυσμα $\sigma(\kappa)^2$

1	2	3	4	5	6	7	8	9
160.274,7	38.242,2	41.155,4	15.336,0	13.787,0	8.010,4	436,1	1.053,0	436,1

Στον τύπο (3.15) που μας δίνει το μέσο τετραγωνικό σφάλμα γίνεται ένας διαχωρισμός. Ως γνωστόν, το μέσο τετραγωνικό σφάλμα είναι ένα μέτρο κινδύνου και μας, στην περίπτωση του Mack Model μας δίνει το σφάλμα της πρόβλεψης που προσπαθούμε να κάνουμε. Για κάθε έτος ατυχήματος αυτό μπορεί να εκφραστεί ως ένα άθροισμα δύο όρων. Ο πρώτος όρος (δηλαδή το πρώτο γινόμενο μέσα στο άθροισμα) είναι η μεταβλητότητα στις ζημιές. Ονομάζεται και *process variance multiplier* . Από την άλλη, ο δεύτερος όρος (το δεύτερο γινόμενο μέσα στο άθροισμα) είναι η μεταβλητότητα των εκτιμώμενων αποθεμάτων , γνωστή και ως *parameter variance multiplier*.

Έτσι, για τα δεδομένα μας, το διάνυσμα για την process variance προκύπτει ως:

Πίνακας 16: Process Variance Multiplier

Process Variance Multiplier

1	2	3	4	5	6	7	8	9
331.576,8	141.490,7	89.983,2	44.116,2	26.043,8	10.388,9	1.877,8	1.424,5	428,5

Ενώ το διάνυσμα για την parameter variance είναι το:

Πίνακας 17: Parameter Variance Multiplier

Parameter Variance Multiplier

1	2	3	4	5	6	7	8	9
0,008353	0,0044	0,003186	0,001901	0,001299	0,000671	0,000247	0,000216	0,00011

Από τα παραπάνω, προκύπτει το μέσο τετραγωνικό σφάλμα για κάθε έτος ατυχήματος, σύμφωνα με τον τύπο (3.40):

Πίνακας 18 MSE ανά Έτος Αναγγελίας Ζημιών

MSE(j)

1	2	3	4	5	6	7	8	9
75.306	117.885	129.809	257.326	408.385	556.748	868.586	968.873	1.361.907

Τέλος, μένει να υπολογιστεί το μέσο τετραγωνικό σφάλμα για το τελικό απόθεμα, από τον τύπο (3.41) που είναι 2.448.277€.

5.7 Mack Method με έκτοπες παρατηρήσεις

Εδώ θα παρουσιάσουμε πώς η μέθοδος Mack επηρεάζεται όταν στα δεδομένα που έχουμε εμφανιστεί έκτοπες παρατηρήσεις. Στη συνέχεια, με τις μεθόδους ανίχνευσης μεγάλων αλμάτων όπως παρουσιάστηκαν στην ενότητα «Ακραίες τιμές» (Κεφάλαιο 3) και έτσι όπως τις πρότειναν οι Busse et al. (2010), θα φιλτράρουμε τα δεδομένα ώστε να δούμε πόσο μεταβάλλεται το απόθεμα, σε σχέση με την περίπτωση όπου δεν είχαμε κάποιο μεγάλο άλμα.

Στη συνέχεια, θα δείξουμε πώς με την ανθεκτικοποίηση του διανύσματος του εκτιμητή σ_k^2 και με τη μεθοδολογία που πρότειναν τόσο οι Verall and England (2006) αλλά και οι Busse et al. (2010) θα προκύψει το νέο μέσο τετραγωνικό σφάλμα (MSE) της ανθεκτικής μεθόδου. Τέλος, και όπως αναφέρθηκε και στο Κεφάλαιο 3, η ανθεκτικοποίση του διανύσματος του εκτιμητή της σ_k^2 προκαλεί μία μεροληψία στον εκτιμητή την οποία και θα παρουσιάσουμε, σχολιάζοντας παράλληλα αν είναι σημαντική ή όχι.

Αρχίζοντας, θεωρούμε τον Πίνακα 4, στην αρχή του Κεφαλαίου, με την εξής παραλλαγή. Στις θέσεις [2010,2] και [2009,5], οι οποίες επιλέχθηκαν τυχαία, βάζουμε τις τιμές που ήδη είχαν πολλαπλασιασμένες με το 1000. (βλ. Πίνακα 9 όπου φαίνεται πώς δημιουργούνται μεγάλα άλματα και βγαίνει μεγάλο το απόθεμα). Πρόκειται για τον πίνακα στην προηγούμενη ενότητα που έχει δύο έκτοπες παρατηρήσεις. Με τον τρόπο αυτό, δημιουργούνται οι προϋποθέσεις ώστε να δούμε πώς μεταβάλλεται το απόθεμα αλλά και γεννάται η ανάγκη για μια ανθεκτικοποιημένη μέθοδο.

Παρουσιάζοντας τη Mack method με 2 έκτοπες παρατηρήσεις βλέπουμε ότι:

Πίνακας 19:Process variance Multiplier με 2 έκτοπες παρατηρήσεις

Process Variance Multiplier με 2 Outliers

1	2	3	4	5	6	7	8	9
227.050.938.868	219.865.334.508	219.861.144.266	219.858.541.405	80.467	25.644	20.174	17.294	2.780

Πίνακας 20:Parameter Variance Multiplier με 2 έκτοπες παρατηρήσεις

Parameter Variance Multiplier με 2 Outliers

1	2	3	4	5	6	7	8	9
40,669778	39,392109	39,391363	39,390898	0,000727	0,000717	0,000716	0,000716	0,000713

Ενώ τέλος, το MSE_j και το συνολικό MSE υπολογίζονται ως:

Πίνακας 21: MSE ανά Έτος Αναγγελίας Ατυχήματος με 2 έκτοπες παρατηρήσεις

MSEj με 2 Outliers

1	2	3	4	5	6	7	8	9
121.285.447	27.708.971	331.449	335.112	559.351	2.051.271.392	1.865.966.990	1.291.255.659	12.073.588.961

Και το συνολικό MSE είναι 11.893.632.417.

5.8 Ανθεκτική Mack Method

Όπως φάνηκε παραπάνω, σε περίπτωση που έχουμε κάποια σημεία όπου εμφανίζονται μεγάλα άλματα, τότε το συνολικό απόθεμα αλλά και το μέσο τετραγωνικό σφάλμα (ως μέτρο καλής εκτίμησης) αλλάζουν σημαντικά. Για να αποφευχθεί αυτή η μεγάλη μεταβολή, χρησιμοποιούμε την Ανθεκτική Mack Method, όπως αυτή παρουσιάστηκε στο Κεφάλαιο 3. Εφαρμόζουμε την αλλαγή στον υπολογισμό του σ_k^2 αντικαθιστώντας το L_{jk} με μία εκτίμηση της μέσης τιμής του $\mathrm{E}\left(L_{jk}\right)$.

Παράλληλα, χρησιμοποιούμε και ένα κριτήριο φιλτραρίσματος ώστε να απομονώσουμε τα σημεία, όπου εμφανίζονται τα μεγάλα άλματα, χωρίς όμως αυτό το κριτήριο να επηρεάζει δραματικά το απόθεμα. Με τη μέθοδο αυτή από τους επόμενους υπολογισμούς αφαιρούνται οι ακραίες τιμές. Έτσι σύμφωνα με τα παραπάνω, έχουμε τα ακόλουθα αποτελέσματα.

Οι συντελεστές εξέλιξης της ανθεκτικής μεθόδου προκύπτουν ως:

Πίνακας 22:Συντελεστές Εξέλιξης Ζημιών Ανθεκτική Mack Method

Development Factors Robust Mack Method

1	2	3	4	5	6	7	8	9
3,190	1,835	1,488	1,158	1,113	1,094	1,060	1,080	1,018

Όπως γίνεται ξεκάθαρο, οι συντελεστές είναι πολύ κοντά σε αυτούς που υπολογίστηκαν με την κλασσική Mack method χωρίς έκτοπη παρατήρηση.

Το απόθεμα που προκύπτει με τη μέθοδο Chain Ladder (επισημαίνουμε πάλι ότι η Chain Ladder είναι κομμάτι της Mack Method άρα και τις ανθεκτικής της μορφής) είναι:

Πίνακας 23: Τρίγωνο ολοκληρωμένο και εύρεση Αποθέματος με Ανθεκτική Mack Method

Full triangle with large jumps

	1	2	3	4	5	6	7	8	9	10	R_j
2008	357.856	1.124.804	1.735.354	2.218.302	2.745.636	3.320.042	3.466.392	3.606.350	3.833.587	3.901.543	0
2009	352.126	1.236.155	2.230.057	3.413.354	3.413.354	3.734.358	4.262.170	4.528.350	4.953.404	5.041.210	87.806
2010	290.515	290.515	1.216.742	2.233.404	2.984.228	3.131.159	3.627.159	3.907.572	4.220.906	4.295.727	388.155
2011	310.616	1.418.874	2.195.071	3.757.479	4.029.969	4.382.030	4.588.324	4.865.728	5.255.892	5.349.061	760.737
2012	443.168	1.136.366	2.128.357	2.897.853	3.402.712	3.873.359	4.239.343	4.495.648	4.856.137	4.942.219	1.068.860
2013	396.140	1.333.233	2.180.739	2.985.784	3.691.752	4.107.132	4.495.205	4.766.979	5.149.225	5.240.503	1.548.751
2014	440.840	1.288.479	2.419.885	3.483.162	4.032.606	4.486.338	4.910.241	5.207.107	5.624.645	5.724.350	2.241.188
2015	359.488	1.421.144	2.864.522	4.262.268	4.934.610	5.489.831	6.008.552	6.371.821	6.882.753	7.004.760	4.140.238
2016	376.694	1.363.310	2.501.345	3.721.877	4.308.977	4.793.805	5.246.760	5.563.972	6.010.126	6.116.664	4.753.354
2017	344.022	1.097.258	2.013.204	2.995.548	3.468.075	3.858.287	4.222.848	4.478.155	4.837.241	4.922.989	4.578.967
										Συνολικό Απόθεμα	19.568.056

Με έντονο κίτρινο χρώμα σημειώνονται τα σημείο, όπου υπήρξαν τα μεγάλα άλματα. Όπως παρατηρείται, το απόθεμα αλλάζει και έρχεται πιο κοντά στο αρχικό απόθεμα (18.363.518). Η εταιρία θα έπρεπε να διακρατήσει περισσότερα κεφάλαια αλλά όχι τόσα όσα προέκυψαν χωρίς την ανθεκτικοποίηση.

Τώρα από τη Ανθεκτική μέθοδο του Mack παίρνουμε τα κάτωθι αποτελέσματα. Ξεκινώντας από το διάνυσμα που μας δίνει το σ_k^2 έχουμε ότι:

Πίνακας 24: Σύγκριση των σ(k)^2

Σύγκριση $\sigma(k)^2$

	1	2	3	4	5	6	7	8	9
$\sigma(k)^2$ w/o outlier	160.275	38.242	41.155	15.336	13.787	8.010	436	1.053	436
$\sigma(k)^2$ with 2 outliers	393.576.463.244	765.884	478.559	1.160.322.413.171	53.858	5.375	2.829	14.263	2.829
rob $\sigma(k)^2$	**731.646**	**316.493**	**152.729**	**84.146**	**30.661**	**18.192**	**4.309**	**3.025**	**1.218**

Βάζοντας το διάνυσμα σ_k^2 χωρίς έκτοπη παρατήρηση και το διάνυσμα σ_k^2 με έκτοπες παρατηρήσεις (πριν και μετά την ανθεκτικοποίηση της Mack) σε έναν πίνακα, βλέπουμε πώς διαφέρουν εξαιτίας των μεγάλων αλμάτων που παρατηρούνται από τις έκτοπες τιμές αλλά και πώς έρχεται πάλι σε επίπεδα πολύ κοντά σε αυτά που υπολογίστηκαν στην αρχή.

Αντίστοιχα, επηρεάζονται τα διανύσματα του process variance multiplier και του parameter variance multiplier, όπως φαίνεται στους πίνακες που ακολουθούν:

Πίνακας 25:Σύγκριση των Process Variance Multiplier

Robust Process Variance Multiplier

	1	2	3	4	5	6	7	8	9
w/o outliers	331.577	141.491	89.983	44.116	26.044	10.389	1.878	1.425	429
with outliers	227.050.938.868	219.865.334.508	219.861.144.266	219.858.541.405	80.467	25.644	20.174	17.294	2.780
rob	**731.646**	**316.493**	**152.729**	**84.146**	**30.661**	**18.192**	**4.309**	**3.025**	**1.218**

Πίνακας 26: Σύγκριση των Parameter Variance Multiplier

Parameter Variance Multiplier

	1	2	3	4	5	6	7	8	9
w/o outliers	0,008353	0,0044	0,003186	0,001901	0,001299	0,000671	0,000247	0,000216	0,00011
with outliers	40,66978	39,39211	39,39136	39,3909	0,000727	0,000717	0,000716	0,000716	0,000713
rob	**0,0184**	**0,0097**	**0,0057**	**0,0037**	**0,0019**	**0,0014**	**0,0006**	**0,0005**	**0,0003**

Τέλος, και σύμφωνα με την μεθοδολογία που παρουσιάστηκε στο Κεφάλαιο 3 θα δείξουμε πώς αλλάζει το MSE_j αλλά και το συνολικό MSE, όταν στο τρίγωνο υπάρχουν σημεία όπου εμφανίζονται έκτοπες παρατηρήσεις. Τα MSE_j και MSE θα συγκριθούν με τα αντίστοιχα για να είναι πιο κατανοητές οι διαφορές. Έτσι έχουμε:

Πίνακας 27:Σύγκριση MSE(j)

MSEj

	1	2	3	4	5	6	7	8	9
w/o outliers	75.306	117.885	129.809	257.326	408.385	556.748	868.586	968.873	1.361.907
with outliers	121.285.447	27.708.971	331.449	335.112	559.351	2.051.271.392	1.865.966.990	1.291.255.659	12.073.588.961
rob	**118.619**	**149.937**	**201.343**	**350.831**	**461.010**	**777.474**	**1.162.428**	**1.515.947**	**2.011.922**

και αντίστοιχα

Πίνακας 28: Σύγκριση Συνολικό MSE

MSE	
w/o outliers	2.448.277
with outliers	11.893.632.417
rob	3.553.764

Από τα παραπάνω, βλέπουμε ότι με την αλλαγή που κάναμε στον παρονομαστή της σχέσης (4.4.2) - αντικαθιστώντας το L_{jk} με το $E\left(L_{jk}\right)$ - οι ζημιές στην κύρια διαγώνιο καθώς και οι συντελεστές εξέλιξης, επηρεάζονται λιγότερο από ακραίες τιμές. Έτσι, ο εκτιμώμενη τιμή $E\left(L_{jk}\right)$ είναι πιο *ελαστική* σε σχέση με την τιμή L_{jk}. Προκύπτει, λοιπόν, ότι ο εκτιμητής $\left(\sigma_k^2\right)^r$ είναι *ανθεκτικός* και μπορεί να αντικαταστήσει τον σ_k^2 στους υπολογισμούς.

Ωστόσο, η ανθεκτικοποίηση της μεθόδου του Mack συνοδεύεται από το τίμημα της μεροληψίας. Δείξαμε στο Κεφάλαιο 3 πώς προκύπτει μία *μη-εξαλειφόμενη* μεροληψία. Παράλληλα, στη σχέση (4.4.15) ακολουθήσαμε μία προσέγγιση των Busse et al. (2010). Θα δείξουμε αρχικά, το σφάλμα αυτής της προσέγγισης και μετά θα παρουσιάσουμε την μεροληψία που προκύπτει, αξιολογώντας την ως προς τη σημαντικότητά της στο αποτέλεσμα.

Το σφάλμα της προσέγγισης που ακολουθήσαμε, φαίνεται στο παρακάτω πίνακα:

Πίνακας 29: Πίνακας Σχετικού Σφάλματος

Relative Error

1	2	3	4	5	6	7	8	9
2,47%	0,42%	0,36%	0,13%	0,13%	0,09%	0,01%	0,02%	0,02%

Όπως παρατηρούμε, το σφάλμα για $N-k \geq 7$ δεν ξεπερνά το 2,5% ενώ για $N-k \leq 6$ το σφάλμα δεν ξεπερνά το 0,13%.

Η μεροληψία $\mathrm{B}_{\left(\hat{\sigma}_k^2\right)^r}$ φαίνεται στον παρακάτω πίνακα:

Πίνακας 30: Μεροληψία ανά σ(k)^2

Bias $(\sigma^2(\kappa))^r$

1	2	3	4	5	6	7	8
-9,00047	-0,16688	-0,13082	-0,01698	-0,01294	-0,0054	-2,3E-05	-0,00026

Όπως γίνεται διακριτό, το μέγεθος της μεροληψίας δεν είναι τέτοιο που να μην την κάνει αποδεκτή. Στο 1^o έτος φαίνεται να έχει μια μεγάλη τιμή, σε σχέση με τα υπόλοιπα έτη εξέλιξης, ωστόσο είναι μικρή και για αυτό

5.9 Bootstrap Method χωρίς έκτοπη παρατήρηση

Στο σημείο αυτό θα παρουσιάσουμε το απόθεμα αλλά και το μέσο τετραγωνικό σφάλμα, όπως αυτά προέκυψαν εφαρμόζοντας τη μέθοδο Bootstrap. Ως μία κατανομή – από την οποία προέρχονται τα δεδομένα μας αλλά και λόγω της βαριάς ουράς που έχει – θεωρούμε την Lognormal. Όπως εξηγήσαμε και στο Κεφάλαιο 3, η Lognormal δέχεται ως ορίσματα ποσότητες που εξαρτώνται από τα προηγούμενα έτη εξέλιξης ζημιών. Με αυτόν τον τρόπο, ζητάμε και πετυχαίνουμε, οι μελλοντικές ζημιές να προέρχονται από αυτήν την κατανομή (η οποία είναι διαφορετική κάθε φορά). Για μεγαλύτερη κατανόηση ο αναγνώστης μπορεί να δει προσεκτικά το Κεφάλαιο 3 όπου εξηγείται αναλυτικά η μέθοδος και ο αλγόριθμός της αλλά και τον κώδικα της R στο Παράρτημα Β. Η επαναληπτική διαδικασία της Bootstrap ορίστηκε, τυχαία- στις 2500 φορές.

Τα αποτελέσματα θα παρουσιαστούν στον ίδιο πίνακα με τα αντίστοιχα της Mack Method ώστε να μπορεί να γίνει και μία σύγκριση. Έτσι, εφαρμόζοντας τις μεθόδους Bootstrap και Mack πήραμε τα ακόλουθα αποτελέσματα:

Πίνακας 31: Σύγκριση Απλών Μεθόδων Mack and Bootstrap

Simple Mack & Bootstrap Method Αποτελέσματα

	Mack Method Res	Mack MSE	Bootstrap Res	Bootstrap MSE
	18.635.518	2.448.277	24.408.811	2.942.910
% of Res	13%		12%	

Όπως παρατηρούμε το απόθεμα που δίνει η Mack και το απόθεμα που δίνει η Bootstrap μέθοδος έχουν μια απόκλιση, η οποία είναι εξηγήσιμη. Όπως και τα αντίστοιχα μέσα τετραγωνικά σφάλματα.

Η διαφορά τους έγκειται στα εξής σημεία:

1. Στην εφαρμογή της Bootstrap ζητήσαμε 2.500 επαναλήψεις. Προσπαθήσαμε να τρέξουμε την διαδικασία των 2.500 επαναλήψεων αρκετές φορές (δηλαδή να αναπαράγουμε πολλές φορές 2.500 επαναλήψεις). Κάθε φορά οι 2.500 επαναλήψεις μας έβγαζαν τόσο διαφορετικό απόθεμα όσο και διαφορετικό μέσο τετραγωνικό σφάλμα. Αυτό που έμενε σταθερό – με μικρέ αποκλίσσεις- είναι η αναλογία του MSE με το απόθεμα. Κυμαίνονταν από 12% -15%.
2. Η Bootstrap μέθοδος σε κάθε επανάληψη και σε κάθε έτος εξέλιξης ζητάει - από μία την κατανομή που προσομοιάζουμε – μία τυχαία τιμή. Αυτό σημαίνει ότι μπορεί να φέρει πολύ μεγάλες έως πολύ μικρές. Επιπλέον, η Bootstrap φαίνεται να δίνει πιο μεγάλες τιμές στα πρώτα χρόνια ατυχήματος, ενώ όσο πλησιάζουμε προς τα τελευταία έτη ατυχήματος δίνει όλο και πιο μικρές τιμές. Αντίστοιχη πορεία ακολουθεί και το μέσο τετραγωνικό σφάλμα ανά έτος ατυχήματος. Οπότε φαίνεται λογικό το συνολικό MSE της Bootstrap να είναι λίγο μεγαλύτερο από της Mack μεθόδου.
3. Η κατανομή που επιλέγουμε είναι βαριάς ουράς. Σε κατανομές με μέτρια ουρά θα είχαμε άλλα αποτελέσματα. Ωστόσο, επιλέξαμε μια κατανομή με βαριά ουρά καθώς οι μεγάλες ζημιές είναι ένα σημείο έντονης μεταβλητότητας για τις ασφαλιστικές και τις μεθόδους αποθεματοποίησης που χρησιμοποιούν.

5.10 Ανθεκτική Bootstrap Method με έκτοπες τιμές

Τέλος, στην ενότητα αυτή, θα δείξουμε αν και πώς αλλάζει το ο υπολογισμός του αποθέματος με τη μέθοδο Bootstrap αν στο τρίγωνο εξέλιξης ζημιών υπάρχουν έκτοπες παρατηρήσεις. Ως τρίγωνο με έκτοπες παρατηρήσεις χρησιμοποιήθηκε το τρίγωνο εξέλιξης ζημιών της ενότητας (*Ανθεκτική Mack Method με 2 έκτοπες παρατηρήσεις*) για να μπορέσουμε να συγκρίνουμε και τα αποτελέσματα. Ο αλγόριθμος της μεθόδου Bootstrap παραμένει ο ίδιος με την μόνη διαφορά να έγκειται στο ότι ανθεκτικοποιήσαμε τη μέθοδο με τον τρόπο που περιγράψαμε στο Κεφάλαιο 3.

Τα αποτελέσματα θα παρουσιαστούν σε ένα πίνακα μαζί με τα αντίστοιχα της ανθεκτικής Mack μεθόδου ώστε να μπορέσουμε να κάνουμε και μία σύγκριση (βλ. Πίνακας 27). Επίσης, θα παρουσιάσουμε σε ξεχωριστό πίνακα τα αποτελέσματα όλων των μεθόδων ώστε να μπορέσει να γίνει πιο κατανοητό πώς οι μέθοδοι Mack και Bootstrap παρουσιάζουν κοντινά αποτελέσματα με τις όποιες ιδιαιτερότητες της κάθε μία να μην επηρεάζουν σημαντικά το αποτέλεσμα (Πίνακας 28).

Πίνακας 32: Αποτελέσματα Ανθεκτικών Μεθόδων σε τρίγωνο με 2 έκτοπες παρατηρήσεις

Robust Mack & Robust Bootstrap Method Αποτελέσματα

	Mack Method Res	Mack MSE	Bootstrap Res	Bootstrap MSE
	19.568.056	3.553.764	25.976.137	4.522.987
% of Res	18%		17%	

Ενώ σε ένα συγκεντρωτικό πίνακα των αποτελεσμάτων έχουμε ότι:

Πίνακας 33: Συνολικά Αποτελέσματα

Συγκεντρωτικά Αποτελέσματα

Method	Reserve / MSE	% of Res
Simple Mack	**18.635.518**	13%
Simple Mack MSE	**2.448.277**	
Simple Bootstrap	**24.408.811**	12%
Simple Bootstrap MSE	**2.942.910**	
Robust Mack	**19.568.056**	18%
Robust Mack MSE	**3.553.764**	
Robust Bootstrap	**25.976.137**	17%
Robust Bootstrap MSE	**4.522.987**	

Όπως βλέπουμε, οι δύο μέθοδοι δίνουν συγγενικά αποτελέσματα, κυρίως όσον αφορά την απλή και την ανθεκτική τους μορφή. Βλέπουμε ότι τόσο η Mack όσο και η Bootstrap στην ανθεκτική τους μορφή δίνουν πολύ κοντινά αποτελέσματα με τη απλή μορφή που έχουν. Η δυνατότητα αυτή καθιστά τις ανθεκτικές μορφές των μεθόδων χρήσιμες για τον ερευνητή αλλά και για μία ασφαλιστική, η οποία θέλει να προβεί σε υπολογισμό των αποθεμάτων της χωρίς να επηρεάζονται οι υπολογισμοί από έκτοπες παρατηρήσεις (είτε αυτές αναφέρονται σε λάθη καταγραφής είτε σε μεγάλες ζημιές που προέκυψαν).

Φυσικά, στη βιβλιογραφία αλλά και στην ευρύτερη ασφαλιστική αγορά χρησιμοποιούνται και άλλες μέθοδοι (πχ. Bornhuetter-Ferguson method, Loss ratio method, Separation Method, Bayesian Method), οι οποίες είναι αποδεκτές τόσο σε επίπεδο μαθηματικής ακρίβειας όσο και σε επίπεδο ρυθμιστικών αρχών (ΤτΕ, EIOPA). Ο ερευνητής ή η ασφαλιστική εταιρία διαλέγει εκείνη τη μέθοδο που πιστεύει ότι εξυπηρετεί καλύτερα τα δεδομένα που διαθέτει αλλά και με βάση την εμπειρία που διαθέτει. Ωστόσο, είναι πολλές οι φορές που χρησιμοποιούνται δύο ή περισσότερες μέθοδοι, ώστε ο ερευνητής να έχει μία πιο ολοκληρωμένη εικόνα του αποθέματος που πρέπει να κρατήσει.

6 Συμπεράσματα

Από την εκπόνηση της εν λόγω διπλωματικής και την εφαρμογή των παραπάνω μεθόδων, προκύπτουν κάποια συμπεράσματα, χρήσιμα για αυτούς που ασχολούνται με τις μεθόδους αποθεματοποίησης και κατανοούν τις όποιες αδυναμίες τους.

1. Η μέθοδος Chain Ladder παρόλη την απλότητά της και την εγκυρότητά της, είναι αρκετά ευαίσθητη στην εμφάνιση ακραίων τιμών ή μεγάλων αλμάτων. Επομένως, γίνεται αδήριτη η ανάγκη επιβεβαίωσης της Chain ladder τόσο με άλλες μεθόδους (όπως μέθοδος BF, μέθοδος Διαχωρισμού, μέθοδος Bootstrap, ή κάποιας άλλης). Κατ' επέκταση, οι όποιες διαφορές προκύψουν πρέπει να μετρηθούν και να αξιολογηθούν στα πλαίσια στατιστικών μεγεθών που δικαιολογούν τις προκύπτουσες αποκλίσεις. (όπως πχ. το MSE). Ωστόσο, όλες οι μέθοδοι επηρεάζονται σημαντικά από μεγάλες ζημιές (που είτε όντως προέκυψαν είτε αποτέλεσαν λάθη) πράγμα που καθιστά απαραίτητη την ανθεκτικοποίηση των μεθόδων. Μάλιστα στη διεθνή βιβλιογραφία, έχουν αναπτυχθεί ανθεκτικές μορφές τόσο για την BF μέθοδο όσο και για τη μέθοδο Διαχωρισμού.
2. Οι ανθεκτικές μορφές της Chain Ladder, είτε στη μορφή που παρουσίασαν οι Verdonck et al. (2009) είτε στη μορφή τής με τη χρήση ανάλυσης παλινδρόμησης (Renshaw, 1989) δίνουν συγγενικά αποτελέσματα. Επίσης, όπως δείξαμε, δίνουν αποτελέσματα πολύ κοντά σε εκείνα που θα είχαμε, αν δεν υπήρχαν έκτοπες παρατηρήσεις. Επομένως, μπορούν να χρησιμοποιηθούν και στην πράξη, μπορεί όχι σαν κύριες μέθοδοι αποθεματοποίησης αλλά με ένα τρόπο ελέγχου και διασταύρωσης των αποτελεσμάτων που έδωσαν οι κλασσικές μέθοδοι.
3. Εξίσου σημαντικό ως συμπέρασμα είναι ότι σε κάθε εκτίμηση του αποθέματός, πρέπει να γίνεται ένας ενδελεχής έλεγχος για την ποιότητα των δεδομένων. Τόσο σε επίπεδο πρακτικής εφαρμογής (βλ. ασφαλιστική αγορά) όσο και σε επίπεδο ερευνητικό, πρέπει να γίνεται λεπτομερής έλεγχος για το αν τα δεδομένα είναι αληθή και έχει γίνει διαχωρισμός των μεγάλων ζημιών από τις συνήθεις και εύκολα διαχειρίσιμες ζημιές. Τυχόν παράληψη τέτοιων ελέγχων μπορεί να επιφέρει κυρώσεις από τις ρυθμιστικές αρχές (αν μιλάμε για μια ασφαλιστική εταιρία) ή σε λάθος συμπεράσματα (αν μιλάμε για ερευνητικό θέμα). Όπως είναι φυσικό, προκύπτει η ανάγκη να επιλεγεί κατάλληλο μοντέλο φιλτραρίσματος των δεδομένων (όπως πχ παρουσιάστηκαν στο Κεφάλαιο 3). Η όποια μέθοδος φιλτραρίσματος των δεδομένων πρέπει να είναι τέτοια που να μην οδηγεί σε εσφαλμένη εκτίμηση αποθέματος (είτε υπερ-αποθεματοποίηση είτε υπό-αποθεματοποίηση, αν μπορούσαν να επιτραπούν οι όροι) και επιπλέον να μην αφαιρεί πληροφορίες, που είναι χρήσιμες, από τα δεδομένα.

4. Ειδικότερα σε επίπεδο ασφαλιστικών εταιριών, πέραν του αναλογιστικού τμήματος υπάρχουν και οι επιμέρους Κλάδοι-Διευθύνσεις (πχ Κλάδος Ζωής, Κλάδος Αυτοκινήτου, Κλάδος Μεταφορών κ.α.), που είναι επιφορτισμένα με την στενή παρακολούθηση των μεγάλων ζημιών των τμημάτων αυτών και οι υπάλληλοί τους διαθέτουν κατάρτιση και εμπειρία για να τις διαχειριστούν. Απαιτείται η στενή συνεργασία, μεταξύ των Κλάδων και του Αναλογιστικού τμήματος ώστε να αποσαφηνιστούν οι μεγάλες ζημιές και να συμφωνηθεί η κατάλληλη διαχείρισή τους και από τις δύο πλευρές.

7 Αναφορές

7.1 Ελληνική

[1] Πιτσέλης Γ., (2018). Θεωρία Αποθεματοποίησης και Αξιοπιστία Χαρτοφυλακίου, Σημειώσεις Μαθήματος.

[2] ΕΑΕΕ, (2016), Φερεγγυότητα ΙΙ.

7.2 Ξενόγλωσση

[1] Bradu D., Mundlak Y., (1970). Estimation in lognormal linear models. J. Amer. Statist. Assoc. 6, 198–211.

[2] Busse M., Müller U., Dacorogna M., (2010). Robust Estimation of Reserve Risk. Astin Bulletin, 40[2]: 459-489.

[3] Christofides S., (1990). Regression Models based on Log-Incremental Payments. Claims Reserving Manual 2. Institute of Actuaries, London.

[4] Efron B., (1979). Bootstrap Methods. Another Look at the Jackknife. The Annals of Statistics, 7 [1]: 1-26.

[5] England P., Verrall R., (1999). Analytic and Bootstrap estimates of prediction errors in claims reserving. Insurance: Mathematics and Economics, 25[3]: 281-293.

[6] England P., Verrall R., (2002). Stochastic claims reserving in general insurance. British Actuarial Journal, 8[3]: 443-518.

[7] England P., Verrall R., (2006). Predictive distributions of outstanding liabilities in general insurance. Annals of Actuarial Science, 2[1]: 221-270.

[8] Finney D.J. (1941). On the distribution of a variate whose logarithm is normally distributed. Supp. J. Royal Stat. Soc. 7[2]:155–161.

[9] Gisler A., Wüthrich V. M., (2008). Credibility for the Chain Ladder Reserving Method. Astin Bulletin, 38[2]: 565-600.

[10] Leong W. K., Wang S.S. & Chen H., (2014). Back-testing the ODP bootstrap of the paid chain-ladder model with actual historical claims data. Variance, 8[2], 182–202.

[11] Hampel F. R., Ronchetti E. M., Rousseeuw P. J., Stahel W. A., (1986). Robust Statistics. The Approach Based on Influence Functions. New York: John Wiley and Sons.

[12] Huber J. P., (1973). Robust Regression Asymptotics, Conjectures and Monte Carlo. The Annals of Statistics, 1: 799-821.

[13] Huber P.J., (1981). Robust Statistics. Wiley, New York.

[14] Huber P.J., Dutter R., (1974). Numerical solutions of robust regression. In: Bruckman, G. (Ed.), COMPSTAT 1974, Proc. in Computational Statistics. Physika Verlag, Vienna.

[15] Kremer E., (1982). IBNR-Claims and the two-way model of ANOVA. Scand. Actuar. J. 1, 47–55.

[16] Mack T., (1993). Distribution-free calculation of the standard error of chain ladder reserve estimate. Astin Bulletin, 23[1]: 214-225.

[17] Mccullagh P., and Nelder J., (1989). Generalized linear models. Chapman and Hall.

[18] Pavlović B., (2012). Unexpired Risk Reserve, Delta Generali osiguranje a.d.o. Beograd.

[19] Pitselis G., Grigoriadou V., Badounas I., (2015). Robust loss reserving in a log-linear model. Insurance: Mathematics and Economics, 64: 14-27.

[20] Renshaw A., (1989). Chain Ladder and Interactive Modelling, Journal of the Institute of Actuaries, 116: 559-587.

[21] Rousseeuw P.J., Leroy A., (1987). Robust Regression and Outlier Detection. Wiley, New York, Wiley

[22] Schnieper R., (1991). Separating True IBNR and IBNER Claims. Astin Bulletin, 21[1]: 111-127.

[23] Taylor G.C., Ashe F.R., (1983). Second moments of outstanding claims. Journal of Econometrics, 23[1]: 37-61.

[24] Tukey J.W., (1977). Explanatory Data Analysis. Addison-Wesley.

[25] Verrall R. J., (1991). On the Estimation of Reserves from Loglinear Models. Insurance: Mathematics and Economics 10: 75–80.

[26] Verdonck T., Van Wouwe M., and Dhaene J., (2009). A Robustification of the Chain-Ladder Method. North American Actuarial Journal, 13[2]: 280-298.

[27] Werner G., Modlin C., (2016). Basic Ratemaking, Fifth Edition, CAS.

[28] Zehnwirth B., (1985). Interactive Claims Reserving Forecasting System. lnsureware P/L, E. St. Kilda Vic 3183, Australia.

ΠΑΡΑΡΤΗΜΑ Α

Ο Αλγόριθμος του Μ-εκτιμητή

Όπως έχουμε αναφέρει πολλές φορές, στην ανάλυση παλινδρόμησης έχουμε σα σκοπό, την ελαχιστοποίηση μιας εκτιμήτριας συνάρτησης που θα μας δώσει τις (εκτιμώμενες) τιμές των αγνώστων παραμέτρων. Έτσι και εδώ στον Μ-εκτιμητή θέλουμε να ελαχιστοποιήσουμε τη σχέση παραπάνω. Προκειμένου να το γίνει αυτό, οι Huber and Dutter (1974) πρότειναν έναν αλγόριθμο που ελαχιστοποιεί τη σχέση αυτή. Ο αλγόριθμος αυτός παρότι είναι απλός στον υπολογισμό του, έχει ένα μειονέκτημα. Συγκλίνει με αργό ρυθμό, πράγμα που δε συνάδει με την ασυμπτωτική συμπεριφορά που θέλουμε. Οφείλουμε να επισημάνουμε ότι ο αλγόριθμος βασίζεται σε επαναληπτικές μεθόδους. Πιο συγκεκριμένα, η πιο συνήθης μέθοδος υπολογισμού Μ-εκτιμητή στηρίζεται στον αλγόριθμο Newton-Raphson.

Υποθέσεις του αλγορίθμου

- Θεωρούμε ένα επίπεδο ανεκτικότητας ε>0
- Η αρχική τιμή για το σ ορίζεται $\sigma^{(0)}$.
- Αν οι προβλεπόμενες τιμές $x_i^T\beta$ είναι γραμμικές, τότε -σαν αρχική τους τιμή- μπορούμε να χρησιμοποιήσουμε την εκτιμήτρια ελαχίστων τετραγώνων.

Παρουσιάζουμε τώρα τον αλγόριθμο των Huber and Dutter (1974):

1. Για το $\beta^{(0)}$ διάνυσμα, θεωρούμε ως αρχικές τιμές, την τιμή του εκτιμητή ελαχίστων τετραγώνων.
2. Υπολογίζουμε τα υπόλοιπα: $r_i^m = y_i - x_i^T\beta^{(m)}$, για $i=1,2,...,n$.
3. Υπολογίζουμε μια νέα τιμή για το σ:

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

με το Α να είναι ένας διορθωτικός παράγοντας, με τον οποίον πετυχαίνουμε αμεροληψία.

4. Υπολογίζουμε τα υπόλοιπα r_i^* :

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

5. Υπολογίζουμε τις μερικές παραγώγους:

$$x_{ik} = \frac{\partial}{\partial\beta}(x_i^T\beta)$$

6. Λύνουμε την εξίσωση $X^\tau X_{\hat{\tau}} = X^T r^*$ ως προς $\hat{\tau}$. Η τελευταία προκύπτει από την λύση της ακόλουθης:

$$\min\sum(r_i - \sum x_{ik}\hat{\tau}_\kappa)^2$$

7. Θέτουμε

$$\beta^{(m+1)} = \beta^{(m)} + q\hat{\tau}$$

με $q \in (0,2)$ να είναι ένας αυθαίρετος σταθερός παράγοντας.

8. Όταν οι εκτιμημένοι συντελεστές συγκλίνουν τότε σταματάει η επανάληψη
9. Σε αντίθετη περίπτωση θέτουμε m:=m+1 και ξεκινάμε πάλι από το βήμα 2.

ΠΑΡΑΡΤΗΜΑ Β

ΚΩΔΙΚΑΣ R

CHAIN LADDER χωρίς έκτοπη τιμή (wo Outlier)

```
Run-off Triangle of Incremental Losses----------------------------------------------------------------------------------

tri<-t(matrix(c(357.856,766.948,610.550,482.948,527.334,574.406,146.350,139.958,227.237,67.956,
352.126,884.029,993.902,1183.297,445.753,321.004,527.812,266.180,425.054,NA,
290.515,1001.807,926.227,1016.662,750.824,146.931,496.00,280.413,NA,NA,
310.616,1108.258,776.197,1562.408,272.490,352.061,206.294,NA,NA,NA,
443.168,693.198,991.991,769.496,504.859,470.647,NA,NA,NA,NA,
396.140,937.093,847.506,805.045,705.968,NA,NA,NA,NA,NA,
440.840,847.639,1131.406,1063.277,NA,NA,NA,NA,NA,NA,
359.488,1061.656,1443.378,NA,NA,NA,NA,NA,NA,NA,
376.694,986.616,NA,NA,NA,NA,NA,NA,NA,NA,
344.022,NA,NA,NA,NA,NA,NA,NA,NA,NA),ncol=10,dimnames=list(dev=1:10,origin=2008:2017)))*1000;tri

#Cumulative Losses---------------------------------------------------------------------------------------------------------------
For (i in 1:10) {
For ( j in 2:10){
tri[i,j]<-tri[i,j]+tri[i,j-1]
}
}
tri



tri
#Development Factors---------------------------------------------------------------------------------------------------------------
m<-c(1:(dim(tri)[1]-1)); m
for( i in 2:10)
{
m[i-1]<-sum(tri[1:(10-i+1),i])/sum(tri[1:(10-i+1),i-1])
}
m

#Fulltri the Runoff triangle to be filled------------------------------------------------------------------------------------
fulltri<-tri;fulltri

# Filling the run-off triangle
for(i in 1:dim(tri)[1])
{
for(j in 2:dim(tri)[1])
{
if(is.na(fulltri[i,j]))
{
fulltri[i,j]<-fulltri[i,j-1]*m[j-1]
}
}
```

```
}
Fulltri

# Res: the vector of future losses
res<-c(1:(dim(tri)[1]-1))

#Sum of vector Res gives the reserve-----------------------------------------------------------------------------------------
for(i in 1:10){
res[i]<-fulltri[i,10]-fulltri[i,10-i+1]
}
res
sum(res)
```

Παρατήρηση: Αλλάζοντας τις τιμές 445,753 σε 445.753 και 1001,807 σε 1001.807 παίρνουμε το τρίγωνο εξέλιξης ζημιών με δύο έκτοπες τιμές (with outliers). Ο κώδικας είναι ίδιος με της Chain Ladder πιο πάνω.

Robust Chain Ladder

```
# Run-off Triangle with incremental Losses------------------------------------------------------------------------------------
tri<-t(matrix(c(357.856,766.948,610.550,482.948,527.334,574.406,146.350,139.958,227.237,67.956,
352.126,884.029,993.902,1183.297,445.7530,321.004,527.812,266.180,425.054,NA,
290.515,1001807,926.227,1016.662,750.824,146.931,496.00,280.413,NA,NA,
310.616,1108.258,776.197,1562.408,272.490,352.061,206.294,NA,NA,NA,
443.168,693.198,991.991,769.496,504.859,470.647,NA,NA,NA,NA,
396.140,937.093,847.506,805.045,705.968,NA,NA,NA,NA,NA,
440.840,847.639,1131.406,1063.277,NA,NA,NA,NA,NA,NA,
359.488,1061.656,1443.378,NA,NA,NA,NA,NA,NA,NA,
376.694,986.616,NA,NA,NA,NA,NA,NA,NA,NA,
344.022,NA,NA,NA,NA,NA,NA,NA,NA,NA),ncol=10,dimnames=list(dev=1:10,origin=2008:2017)))*1000;tri

#Store incremental triangle in another matrix---------------------------------------------------------------------------------
tri1<-tri


#Cumulative Payments---------------------------------------------------------------------------------------------------------
for(i in 1:10) {
for( j in 2:10){
tri[i,j]<-tri[i,j]+tri[i,j-1]
}
}
tri

#Year-to-year development factors-------------------------------------------------------------------------------------------
fy<-matrix(c(NA,NA,NA,NA,NA,NA,NA,NA,NA), nrow=9,ncol=9)
for( k in 1:9)
{
for(j in 1:(10-k))
{
fy[k,j]<-tri[k,j+1]/tri[k,j]
}
}
fy
```

```
#Development factors using the median. Also used for residuals and E (see below)
m<-c(1:(dim(tri)[1]-1))
m
for( i in 1:9)
{
m[i]<-median(fy[(1:(10-i)),i])
}
m

#ckeck---------------------------------------------------------------------------------------------------------------------------
median(fy[(1:8),2])

#Fitted culumative losses run off triangle---------------------------------------------------------------------------------------
E<-matrix(c(NA,NA,NA,NA,NA,NA,NA,NA,NA,NA),ncol=10,nrow=10)
for( i in 1:10)
{
E[i,11-i]<-tri[i,11-i]
}
E

# Fitted cumulative losses for the past years------------------------------------------------------------------------------------
for( i in 1:9)
{
for( j in (10-i):1)
{
E[i,j]<-E[i,j+1]/m[j]
}
}
E

# Incremental of fitted values---------------------------------------------------------------------------------------------------
for (i in 1:10)
{
for( j in 10:2)
{
E[i,j]<-E[i,j]-E[i,j-1]
}
}
E<-round(E,0);E

tri1
# Pearson Residuals--------------------------------------------------------------------------------------------------------------
red<-matrix(c(NA,NA,NA,NA,NA,NA,NA,NA,NA,NA), nrow=10,ncol=10)

for( k in 1:10)
{
for( j in 1:(11-k))
{
red[j,k]<-sqrt(55/36)*(tri1[j,k]-E[j,k])/sqrt(E[j,k])
}
}
red
```

```
#Boxplot for residuals------------------------------------------------------------------------------------------------------------
boxplot(red)

Normality Test--------------------------------------------------------------------------------------------------------------------
shapiro.test(red)

l<-as.vector(red)
boxplot(l)
summary(l)

#Threshold for outliers------------------------------------------------------------------------------------------------------------
a<-85.07-(-233.63);a
dist<-1.5*(a);dist

#turn NA in 0----------------------------------------------------------------------------------------------------------------------

for( i in 1:10)
{
for(j in 1:10)
{
if(is.na(red[i,j]))
{
red[i,j]<-0
}
}
}
Red

#Indicator matrix all for outliers-------------------------------------------------------------------------------------------------
all<-matrix(c(1,1,1,1,1,1,1,1,1,1),ncol=10,nrow=10)
for( i in 1:10)
{
for( j in 1:10)
{
if(red[i,j]>=85.07+dist || red[i,j]<=-233.63-dist)
{
all[i,j]<-0
}
}
}
all

#Detecting outlier's position------------------------------------------------------------------------------------------------------

for( i in 1:10)
{
if(all[i,1]==0)
{
tri[i,1]<-median(tri[,1])
}
}
tri

#Replace first column from incremental losses with column from new run off triangle-----------------------------
tri1
```

```
tri1[,1]<-tri[,1];tri1

#Development factors using incremental losses of first column median-----------------------------------------------

#fy1 year-to-year development factos using first column of incremental losses (tri1) --------------------------------
fy1<-matrix(c(NA,NA,NA,NA,NA,NA,NA,NA,NA), nrow=9,ncol=9)
for( k in 1:9)
{
for(j in 1:(10-k))
{
fy1[k,j]<-tri1[k,j+1]/tri1[k,1]
}
}
fy1

#new development factors using median and matrix fy1---------------------------------------------------------------
m1<-c(1:(dim(tri)[1]-1))
m1
for( i in 1:9)
{
m1[i]<-median(fy1[(1:(10-i)),i])
}
m1

# tri2 matrix using first column of tri1 matrix and development factors---------------------------------------------
tri2<-matrix(c(rep(NA,10)), nrow=10,ncol=10);tri2
tri2[,1]<-tri1[,1];tri2

for(i in 1:9)
{
for(j in 2:(10-i+1))
{
tri2[i,j]<-tri2[i,1]*m1[j-1]
}
}
tri2

#Pearson residuals---------------------------------------------------------------------------------------------------
red2<-matrix(c(NA,NA,NA,NA,NA,NA,NA,NA,NA,NA), nrow=10,ncol=10)

for( k in 1:10)
{
for( j in 1:(11-k))
{
red2[j,k]<-sqrt(55/36)*(tri1[j,k]-tri2[j,k])/sqrt(tri2[j,k])
}
}
red2

l2<-as.vector(red2)

#boxplot for l2-----------------------------------------------------------------------------------------------------
boxplot(l2)
summary(l2)
```

```
Normality test------------------------------------------------------------------------------------------------------------------------
shapiro.test(red2)
a2<-99.6-(-166.8);a

dist2<-1.5*a;dist

#turn NA's into 0 red2------------------------------------------------------------------------------------------------------------

for( i in 1:10)
{
for(j in 1:10)
{
if(is.na(red2[i,j]))
{
red2[i,j]<-0
}
}
}
red2

#check for outliers using an indicator matrix ------------------------------------------------------------------------------
all2<-matrix(c(1,1,1,1,1,1,1,1,1,1),ncol=10,nrow=10)

for( i in 1:10)
{
for( j in 1:10)
{
if(red2[i,j]>=99.6+dist2 || red2[i,j]<=-166.8-dist2)
{
all2[i,j]<-0
}
}
}
all2

#turn 0's to NA's----------------------------------------------------------------------------------------------------------------

for( i in 10:2)
{
for (j in 10:(10-i+2))
{
if( red2[i,j]==0)
{
red2[i,j]<-NA
}
}
}
red2

median(red2,na.rm=TRUE)

# Replace outliers with median from red2---------------------------------------------------------------------------------
for ( i in 1:10)
{
for( j in 1:10)
```

```
{
if(all2[i,j]==0)
{
red2[i,j]<-median(red2,na.rm=TRUE)
}
}
}
red2

#backtransofrm residuals in a new run off triangle tri3-------------------------------------------------------------------

tri3<-matrix(c(rep(NA,10)), nrow=10,ncol=10);tri3

for( i in 1:10)
{
for( j in 1:10)
{
if(is.na(red2[i,j]))
{
tri3[i,j]<-tri2[i,j]
}
else
{
tri3[i,j]<-sqrt(36/55)*red2[i,j]*sqrt(tri2[i,j])+ tri2[i,j]
}
}
}
tri3

#Apply Classic Chain Ladder-------------------------------------------------------------------------------------------------------
#cumulative tri3----------------------------------------------------------------------------------------------------------------------
for(i in 1:10) {
for( j in 2:10){
tri3[i,j]<-tri3[i,j]+tri3[i,j-1]
}
}
tri3

fy3<-matrix(c(NA,NA,NA,NA,NA,NA,NA,NA,NA), nrow=9,ncol=9)
for( k in 1:9)
{
for(j in 1:(10-k))
{
fy3[k,j]<-tri3[k,j+1]/tri3[k,j]
}
}
fy3

m3<-c(1:(dim(tri)[1]-1))
m3
for( i in 1:9)
{
m3[i]<-median(fy3[(1:(10-i)),i])
}
m3
```

```
Reserve of tri3------------------------------------------------------------------------------------------------------------------------
fulltri3<-tri3;fulltri3

for(i in 1:dim(tri3)[1])
{
for(j in 2:dim(tri3)[1])
{
if(is.na(fulltri3[i,j]))
{
fulltri3[i,j]<-fulltri3[i,j-1]*m3[j-1]
}
}
}
fulltri3
res<-c(1:(dim(tri3)[1]-1))

for(i in 1:10){

res[i]<-fulltri3[i,10]-fulltri3[i,10-i+1]
}
res
sum(res)
```

Log-incremental Chain Ladder

```
#Run off triangle of incremental losses---------------------------------------------------------------------------------------------
tri<-t(matrix(c(357.856,766.948,610.550,482.948,527.334,574.406,146.350,139.958,227.237,67.956,
352.126,884.029,993.902,1183.297,445.753,321.004,527.812,266.180,425.054,NA,
290.515,1001.807,926.227,1016.662,750.824,146.931,496.00,280.413,NA,NA,
310.616,1108.258,776.197,1562.408,272.490,352.061,206.294,NA,NA,NA,
443.168,693.198,991.991,769.496,504.859,470.647,NA,NA,NA,NA,
396.140,937.093,847.506,805.045,705.968,NA,NA,NA,NA,NA,
440.840,847.639,1131.406,1063.277,NA,NA,NA,NA,NA,NA,
359.488,1061.656,1443.378,NA,NA,NA,NA,NA,NA,NA,
376.694,986.616,NA,NA,NA,NA,NA,NA,NA,NA,
344.022,NA,NA,NA,NA,NA,NA,NA,NA,NA),ncol=10,dimnames=list(dev=0:9,origin=2008:2017)))*1000; tri

m<-dim(tri)[1]; n<-dim(tri)[2];m;n

# claims as data-frame------------------------------------------------------------------------------------------------------------------
data<-data.frame(acc=rep(0:(n-1),each=n),dev=(rep(0:(m-1),m)),claims=as.vector(t(tri)));data

#setting logclaims------------------------------------------------------------------------------------------------------------------------
data<-with(data,data.frame(acc,dev,claims,logclaims=round(log(claims),3),accf=factor(acc)));data

rownames(data)<-with(data,paste(acc,dev,sep="-"));data

data<-
with(data,data.frame(acc,dev,claims,logclaims=round(log(claims),3),accf=factor(acc),devf=paste(dev)));data
```

```
dat<-data;dat

#apply linear regression model------------------------------------------------------------------------------------------------------
fit<-lm(logclaims~accf+devf,data=dat);fit
summary(fit)
fit
formula(fit)
model.frame(fit)

#model matrix-----------------------------------------------------------------------------------------------------------------------
dm<-model.matrix(formula(fit),dat=model.frame(fit))
dm
fdm<-model.matrix(~accf+devf,data=dat[is.na(dat$claims),])
fdm

coeffs<-as.matrix(summary(fit)$coefficients[,1]);coeffs

vcov(fit)

#Variance-Covariance matrix---------------------------------------------------------------------------------------------------------
varcovar<-fdm%*%vcov(fit)%*%t(fdm)

round(varcovar,4)

sigma<-summary(fit)$sigma
sigma

Var<-varcovar+sigma^2;Var

# future log claims-----------------------------------------------------------------------------------------------------------------
Z<-fdm%*%coef(fit)
Z
VarZ<-Var[row(Var)==col(Var)]
VarZ

#future cumulative claims backtransforming log claims
Y<-exp(Z+VarZ/2);Y

#variance of future claims---------------------------------------------------------------------------------------------------------
VarY<-Y^2*(exp(VarZ)-1);VarY

seY<-sqrt(VarY);seY

#Results----------------------------------------------------------------------------------------------------------------------------
i<-fdm%*%c((1:m)-1,rep(0,(n-1)))
i
j<-fdm%*%c(rep(0,(m-1)),(1:n)-1)

Results<-data.frame(i,j,Z,VarZ,Y,VarY,seY); Results

#total reserves---------------------------------------------------------------------------------------------------------------------
LLM.reserves<-sum(Y);LLM.reserves
```

Παρατήρηση: Η **Robust Chain Ladder** με χρήση του Μ-εκτιμητή του Huber προκύπτει ως εξής αντικαθιστώντας στην εφαρμογή του μοντέλου την συνάρτηση lm με rlm και method='M'.

Mack method -χωρίς έκτοπη Παρατήρηση

```
tri<-t(matrix(c(357.856,766.948,610.550,482.948,527.334,574.406,146.350,139.958,227.237,67.956,
352.126,884.029,993.902,1183.297,445.753,321.004,527.812,266.180,425.054,NA,
290.515,1001.807,926.227,1016.662,750.824,146.931,496.00,280.413,NA,NA,
310.616,1108.258,776.197,1562.408,272.490,352.061,206.294,NA,NA,NA,
443.168,693.198,991.991,769.496,504.859,470.647,NA,NA,NA,NA,
396.140,937.093,847.506,805.045,705.968,NA,NA,NA,NA,NA,
440.840,847.639,1131.406,1063.277,NA,NA,NA,NA,NA,NA,
359.488,1061.656,1443.378,NA,NA,NA,NA,NA,NA,NA,
376.694,986.616,NA,NA,NA,NA,NA,NA,NA,NA,
344.022,NA,NA,NA,NA,NA,NA,NA,NA,NA),ncol=10,dimnames=list(dev=1:10,origin=2008:2017)))*1000;tri

#Cumulative Losses------------------------------------------------------------------------------------------------------------
for(i in 1:10) {
for( j in 2:10){
tri[i,j]<-tri[i,j]+tri[i,j-1]
}
}
Tri

# Development Factors------------------------------------------------------------------------------------------------------------
m<-c(1:(dim(tri)[1]-1))
m
for( i in 2:10)
{
m[i-1]<-sum(tri[1:(10-i+1),i])/sum(tri[1:(10-i+1),i-1])
}
m

# Run off Triangle to be filtered------------------------------------------------------------------------------------------------
fulltri<-tri;fulltri

for(i in 1:dim(tri)[1])
{
for(j in 2:dim(tri)[1])
{
if(is.na(fulltri[i,j]))
{
fulltri[i,j]<-fulltri[i,j-1]*m[j-1]
}
}
}
fulltri
res<-c(1:(dim(tri)[1]-1))

for(i in 1:10){

res[i]<-fulltri[i,10]-fulltri[i,10-i+1]
}
res
sum(res)
```

```
#Calculating process variance multiplier------------------------------------------------------------------------------------------------
f1<-c(0,0,0,0,0,0,0,0,0,1)
f<-c(0,0,0,0,0,0,0,0,0)
f
for( i in 9:1)
{
f1[i]<-f1[i+1]*m[i]
}
f1
f<-f1[1:9]
f

#Calculating σ(κ)^2-----------------------------------------------------------------------------------------------------------------

s2<-c(0,0,0,0,0,0,0,0,0)
s1<-c(0,0,0,0,0,0,0,0,0,0);s1
for(k in 1:(dim(tri)[1]-2))
{
for (j in 1:(dim(tri)[1]-k))
{
s1[k]<- s1[k]+ fulltri[j,k]*((fulltri[j,k+1]/fulltri[j,k])-m[k])^2
}
}
s1
for (i in 1: (length(s2)-1))
{
s2[i]<-(1/(dim(tri)[1]-i-1))*s1[i]
}
s2
s2[length(s2)]<-min((s2[length(s2)-1]^2/s2[length(s2)-2]),min(s2[length(s2)-2],s2[length(s2)-1]))
s2



s21<-c(0,0,0,0,0,0,0,0,0,0)
for( i in 1:9)
{
s21[i]<-f[i]*s2[i]/(m[i]^2)
}
s21

s22<-c(0,0,0,0,0,0,0,0,0,0)
s22[9]<-s21[9]
s22
for( i in (8:1))
{
s22[i]<-s21[i]+s22[i+1]
}
s22

#Calculating parameter multiplier----------------------------------------------------------------------------------------------------
s33<-c(0,0,0,0,0,0,0,0,0,0)
s3<-c(0,0,0,0,0,0,0,0,0,0)
for( i in (1:10))
{
```

```
s33[i]<-(s2[11-i]/(m[11-i]^2))*(1/sum(fulltri[1:(i-1),(11-i)]))
}
s33
s33<-rev(s33)
s3<-s33[1:9]
s3
for( i in (8:1))
{
s3[i]<-s3[i]+s3[i+1]
}
s3


#Calculating MSE---------------------------------------------------------------------------------------------------------------
r1<-c(0,0,0,0,0,0,0,0,0)
r2<-matrix(c(0,0,0,0,0,0,0,0,0), nrow=9,ncol=9)
for( j in (1:10))
{
r1[j]<-sqrt(((fulltri[j,10])*(s22[11-j])))
}
r1

#Calculating Var-Covar matrix--------------------------------------------------------------------------------------------------
s333<-matrix(c(0,0,0,0,0,0,0,0,0),nrow=9,ncol=9);s333

# Filling the half and down var-covar matrix------------------------------------------------------------------------------------
for (i  in (1:9))
{
for( j in (i:1))
{
s333[i,j]<-s3[10-j]
}
}
s333
Filling the half and up var-covar matrix--------------------------------------------------------------------------------------
for ( i in (1:8))
{
for (j in ((i+1):9))
{
s333[i,j]<-s3[10-i]
}
}
s333

Calculating MSE---------------------------------------------------------------------------------------------------------------
for(i in (2:9))
{
for( j in (2:9))
{
r2[i,j]<-fulltri[i,10]*fulltri[j,10]*s333[i,j]
}
}
r2
r22<-r2[2:9,2:9]
r22
```

```
sum(r22)
sqrt(sum(r22))
sqrt(v1+sum(r22))
r11<-c(r1[2:10]);r11

r11<-r11^2;r11
v1<-sum(r11)^0.5;v1
v2<-(sum(r22))^0.5;v2
sqrt((v1^2)+(v2^2))
sqrt((v1^2)+sum(r22))

#Calculating sd per accident year------------------------------------------------------------------------------------------------
pm<-c(rep(0,9));pm
for ( i in 1:9)
{
pm[i]<-fulltri[i+1,10]*sqrt(s3[10-i])
}
pm

total sd
sdt<-c(rep(0,9));sdt
for ( i in 1:9)
{
sdt[i]<-sqrt(r1[i+1]^2 + pm[i]^2)
}
sdt
Robust Mack Method- με έκτοπη παρατήρηση
Using run off triangle as above------------------------------------------------------------------------------------------------------

tri1<-tri
#Cumulative losses-------------------------------------------------------------------------------------------------------------------
for(i in 1:10) {
for( j in 2:10){
tri[i,j]<-tri[i,j]+tri[i,j-1]
}
}
Tri

#Development factors (unfiltered)-----------------------------------------------------------------------------------------------
m<-c(1:(dim(tri)[1]-1))
m
for( i in 2:10)
{
m[i-1]<-sum(tri[1:(10-i+1),i])/sum(tri[1:(10-i+1),i-1])
}
m

#Run off triangle to be filtered------------------------------------------------------------------------------------------------------
fulltri<-tri;fulltri

for(i in 1:dim(tri)[1])
{
for(j in 2:dim(tri)[1])
{
if(is.na(fulltri[i,j]))
```

```
{
fulltri[i,j]<-fulltri[i,j-1]*m[j-1]
}
}
}
fulltri
res<-c(1:(dim(tri)[1]-1))

for(i in 1:10){

res[i]<-fulltri[i,10]-fulltri[i,10-i+1]
}
res
sum(res)

# Detecting outliers Using Threshold b----------------------------------------------------------------------------------------
b<-10

# Year to year development factors--------------------------------------------------------------------------------------------
fm<-matrix(c(rep(0,9)),nrow=9,ncol=9);fm

for( i in 1:9)
{
for( j in 2:10)
{
fm[i,j-1]<-tri[i,j]/tri[i,j-1]
}
}
fm

for ( i in 1:9)
{
for(j in 2:9)
{
if(is.na(fm[i,j]))
{
fm[i,j]<-0
}
}
}
fm

#Indicator Matrix for outliers-----------------------------------------------------------------------------------------------------
all<-matrix(c(1,1,1,1,1,1,1,1,1,1),ncol=10,nrow=10)

for( i in 1:9)
{
for( j in 1:9)
{
if(fm[i,j]>=5*m[j])
{
all[i,j+1]<-0
}
}
}
```

```
all

# Filtering first run off triangle from outliers and calculating reserve-------------------------------------------------------
tri1<-tri1*all;tri1

# Cumulative losses--------------------------------------------------------------------------------------------------------------------
for(i in 1:10) {
for( j in 2:10){
tri1[i,j]<-tri1[i,j]+tri1[i,j-1]
}
}
tri1

#Development factors (filtered)------------------------------------------------------------------------------------------------------
m1<-c(1:(dim(tri)[1]-1))
m
for( i in 2:10)
{
m1[i-1]<-sum(tri1[1:(10-i+1),i])/sum(tri1[1:(10-i+1),i-1])
}
m1

fulltri1<-tri1;fulltri1

for(i in 1:dim(tri)[1])
{
for(j in 2:dim(tri)[1])
{
if(is.na(fulltri1[i,j]))
{
fulltri1[i,j]<-fulltri1[i,j-1]*m1[j-1]
}
}
}
fulltri1
res<-c(1:(dim(tri)[1]-1))

for(i in 1:10){

res[i]<-fulltri1[i,10]-fulltri1[i,10-i+1]
}
res
sum(res)

#Calculating fitted losses-------------------------------------------------------------------------------------------------------------
E<-matrix(c(NA,NA,NA,NA,NA,NA,NA,NA,NA,NA),ncol=10,nrow=10)
for( i in 1:10)
{
E[i,11-i]<-tri1[i,11-i]
}
E

for( i in 1:9)
{
for( j in (10-i):1)
```

```
{
E[i,j]<-E[i,j+1]/m1[j]
}
}
E

for(i in 1:dim(tri)[1])
{
for(j in 2:dim(tri)[1])
{
if(is.na(E[i,j]))
{
E[i,j]<-E[i,j-1]*m1[j-1]
}
}
}
E

#Calculating process variance multiplier------------------------------------------------------------------------------------------
f1<-c(0,0,0,0,0,0,0,0,0,1)
f<-c(0,0,0,0,0,0,0,0,0)
f
for( i in 9:1)
{
f1[i]<-f1[i+1]*m1[i]
}
f1
f<-f1[1:9]
f


#Calculating σ(κ)^2----------------------------------------------------------------------------------------------------------------

s2<-c(0,0,0,0,0,0,0,0,0)
s1<-c(0,0,0,0,0,0,0,0,0,0);s1
m1

for(k in 1:(dim(tri)[1]-2))
{
for (j in 1:(dim(tri)[1]-k))
{
s1[k]<- s1[k]+ (1/E[j,k])*(fulltri1[j,k+1]-fulltri1[j,k]*m1[k])^2
}
}
s1
fulltri1
for (i in 1: (length(s2)-1))
{
s2[i]<-(1/(dim(tri)[1]-i-1))*s1[i]
}
s2

s2[length(s2)]<-min((s2[length(s2)-1]^2/s2[length(s2)-2]),min(s2[length(s2)-2],s2[length(s2)-1]))
s2
s21<-c(0,0,0,0,0,0,0,0,0)
```

```
for( i in 1:9)
{
s21[i]<-f[i]*s2[i]/(m1[i]^2)
}
s21
s22<-c(0,0,0,0,0,0,0,0,0)
s22[9]<-s21[9]
s22
#s22 einai o process multiplier, pou tha ton vrw amesws twra
for( i in (8:1))
{
s22[i]<-s21[i]+s22[i+1]
}
s22

#Calculating parameter multiplier------------------------------------------------------------------------------------------------
s33<-c(0,0,0,0,0,0,0,0,0)
s3<-c(0,0,0,0,0,0,0,0,0)
for( i in (1:10))
{
s33[i]<-(s2[11-i]/(m1[11-i]^2))*(1/sum(fulltri1[1:(i-1),(11-i)]))
}
s33
s33<-rev(s33)
s3<-s33[1:9]
s3
for( i in (8:1))
{
s3[i]<-s3[i]+s3[i+1]
}
s3

#Calculating MSE
r1<-c(0,0,0,0,0,0,0,0,0)
r2<-matrix(c(0,0,0,0,0,0,0,0,0), nrow=9,ncol=9)
for( j in (1:10))
{
r1[j]<-sqrt(((fulltri1[j,10])*(s22[11-j])))
}
r1

# Calculating var-covar matrix
s333<-matrix(c(0,0,0,0,0,0,0,0,0),nrow=9,ncol=9);s333

for (i  in (1:9))
{
for( j in (i:1))
{
s333[i,j]<-s3[10-j]
}
}
s333

for ( i in (1:8))
{
```

```
for (j in ((i+1):9))
{
s333[i,j]<-s3[10-i]
}
}
s333

for(i in (2:9))
{
for( j in (2:9))
{
r2[i,j]<-fulltri1[i,10]*fulltri1[j,10]*s333[i,j]
}
}
r2
r22<-r2[2:9,2:9]
r22
sum(r22)
sqrt(sum(r22))
sqrt(v1+sum(r22))
r11<-c(r1[2:10]);r11
r11<-r11^2;r11
v1<-sum(r11)^0.5;v1
v2<-(sum(r22))^0.5;v2
sqrt((v1^2)+(v2^2))
sqrt((v1^2)+sum(r22))
mset<-sqrt((v1^2)+sum(r22));mset

# Calculating total sd------------------------------------------------------------------------------------------------------------
pm<-c(rep(0,9));pm
for ( i in 1:9)
{
pm[i]<-fulltri1[i+1,10]*sqrt(s3[10-i])
}
pm
sdt<-c(rep(0,9));sdt
for ( i in 1:9)
{
sdt[i]<-sqrt(r1[i+1]^2 + pm[i]^2)
}
sdt
sum(sqrt(sdt))


# Calculating error function-------------------------------------------------------------------------------------------------

error<-c(rep(0,9))
for( i in 1:9)
{
error[i]<-(2/(10-i))*s2[i]/(mean(fulltri[,i+1])+ s2[i])
}
error

for (i in 1: (length(s2)-1))
{
```

```
s2[i]<-(1/(dim(tri)[1]-i-1))*s1[i]
}
s2
s2[length(s2)]<-min((s2[length(s2)-1]^2/s2[length(s2)-2]),min(s2[length(s2)-2],s2[length(s2)-1]))
s2

# Calculating Bias for Robust Mack Model--------------------------------------------------------------------------------------
b<-c(0,0,0,0,0,0,0,0,0)
for( k in (1:9))
{
b[k]<-(-s2[k]/(10-k-1))*(sum(s2[k]*E[(1:(10-k)),k]/(E[(1:(10-k)),k]*sum(E[1:(10-k),k]))))
}
b
b[1:8]

# Diving eb matrix with vb matrix to find franction fr-------------------------------------------------------------------------
eb<-matrix(c(rep(0,10)), nrow=10, ncol=10);eb
vb<-matrix(c(rep(0,10)), nrow=10, ncol=10);vb
for (i in 10:2)
{
for( j in (10-i+2):10)
{
eb[i,j]<-fulltri[i,j]
vb[i,j]<-fulltri[i,j]*s2[j-1]
}
}
eb
vb
eb/vb

fr<-0.85

# Total Bias----------------------------------------------------------------------------------------------------------------------------
frbias<-c(rep(0,9));frbias
for( k in 1:8)
{
frbias[k]<-abs(b[k])/s2[k]
}
frbias

frbias2<-c(rep(0,9))
for( k in 1:8)
{
frbias2[k]<-fr^2/(10-k-1)
}
frbias2
```

Bootstrap Method – χωρίς έκτοπη παρατήρηση

Παρατήρηση: Η Bootstrap method χρησιμοποιεί πολλά στοιχεία της Mack. Θεωρείται ότι έχει τρέξει η κλασσική Mack Method. Έτσι, εδώ θα παρατεθεί η διαδικασία της Bootstrap, αφότου έχει τρέξει η Mack Method.

```
#Finding year-to year development factors-----------------------------------------------------------------------------------
```

```
fy<-matrix(c(NA,NA,NA,NA,NA,NA,NA,NA,NA), nrow=9,ncol=9)
for( k in 1:9)
{
for(j in 1:(10-k))
{
fy[k,j]<-tri[k,j+1]/tri[k,j]
}
}
Fy

# Calculating Pearson residuals----------------------------------------------------------------------------------------------------
red<-matrix(c(NA,NA,NA,NA,NA,NA,NA,NA,NA), nrow=9,ncol=9)
for( k in 1:9)
{
for( j in 1:(10-k))
{
red[j,k]<-sqrt(55/36)*sqrt(tri[j,k])*(fy[j,k]-m[k])/sqrt(s2[k])
}
}
red

#Storing results from Bootstrap method, uing a matrix-------------------------------------------------------------------
rs<-matrix(c(rep(0,50)), nrow=50,ncol=50)
rs

# Extra Comments about Bootstap Method-----------------------------------------------------------------------------------
```

- #rs1 is resampling triangle.
- #fyn new yeat-to-year development factors
- #fn vector of new development factors
- #btri is filled with random values using lognormal distribution. Substracting first diagonal from last column, gives the reserve for each application of the loop.
- # Storing reserves in rs matrix

```
set.seed(2500)
for ( i in 1:50)
{
for( j in 1:50)
{
rs1<-0
rs1<-red
rs1[!is.na(rs1)]<-sample(x=rs1[!is.na(rs1)],replace=TRUE)
rs1

fyn<-matrix(c(rep(NA,9)),nrow=9,ncol=9)
for( k in 1:9)
{
for(l in 1:9)
{
fyn[l,k]<-sqrt(36/55)*rs1[l,k]*(sqrt(s2[k])/sqrt(tri[l,k])) + m[k]
}
}
fyn
fn<-c(rep(0,9))
for( n in 1:9)
{
```

```
fn[n]<-sum(tri[1:(10-n),n]*fyn[1:(10-n),n])/sum(tri[1:(10-n),n])
}
fn

btri<-tri

for( p in 2:10)
{
x<-fn[10-p+1]*btri[p,10-p+1]
y<-s2[10-p+1]*btri[p,10-p+1]
btri[p,10-p+2]<-rlnorm(1, log(x^2/sqrt(x^2+y^2)),sqrt(log(1+(y^2/x^2))))
}
btri
for( p in 10:3)
{
for(q in (10-p+3):10)
{
a<-fn[q-1]*btri[p,q-1]
c<-s2[q-1]*btri[p,q-1]
btri[p,q]<-rlnorm(1, log(a^2/sqrt(a^2+c^2)),sqrt(log(1+(c^2/a^2))))
}
}
btri

sm<-0
for(p in 2:10)
{
sm<-sm + sum(btri[p:10,10-p+2])
}

rs[i,j]<-sm
}
}
apothemaboot<-mean(rs);apothemaboot

#pred. error of Bootstrap-------------------------------------------------------------------------------------------------------------
pesb<-sqrt(mean(rs)+(sd(rs)/sqrt(2500))^2)
pesb
```

Robust Bootstrap – με έκτοπη παρατήρηση

Παρατήρηση: Όπως και η κλασσική Bootstap Method έτσι και η Ανθεκτική χρησιμοποιεί πολλά στοιχεία της Ανθεκτικής Mack method. Θεωρείται ότι για να τρέξει η Robust Bootstrap πρέπει πρώτα να έχει τρέξει η Robust Mack Method. H Robust Bootstrap Method ως προς τη μεθοδολογία δεν διαφέρει σε τίποτα με την κλασσική μορφή της Bootstrap Method. Εφαρμόζεται πρώτα ο κώδικας της robust Mack method και στη συνέχεια εκείνος της Bootstrap Method. Συνιστάται προσοχή στα ονόματα των διανυσμάτων και των πινάκων.