



ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΕΙΡΑΙΩΣ

ΤΜΗΜΑ ΣΤΑΤΙΣΤΙΚΗΣ ΚΑΙ ΑΣΦΑΛΙΣΤΙΚΗΣ ΕΠΙΣΤΗΜΗΣ
Π.Μ.Σ. ΣΤΗΝ ΑΝΑΛΟΓΙΣΤΙΚΗ ΕΠΙΣΤΗΜΗ ΚΑΙ ΔΙΟΙΚΗΤΙΚΗ
ΚΙΝΔΥΝΟΥ

ΜΕΘΟΔΟΙ ΑΠΟΘΕΜΑΤΟΠΟΙΗΣΗΣ ΜΕ ΧΡΗΣΗ ΜΕΘΟΔΩΝ MCMC
ΚΑΙ BOOTSTRAP

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ
ΠΑΣΣΑΣ ΚΩΝΣΤΑΝΤΙΝΟΣ

ΕΠΙΒΛΕΠΩΝ ΚΑΘΗΓΗΤΗΣ: ΠΙΤΣΕΛΗΣ ΓΕΩΡΓΙΟΣ

ΠΕΙΡΑΙΑΣ 2018

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

Τα μέλη της Επιτροπής ήταν:

- Πιτσέλης Γεώργιος (Επιβλέπων)
- Σεβρόγλου Βασίλειος
- Τζαβελάς Γεώργιος

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



UNIVERSITY OF PIRAEUS

DEPARTMENT OF STATISTICS AND INSURANCE SCIENCE

M.Sc. IN ACTUARIAL SCIENCE AND RISK MANAGEMENT

LOSS RESERVING USING MCMC METHODS.

MASTER THESIS

PASSAS KONSTANTINOS

ADVISOR PROFESSOR: PITSELIS GEORGIOS

PIRAEUS 2018

*Στους γονείς μου, Τσαμπίκα και Γιώργο,
και στην Φωτεινή*

ΕΥΧΑΡΙΣΤΙΕΣ

Αρχικά, θα ήθελα να ευχαριστήσω τον επιβλέποντα καθηγητή της μεταπτυχιακής μου εργασίας, Επίκουρο Καθηγητή Γεώργιο Πιτσέλη, του τμήματος Στατιστικής και Ασφαλιστικής Επιστήμης του Πανεπιστημίου Πειραιώς, τόσο για την ανάθεση του θέματος, όσο και για τη συνεχή στήριξη, ενθάρρυνση και βοήθεια που μου προσέφερε σε όλο το διάστημα αυτής μου της προσπάθειας. Επίσης, θα ήθελα να ευχαριστήσω τον Αναπληρωτή Καθηγητή Βασίλειο Σεβρόγλου και τον Επίκουρο Καθηγητή Γεώργιο Τζαβελά, για την τιμή που μου έκαναν να είναι μέλη της τριμελούς μου επιτροπής. Αισθάνομαι, επίσης, υποχρεωμένος να ευχαριστήσω τον Υποψήφιο Διδάκτορα Απόστολο Μποζίκα για τις χρήσιμες συμβουλές του και την βοήθεια του στα τεχνικά ζητήματα της διπλωματικής. Τέλος, θα ήθελα να ευχαριστήσω την οικογένεια μου για την αμέριστη συμπαράσταση και υποστήριξη τους σε κάθε μου προσπάθεια.

ΠΕΡΙΛΗΨΗ

Σκοπός της παρούσας διπλωματικής εργασίας είναι η εφαρμογή των μεθόδων επαναδειγματοληψίας Bootstrap και MCMC, επικεντρώνοντας στα τρίγωνα εξέλιξης ζημιών. Κύριος στόχος είναι η κατανόηση της διαδικασίας που ακολουθείται σε κάθε μια από τις μεθόδους, η γνώση των ιδιομορφιών τους αλλά και τα πλεονεκτήματα/μειονεκτήματα τους. Οι παραπάνω στοχαστικές μέθοδοι αποθεματοποίησης συμβάλουν στον υπολογισμό του κινδύνου που σχετίζεται με την αποθεματοποίηση ζημιών (κίνδυνος αποθεμάτων), αφού σε αντίθεση με τις ντετερμινιστικές μεθόδους αποθεματοποίησης μπορούμε πλέον να εξάγουμε την κατανομή πρόβλεψης των μελλοντικών υποχρεώσεων και κατ' επέκταση να υπολογίσουμε διάφορα μέτρα κινδύνου (π.χ. Ποσοστημόρια (quantiles), Αξία σε Κίνδυνο (Value at Risk), Δεσμευμένη Αξία σε Κίνδυνο (Conditional Value at Risk) κ.λπ.). Εκτός αυτών, με τον υπολογισμό του τυπικού σφάλματος πρόβλεψης, είμαστε σε θέση να εξάγουμε συμπεράσματα για την ακρίβεια της εκτίμησης του αποθέματος. Βασιζόμενοι στα παραπάνω, παρουσιάζονται αναλυτικά παραδείγματα μαζί με γραφικές παραστάσεις για την καλύτερη κατανόηση του αναγνώστη.

ABSTRACT

The purpose of the present thesis is the implementation of the Bootstrap and MCMC resampling methods, focusing on the loss development triangles. The main objective is the understanding of the procedure followed in each method, getting to know their distinctive features as well as their advantages/disadvantages. The abovementioned stochastic loss reserving methods contribute to the calculation of the risk associated with loss reserving (reserve risk), since, in contrast with the deterministic reserving methods we can derive the future liabilities prediction distribution and further calculate various risk measures (e.g. distribution Quantiles, Value at Risk, Conditional Value at Risk etc.). Besides those, by calculating the standard prediction error, we are in a position to draw conclusions on how accurate the reserve estimate is. Based on the above, detailed examples along with graphs are presented.

Περιεχόμενα

Εισαγωγή	16
1 Εισαγωγή στην Αποθεματοποίηση Ζημιών	18
1.1 Εισαγωγή	18
1.2 Ορισμοί Αποθεμάτων Ζημιών	18
1.3 Βασική Ορολογία	19
1.4 Στρατηγική Εκτίμησης Αποθέματος	20
2 Ντετερμινιστικές Μέθοδοι Αποθεματοποίησης	22
2.1 Εισαγωγή	22
2.2 Μέθοδος Τριγώνου Εξέλιξης Ζημιών (Chain-Ladder)	23
2.3 Προσαρμοσμένες Τιμές	25
2.4 Παρουσίαση Δεδομένων	26
2.5 Υπολογισμός Αποθέματος IBNR	27
3 Στοχαστικές Μέθοδοι Αποθεματοποίησης	29
3.1 Εισαγωγή	29
3.2 Μέτρα Αβεβαιότητας των Αποθεμάτων	29
3.3 Το μοντέλο του Mack	32
3.3.1 Υπολογισμός Αποθέματος IBNR	38
3.4 Γενικευμένα Γραμμικά Μοντέλα (Generalized Linear Models - GLM)	39
3.4.1 Μοντέλο Υπερδιασκορπισμένη (Over-Dispersed) Poisson	40
3.4.2 Μοντέλο Γάμμα (Gamma)	42
3.4.3 Λογαριθμοκανονικό (Log-normal) Μοντέλο	43
3.5 Μέθοδος Bootstrap	44
3.5.1 Περιπτώσεις Χρήσης της Μεθόδου	46

3.6	Μέθοδος Bootstrap στα Τρίγωνα Εξέλιξης Ζημιών	47
3.6.1	Κατάλοιπα (Residuals)	49
3.6.2	Τα βήματα της διαδικασίας	51
3.6.3	Μοντέλο Υπερδιασκορπισμένη (Over-Dispersed) Poisson για Bootstrap	53
3.6.4	Μοντέλο Γάμμα (Gamma) για Bootstrap	66
4	Μπεϋζιανά MCMC Μοντέλα	69
4.1	Εισαγωγή	69
4.2	Μπεϋζιανή Αποθεματοποίηση Ζημιών	71
4.3	Μοντέλο Poisson-Γάμμα	73
4.4	Μέθοδος MCMC	74
4.5	Λογαριθμοκανονικό Μοντέλο για MCMC	77
4.5.1	Υπολογισμός των εκ των προτέρων κατανομών	78
4.5.2	Υπολογισμός Αποθέματος IBNR	80
4.6	Μοντέλο Over-Dispersed Poisson για MCMC	87
4.7	Σύγκριση των Μοντέλων (DIC)	91
4.8	Μοντέλο Chain-Ladder για MCMC	93
	Επίλογος	95
	Αναφορές	96
	Παραρτήματα	100
	Παράρτημα Α: Κώδικας R	100
	A-1: Μέθοδος Chain-Ladder	100
	A-2: Mack Chain-Ladder	101
	A-3: Μέθοδος Bootstrap - Μοντέλο Over-Dispersed Poisson	102
	A-4: Μέθοδος Bootstrap - Μοντέλο Γάμμα (Gamma)	106

Παράρτημα Β: Κώδικας WinBUGS	107
B-1: MCMC - Το Λογαριθμοκανονικό Μοντέλο	107
B-2: MCMC - Μοντέλο Over-Dispersed Poisson	108
B-3: MCMC - Μέθοδος Chain-Ladder	109

Εισαγωγή

Η αποθεματοποίηση ζημιών είναι μια από τις πιο σημαντικές διαδικασίες στον ασφαλιστικό κλάδο. Στο παρελθόν, προτάθηκαν αρκετές μέθοδοι για την επίλυση του συγκεκριμένου θέματος, και από τότε συνιστούν σημαντικά εργαλεία για τον υπολογισμό των εκκρεμών υποχρεώσεων μια ασφαλιστικής εταιρίας. Παρόλα αυτά, οι περισσότερες από αυτές τις μεθόδους είναι ντετερμινιστικές και οι δυνατότητες τους περιορίζονται στην εκτίμηση των μελλοντικών απαιτήσεων. Η συγκεκριμένη τιμή όμως δεν είναι επαρκής, καθώς μια ασφαλιστική εταιρία θέλει να γνωρίζει πόσο σωστή είναι η παραπάνω εκτίμηση, αλλά και πως επηρεάζεται από διάφορα τυχαία γεγονότα.

Ερχόμενοι αντιμέτωποι με αυτό το γεγονός, αναπτύχθηκαν αρκετές στοχαστικές μέθοδοι αποθεματοποίησης, οι οποίες μας έδωσαν την δυνατότητα να προσδιορίσουμε όχι μόνο την εκτίμηση των μελλοντικών απαιτήσεων, αλλά και την κατανομή τους μαζί με τα ποσοστημόρια της. Ως συνέπεια αυτού, μπορούμε να υπολογίσουμε μέτρα όπως την Αξία σε Κίνδυνο (Value at Risk – VaR και την υπό δέσμευση Αξία σε Κίνδυνο (Conditional Value at Risk – CvaR), τα οποία μπορούν να χρησιμοποιηθούν στον υπολογισμό του κινδύνου των αποθεμάτων.

Σκοπός της παρούσας διπλωματικής εργασίας είναι η παρουσίαση διαφόρων στοχαστικών μεθόδων αποθεματοποίησης ζημιών, με έμφαση στην μέθοδο Bootstrap, στην μέθοδο Markov Chain Monte Carlo – MCMC. Η δομή που ακολουθείται είναι η εξής:

Στο πρώτο Κεφάλαιο της εργασίας παρουσιάζονται κάποιοι βασικοί ορισμοί και ορολογίες που θα χρησιμοποιηθούν καθ' όλη την έκταση της. Πιο συγκεκριμένα, αναλύεται ο ορισμός τους αποθέματος εκκρεμών ζημιών και οι επιμέρους κατηγορίες του, καθώς και ορολογίες που χρησιμοποιούνται τόσο στις ντετερμινιστικές, όσο και στις στοχαστικές μεθόδους αποθεματοποίησης ζημιών (έτος εξέλιξης, έτος ατυχήματος κ.λπ.). Τέλος, αναφέρονται τα τέσσερα στάδια στρατηγικής εκτίμησης του αποθέματος.

Στο δεύτερο Κεφάλαιο αναλύεται η σημαντικότερη ντετερμινιστική μέθοδος αποθεματοποίησης, που θα αποτελέσει και την βάση σύγκρισης με τις στοχαστικές μεθόδους, η μέθοδος Τριγώνου Εξέλιξης Ζημιών (Chain-Ladder). Έστερα από την ανάλυση ακολουθεί η παρουσίαση των δεδομένων που θα χρησιμοποιηθούν σε όλη την έκταση της διπλωματικής εργασίας καθώς και ένα αριθμητικό παράδειγμα της μεθόδου Chain-Ladder.

Στο τρίτο Κεφάλαιο αναλύονται οι στοχαστικές μέθοδοι αποθεματοποίησης ζημιών. Αρχικά γίνεται μια αναφορά στα διάφορα είδη σφαλμάτων και μέτρων αβεβαιότητας όπως για παράδειγμα το Μέσο Τετραγωνικό Σφάλμα Πρόβλεψης (Mean Square Error of Prediction – MSE). Έπειτα παρουσιάζεται το μοντέλο του Mack που δίνει ακριβώς τις ίδιες τιμές με την μέθοδο Chain-Ladder, το οποίο επιβεβαιώνεται μέσω ενός σύντομου αριθμητικού παραδείγματος με την χρήση του λογισμικού προγράμματος *R*. Στην συνέχεια αναφέρονται τα Γενικευμένα Γραμμικά Μοντέλα (GLM's) και πιο συγκεκριμένα το μοντέλο Over-dispersed Poisson (ODP), το Μοντέλο Γάμμα (Gamma) και το Λογαριθμοκανονικό (Log-normal) μοντέλο. Τα GLM's θα συνεισφέρουν τόσο στην παρουσίαση της μεθόδου Bootstrap, όσο και της μεθόδου MCMC. Στην συνέχεια παρουσιάζεται στα γενικά της πλαίσια η μέθοδος Bootstrap και οι περιπτώσεις χρήσης της μεθόδου. Τέλος, αναλύεται η μέθοδος Bootstrap στα Τρίγωνα Εξέλιξης Ζημιών στο μοντέλο Over-dispersed Poisson (ODP) και στο μοντέλο Γάμμα (Gamma), με την χρήση του λογισμικού προγράμματος *R*.

Στο τέταρτο Κεφάλαιο παρουσιάζεται αρχικά η Μπεύζιανή μοντελοποίηση και η Μπεύζιανή αποθεματοποίηση ζημιών παρουσιάζοντας και ένα θεωρητικό παράδειγμα προς κατανόηση του αναγνώστη. Στην συνέχεια αναλύεται η μέθοδος MCMC και πιο συγκεκριμένα ο αλγόριθμος του Metropolis-Hastings και του Gibbs. Έπειτα δίνεται ένα παράδειγμα του Λογαριθμοκανονικού και του ODP μοντέλου, με την χρήση του λογισμικού προγράμματος WinBUGS, ενώ γίνεται και σύγκριση των αποτελεσμάτων τους. Τέλος παρατίθενται τα αποτελέσματα του μοντέλου Chain-Ladder, που προέκυψαν από το πρόγραμμα WinBUGS, προς σύγκριση με αυτά του μοντέλου ODP.

1. Εισαγωγή στην Αποθεματοποίηση Ζημιών

1.1. Εισαγωγή

Ο αναλογιστής προσεγγίζει την εκτίμηση του αποθέματος ζημιών με πολύ διαφορετικό τρόπο από τον ρυθμιστή απαιτήσεων. Ο αναλυτής πρέπει να γνωρίζει τις απαιτήσεις και τις λογιστικές προοπτικές του συνολικού αποθέματος ζημιών αλλά θα πρέπει συχνότερα να ασχολείται με θέματα εγγενή στις αναλογιστικές προσεγγίσεις του συλλογικού αποθέματος ζημιών. Η καταγραφή του πρέπει να διανέμεται σε αρχεία απαιτήσεων προερχόμενα από μια χρονική περίοδο με ακριβή ημερομηνία έναρξης και λήξης. Οι τελευταίες πρέπει να σχετίζονται με τις χαρακτηριστικές ημερομηνίες στην διάρκεια ενός αρχείου απαιτήσεων. Χαρακτηριστικές ημερομηνίες είναι οι ημερομηνίες έκδοσης, απώλειας, έναρξης συμβολαίου κ.ο.κ. δηλαδή αυτές στις οποίες συνέβη ένα σημαντικό γεγονός κατά τη διάρκεια της απαίτησης. Επίσης η καταγραφή πρέπει να καθορίζεται στους όρους των απαιτήσεων από συγκεκριμένη γεωγραφική τοποθεσία και πολιτική κάλυψης του ασφαλιστικού συμβολαίου για λογούς ομοιογένειας και αξιοπιστίας.

1.2. Ορισμοί Αποθεμάτων Ζημιών

Συνήθως υπάρχει μια χρονο-καθυστέρηση μεταξύ της αναγγελίας ενός περιστατικού και της διευθέτησής του. Οι κύριοι λόγοι είναι ότι μια απαίτηση/ζημιά μπορεί να αναφερθεί με καθυστέρηση, να χρειάζεται χρόνο να διευθετηθεί ή να αναβιώσει στο μέλλον. Για όλα τα παραπάνω στοιχεία, μια ασφαλιστική εταιρία ζημιών (non-life insurance) πρέπει να έχει στο χαρτοφυλάκιο της το κατάλληλο ποσό αποθεμάτων έτσι ώστε να είναι σε θέση να ανταπεξέλθει σε μελλοντικές της υποχρεώσεις απέναντι στους ασφαλισμένους της. Τα αποθέματα ζημιών διαιρούνται σε δύο μεγάλες κατηγορίες (βλέπε Πιτσέλης (2017)). Η πρώτη είναι το σύνολο που πρέπει να πληρωθεί για τις αναγγελθείσες αλλά όχι διευθετημένες απαιτήσεις (Report But Not Settled claims, RBNS). Η δεύτερη είναι το σύνολο των χρημάτων που θα πληρωθούν για τις ζημιές που έγιναν αλλά δεν έχουν ακόμα αναφερθεί (Incurred claims But Not Reported yet, IBNR), καθώς και για τις ζημιές που έχουν αναγγελθεί στην εταιρία αλλά δεν έχουν υπολογιστεί. Το απόθεμα περιλαμβάνει επίσης μια εκτίμηση του κόστους της διευθέτησης της απαίτησης. Το απόθεμα γνωστών απαιτήσεων/ζημιών (RBNS) συνίσταται από:

- Τα αποθέματα που αναφέρονται σε συγκεκριμένες απαιτήσεις/ζημιές.
- Την πρόβλεψη και μελλοντική εξέλιξη γνωστών απαιτήσεων/ζημιών.
- Την πρόβλεψη απαιτήσεων/ζημιών που έχουν κλείσει και ο φάκελος τους έχει ανοίξει ξανά,

Το παραπάνω απόθεμα μπορεί να εκτιμηθεί από κάποιον ειδικό στην αποθεματοποίηση ζημιών, ενώ η τιμή του ποσού των IBNR απαιτήσεων χρειάζεται να προσεγγιστεί από μαθηματικές μεθόδους, χρησιμοποιώντας ιστορικά στοιχεία για την εξέλιξη των απαιτήσεων.

1.3. Βασική Ορολογία

Είναι χρήσιμο να ορίσουμε κάποια βασική ορολογία του αποθέματος ζημιών, η οποία θα χρησιμοποιείται στην ανάλυση της διαδικασίας εκτίμησης του (βλέπε Πιτσέλης (2017)):

- **Ημερομηνία ατυχήματος:** είναι η ημερομηνία που γίνεται το ατύχημα.
- **Ημερομηνία αναγγελίας:** είναι η ημερομηνία που αναγγέλλεται το ατύχημα.
- **Λογιστική ημερομηνία:** είναι η ημερομηνία κατάρτισης των λογιστικών καταστάσεων μέχρι την οποία υπολογίζονται οι οφειλόμενες απαιτήσεις για το απόθεμα ζημιών.
- **Ημερομηνία αποτίμησης:** είναι η μέρα που γίνεται η εκτίμηση του οφειλόμενου αποθέματος. Η εκτίμηση αυτή πρέπει να γίνεται στο τέλος μιας οικονομικής περιόδου. Συνήθως για να ελέγχουμε την ακρίβεια των εκτιμήσεών μας κάνουμε εκτίμηση του αποθέματος ανά τρίμηνο. Επομένως η ημερομηνία αποτίμησης δεν συμπίπτει πάντα με τη λογιστική, η οποία είναι στο τέλος του έτους.

Έπειτα είναι σημαντικό να αναφέρουμε μια συμβατική ορολογία για να συζητάμε τα αποτελέσματα της διαδικασίας εκτίμησης του αποθέματος ζημιών καθώς αυτό μεταβάλλεται στις διαδοχικές ημερομηνίες αποτίμησης:

- **Απαιτούμενο απόθεμα:** είναι το ποσό που πρέπει να πληρωθεί τελικά για να τακτοποιήσει όλες τις οφειλόμενες απαιτήσεις. Το απαιτούμενο απόθεμα είναι ένας σταθερός αριθμός που δεν αλλάζει σε διαφορετικές ημερομηνίες αποτιμήσεις και γίνεται γνωστός μόνο όταν όλες οι απαιτήσεις έχουν τελικά τακτοποιηθεί.
- **Ενδεικτικό απόθεμα:** είναι η εκτίμηση ενός αναλυτή για το σύνολο του απαιτούμενου αποθέματος ζημιών σε μια συγκεκριμένη ημερομηνία αποτίμησης.
- **Μεταφερόμενο απόθεμα:** είναι το σύνολο των απλήρωτων απαιτούμενων οφειλών που εμφανίζεται στις οικονομικές καταστάσεις.
- **Περιθώριο απόθεμα:** είναι η διαφορά ανάμεσα στο μεταφερόμενο και το απαιτούμενο απόθεμα. Το περιθώριο δεν μπορεί να είναι μηδέν γιατί είναι απίθανο το μεταφερόμενο απόθεμα να είναι ίσο με το ενδεικτικό ή το απαιτούμενο απόθεμα. Το απόθεμα ζημιών μπορεί να θεωρηθεί ότι αποτελείται από το απόθεμα για γνωστές απαιτήσεις και το απόθεμα για άγνωστες απαιτήσεις.
- **Έξοδα προσαρμογής των αποθεμάτων:** είναι το απαιτούμενο σύνολο για να καλύψει τα μελλοντικά έξοδα που απαιτούνται για την διερεύνηση και την εξόφληση των απαιτήσεων σε μια δεδομένη χρονική περίοδο. Τα έξοδα προσαρμογής διακρίνονται σε κατανεμημένα και μη, επειδή μπορούν να επιβαρυνθούν με αποθέματα συγκεκριμένων αλλά και γενικών απαιτήσεων. Τα κατανεμημένα είναι τα έξοδα προσαρμογής που προορίζονται για συγκεκριμένες απαιτήσεις (λογαριασμοί, νόμιμα έξοδα). Μη κατανεμημένα έξοδα είναι οι μισθοί, τα ενοίκια το ρεύμα κ.τ.λ.

1.4. Στρατηγική Εκτίμησης Αποθέματος

Χωρίζουμε την συνολική προσέγγιση ενός προβλήματος εκτίμησης αποθέματος σε 4 φάσεις (βλέπε Πιτσέλης (2017)):

1. Ο αναλυτής θα πρέπει να επανεξετάσει τα δεδομένα του έτσι ώστε να εντοπίσει τυχόν ανωμαλίες, αν υπάρχουν, και να αναγνωρίσει τα βασικά χαρακτηριστικά τους. Επίσης είναι σημαντικό ο αναλυτής να συγκρίνει τα δεδομένα του με άλλες έγκυρες πηγές.
2. Ο αναλυτής εφαρμόζει έπειτα την κατάλληλη τεχνική εκτίμησης αποθέματος.
3. Στην συνέχεια πρέπει να υπολογίσει τα αποτελέσματα που δεν συμφωνούν από τη χρήση των διαφορών μεθόδων εκτίμησης αποθέματος και να τα εξηγήσει ή να τα συμβιβάσει.
4. Τέλος πρέπει να παρουσιάσει και να προετοιμάσει την εξέλιξη του αποθέματος όπως αυτή διαγράφεται στα μεταγενέστερα ημερολογιακά έτη.

2. Ντετερμινιστικές Μέθοδοι Αποθεματοποίησης

2.1. Εισαγωγή

Τα ιστορικά δεδομένα που χρησιμοποιεί μια ασφαλιστική εταιρία προκειμένου να υπολογίσει τα αποθέματα που απαιτούνται για να καλύψει τις μελλοντικές τις υποχρεώσεις παρουσιάζονται κατά το πλείστον στα λεγόμενα Τρίγωνα Εξέλιξης Ζημιών (Run-off Triangles). Τα δεδομένα μας χωρίζονται σε δύο άξονες. Ο κατακόρυφος άξονας ορίζει το έτος ατυχήματος i (το έτος που συνέβη η ζημιά) και ο οριζόντιος ορίζει το έτος εξέλιξης j (το πως εξελίσσεται η ζημιά με την πάροδο των χρόνων). Ένα τέτοιο τρίγωνο παρουσιάζεται παρακάτω:

Πίνακας 2.1: Τρίγωνο εξέλιξης ζημιών (Run-off triangle)

i/j	1	2	3	n-1	n
1										
2										
3										
⋮										
⋮										
⋮										
⋮										
⋮										
⋮										
n-1										
n										

Στην συγκεκριμένη εργασία θα υποθέσουμε ότι έχουμε τόσα έτη εξέλιξης όσα και έτη ατυχήματος, δηλαδή $1 \leq i \leq n$ και $1 \leq j \leq n$. Τα στοιχεία X_{ij} αναφέρονται στις προσαυξητικές απαιτήσεις για το έτος εξέλιξης j για ζημιές που συνέβησαν το έτος i ενώ τα C_{ij} στις συσσωρευμένες (αθροιστικές) απαιτήσεις για το i έτος ατυχήματος και j έτος εξέλιξης. Οι αθροιστικές απαιτήσεις δίνονται από τον τύπο:

$$C_{ij} = \sum_{k=1}^j X_{ik}, \quad (2.1)$$

Η τιμή \hat{C}_{in} ($\hat{C}_{1n} = C_{1n}$) αναφέρεται ως τελικό (ultimate) ποσό απαίτησης του έτους ατυχήματος i . Αρκετές φορές το X_{ij} δηλώνει το πλήθος των απαιτήσεων που συνέβησαν στο έτος i με μια καθυστέρηση j ετών στην αναφορά τους. Σε αυτή την περίπτωση, το C_{ij} δηλώνει τον συνολικό αριθμό απαιτήσεων για το έτος ατυχήματος i που αναφέρθηκαν έως και j χρόνια μετά, και το C_{in} τον συνολικό αριθμό ατυχημάτων που έλαβαν χώρα το έτος i .

Όσον αφορά το τρίγωνο εξέλιξης ζημιών (Run-off triangle) οι παρατηρήσεις με $i + j \leq n + 1$ αποτελούν τα δεδομένα μας, ενώ το σύνολο των παρατηρήσεων με $i + j > n + 1$ πρέπει να εκτιμηθούν με διάφορες μεθόδους που θα αναλύσουμε στην συνέχεια της εργασίας.

Αυτό που έχει ιδιαίτερο ενδιαφέρον και απασχολεί κυρίως τις ασφαλιστικές είναι η πρόβλεψη του αποθέματος των απαιτήσεων που έχουν γίνει και δεν έχουν ακόμη αναγγελθεί στην εταιρία (Incurred But not Reported (IBNR) Reserve). Θα συμβολίσουμε τα παραπάνω αποθέματα για κάθε έτος ατυχήματος με \hat{R}_i και το συνολικό απόθεμα με \hat{R} . Οι τύποι για τα παραπάνω δίνονται αντίστοιχα:

$$\hat{R}_i = \hat{C}_{in} - C_{i,n-i+1},$$

$$\hat{R} = \hat{R}_2 + \hat{R}_3 + \dots + \hat{R}_n.$$

Στην συνέχεια θα παρουσιάσουμε μερικές από τις βασικές μεθόδους αποθεματοποίησης ζημιών.

2.2. Μέθοδος Τριγώνου Εξέλιξης Ζημιών (Chain-Ladder)

Η μέθοδος Chain-Ladder είναι μια από τις πιο δημοφιλείς τεχνικές αποθεματοποίησης ζημιών. Αν και ντετερμινιστική μέθοδος, η ανάγκη για τον υπολογισμό του σφάλματος πρόβλεψης, δηλαδή της αβεβαιότητας για το κατά πόσο η εκτίμηση μας είναι σωστή, οδήγησε στην ανάπτυξη αρκετών στοχαστικών μοντέλων (βλέπε Mack (1993), Taylor (2000), Verrall (2000)) που

δίνουν ίδια αποτελέσματα με την παραπάνω (θα αναλύσουμε πολλά από τα μοντέλα αυτά στο επόμενο κεφάλαιο της εργασίας).

Η ντετερμινιστική (ή distribution-free) Chain-Ladder μέθοδος βασίζεται στις αθροιστικές απαιτήσεις C_{ij} , $1 \leq i \leq n$, $1 \leq j \leq n$, για τις οποίες υποθέτουμε πως είναι τυχαίες μεταβλητές με γνωστές τιμές μόνο για $i + j \leq n + 1$ όπως αναφέρθηκε στην εισαγωγή του κεφαλαίου. Σκοπός της μεθόδου είναι η εκτίμηση των τελικών (ultimate) απαιτήσεων \hat{C}_{in} και των αποθεμάτων \hat{R}_i , για $i = 2, 3, \dots, n$.

Πίνακας 2.2: Τρίγωνο εξέλιξης ζημιών για αθροιστικά δεδομένα

i/j	1	2	3	n
1	C_{11}	C_{12}	C_{13}	C_{1j}	C_{1n}
2	C_{21}	C_{22}	C_{23}	C_{2j}	C_{2n}	
⋮	⋮	⋮	⋮			⋮		⋮		
⋮	⋮	⋮	⋮			⋮	⋮			
⋮	C_{i1}	C_{i2}	C_{i3}	$C_{i,n-i+1}$				
⋮	⋮	⋮	⋮			⋮				
⋮	⋮	⋮	⋮	⋮						
⋮	⋮	⋮	$C_{n-2,3}$							
n-1	$C_{n-1,1}$	$C_{n-1,2}$								
n	C_{n1}									

Για το μοντέλο της Chain-Ladder κάνουμε τις εξής υποθέσεις:

- Οι αθροιστικές απαιτήσεις C_{ij} είναι ανεξάρτητες για διαφορετικά έτη ατυχήματος. Δηλαδή, τα σύνολα $\{C_{i1}, \dots, C_{in}\}$, $\{C_{m1}, \dots, C_{mn}\}$ για $i \neq m$ είναι ανεξάρτητα.
- Υπολογίζουμε τους συντελεστές εξέλιξης ζημιών f_2, f_3, \dots, f_n τέτοιοι ώστε:

$$E(C_{ij} | C_{i1}, \dots, C_{i,j-1}) = E(C_{ij} | C_{i,j-1}) = f_j C_{i,j-1},$$

για $1 \leq i \leq n$ και $2 \leq j \leq n$.

Οι υποθέσεις που κάναμε είναι αρκετές προκειμένου να περιγράψουμε τον αλγόριθμο της μεθόδου Chain-Ladder, ξεκινώντας από την εκτίμηση των παραγόντων εξέλιξης f_j με:

$$\hat{f}_j = \frac{\sum_{i=1}^{n-j+1} C_{ij}}{\sum_{i=1}^{n-j+1} C_{i,j-1}}, \quad 2 \leq j \leq n, \quad (2.2)$$

και στην συνέχεια υπολογίζουμε τους εκτιμητές Chain-Ladder για κάθε έτος ατυχήματος i και έτος εξέλιξης m , με $i + m > n + 1$, από τον τύπο (βλέπε Wüthrich and Merz (2008)):

$$\hat{C}_{im}^{CL} = C_{i,n-i+1} \prod_{j=n-i+2}^m \hat{f}_j, \quad m > n - i + 1. \quad (2.3)$$

Οι Εξισώσεις (2.2) και (2.3) αποτελούν την βάση της τεχνικής Chain-Ladder καθώς αυτές οι δύο επαρκούν για να υπολογίσει κανείς τις τελικές απαιτήσεις \hat{C}_{in}^{CL} , τα αποθέματα για κάθε έτος ατυχήματος \hat{R}_i^{CL} , καθώς και το συνολικό απόθεμα \hat{R}^{CL} (για $2 \leq i \leq n$). Αντίστοιχα:

$$\hat{C}_{in}^{CL} = C_{i,n-i+1} \prod_{j=n-i+2}^n \hat{f}_j, \quad (2.4)$$

$$\hat{R}_i^{CL} = \hat{C}_{in}^{CL} - C_{i,n-i+1}, \quad (2.5)$$

$$\hat{R}^{CL} = \hat{R}_2^{CL} + \hat{R}_3^{CL} + \dots + \hat{R}_n^{CL}. \quad (2.6)$$

2.3. Προσαρμοσμένες Τιμές

Δουλεύοντας αναδρομικά προς τα πίσω και ξεκινώντας από τα τελευταία διαγώνια στοιχεία του τριγώνου εξέλιξης ζημιών, υπολογίζουμε της προσαρμοσμένες τιμές (fitted values) \hat{C}_{ij} για $i + j \leq n + 1$ των αθροιστικών απαιτήσεων από τον τύπο:

$$\hat{C}_{i,j-1} = \frac{\hat{C}_{ij}}{\hat{f}_j}, \quad (2.7)$$

παίρνοντας ως αρχική τιμή

$$\hat{C}_{i,n-i+1} = C_{i,n-i+1}. \quad (2.8)$$

Επιπλέον, οι προσαυξητικές απαιτήσεις προκύπτουν από την διαφορά:

$$\hat{X}_{ij} = \hat{C}_{ij} - \hat{C}_{i,j-1}, \quad i = 1, 2, \dots, n, \quad j = 2, 3, \dots, n, \quad (2.9)$$

$$\hat{X}_{i1} = \hat{C}_{i1}, \quad i = 1, 2, \dots, n. \quad (2.10)$$

Οι παραπάνω τιμές θα μας χρειαστούν ιδιαίτερα στα επόμενα κεφάλαια, και συγκεκριμένα στην αποθεματοποίηση με την μέθοδο Bootstrap.

2.4. Παρουσίαση Δεδομένων

Για τον σκοπό της διπλωματικής εργασίας θα χρησιμοποιηθούν τα δεδομένα όπως παρουσιάστηκαν από τους Taylor and Ashe (1983). Στον Πίνακα 2.3 παρουσιάζονται οι προσαυξητικές απαιτήσεις, τις οποίες συμβολίσαμε παραπάνω με X_{ij} .

Πίνακας 2.3: Προσαυξητικές Απαιτήσεις X_{ij}

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

Οι μέθοδοι αποθεματοποίησης χρησιμοποιούνται για να εκτιμηθεί το κάτω δεξιά τρίγωνο του πίνακα (για $i + j \leq n + 1$), του οποίου το άθροισμα των κελιών είναι το απόθεμα για τις ζημιές που έγιναν αλλά δεν έχουν ακόμα αναγγελθεί στην ασφαλιστική (IBNR Claims Reserve ή απλά IBNR). Στο παραπάνω τρίγωνο εξέλιξης ζημιών έχουμε 10 έτη ατυχήματος, όσα και τα έτη εξέλιξης. Τα στοιχεία της τελευταίας διαγωνίου αποτελούν τις απαιτήσεις του τρέχοντος έτους.

Οι αθροιστικές απαιτήσεις C_{ij} που υπολογίζονται από τη σχέση (2.1) παρουσιάζονται στον Πίνακα 2.4.

Στην συνέχεια θα εφαρμοστεί την μέθοδο Chain-Ladder προκειμένου να υπολογιστεί το απόθεμα IBNR. Για τον σκοπό αυτό, θα παρουσιάσουμε αναλυτικά την όλη διαδικασία χρησι-

Πίνακας 2.4: Αθροιστικές Απαιτήσεις C_{ij}

i/j	1	2	3	4	5	6	7	8	9	10
1	357848	1124788	1735330	2218270	2745596	3319994	3466336	3606286	3833515	3901463
2	352118	1236139	2170033	3353322	3799067	4120063	4647867	4914039	5339085	
3	290507	1292306	2218525	3235179	3985995	4132918	4628910	4909315		
4	310608	1418858	2195047	3757447	4029929	4381982	4588268			
5	443160	1136350	2128333	2897821	3402672	3873311				
6	396132	1333217	2180715	2985752	3691712					
7	440832	1288463	2419861	3483130						
8	359480	1421128	2864498							
9	376686	1363294								
10	344014									

μοποιώντας το λογισμικό R και συγκεκριμένα το πακέτο ChainLadder έχοντας πάντα υπόψη τους τύπους που περιγράψαμε στο Υποκεφάλαιο 2.2 (βλέπε Παράρτημα).

2.5. Υπολογισμός Αποθέματος IBNR

Ξεκινώντας από το τρίγωνο αθροιστικών απαιτήσεων (Πίνακας 2.4), υπολογίζουμε τις εκτιμήσεις των συντελεστών εξέλιξης ζημιών \hat{f}_j για $2 \leq j \leq 10$ σύμφωνα με τη σχέση (2.2) και παίρνουμε τα εξής αποτελέσματα:

Πίνακας 2.5: Συντελεστές εξέλιξης ζημιών

\hat{f}_j	3.490607	1.747333	1.457413	1.173852	1.103824	1.086269	1.053874	1.076555	1.017725
-------------	----------	----------	----------	----------	----------	----------	----------	----------	----------

Έπειτα χρησιμοποιώντας τη σχέση (2.3) εκτιμάμε τις αθροιστικές εκτιμήσεις για το κάτω τρίγωνο του πίνακα αθροιστικών απαιτήσεων, δηλαδή για $i + j > 11$ και εξάγουμε τον συμπληρωμένο πλέον 10×10 Πίνακα 2.6:

Τέλος, υπολογίζουμε τις εκτιμημένες προσαυξητικές απαιτήσεις σύμφωνα με τον Τύπο (2.6) και παράγουμε τον Πίνακα 2.7 από τον οποίο υπολογίζεται και το απόθεμα IBNR τόσο κάθε έτους όσο και το συνολικό, με την βοήθεια των Σχέσεων (2.5) και (2.6):

Πίνακας 2.6: Πλήρες Τρίγωνο Εξέλιξης Αθροιστικών Ζημιών

i/j	1	2	3	4	5	6	7	8	9	10
1	357848	1124788	1735330	2218270	2745596	3319994	3466336	3606286	3833515	3901463
2	352118	1236139	2170033	3353322	3799067	4120063	4647867	4914039	5339085	5433719
3	290507	1292306	2218525	3235179	3985995	4132918	4628910	4909315	5285148	5378826
4	310608	1418858	2195047	3757447	4029929	4381982	4588268	4835458	5205637	5297906
5	443160	1136350	2128333	2897821	3402672	3873311	4207459	4434133	4773589	4858200
6	396132	1333217	2180715	2985752	3691712	4074999	4426546	4665023	5022155	5111171
7	440832	1288463	2419861	3483130	4088678	4513179	4902528	5166649	5562182	5660771
8	359480	1421128	2864498	4174756	4900545	5409337	5875997	6192562	6666635	6784799
9	376686	1363294	2382128	3471744	4075313	4498426	4886502	5149759	5544000	5642266
10	344014	1200818	2098228	3057984	3589620	3962307	4304132	4536015	4883270	4969825

Πίνακας 2.7: Πλήρες Τρίγωνο Εξέλιξης Προσαυξητικών Ζημιών

i/j	1	2	3	4	5	6	7	8	9	10	Απόθεμα
1	357848	766940	610542	482940	527326	574398	146342	139950	227229	67948	0
2	352118	884021	933894	1183289	445745	320996	527804	266172	425046	94634	94634
3	290507	1001799	926219	1016654	750816	146923	495992	280405	375833	93678	469511
4	310608	1108250	776189	1562400	272482	352053	206286	247190	370179	92268	709638
5	443160	693190	991983	769488	504851	470639	334148	226674	339456	84611	984889
6	396132	937085	847498	805037	705960	383287	351548	238477	357132	89016	1419459
7	440832	847631	1131398	1063269	605548	424501	389349	264121	395534	98588	2177641
8	359480	1061648	1443370	1310258	725788	508792	466660	316566	474073	118164	3920301
9	376686	986608	1018834	1089616	603569	423113	388076	263257	394241	98266	4278972
10	344014	856804	897410	959756	531636	372687	341826	231882	347255	86555	4625811

Στην πράξη, για το απόθεμα κάθε έτους \hat{R}_i (για $i = 2, \dots, 10$) προσθέτουμε τα κελιά δεξιά της κύριας (μπλε) διαγωνίου, ενώ για το συνολικό αθροίζουμε όλα τα κελιά κάτω από αυτή (ή προσθέτουμε το απόθεμα κάθε έτους, σχέση 2.6). Οπότε το συνολικό απόθεμα είναι 18680856.

3. Στοχαστικές Μέθοδοι Αποθεματοποίησης

3.1. Εισαγωγή

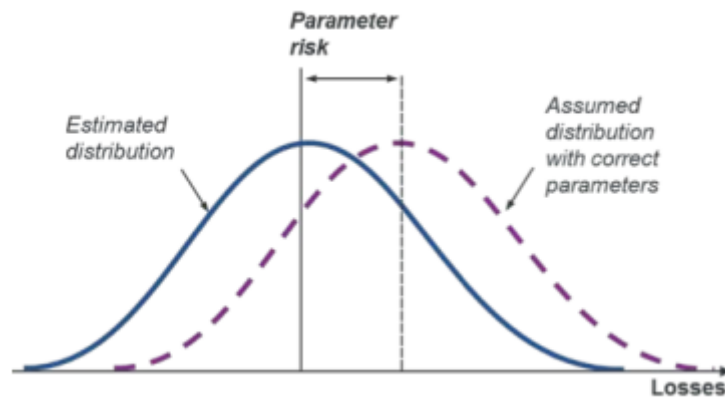
Παρόλο που τα τελευταία χρόνια το ενδιαφέρον για τις στοχαστικές μεθόδους αποθεματοποίησης έχει αυξηθεί αισθητά, λίγοι είναι αυτοί που τις χρησιμοποιούν στην πράξη. Μερικοί από τους λόγους που συμβαίνει το παραπάνω, είναι η έλλειψη κατανόησης των μεθόδων, η έλλειψη ευελιξίας στις μεθόδους κ.ο.κ. Ο κύριος λόγος όμως είναι πως οι υπάρχουσες κλασσικές ντετερμινιστικές μέθοδοι επαρκούν για τον υπολογισμό μια καλής εκτίμησης του αποθέματος IBNR. Ύστερα από την εφαρμογή του πλαισίου της Φερεγγυότητας II (Solvency II), οι ασφαλιστικές είναι υποχρεωμένες περά από την εκτίμηση του αποθέματος να παρουσιάζουν αναλυτικά μια πιθανή πτώση αυτού. Για τον συγκεκριμένο σκοπό, είναι απαραίτητο να υπολογιστεί η μεταβλητότητα των αποθεμάτων ζημιών, και ιδανικά, η κατανομή πρόβλεψης τους, από την οποία μπορούμε να εξάγουμε διάφορα μέτρα για την ποσοτικοποίηση του κινδύνου πτώσης των αποθεμάτων.

3.2. Μέτρα Αβεβαιότητας των Αποθεμάτων

Είναι αναγκαίο, πριν ξεκινήσει η παρουσίαση των στοχαστικών μεθόδων αποθεματοποίησης, να γίνει μια μικρή αναφορά στα είδη σφαλμάτων που ενέχει η κάθε διαδικασία (βλέπε Carrato, McGuire and Scarth (2016)):

- **Σφάλμα παραμέτρου ή σφάλμα εκτίμησης (Parameter or Estimation error):** Αντιπροσωπεύει την αβεβαιότητα στην εκτίμηση των παραμέτρων, υποθέτοντας ότι το μοντέλο έχει προσδιοριστεί σωστά.

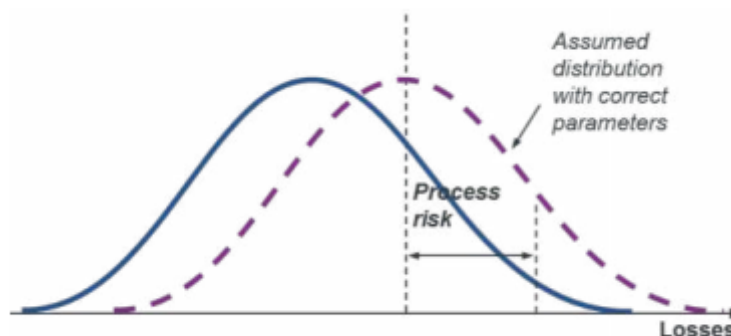
Σχήμα 3.1: Σφάλμα παραμέτρου



Είναι η διαφορά των μεσών μεταξύ της εκτιμημένης κατανομής και της θεωρητικής κατανομής με τις σωστές παραμέτρους (Σχήμα 3.1).

- **Σφάλμα διαδικασίας (Process error):** Αντιπροσωπεύει την τυχαιότητα των μελλοντικών αποτελεσμάτων. Ακόμα και αν το μοντέλο έχει προσδιοριστεί σωστά και οι παράμετροι έχουν εκτιμηθεί σωστά, η ευαισθησία που συνδέεται με την ασφαλιστική διαδικασία είναι πιθανόν να δώσει διαφορετικά αποτελέσματα από τα αναμενόμενα. Με άλλα λόγια, είναι η διασπορά μιας τυχαίας μεταβλητής.

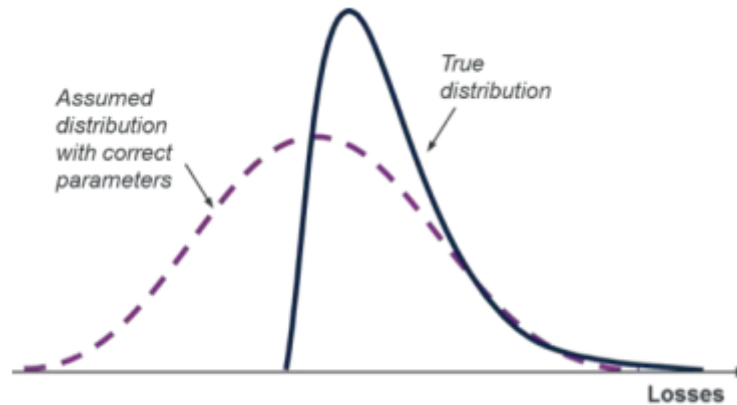
Σχήμα 3.2: Σφάλμα διαδικασίας



Όπως φαίνεται στο Σχήμα 3.2, είναι το σφάλμα που προέρχεται από την διαφορά της πραγματικής μελλοντικής τιμής από τον μέσο της θεωρητικής κατανομής με τις πραγματικές παραμέτρους.

- **Σφάλμα μοντέλου (Model error):** Ποσοτικοποιεί την διαφορά μεταξύ της παραμετρικής μορφής που υποθέσαμε για το μοντέλο, με την άγνωστη αλλά πραγματική παραμετρική του μορφή, είναι δηλαδή το σφάλμα που προέρχεται από την επιλογή του μοντέλου (Σχήμα 3.3).

Σχήμα 3.3: Σφάλμα μοντέλου



Ένα σύνηθες μέτρο τους λάθους πρόβλεψης είναι το Μέσο Τετραγωνικό Σφάλμα Πρόβλεψης (Mean Square Error of Prediction, $MSEP$). Στην περίπτωση της αποθεματοποίησης ζημιών και συγκεκριμένα στην εκτίμηση του αποθέματος, το $MSEP$ είναι η αναμενόμενη τιμή του τετραγώνου του συνολικού σφάλματος του αποθέματος, δηλαδή:

$$MSEP(\hat{R}) = E[(R - \hat{R})^2].$$

Όταν το σφάλμα παραμέτρου και σφάλμα διαδικασίας είναι (στοχαστικά) ανεξάρτητα, τότε το $MSEP$ μπορεί να γραφεί σαν το άθροισμα των τετραγώνων των δύο.

Είναι επίσης συνηθισμένο να χρησιμοποιείται ένας συντελεστής διασποράς (coefficient of variation, CoV) για να εκφράσουμε την αβεβαιότητα των αποθεμάτων. Ορίζεται από τον τύπο:

$$\frac{\sqrt{MSEP(\hat{R})}}{\hat{R}}.$$

3.3. Το μοντέλο του Mack

Μια πρώτη απάντηση στην ερώτηση ποιο στοχαστικό μοντέλο αποτελεί την βάση της μεθόδου Chain-Ladder δόθηκε από τον Mack (1993), ο οποίος χρησιμοποίησε τις αθροιστικές απαιτήσεις ως τυχαίες μεταβλητές και κατάφερε να αναπτύξει ένα μοντέλο που να δίνει ακριβώς ίδιες εκτιμήσεις αποθεμάτων με την μέθοδο Chain-Ladder. Οι υποθέσεις αυτού του μοντέλου είναι οι εξής (βλέπε Wüthrich and Merz (2008)):

- Οι αθροιστικές απαιτήσεις C_{ij} είναι ανεξάρτητες για διαφορετικά χρόνια ατυχήματος i .
- Οι αθροιστικές απαιτήσεις $(C_{ij})_{j>0}$ σχηματίζουν μια Μαρκοβιανή αλυσίδα. Εκεί υπάρχουν οι παράγοντες $f_2, \dots, f_n > 0$ και οι παράμετροι της διασποράς $\sigma_2^2, \dots, \sigma_n^2 > 0$, τέτοιοι ώστε για κάθε $1 \leq i \leq n$ και $2 \leq j \leq n$ έχουμε:

$$E(C_{ij} | C_{i,j-1}) = f_j C_{i,j-1}, \quad (3.1)$$

$$V(C_{ij} | C_{i,j-1}) = \sigma_j^2 C_{i,j-1}. \quad (3.2)$$

Οι εκτιμητές για τις άγνωστες παραμέτρους f_i και σ_j^2 δίνονται από τις σχέσεις:

$$\hat{f}_j = \frac{\sum_{i=1}^{n-j+1} w_{ij} \lambda_{ij}}{\sum_{i=1}^{n-j+1} w_{ij}},$$

$$\hat{\sigma}_j^2 = \frac{\sum_{i=1}^{n-j+1} w_{ij} (\lambda_{ij} - \hat{f}_j)}{n - j},$$

όπου $w_{ij} = C_{i,j-1}$ και $\lambda_{ij} = \frac{C_{ij}}{C_{i,j-1}}$.

Επιπρόσθετα, σύμφωνα με τον Mack, οι εκτιμητές \hat{f}_j είναι αμερόληπτοι εκτιμητές των συντελεστών εξέλιξης απαιτήσεων f_j που υπολογίζονται με την μέθοδο Chain-Ladder και είναι ανά

δύο ασυσχέτιστοι. Οι εκτιμώμενες αθροιστικές απαιτήσεις \hat{C}_{im} υπολογίζονται ακριβώς όπως με την μέθοδο Chain-Ladder με τον γνωστό τύπο:

$$\hat{C}_{im} = C_{i,n-i+1} \prod_{j=n-i+2}^m \hat{f}_j, \quad m > n - i + 1.$$

Ενώ η εκτίμηση του αποθέματος για κάθε έτος ατυχήματος και η εκτίμηση του συνολικού αποθέματος δίνεται αντίστοιχα από τους τύπους:

$$\hat{R}_i = \hat{C}_{in} - C_{i,n-i+1},$$

$$\hat{R} = \hat{R}_2 + \hat{R}_3 + \dots + \hat{R}_n.$$

Δεδομένης της ανεξαρτησίας μεταξύ των ετών ατυχημάτων, το μέσο τετραγωνικό σφάλμα για τις εκτιμώμενες αθροιστικές απαιτήσεις συμβολίζεται με $MSEP(\hat{C}_{ij})$ (Mean Square Error of Prediction) και υπολογίζεται από τον τύπο:

$$\begin{aligned} MSEP[\hat{C}_{ij}] &= E[(C_{ij} - \hat{C}_{ij})^2] \\ &= E[(C_{ij} - E(C_{ij}))^2] + E[(\hat{C}_{ij} - E(\hat{C}_{ij}))^2], \end{aligned}$$

όπου ο πρώτος όρος συμβολίζει την διασπορά της διαδικασίας (process variance) και ο δεύτερος όρος την διασπορά της εκτίμησης (estimation variance). Έστω τώρα το σύνολο \mathcal{M} να είναι το τρίγωνο των γνωστών δεδομένων μας, δηλαδή $\mathcal{M} = \{C_{ij} : i + j \leq n + 1\}$. Αποδεικνύεται (βλέπε Mack (1993)) ότι:

$$\begin{aligned} MSEP(\hat{R}_i) &= E[(\hat{R}_i - R_i)^2 \mid \mathcal{M}] \\ &= E[(\hat{C}_{in} - C_{in})^2 \mid \mathcal{M}] \\ &= MSEP(\hat{C}_{in}). \end{aligned}$$

από το οποίο προκύπτει το κάτωθι θεώρημα:

Θεώρημα 3.1. Το Μέσο Τετραγωνικό Σφάλμα Πρόβλεψης του αποθέματος κάθε έτους δίνεται από τον τύπο:

$$MSE\widehat{P}(\hat{R}_i) = \hat{C}_{in}^2 \sum_{k=n-i+1}^{n-1} \frac{\hat{\sigma}_k^2}{\hat{f}_k^2} \left(\frac{1}{\hat{C}_{ik}} + \frac{1}{\sum_{j=1}^{n-k} C_{jk}} \right), \quad (3.3)$$

όπου $\hat{C}_{ik} = C_{i,n-i+1} \hat{f}_{n-i+1} \cdots \hat{f}_{k-1}$, $k > n - i + 1$ και $\hat{C}_{i,n-i+1} = C_{i,n-i+1}$.

Απόδειξη. Με βάση τον Mack (1991) έχουμε:

Χρησιμοποιώντας τους τύπους

$$E_i(X) = E(X | C_{i1}, \dots, C_{i,n+1-i}),$$

$$V_i(X) = V(X | C_{i1}, \dots, C_{i,n+1-i}).$$

Ξεκινώντας από τον τύπο

$$MSE\widehat{P}(\hat{R}_i) = V(C_{in} | \mathcal{M}) + (E(C_{in} | \mathcal{M}) - \hat{C}_{in})^2.$$

Επαναλαμβάνοντας τις βασικές υποθέσεις της μεθόδου Chain-Ladder (3.1) και (3.2) έχουμε:

$$\begin{aligned} V_i(C_{in} | \mathcal{M}) &= V_i(C_{in}) \\ &= E_i(V(C_{in} | C_{i1}, \dots, C_{i,n-1})) + V_i(E(C_{in} | C_{i1}, \dots, C_{i,n-1})) \\ &= E_i(C_{i,n-1})\sigma_{n-1}^2 + V_i(C_{i,n-1})f_{n-1}^2 \\ &= E_i(C_{i,n-2})f_{n-2}\sigma_{n-1}^2 + E_i(C_{i,n-2})\sigma_{n-2}^2 f_{n-1}^2 + V_i(C_{i,n-2})f_{n-2}^2 f_{n-1}^2 \\ &= \dots \\ &= C_{i,n+1-i} \sum_{k=n-1-i}^{n-1} f_{n+1-i} \cdots f_{k-1} \sigma_k^2 f_{k+1}^2 \cdots f_{n-1}^2, \end{aligned}$$

αφού $V_i(C_{i,n+1-i}) = 0$.

Στην συνέχεια από την σχέση $E(C_{in} | \mathcal{M}) = C_{i,n+1-i} f_{n+1-i} \cdots f_{n-1}$ έχουμε:

$$(E(C_{in} | \mathcal{M}) - \hat{C}_{in})^2 = C_{i,n+1-i}^2 (f_{n+1-i} \cdots f_{n-1} - \hat{f}_{n+1-i} \cdots \hat{f}_{n-1})^2.$$

Στην πράξη, πρέπει να βρούμε τους εκτιμητές των δύο όρων του $MSEP(\hat{R}_i)$. Για τον πρώτο όρο αυτό επιτυγχάνεται αντικαθιστώντας τις άγνωστες παραμέτρους f_k και σ_k^2 με τους εκτιμητές τους \hat{f}_k και $\hat{\sigma}_k^2$, δηλαδή:

$$\begin{aligned} V_i(\widehat{C_{in}} | \mathcal{M}) &= C_{i,n+1-i} \sum_{k=n+1-i}^{n-1} f_{n+1-i} \cdots \hat{f}_{k-1} \hat{\sigma}_k^2 \hat{f}_{k+1}^2 \cdots \hat{f}_{n-1}^2, \\ &= \hat{C}_{in}^2 \sum_{k=n+1-i}^{n-1} \frac{\hat{\sigma}_k^2}{\hat{f}_k^2} \frac{1}{\hat{C}_{ik}}. \end{aligned}$$

Για τον δεύτερο όρο δεν μπορούμε απλά να αντικαταστήσουμε την f_k με \hat{f}_k γιατί αυτό θα έκανε 0. Συνεπώς χρησιμοποιούμε μια διαφορετική προσέγγιση και έχουμε:

$$\begin{aligned} F &= f_{n+1-i} \cdots f_{n-1} - \hat{f}_{n+1-i} \cdots \hat{f}_{n-1} \\ &= S_{n+1-i} + \cdots + S_{n-1}, \end{aligned}$$

όπου

$$S_k = \hat{f}_{n+1-i} \cdots \hat{f}_{k-1} (f_k - \hat{f}_k) f_{k+1} \cdots f_{n-1}$$

και συνεπώς

$$\begin{aligned} F^2 &= (S_{i+1-i} + \cdots + S_{n-1})^2 \\ &= \sum_{k=n+1-i}^{n-1} S_k^2 + 2 \sum_{j < k} S_j S_k. \end{aligned}$$

Έστω τώρα $B_k = \{C_{ij} \leq k, i + j \leq n + 1, 1 \leq k \leq n\}$. Αντικαθιστούμε το S_k^2 με $E(S_k^2 | B_k)$ και $S_j S_k$, $j < k$, με $E(S_j S_k | B_k)$. Επειδή η $E(f_k - \hat{f}_k | B_k) = 0$ έχουμε $E(S_j S_k | B_k) = 0$ για $j < k$. Επειδή:

$$\begin{aligned} E((f_k - \hat{f}_k)^2 | B_k) &= V(\hat{f}_k | B_k) \\ &= \sum_{j=1}^{n-k} V(C_{j,k+1} | B_k) \left(\frac{1}{\sum_{j=1}^{n-k} C_{jk}} \right)^2 \\ &= \frac{\sigma_k^2}{\sum_{j=1}^{n-k} C_{ij}}, \end{aligned}$$

παίρνουμε

$$E(S_k^2 | B_k) = \frac{\hat{f}_{n+1-i}^2 \cdots \hat{f}_{k-1}^2 \sigma_k^2 f_{k+1}^2 \cdots f_{n-1}^2}{\sum_{i=1}^{n-k} C_{ik}}.$$

Έχοντας όλα όσα χρειαζόμαστε, αντικαθιστούμε το $F^2 = (\sum S_k)^2$ με $\sum_k E(S_k^2 | B_k)$ και επειδή όλοι οι όροι του αθροίσματος είναι θετικοί μπορούμε να αντικαταστήσουμε όλες τις άγνωστες παραμέτρους f_k, σ_k^2 με τους αμερόληπτους εκτιμητές τους $\hat{f}_k, \hat{\sigma}_k^2$. Συνοψίζοντας, εκτιμάμε το

$$F^2 = (f_{n+1-i} \cdots f_{n-1} - \hat{f}_{n+1-i} \cdots \hat{f}_{n-1}),$$

με

$$\begin{aligned} & \sum_{k=n+1-i}^{n-1} (\hat{f}_{n+1-i}^2 \cdots \hat{f}_{k-1}^2 \hat{\sigma}_k^2 \hat{f}_{k+1}^2 \cdots \hat{f}_{n-1}^2 \frac{1}{\sum_{j=1}^{n-k} C_{jk}}) \\ &= \hat{f}_{n+1-i}^2 \cdots \hat{f}_{n-1}^2 \sum_{k=n+1-i}^{n-1} \frac{\hat{\sigma}_k^2}{\hat{f}_k^2} \frac{1}{\sum_{j=1}^{n-k} C_{ij}}, \end{aligned}$$

το οποίο οδηγεί στον εκτιμητή του Μέσου Τετραγωνικού Σφάλματος Πρόβλεψης που αναφέρεται στο θεώρημα.

□

Θεώρημα 3.2. Το $MSEP$ για το συνολικό απόθεμα \hat{R} δίνεται από:

$$MSEP(\hat{R}) = \sum_{i=2}^n \left(SE(\hat{R}_i)^2 + \hat{C}_{in} \left(\sum_{j=i+1}^n \hat{C}_{jn} \right) \sum_{k=n-i+1}^{n-1} \frac{2\hat{\sigma}_k^2}{\hat{f}_k^2 \sum_{n=1}^{n-k} C_{nk}} \right), \quad (3.4)$$

όπου $SE(\hat{R}_i)$ το τυπικό σφάλμα (*standard error*) της εκτίμησης του αποθέματος.

Απόδειξη. Με βάση τον Mack (1991) έχουμε:

$$\begin{aligned} MSEP(\sum_{i=2}^n \hat{R}_i) &= E\left(\left(\sum_{i=2}^n \hat{R}_i - \sum_{i=2}^n R_i\right)^2 \mid \mathcal{M}\right) \\ &= E\left(\left(\sum_{i=2}^n \hat{C}_{in} - \sum_{i=2}^n C_{in}\right)^2 \mid \mathcal{M}\right) \\ &= V\left(\sum_{i=2}^n C_{in} \mid \mathcal{M}\right) + \left(E\left(\sum_{i=2}^n C_{in} \mid \mathcal{M}\right) - \sum_{i=2}^n \hat{C}_{in}\right)^2. \end{aligned}$$

Λόγω της ανεξαρτησίας των ετών ατυχημάτων έχουμε:

$$V\left(\sum_{i=2}^n C_{in} \mid \mathcal{M}\right) = \sum_{i=2}^n V(C_{in} \mid \mathcal{M}).$$

Επιπλέον

$$\begin{aligned} \left(E\left(\sum_{i=2}^n C_{in} \mid \mathcal{M}\right) - \sum_{i=2}^n \hat{C}_{in}\right)^2 &= \left(\sum_{i=2}^n (E(C_{in} \mid \mathcal{M}) - \hat{C}_{in})\right)^2 \\ &= \sum_{i,j} (E(C_{in} \mid \mathcal{M}) - \hat{C}_{in})(E(C_{jn} \mid \mathcal{M}) - \hat{C}_{jn}) \\ &= \sum_{i,j} C_{i,n+1-i} C_{j,n+1-j} F_i F_j, \end{aligned}$$

με

$$F_i = f_{n+1-i} \cdots f_{n-1} - \hat{f}_{n+1-i} \cdots \hat{f}_{n-1}.$$

Επειδή

$$MSEP(\hat{R}_i) = V(C_{in} \mid \mathcal{M}) + (C_{i,n+1-i} F_i)^2,$$

έχουμε

$$MSEP\left(\sum_{i=2}^n \hat{R}_i\right) = \sum_{i=2}^n MSEP(\hat{R}_i) + \sum_{2 \leq i < j \leq n} 2C_{i,n+1-i} C_{j,n+1-j} F_i F_j.$$

Με μια ανάλογη διαδικασία για την απόδειξη του F^2 του προηγούμενου θεωρήματος παίρνουμε τον εκτιμητή του $F_i F_j$ για $i < j$:

$$\frac{\sum_{k=n+1-i}^{n-1} \hat{f}_{n+1-j} \cdots \hat{f}_{n-i} \hat{f}_{n+1-i}^2 \cdots \hat{f}_{k-1}^2 \hat{\sigma}_k^2 \hat{f}_{k+1}^2 \cdots \hat{f}_{n-1}^2}{\sum_{n=1}^{n-k} C_{nk}}.$$

Έτσι ολοκληρώνεται και η απόδειξη.

□

3.3.1. Υπολογισμός Αποθέματος IBNR

Όπως είδαμε και στο προηγούμενο κεφάλαιο, όταν ισχύουν οι υποθέσεις του Μοντέλου του Mack (βλέπε Wüthrich and Merz (2008)), παίρνουμε εκτιμήσεις αποθεμάτων ίδιες με την μέθοδο Chain-Ladder:

Πίνακας 3.1: Αποτελέσματα Μεθόδου Mack

Έτος	Τιμή Κύριας Διαγωνίου (Latest Claim)	Εξέλιξη μέχρι Σήμερα (Development to date)	Τελική Απαίτηση (Ultimate Claim)	IBNR	Τυπικό Σφάλμα Mack (Mack Standard Error)	Συντελεστής της Διασποράς (Coefficient of Variance)
1	3901463	1	3901463	0	0	-
2	5339085	0.98	5433719	94633.81	75535.04	0.80
3	4909315	0.91	5378826	469511.29	121698.56	0.26
4	4588268	0.87	5297906	709637.82	133548.85	0.19
5	3873311	0.80	4858200	984888.64	261406.45	0.27
6	3691712	0.72	5111171	1419459.46	411009.70	0.29
7	3483130	0.62	5660771	2177640.62	558316.86	0.26
8	2864498	0.42	6784799	3920301.01	875327.51	0.22
9	1363294	0.24	5642266	4278972.26	971257.81	0.23
10	344014	0.07	4969825	4625810.69	1363154.91	0.29
Σύνολο	34358090	0.65	53038946	18680855.61	2447094.86	0.13

Στον παραπάνω πίνακα έχουμε:

- **Τιμή Κύριας Διαγωνίου (Latest Claim):** Είναι η τελευταία απαίτηση που πληρώθηκε από την ασφαλιστική. Δηλαδή τα στοιχεία της κύριας διαγωνίου του Πίνακα 2.4.
- **Τελική Απαίτηση (Ultimate Claim):** Είναι οι τελικές απαιτήσεις που εκτιμηθήκαν για τα έτη εξέλιξης που διαθέτουμε. Με λίγα λόγια, τα \hat{C}_{i10} για $2 \leq i \leq 10$ και το $C_{1,10}$ (τελευταία στήλη του 10×10 πίνακα).
- **Εξέλιξη μέχρι Σήμερα (Development to date):** Ορίζεται ως $\frac{Latest}{Ultimate}$.
- **IBNR:** Το απόθεμα για τις ζημιές που έγιναν αλλά δεν έχουν ακόμα αναγγελθεί στην ασφαλιστική.
- **Τυπικό Σφάλμα Mack (Mack Standard Error):** Ορίζεται ως η τετραγωνική ριζά του μέσου τετραγωνικού σφάλματος πρόβλεψης (MSEP), υπολογισμένο με τη σχέση (3.3) για κάθε έτος ατυχήματος, και με τη σχέση (3.4) για το συνολικό απόθεμα.

- **Συντελεστής της Διασποράς (Coefficient of Variance):** Συντελεστής της διασποράς (Coefficient of Variance), ορίζεται ως $\frac{Mack.S.E.}{IBNR}$.

Το συνολικό απόθεμα (IBNR) είναι 18680856, ίσο δηλαδή με την εκτίμηση της μεθόδου Chain-Ladder.

3.4. Γενικευμένα Γραμμικά Μοντέλα (Generalized Linear Models - GLM)

Τα Γενικευμένα Γραμμικά Μοντέλα (GLM) αφορούν τη σχέση μεταξύ μεταβλητών. Γενικεύουν το κλασικό κανονικό γραμμικό μοντέλο, κάνοντας χρήση λιγότερο αυστηρών περιορισμών, παρέχοντάς μας μεθόδους για την ανάλυση δεδομένων που δεν ακολουθούν την κανονική κατανομή. Τα GLM διαδραματίζουν έναν σημαντικό ρόλο στον τομέα των ασφαλίσεων, και συγκεκριμένα των αποθεμάτων, για το λόγο ότι τα δεδομένα σε μια ασφαλιστική (π.χ. ζημιές) τις περισσότερες των περιπτώσεων δεν ακολουθούν κανονική κατανομή. Σύμφωνα με τους Renshaw and Verrall (1998), Pinheiro, Andrade e Silva and Centeno (2003), Liivika, Meelis and Raunos (2017), η δομή των GLM είναι η εξής:

1. Οι προσαυξητικές απαιτήσεις X_{ij} ανήκουν στην εκθετική οικογένεια κατανομών.
2. $E(X_{ij}) = \mu_{ij}$.
3. $\eta_{ij} = g(\mu_{ij})$, όπου η συνάρτηση $g(\cdot)$ είναι η συνάρτηση σύνδεσης και η_{ij} ένας γραμμικός εκτιμητής.
4. $\eta_{ij} = c + a_i + b_j$.

Σύμφωνα λοιπόν με το μοντέλο που αναπτύξαμε παραπάνω μπορούμε να παράγουμε μερικές τεχνικές αποθεματοποίησης, ανάλογα με την υπόθεση μας για την κατανομή των προσαυξητικών απαιτήσεων X_{ij} . Είναι σύνηθες στην αποθεματοποίηση ζημιών να χρησιμοποιούμε την υπερδιασκορπισμένη (over-dispersed) κατανομή Poisson, την κατανομή Γάμμα (Gamma) ή την Λογαριθμοκανονική (Log-normal) κατανομή. Θα αναλύσουμε κάθε ένα μοντέλο παρακάτω.

3.4.1. Μοντέλο Υπερδιασκορπισμένη (Over-Dispersed) Poisson

Ήδη το 1975 προτάθηκε ένα στοχαστικό μοντέλο για GLM, το μοντέλο Poisson, το οποίο οδηγεί στην τεχνική Chain-Ladder. Το συγκεκριμένο δουλεύει με τις προσαυξητικές απαιτήσεις X_{ij} από μια κατανομή Poisson, όπου η μέση τιμή είναι $E(X_{ij}) = \alpha_i \beta_j$ με άγνωστες παραμέτρους α_i και β_j . Εδώ η παράμετρος α_i είναι οι αναμενόμενες τελικές (ultimate) απαιτήσεις (μέχρι το τελευταίο έτος εξέλιξης που έχει παρατηρηθεί), και η β_j είναι το ποσοστό των τελικών απαιτήσεων που αντιστοιχούν σε κάθε έτος εξέλιξης, με $\sum_{j=1}^n \beta_j = 1$. Ο παραπάνω περιορισμός προκύπτει από το γεγονός ότι τα β_j εισάγονται ως τα ποσοστά των απαιτήσεων που αναφέρονται σε κάθε έτος εξέλιξης j . Προφανώς, το συνολικό ποσοστό για όλα τα έτη εξέλιξης πρέπει να ισούται με 1.

Ένας τρόπος να εκτιμήσουμε τις άγνωστες παραμέτρους α_i και β_j είναι με την μέθοδο μέγιστης πιθανοφάνειας (Maximum Likelihood Estimation, MLE) χρησιμοποιώντας το τρίγωνο με τα δεδομένα μας. Στην συνέχεια, θα χρησιμοποιήσουμε τον συμβολισμό Δ για το τρίγωνο των δεδομένων μας, δηλαδή το σύνολο των κελιών (i, j) όπου τα X_{ij} είναι γνωστά (δηλαδή για $i + j \leq n + 1$). Επίσης έχουμε το σύνολο $\Delta_i = \{j : (i, j) \in \Delta\}$ και το $\Delta^j = \{i : (i, j) \in \Delta\}$. Η όλη διαδικασία δίνεται αναλυτικά στον Mack (1991) από όπου προκύπτουν οι εξής εκτιμητές:

$$\hat{\alpha}_i = \frac{\sum_{j \in \Delta_i} X_{ij}}{\sum_{j \in \Delta_i} \beta_j}, \quad i = 1, \dots, n$$

και:

$$\hat{\beta}_j = \frac{\sum_{i \in \Delta^j} X_{ij}}{\sum_{i \in \Delta^j} \alpha_i}, \quad j = 1, \dots, n.$$

Συνεπώς, οι παράγοντες β_j εκφράζουν την αναλογία του αθροίσματος των παρατηρούμενων προσαυξητικών τιμών για ένα συγκεκριμένο έτος εξέλιξης j που αντιστοιχούν σε συγκεκριμένες τελικές απαιτήσεις, δηλαδή τα β_j είναι η αναλογία των απαιτήσεων που αναφέρθηκαν στο έτος εξέλιξης j . Από την άλλη, η παράμετρος α_i , αναφέρεται στην αναλογία του αθροίσματος των παρατηρούμενων προσαυξητικών απαιτήσεων για ένα συγκεκριμένο έτος ατυχήματος i . Η διαδικασία παρουσιάζει αρκετές ομοιότητες με την μέθοδο Chain-Ladder, όπου και εκεί οι συντελεστές εξέλιξης f_j είναι αποτέλεσμα συγκεκριμένων αναλογιών. Για το μοντέλο Poisson, διαλέγοντας ως συνάρτηση σύνδεσης (link function) τη λογαριθμική, έτσι ώστε

$$\eta_{ij} = \ln(\mu_{ij}),$$

έχουμε τις εξής υποθέσεις:

$$E(X_{ij}) = \mu_{ij} = \exp(\ln(\alpha_i) + \ln(\beta_j))$$

ή ισοδύναμα

$$\ln(E(X_{ij})) = \eta_{ij} = \ln(\alpha_i) + \ln(\beta_j), \quad (3.5)$$

και με μια αναπαραμετροποίηση (reparametrizing) παίρνουμε τον τελικό τύπο για το μοντέλο μας:

$$\eta_{ij} = c + a_i + b_j, \quad (3.6)$$

όπου $a_i = \ln(\alpha_i)$, $b_j = \ln(\beta_j)$ και η παράμετρος c είναι σταθερή και αντιστοιχεί στο προσαυξητικό ποσό του κελιού $(1, 1)$, λόγω του περιορισμού $a_1 = b_1 = 0$ για να αποφύγουμε τις παραπάνω παραμέτρους από όσες χρειαζόμαστε (over-parametrization). Επιπλέον λεπτομέρειες για το Poisson μοντέλο μπορεί κανείς να βρει στους Kuang and Nielsen (2009). Επιτρέποντας στα δεδομένα μας να έχουν μια μικρότερη ή μεγαλύτερη διασκόρπιση (dispersion) το μοντέλο Poisson με $V(X_{ij}) = E(X_{ij})$ μπορεί να γενικευτεί σε ένα Over-Dispersed Poisson (ODP) μοντέλο με $V(X_{ij}) = \phi E(X_{ij})$. Στον Schmidt (2002) αποδεικνύεται πως κάθε μοντέλο Poisson μπορεί να μετατραπεί σε Over-Dispersed Poisson διαιρώντας κάθε X_{ij} με μια συγκεκριμένη παράμετρο. Η γενική μορφή του ODP μοντέλου είναι:

$$E(X_{ij}) = \mu_{ij} = \alpha_i \beta_j,$$

$$V(X_{ij}) = \phi \alpha_i \beta_j,$$

όπου $\sum_{j=1}^n \beta_j = 1$.

Η υπερδιασκόρπιση (over-dispersion) εισάγεται από την παράμετρο ϕ , η οποία είναι άγνωστη και πρέπει να εκτιμηθεί από τα δεδομένα. Έχοντας λοιπόν μια προσαυξητική ζημιά, είμαστε σε

θέση να εκτιμήσουμε τις μελλοντικές πληρωμές εισάγοντας τα εκτιμημένα $\hat{\alpha}_i$ και $\hat{\beta}_j$ στη σχέση (3.5), με αποτέλεσμα:

$$\hat{X}_{ij} = \hat{\alpha}_i \hat{\beta}_j = \exp(\hat{\eta}_{ij}). \quad (3.7)$$

Αφού εκτιμήσουμε όλες τις μελλοντικές προσαυξητικές απαιτήσεις, είναι εύκολο να εκτιμήσουμε τα αποθέματα για οποιοδήποτε έτος ατυχήματος με τη σχέση:

$$\hat{R}_i = \hat{\alpha}_i \hat{\beta}_{n+2-i} + \dots + \hat{\alpha}_i \hat{\beta}_n, \quad i = 2, \dots, n \quad (3.8)$$

και την συνολική εκτίμηση αποθέματος από τον τύπο:

$$\hat{R} = \sum_{i=2}^n \hat{R}_i = \sum_{i=2}^n (\hat{\alpha}_i \hat{\beta}_{n+2-i} + \dots + \hat{\alpha}_i \hat{\beta}_n), \quad i = 2, \dots, n. \quad (3.9)$$

Το Αρνητικό Διωνυμικό (Negative Binomial) μοντέλο μπορεί να παραχθεί από το μοντέλο Poisson, οπότε αυτά τα δύο μοντέλα είναι παρόμοια, αλλά με διαφορετική παραμετροποίηση (βλέπε Verrall (2000)). Η κατανομή πρόβλεψης και για τα δύο μοντέλα είναι σχεδόν η ίδια και δίνει ταυτόσημες προβλέψεις για τα αποθέματα.

3.4.2. Μοντέλο Γάμμα (Gamma)

Ένα ακόμα μοντέλο προτάθηκε από τον Mack (1991) με πολλαπλασιαστική παραμετρική δομή για τον μέσο των προσαυξητικών απαιτήσεων, οι οποίες ακολουθούν πλέον μια κατανομή Γάμμα. Η ίδια μοντελοποίηση όπως αυτή στο ODP μοντέλο είναι δυνατή, με την ίδια συνάρτηση σύνδεσης (λογαριθμική), κάνοντας άπλα μια αλλαγή στην εξίσωση $V(X_{ij}) = \phi \alpha_i \beta_j$, παίρνοντας εκτιμήσεις για τα αποθέματα που είναι πολύ κοντά με αυτές της Chain-Ladder μεθόδου. Το μοντέλο Γάμμα δίνεται με μέση τιμή:

$$E(X_{ij}) = \mu_{ij} = \alpha_i \beta_j,$$

και διασπορά:

$$V(X_{ij}) = \phi (E(X_{ij}))^2 = \phi \mu_{ij}^2,$$

Ο γραμμικός εκτιμητής η_{ij} (3.6) συνδέεται πάλι με την μέση τιμή μ_{ij} μέσω της λογαριθμικής συνάρτησης σύνδεσης, δηλαδή:

$$\eta_{ij} = \ln(\mu_{ij}) = c + a_i + b_j,$$

με $a_1 = b_1 = 0$.

Προκειμένου να εξάγουμε τις εκτιμήσεις για τα αποθέματα κάθε έτους ατυχήματος και το συνολικό απόθεμα χρησιμοποιούμε τις σχέσεις (3.8) και (3.9) όπως στο μοντέλο ODP. Ο περιορισμός και στα δύο μοντέλα είναι πως οι προσαυξητικές απαιτήσεις πρέπει να είναι μη-αρνητικές.

3.4.3. Λογαριθμοκανονικό (Log-normal) Μοντέλο

Υποθέτοντας πως οι προσαυξητικές απαιτήσεις X_{ij} ακολουθούν την λογαριθμοκανονική κατανομή (log-normal), μπορούμε να χρησιμοποιήσουμε τις υποθέσεις που κάναμε για τα γενικευμένα γραμμικά μοντέλα για τον λογάριθμο των X_{ij} (βλέπε Kremer (1982)). Το μοντέλο μας (όπως παρουσιάστηκε πρώτα από τον Kremer (1982) και χρησιμοποιήθηκε από τους Renshaw (1989), Verrall (1989,1991a,1991b), Zehnirith (1989,1991) and Christofides (1990)) δίνεται από τις εξής υποθέσεις:

$$\ln(X_{ij}) \sim \mathcal{N}(\mu_{ij}, \sigma^2),$$

δηλαδή,

$$E(\ln(X_{ij})) = \mu_{ij}, \quad V(\ln(X_{ij})) = \sigma^2.$$

Εδώ, γίνεται χρήση της ταυτοτικής συνάρτησης σύνδεσης (λογαριθμική) και οι προσαυξητικές τιμές σε λογαριθμική κλίμακα $\ln(X_{ij})$ μπορούν να αναλυθούν σε δυο συνιστώσες, μια ντετερμινιστική μη τυχαία μεταβλητή με μέσο $\eta_{ij} = \mu_{ij}$ και ένα τυχαίο σφάλμα ϵ_{ij} που ακολουθεί την κανονική κατανομή με μέση τιμή 0 και διασπορά σ^2 :

$$\ln(X_{ij}) = \mu_{ij} + \epsilon_{ij}.$$

Ακολουθώντας τους Verrall and England (2002), οι προσαρμοσμένες τιμές σε λογαριθμική κλίμακα $\ln(\hat{X}_{ij})$, μπορούν να υπολογιστούν εκτιμώντας τους παράγοντες η_{ij} και τη διασπορά σ^2 :

$$\begin{aligned}\ln(\hat{X}_{ij}) &= \hat{\eta}_{ij} + \frac{1}{2}\hat{\sigma}_{ij}^2 \\ \Leftrightarrow \hat{X}_{ij} &= \exp(\hat{\eta}_{ij} + \frac{1}{2}\hat{\sigma}_{ij}^2),\end{aligned}$$

που είναι και η μορφή της αναμενόμενης τιμής μιας λογαριθμοκανονικής κατανομής, όπου:

$$\hat{\sigma}_{ij}^2 = V(\hat{\eta}_{ij}) + \hat{\sigma}^2,$$

είναι η διασπορά πρόβλεψης του γραμμικού εκτιμητή η_{ij} , (3.6).

Με παρόμοιο τρόπο όπως προηγουμένως (Υποκεφάλαιο 3.4.1), θα συμβολίσουμε με D το σύνολο των κελιών (i, j) από το τρίγωνο των εκτιμημένων απαιτήσεων (δηλαδή για $i+j > n+1$) από τις οποίες υπολογίζουμε τα απόθεμα κάθε έτους, αλλά και το συνολικό. Το απόθεμα για το έτος ατυχήματος i είναι το άθροισμα των στοιχείων του D στην αντίστοιχη γραμμή και δίνεται από τον τύπο

$$\hat{R}_i = \sum_{j \in D} \hat{C}_{ij}.$$

Τέλος, το συνολικό απόθεμα είναι το:

$$\hat{R} = \sum_{i,j \in D} \hat{C}_{ij} = \sum_{i=2}^n \hat{R}_i.$$

3.5. Μέθοδος Bootstrap

Η στατιστική τεχνική Bootstrap είναι μια μέθοδος δειγματοληψίας με επανάθεση, με την βοήθεια της οποίας μπορούμε να υπολογίσουμε την κατανομή πρόβλεψης και το σφάλμα μιας εκτίμησης. Το πλεονέκτημα αυτής της μεθόδου είναι η απλότητά της στον υπολογισμό αυτών των σφαλμάτων, χωρίς να κάνουμε κάποια υπόθεση για την κατανομή από την οποία έγινε η δειγματοληψία.

Ας υποθέσουμε λοιπόν ότι έχουμε μια κατανομή F και n το πλήθος παρατηρήσεις $\vec{x} = (x_1, x_2, \dots, x_n)$ από τις οποίες θέλουμε να εκτιμήσουμε ένα στατιστικό στοιχείο $\hat{\theta}$ το οποίο

προέρχεται εφαρμόζοντας μια συνάρτηση $s(\cdot)$, δηλαδή $\hat{\theta} = s(\bar{x})$. Η συνάρτηση $s(\cdot)$ μπορεί να είναι για παράδειγμα ο μέσος, η διάμεσος, η τυπική απόκλιση κ.ο.κ.

Ο σκοπός μας λοιπόν είναι να απαντήσουμε στην ερώτηση: Πόσο σωστή είναι η εκτίμηση μας $\hat{\theta}$ για το θ ; Υποθέτοντας ότι τα x_i είναι ανεξάρτητα μεταξύ τους και προέρχονται από την ίδια κατανομή F χρησιμοποιούμε την μέθοδο Bootstrap και συγκεκριμένα την επαναδειγματοληψία μέσω της Bootstrap για να υπολογίσουμε το παραπάνω σφάλμα.

Για να ξεκαθαρίσουμε τον όρο επαναδειγματοληψία, ξεκινάμε εξάγοντας ένα διάνυσμα \bar{x}^* από τα στοιχεία του αρχικού διανύσματος \bar{x} με τυχαία σειρά και επανάληψη (κάποιο στοιχείο μπορεί να μην επιλεγεί καμία φορά). Για παράδειγμα το \bar{x}^* μπορεί να είναι το διάνυσμα $(x_2, x_3, x_5, x_2, \dots, x_1)$ μήκους n όπως το αρχικό. Από το \bar{x}^* που προέκυψε με την μέθοδο Bootstrap μπορούμε να υπολογίσουμε μια νέα εκτιμήτρια του θ , την $\hat{\theta}^* = s(\bar{x}^*)$, όπου $s(\cdot)$ η ίδια συνάρτηση που χρησιμοποιήθηκε προηγουμένως.

Επαναλαμβάνοντας την παραπάνω δειγματοληψία αρκετές φορές, είμαστε σε θέση να χτίσουμε μια κατανομή των εκτιμητριών $\hat{\theta}$ η οποία μπορεί να μας δώσει πληροφορίες για το τυπικό σφάλμα (standard error) που κάναμε υπολογίζοντας την εκτιμήτρια $\hat{\theta}$.

Πιο αναλυτικά, ο αλγόριθμος της μεθόδου Bootstrap για την εκτίμηση του τυπικού σφάλματος (standard error) όπως παρουσιάστηκε από τους Efron and Tibshirani (1993) είναι ο εξής:

- Δημιουργούμε B ανεξάρτητα δείγματα Bootstrap ή αλλιώς ψευδο-δείγματα $\bar{x}^{*1}, \bar{x}^{*2}, \dots, \bar{x}^{*B}$, μήκους n από τα στοιχεία του αρχικού διανύσματος \bar{x} με επανάληψη.
- Υπολογίζουμε την εκτίμηση $\hat{\theta}$ που αντιστοιχεί σε κάθε δείγμα:

$$\hat{\theta}^*(b) = s(\bar{x}^{*b}), \quad b = 1, 2, \dots, B.$$

- Εκτιμάμε το τυπικό σφάλμα (standard error) $se_F(\hat{\theta})$ από την δειγματική τυπική απόκλιση των B ψευδο-δειγμάτων. Το αποτέλεσμα ονομάζεται Bootstrap τυπικό σφάλμα (\hat{se}_B) και δίνεται από τη σχέση:

$$\hat{se}_B = \sqrt{\frac{\sum_{b=1}^B [\hat{\theta}^*(b) - \hat{\theta}^*(\cdot)]^2}{B - 1}},$$

$$\text{όπου } \hat{\theta}^*(\cdot) = \frac{\sum_{b=1}^B \hat{\theta}^*(b)}{B}.$$

Όσον αφορά τον αριθμό επαναλήψεων B έχουμε τους εξής δυο κανόνες (βλέπε Efron and Tibshirani (1993)):

1. Για την εκτίμηση του τυπικού σφάλματος το B κυμαίνεται μεταξύ 25 και 200. Παρόλα αυτά, συνήθως οι 25 επαναλήψεις δεν δίνουν αρκετές πληροφορίες ώστε να εκτιμήσουμε το $se_F(\hat{\theta})$. Όταν το $B = 50$ τότε έχουμε μια καλή εκτίμηση του παραπάνω.
2. Για τον υπολογισμό διαστημάτων εμπιστοσύνης χρειαζόμαστε πολύ μεγαλύτερο αριθμό επαναλήψεων B .

Μια άλλη προσέγγιση για τον αριθμό των επαναλήψεων B αναφέρεται στους Davidson and MacKinnon (2000), που εφαρμόζεται επίσης και στους ελέγχους Monte Carlo, με τους οποίους η μέθοδος Bootstrap είναι ισοδύναμη εάν η κατανομή του προς μελέτη στατιστικού στοιχείου δεν εξαρτάται από άγνωστες παραμέτρους (pivotal). Εάν θέλουμε ο έλεγχος να είναι ακριβής, τότε για α επίπεδο σημαντικότητας ισχύει η επόμενη σχέση:

$$\alpha(1 + B) \in Z.$$

Παρόλα αυτά οι πλειοψηφία των συγγραφέων, δεδομένου της ανάπτυξης της τεχνολογίας και συγκεκριμένα των ηλεκτρικών υπολογιστών, προτείνουν η διαδικασία να επαναλαμβάνεται 1000 ή 10000 φορές για την εκτίμηση της κατανομής πρόβλεψης (predictive distribution) οποιουδήποτε στατιστικού στοιχείου ή εκτιμήτριας.

3.5.1. Περιπτώσεις Χρήσης της Μεθόδου

Σύμφωνα με τους Adèr, Mellenbergh and Hand (2008), η μέθοδος Bootstrap χρησιμοποιείται στις εξής περιπτώσεις:

- Όταν η θεωρητική κατανομή του στατιστικού στοιχείου που μας ενδιαφέρει είναι πολύπλοκη ή άγνωστη. Από την στιγμή που η διαδικασία Bootstrap είναι ανεξάρτητη κατανομής,

μας παρέχει μια έμμεση μέθοδο για να αξιολογήσουμε τις ιδιότητες της κατανομής που “κρύβεται” στο δείγμα αλλά και τις παραμέτρους που εξάγονται από αυτή την κατανομή.

- Όταν το μέγεθος του δείγματος δεν είναι επαρκές για σαφή συμπεράσματα. Εάν η “κρυφή” κατανομή είναι γνωστή, η μέθοδος Bootstrap μας παρέχει έναν τρόπο να υπολογίσουμε τις διασπρεβλώσεις που προκαλούνται από το συγκεκριμένο δείγμα, το οποίο είναι πιθανόν να μην αντιπροσωπεύει πλήρως τον πληθυσμό.
- Όταν εκτελούνται δυναμικοί υπολογισμοί και έχουμε ένα μικρό δείγμα-πilotο (το αρχικό μας δείγμα). Οι περισσότεροι δυναμικοί υπολογισμοί και οι υπολογισμοί του μεγέθους του δείγματος εξαρτώνται σε μεγάλο βαθμό από την τυπική απόκλιση του στατιστικού στοιχείου ενδιαφέροντος. Εάν η εκτίμηση που χρησιμοποιήθηκε είναι λανθασμένη, το ίδιο θα είναι και το απαιτούμενο μέγεθος του δείγματος. Μια μέθοδος για να πάρουμε μια γεύση από την διασπορά του στατιστικού στοιχείου, είναι να χρησιμοποιήσουμε μικρό πιλοτικό δείγμα και να εκτελέσουμε την διαδικασία Bootstrap.

3.6. Μέθοδος Bootstrap στα Τρίγωνα Εξέλιξης Ζημιών

Στο Υποκεφάλαιο 3.4 μελετήσαμε τα γενικευμένα γραμμικά μοντέλα (GLM) στον τομέα της αποθεματοποίησης ζημιών. Περιγράψαμε τους τρόπους με τους οποίους μπορούμε να υπολογίσουμε τις παραμέτρους του μοντέλου, και ως εκ τούτου, τις εκτιμήσεις των αποθεμάτων που χρειάζονται για να καλύψουν τις μελλοντικές υποχρεώσεις μιας ασφαλιστικής. Επιπλέον, είναι δυνατόν να υπολογιστεί το Μέσο Τετραγωνικό Σφάλμα Πρόβλεψης (Mean Square Error of Prediction - MSEP), δηλαδή η αβεβαιότητα στις εκτιμήσεις μας. Ο τύπος για το MSEP είναι:

$$\begin{aligned}
 MSEP &= E[(X_{ij} - \hat{X}_{ij})^2] \\
 &= E[(X_{ij} - E(X_{ij})) - (\hat{X}_{ij} - E(X_{ij}))]^2 \\
 &= E[(X_{ij} - E(X_{ij}))^2] + E[(\hat{X}_{ij} - E(\hat{X}_{ij}))^2],
 \end{aligned}
 \tag{3.10}$$

που ισχύει από την ανεξαρτησία μεταξύ των παρατηρήσεων μας. Όπως φαίνεται το MSEP αποτελείται από δύο συνιστώσες:

- την διασπορά της διαδικασίας (Process Variance),

- και την διασπορά της εκτίμησης (Estimation Variance).

Όμως, η συγκεκριμένη προσέγγιση έχει το μειονέκτημα ότι η (3.10), τις περισσότερες των περιπτώσεων, είναι δύσκολη στην επίλυση της. Επιπρόσθετα, είναι χρήσιμο εκτός από τις πρώτες δύο ροπές της κατανομής των προβλέψεων (μέση τιμή και διασπορά) να είμαστε σε θέση να υπολογίσουμε ολόκληρη την κατανομή (predictive distribution). Ο υπολογισμός αυτός δύναται να είναι περίπλοκος όταν τα δεδομένα μας είναι σε μορφή τριγώνων εξέλιξης (development triangles), όπου χρειάζεται να εξάγουμε την κατανομή αθροίσματος τυχαίων μεταβλητών.

Σε μια τέτοια περίπτωση, καταφεύγουμε στην μέθοδο Bootstrap. Η μέθοδος αυτή είναι ιδιαίτερα γνωστή στον τομέα της στοχαστικής αποθεματοποίησης ζημιών, λόγω της απλότητας και της ευελιξίας στην προσέγγιση της. Χρησιμοποιούμε την Bootstrap για να υπολογίσουμε το σφάλμα πρόβλεψης (prediction error), για να προσεγγίσουμε την κατανομή πρόβλεψης, και ως αποτέλεσμα αυτών, να ποσοτικοποιήσουμε τον κίνδυνο μιας ασφαλιστικής να μην μπορέσει να ανταπεξέλθει στις μελλοντικές της υποχρεώσεις. Λέγοντας ποσοτικοποίηση του κινδύνου, εννοούμε πως με την μέθοδο Bootstrap μπορούμε να υπολογίσουμε για παράδειγμα τα ποσοστημόρια (quantiles), την Αξία σε Κίνδυνο (Value at Risk - VaR), την Δεσμευμένη Αξία σε Κίνδυνο (Conditional Value at Risk - CVaR).

Η μέθοδος bootstrap παρουσιάζεται αναλυτικά στους Efron and Tibshirani (1993) και βασίζεται στην επαναδειγματοληψία (resampling) των δεδομένων. Υπάρχουν δυο είδη μεθόδων Bootstrap:

- **Paired Bootstrap:** Η επαναδειγματοληψία γίνεται απευθείας από τα δεδομένα που εξετάζουμε, τα οποία έχουμε υποθέσει πως είναι ανεξάρτητα και ισόνομα.
- **Bootstrap καταλοίπων (Residual Bootstrap):** Η επαναδειγματοληψία γίνεται στα κατάλοιπα του μοντέλου, με τις παρατηρήσεις να θεωρούνται ισόνομες, αλλά όχι απαραίτητα ανεξάρτητες.

Παρόλο που η Paired Bootstrap είναι πιο ανθεκτική (robust) από την Bootstrap καταλοίπων, δεν χρησιμοποιείται όσον αφορά τα GLM's στον τομέα της αποθεματοποίησης ζημιών, καθώς υπάρχει εξάρτηση τόσο στα δεδομένα μας, όσο και στις εκτιμήτριες των παραμέτρων.

3.6.1. Κατάλοιπα (Residuals)

Όπως αναφέραμε παραπάνω, ο τρόπος με τον οποίο εξάγεται η κατανομή των προβλέψεων (predictive distribution) είναι η μέθοδος Bootstrap, είτε παραμετρική είτε απαραμετρική. Στην συγκεκριμένη περίπτωση, δηλαδή στην αποθεματοποίηση ζημιών, θα χρησιμοποιήσουμε την απαραμετρική Bootstrap καταλοίπων. Για τα GLM, ο λόγος που δεν εξετάζουμε τα απλά κατάλοιπα $r_{ij} = X_{ij} - \hat{X}_{ij}$, είναι η δυσκολία στον έλεγχο της εγκυρότητας της σχέσης μεταξύ μέση τιμής και διασποράς σε αυτά. Υπάρχουν 3 ειδών κατάλοιπα που χρησιμοποιούνται στην μέθοδο bootstrap (England and Verrall (1999)):

- **Κατάλοιπα Pearson:**

$$r_{ij}^P = \frac{X_{ij} - \hat{X}_{ij}}{\sqrt{V(\hat{X}_{ij})}}. \quad (3.11)$$

- **Κατάλοιπα Απόκλισης (Deviance):**

$$r_{ij}^D = \text{sign}(X_{ij} - \hat{X}_{ij})\sqrt{d_i}, \quad (3.12)$$

όπου d_i η απόκλιση (deviance) για μία μονάδα, με $\sum_{i=1}^n d_i = D$, όπου D η συνολική απόκλιση του μοντέλου.

- **Κατάλοιπα Anscombe:**

$$r_{ij}^A = \frac{A(X_{ij}) - A(\hat{X}_{ij})}{A'(\hat{X}_{ij})\sqrt{V(\hat{X}_{ij})}}, \quad (3.13)$$

όπου η συνάρτηση $A(\cdot)$ επιλέγεται με τέτοιο τρόπο έτσι ώστε να κάνει την κατανομή των $A(X_{ij})$ όσο το δυνατόν “πιο κανονική”.

Είναι φανερό πως στην διαδικασία Bootstrap χρησιμοποιούνται κυρίως τα κατάλοιπα Pearson λόγω της ευκολίας στην επίλυση της εξίσωσης:

$$X_{ij}^* = r_{ij}^{*P}\sqrt{V(\hat{X}_{ij})} + \hat{X}_{ij}, \quad (3.14)$$

όπου X_{ij}^* τα νέα δεδομένα και r_{ij}^{*P} τα κατάλοιπα που προέκυψαν από την μέθοδο bootstrap.

Προκειμένου λοιπόν αυτά τα κατάλοιπα να έχουν (περίπου) ίδια διασπορά είναι αναγκαίο να προσαρμοστούν (adjusted residuals). Υπάρχουν αρκετές προτάσεις από αρκετούς συγγραφείς όπως είναι οι England and Verrall (1999,2002), οι οποίοι πρότειναν έναν τρόπο προσαρμογής των καταλοίπων του Pearson πολλαπλασιάζοντας τα με έναν παράγοντα διόρθωσης:

$$r_{ij}^{PE} = \sqrt{\frac{N}{N-p}} r_{ij}^P, \quad (3.15)$$

όπου $N = \left(\frac{n(n+1)}{2}\right)$ είναι ο αριθμός των παρατηρήσεων και $p = 2n - 1$ είναι ο αριθμός των παραμέτρων που εκτιμηθήκαν. Μια ακόμα πρόταση σχετικά με το κλασσικό γραμμικό μοντέλο είναι να χρησιμοποιείται ο πίνακα επιρροής H (Hat Matrix) για να τυποποιήσουμε τα κατάλοιπα (standardized residuals):

$$r_{ij}^{*PS} = \frac{r_{ij}^P}{\sqrt{(1-h_{ij})}}, \quad (3.16)$$

όπου ο παράγοντας h_{ij} το αντίστοιχο διαγώνιο στοιχείο του πίνακα επιρροής H (βλέπε McCullagh and Nelder (1989)). Ο τύπος για τον πίνακα επιρροής για το κλασσικό μοντέλο είναι ο:

$$H = X(X^T X)^{-1} X^T,$$

και για το γενικευμένο γραμμικό μοντέλο (GLM) δίνεται από τη σχέση:

$$H = X(X^T W X)^{-1} X^T W,$$

όπου X είναι ο πίνακας σχεδιασμού και W ο διαγώνιος πίνακας τα στοιχεία του οποίου ικανοποιούν την σχέση:

$$w_{ii} = (V(\mu_{ij}) \left(\frac{\partial \eta_{ij}}{\partial \mu_{ij}}\right)^2)^{-1}. \quad (3.17)$$

Στο σημείο αυτό, και για τον λόγο ότι η διασπορά της εμπειρικής κατανομής είναι πολύ μικρή (βλέπε Wüthrich and Merz (2008)), είναι αναγκαία η χρήση μιας παραμέτρου κλίμακας

(scale parameter). Για τα κατάλοιπα του Pearson συμβολίζεται με $\hat{\phi}_p$ και υπολογίζεται από την σχέση:

$$\hat{\phi}_p = \frac{\sum_{i+j \leq (n+1)} (r_{ij}^P)^2}{N - p}, \quad (3.18)$$

όπου N και p είναι ο αριθμός των παρατηρήσεων και παραμέτρων προς εκτίμηση αντίστοιχα.

3.6.2. Τα βήματα της διαδικασίας

Στην συνέχεια παρουσιάζονται αναλυτικά τα βήματα της μεθόδου Bootstrap με στόχο να εξάγουμε την κατανομή των προβλέψεων για τα αποθέματα μας (βλέπε England and Verall (2002)):

1. Τοποθετούμε τις αθροιστικές απαιτήσεις C_{ij} σε ένα τρίγωνο.
2. Υπολογίζουμε τους συντελεστές εξέλιξης \hat{f}_j με την μέθοδο Chain-Ladder, σύμφωνα με την σχέση (2.1).
3. Υπολογίζουμε αναδρομικά και προς τα πίσω τις προσαρμοσμένες fitted συσσωρευμένες απαιτήσεις \hat{C}_{ij} , σύμφωνα με τις σχέσεις (2.7) και (2.8).
4. Υπολογίζουμε τις προσαρμοσμένες αυξητικές απαιτήσεις \hat{X}_{ij} , από τις σχέσεις (2.9) και (2.10).
5. Υπολογίζουμε τα κατάλοιπα Pearson σύμφωνα με την σχέση (3.11).
6. Στην συνέχεια υπολογίζουμε την παράμετρο κλίμακας $\hat{\phi}_P$, (3.18).
7. Υπολογίζουμε τα προσαρμοσμένα κατάλοιπα του Pearson (adjusted Pearson residuals) σύμφωνα με την σχέση (3.15).

Τα Βήματα “8-15” αφορούν μια κυκλική διαδικασία (loop process) που επαναλαμβάνεται B φορές:

8. Στα προσαρμοσμένα κατάλοιπα εφαρμόζουμε επαναδειγματοληψία με επανάληψη, δημιουργώντας ένα νέο τρίγωνο από κατάλοιπα. Στο συγκεκριμένη βήμα παρατηρείται πως το κατάλοιπο της τελευταίας γραμμής, πρώτης στήλης και εκείνο της τελευταίας στήλης, πρώτης γραμμής είναι πάντα 0. Αυτά τα δυο κατάλοιπα δεν πρέπει να συμπεριληφθούν στην διαδικασία επαναδειγματοληψίας καθώς στην μέθοδο Bootstrap οι τυχαίες μεταβλητές πρέπει να είναι ανεξάρτητες και ισόνομες, αλλά αν ληφθούν υπόψη θα υπάρχουν δύο μη τυχαίες μεταβλητές που είναι ίσες με μηδέν πάντα. ($\hat{X}_{1n} - X_{1n} = 0$ και $\hat{X}_{n1} - X_{n1} = 0$).
9. Σε κάθε κελί στο νέο τρίγωνο τοποθετούμε ένα X_{ij}^* από τη σχέση (3.14).
10. Υπολογίζουμε τις νέες ψευδο-αθροιστικές απαιτήσεις, C_{ij}^* με τη σχέση:

$$C_{ij}^* = \sum_{k=1}^j X_{ik}^*.$$

11. Εφαρμόζουμε το κλασικό μοντέλο Chain-Ladder στα ψευδο-δεδομένα μας και εκτιμάμε τις μελλοντικές αθροιστικές απαιτήσεις με την βοήθεια της σχέσης (2.3).
12. Υπολογίζουμε τις προσαυξητικές απαιτήσεις \hat{X}_{ij}^* , $i + j > n + 1$, όπου έπειτα στην προσομοίωση θα χρησιμοποιηθούν ως μέση τιμή.
13. Σε κάθε κελί (i, j) με $i + j > n + 1$ στο μελλοντικό τρίγωνο προσομοιώνουμε μια πληρωμή από την κατανομή της διαδικασίας (process distribution) με μέση τιμή $E(\hat{X}_{ij})$ και διασπορά $V(\hat{X}_{ij})$.
14. Προσθέτουμε τις προσομοιωμένες τιμές στο μελλοντικό τρίγωνο για κάθε έτος προσέλευσης, και γενικά, για να πάρουμε ένα απόθεμα Bootstrap για κάθε έτος ατυχήματος και το συνολικό αντίστοιχα.
15. Αποθηκεύουμε τα αποτελέσματα και επαναλαμβάνουμε τα Βήματα “8-15” B φορές.

Τα αποτελέσματα που αποθηκεύσαμε σχηματίζουν την κατανομή πρόβλεψης των μελλοντικών αποθεμάτων (predictive distribution). Ο μέσος της κατανομής πρόβλεψης πρέπει να

συγκριθεί με την εκτίμηση του αποθέματος από την μέθοδο Chain-Ladder. Η τυπική απόκλιση της δίνει μία εκτίμηση του σφάλματος πρόβλεψης.

Η παραπάνω κατανομή συμπεριλαμβάνει την εκτίμηση Bootstrap για το σφάλμα εκτίμησης (estimation error) καθώς επίσης και την διασπορά της διαδικασίας (process variance). Η τελευταία είναι αποτέλεσμα του βήματος “13”.

Αυτή η προσέγγιση της μεθόδου Bootstrap στα τρίγωνα εξέλιξης παρουσιάστηκε από τους England and Verrall (2002), οι οποίοι πρότειναν τον πολλαπλασιασμό της διασποράς της διαδικασίας (process variance) με τον παράγοντα $\frac{N}{N-p}$, για τον λόγο ότι τα κατάλοιπα χρησιμοποιούνται στην διαδικασία δίχως να μας ενδιαφέρει το πως αποκτήθηκαν και προκειμένου να ληφθούν υπόψη ο αριθμός των παραμέτρων που εκτιμηθήκαν. Το MSEP υπολογίζεται από τον τύπο (βλέπε Liivika, Meelis and Raunos (2017)):

$$MSEP(\hat{R}) = V(\hat{R}) + \frac{N}{N-p} V_{bs}(\hat{R}). \quad (3.19)$$

Ο πρώτος όρος του δεξιού σκέλους της εξίσωσης είναι η εκτίμηση της διασποράς της διαδικασίας (θα την υπολογίσουμε στην συνέχεια για κάθε μοντέλο ξεχωριστά). Ο δεύτερος όρος είναι η Bootstrap διασπορά της εκτίμησης και υπολογίζεται από την σχέση:

$$V_{bs}(\hat{R}) = \frac{1}{B} \sum_{b=1}^B (\hat{R}_{(b)}^* - \hat{R})^2, \quad (3.20)$$

όπου B είναι ο αριθμός των επαναλήψεων της διαδικασίας, \hat{R} η εκτίμηση του αποθέματος με την μέθοδο Chain-Ladder και $\hat{R}_{(b)}^*$ η εκτίμηση του αποθέματος για κάθε επανάληψη της διαδικασίας.

Στην συνέχεια θα παρουσιάσουμε την μέθοδο Bootstrap στα μοντέλα ODP και Γάμμα που αναφέραμε προηγουμένως.

3.6.3. Μοντέλο Υπερδιασκορπισμένη (Over-Dispersed) Poisson για Bootstrap

Το πρώτο μοντέλο που θα μελετήσουμε όσον αφορά την μέθοδο Bootstrap είναι το ODP μοντέλο που περιγράψαμε προηγουμένως σύμφωνα με τις υποθέσεις:

$$E(X_{ij}) = \mu_{ij} = \alpha_i \beta_j,$$

$$V(X_{ij}) = \phi\alpha_i\beta_j,$$

$$\eta_{ij} = \ln(\mu_{ij}) = c + a_i + b_j,$$

όπου $a_1 = b_1 = 0$.

Χρησιμοποιώντας λοιπόν την λογαριθμική συνάρτηση σύνδεσης έχουμε το έξης αποτέλεσμα για τις προσαρμοσμένες (fitted) τιμές του μοντέλου:

$$\hat{\mu}_{ij} = \exp(c + \hat{a}_i + \hat{b}_j), \quad (3.21)$$

τις οποίες χρειαζόμαστε για να υπολογίσουμε τα κατάλοιπα Pearson όπως οριστήκαν από την σχέση (3.11). Επομένως σύμφωνα με τις υποθέσεις τις μέσης τιμής και της διασποράς για το μοντέλο ODP έχουμε:

$$r_{ij}^P = \frac{X_{ij} - \hat{\mu}_{ij}}{\sqrt{\hat{\phi}\hat{\mu}_{ij}}}, \quad i + j \leq n + 1. \quad (3.22)$$

Για τον λόγο ότι η παράμετρο κλίμακας ϕ είναι σταθερή για όλα τα δεδομένα μας, μπορούμε να την παραλείψουμε παίρνοντας τον τελικό τύπο που θα χρησιμοποιηθεί στην διαδικασία:

$$r_{ij}^P = \frac{X_{ij} - \hat{\mu}_{ij}}{\sqrt{\hat{\mu}_{ij}}}, \quad i + j \leq n + 1. \quad (3.23)$$

Σύμφωνα με το Βήμα “13” της διαδικασίας Bootstrap, προκειμένου να εκτιμήσουμε τις μελλοντικές απαιτήσεις (για $i + j > n + 1$) προσομοιώνουμε μια τιμή από την κατανομή υπερδιασκορπισμένη Poisson (process distribution) με μέση τιμή $\hat{\mu}_{ij}$ και διασπορά $\hat{\phi}\hat{\mu}_{ij}$ έτσι ώστε να συμπεριλάβουμε την διασπορά της διαδικασίας (process variance) στην κατανομή πρόβλεψης (predictive distribution).

Το μέσο τετραγωνικό σφάλμα πρόβλεψης (MSEP) υπολογίζεται πλέον από την σχέση:

$$MSEP(\hat{R}) = \hat{\phi}_p \hat{R} + \frac{N}{N-p} V_{bs}(\hat{R}), \quad (3.24)$$

οπού η $V_{bs}(\hat{R})$ είναι η διασπορά της εκτίμησης και δίνεται από τον σχέση (3.20), και $\hat{\phi}_p \hat{R}$ είναι η διασπορά της διαδικασίας.

Θέτοντας σε εφαρμογή τα όλα τα παραπάνω, είμαστε σε θέση να παρουσιάσουμε αναλυτικά βήμα προς βήμα ένα παράδειγμα της μεθόδου Bootstrap σύμφωνα με το ODP μοντέλο χρησιμοποιώντας όπως και πριν τα δεδομένα του Taylor and Ashe (1983) (βλέπε Παράρτημα A-3):

1. Τοποθετούμε τις αθροιστικές απαιτήσεις μας σε ένα τρίγωνο:

Πίνακας 3.2: Αθροιστικές Απαιτήσεις C_{ij}

i/j	1	2	3	4	5	6	7	8	9	10
1	357848	1124788	1735330	2218270	2745596	3319994	3466336	3606286	3833515	3901463
2	352118	1236139	2170033	3353322	3799067	4120063	4647867	4914039	5339085	
3	290507	1292306	2218525	3235179	3985995	4132918	4628910	4909315		
4	310608	1418858	2195047	3757447	4029929	4381982	4588268			
5	443160	1136350	2128333	2897821	3402672	3873311				
6	396132	1333217	2180715	2985752	3691712					
7	440832	1288463	2419861	3483130						
8	359480	1421128	2864498							
9	376686	1363294								
10	344014									

2. Υπολογίζουμε τους συντελεστές εξέλιξης \hat{f}_j με την μέθοδο Chain-Ladder, σύμφωνα με τη σχέση (2.1) (Πίνακας 3.3):

Πίνακας 3.3: Συντελεστές εξέλιξης ζημιών

\hat{f}_j	3.490607	1.747333	1.457413	1.173852	1.103824	1.086269	1.053874	1.076555	1.017725
-------------	----------	----------	----------	----------	----------	----------	----------	----------	----------

3. Υπολογίζουμε αναδρομικά και προς τα πίσω τις προσαρμοσμένες (fitted) συσσωρευμένες απαιτήσεις \hat{C}_{ij} , σύμφωνα με τις σχέσεις (2.7) και (2.8).

Πίνακας 3.4: Προσαρμοσμένες Αθροιστικές Απαιτήσεις

i/j	1	2	3	4	5	6	7	8	9	10
1	270061	942678	1647172	2400610	2817960	3110531	3378874	3560909	3833515	3901463
2	376125	1312904	2294081	3343423	3924682	4332157	4705889	4959416	5339085	
3	372325	1299641	2270905	3309647	3885035	4288393	4658349	4909315		
4	366724	1280089	2236741	3259856	3826587	4223877	4588268			
5	336287	1173846	2051100	2989300	3508995	3873311				
6	353798	1234970	2157903	3144956	3691712					
7	391842	1367765	2389941	3483130						
8	469648	1639355	2864498							
9	390561	1363294								
10	344014									

4. Υπολογίζουμε τις προσαρμοσμένες αυξητικές απαιτήσεις \hat{X}_{ij} με τις σχέσεις (2.9) και (2.10):

Πίνακας 3.5: Προσαρμοσμένες Προσαυξητικές Απαιτήσεις

i/j	1	2	3	4	5	6	7	8	9	10
1	270061	672617	704494	753438	417350	292571	268344	182035	272606	67948
2	376125	936779	981176	1049342	581260	407474	373732	253527	379669	
3	372325	927316	971264	1038741	575388	403358	369957	250966		
4	366724	913365	956652	1023114	566731	397290	364391			
5	336287	837559	877254	938200	519695	364316				
6	353798	881172	922933	987053	546756					
7	391842	975923	1022175	1093189						
8	469648	1169707	1225143							
9	390561	972733								
10	344014									

5. Υπολογίζουμε τα κατάλοιπα Pearson σύμφωνα με τη σχέση (3.11):

Πίνακας 3.6: Κατάλοιπα Pearson

i/j	1	2	3	4	5	6	7	8	9	10
1	169	115	-112	-312	170	521	-236	-99	-87	0
2	-39	-55	-48	131	-178	-135	252	25	74	
3	-134	77	-46	-22	231	-404	207	59		
4	-93	204	-185	533	-391	-72	-262			
5	184	-158	122	-174	-21	176				
6	71	60	-79	-183	215					
7	78	-130	108	-29						
8	-161	-100	197							
9	-22	14								
10	0									

6. Υπολογίζουμε την παράμετρο κλίμακας $\hat{\phi}_P$, (3.18):

$$\hat{\phi}_P = \frac{\sum_{i+j \leq (n+1)} (r_{ij}^P)^2}{N - p} = \frac{1894178}{55 - 19} = 34420.6.$$

7. Υπολογίζουμε τα προσαρμοσμένα κατάλοιπα του Pearson (adjusted Pearson residuals) σύμφωνα με τη σχέση (3.15):

Πίνακας 3.7: Προσαρμοσμένα Κατάλοιπα Pearson

i/j	1	2	3	4	5	6	7	8	9	10
1	209	142	-138	-386	210	644	-292	-122	-108	0
2	-48	-68	-59	162	-220	-167	311	31	91	
3	-166	95	-57	-27	286	-499	256	73		
4	-115	252	-229	659	-483	-89	-324			
5	227	-195	151	-215	-26	218				
6	88	74	-98	-226	266					
7	96	-161	133	-36						
8	-199	-124	243							
9	-27	17								
10	0									

Τα επόμενα Βήματα “8-15” αφορούν μια κυκλική διαδικασία (loop process) που επαναλαμβάνεται B φορές:

8. Στα προσαρμοσμένα κατάλοιπα εφαρμόζουμε επαναδειγματοληψία με επανάληψη, δημιουργώντας ένα νέο τρίγωνο από κατάλοιπα (δεν λαμβάνουμε υπόψη τα μηδενικά κατάλοιπα):

Πίνακας 3.8: Επαναδειγματοληψία στα Προσαρμοσμένα Κατάλοιπα Pearson

i/j	1	2	3	4	5	6	7	8	9	10
1	31	142	-166	-26	252	-229	-220	-220	133	209
2	-324	-36	142	243	91	151	-59	-195	-115	
3	266	-108	-499	142	-108	311	-124	-48		
4	-27	311	-229	-229	-124	644	-166			
5	644	210	-483	-292	-27	286				
6	218	286	-483	-52	-68					
7	73	-26	-483	-108						
8	-483	133	256							
9	-161	-215								
10	286									

9. Σε κάθε κελί στο νέο τρίγωνο τοποθετούμε ένα X_{ij}^* από τη σχέση:

$$X_{ij}^* = r_{ij}^{*P} \sqrt{V(\hat{X}_{ij})} + \hat{X}_{ij}$$

Πίνακας 3.9: Ψευδο-προσαυξητικές Απαιτήσεις

i/j	1	2	3	4	5	6	7	8	9	10
1	286171	789076	565163	730870	580149	168705	154380	88171	342048	122428
2	177419	901936	1121833	1298265	650639	503863	337663	155342	308809	
3	534634	82331	479486	1183465	493465	600876	294535	226920		
4	350373	1210588	732670	791483	473382	803209	264185			
5	709744	1029747	424867	655367	500231	536942				
6	483466	1149642	458918	935391	496475					
7	437538	950238	533849	980269						
8	138644	1313550	1508500							
9	289944	760684								
10	511761									

10. Υπολογίζουμε τις νέες ψευδο-αθροιστικές απαιτήσεις, C_{ij}^* με τη σχέση $C_{ij}^* = \sum_{k=1}^j X_{ik}^*$:

Πίνακας 3.10: Ψευδο-αθροιστικές Απαιτήσεις

i/j	1	2	3	4	5	6	7	8	9	10
1	286171	1075247	1640410	2371280	2951429	3120134	3274514	3362685	3704733	3827161
2	177419	1079355	2201188	3499453	4150092	4653955	4991618	5146960	5455769	
3	534634	1357949	1837435	3020900	3514365	4115241	4409776	4636696		
4	350373	1560961	2293631	3085114	3558496	4361705	4625890			
5	709744	1739491	2164358	2819725	3319956	3856898				
6	483466	1633108	2092026	3027417	3523892					
7	437538	1387776	1921625	2901894						
8	138644	1452194	2960694							
9	289944	1050628								
10	511761									

11. Εφαρμόζουμε το κλασσικό μοντέλο Chain-Ladder στα ψευδο-δεδομένα μας και εκτιμάμε τις μελλοντικές αθροιστικές απαιτήσεις:

- Υπολογίζουμε τους νέους συντελεστές εξέλιξης από τα ψευδο-δεδομένα μας:

Πίνακας 3.11: Ψευδο-συντελεστές εξέλιξης ζημιών

\hat{f}_j	3.619998	1.516148	1.464299	1.179267	1.149397	1.064658	1.037112	1.076485	1.033046
-------------	----------	----------	----------	----------	----------	----------	----------	----------	----------

- Στην συνέχεια εκτιμάμε τις ψευδο-αθροιστικές απαιτήσεις για $i + j > 11$ (δεξιά της μπλε διαγωνίου) σύμφωνα με την σχέση (2.3):

Πίνακας 3.12: Πλήρες Τρίγωνο Εξέλιξης Ψευδο-αθροιστικών Ζημιών

i/j	1	2	3	4	5	6	7	8	9	10
1	286171	1075247	1640410	2371280	2951429	3120134	3274514	3362685	3704733	3827161
2	177419	1079355	2201188	3499453	4150092	4653955	4991618	5146960	5455769	5636062
3	534634	1357949	1837435	3020900	3514365	4115241	4409776	4636696	4991332	5156277
4	350373	1560961	2293631	3085114	3558496	4361705	4625890	4797568	5164508	5335176
5	709744	1739491	2164358	2819725	3319956	3856898	4106278	4258672	4584395	4735892
6	483466	1633108	2092026	3022449	3518924	4044639	4306159	4465970	4807548	4966420
7	437538	1387776	1921625	2901894	3422107	3933359	4187682	4343097	4675277	4829778
8	138644	1452194	2960694	4335341	5112524	5876317	6256270	6488455	6984722	7215541
9	289944	1050628	1592907	2332492	2750631	3161566	3365987	3490907	3757908	3882093
10	511761	1852574	2808775	4112887	4850191	5574793	5935249	6155520	6626323	6845299

12. Υπολογίζουμε τις προσαυξητικές απαιτήσεις \hat{X}_{ij}^* , $i + j > n + 1$, όπου έπειτα στην προσομοίωση θα χρησιμοποιηθούν ως μέση τιμή:

Πίνακας 3.13: Πλήρες Τρίγωνο Εξέλιξης Ψευδο-προσαυξητικών Ζημιών

i/j	1	2	3	4	5	6	7	8	9	10
1	286171	789076	565163	730870	580149	168705	154380	88171	342048	122428
2	177419	901936	1121833	1298265	650639	503863	337663	155342	308809	180293
3	534634	823315	479486	1183465	493465	600876	294535	226920	354636	164945
4	350373	1210588	732670	791483	473382	803209	264185	171678	366940	170668
5	709744	1029747	424867	655367	500231	536942	249380	152394	325723	151497
6	483466	1149642	458918	930423	496475	525715	261520	159811	341578	158872
7	437538	950238	533849	980269	520213	511252	254323	155415	332180	154501
8	138644	1313550	1508500	1374647	777183	763793	379953	232185	496267	230819
9	289944	760684	542279	739585	418139	410935	204421	124920	267001	124185
10	511761	1340813	956201	1304112	737304	724602	360456	220271	470803	218976

13. Σε κάθε κελί (i, j) με $i + j > n + 1$ στο μελλοντικό τρίγωνο (δεξιά της μπλε διαγωνίου) προσομοιώνουμε μια πληρωμή από την κατανομή της διαδικασίας (process distribution), υπερδιασκορπισμένη Poisson με μέση τιμή $E(\hat{X}_{ij}) = \hat{\mu}_{ij}$ και διασπορά $V(\hat{X}_{ij}) = \hat{\phi}_p \hat{\mu}_{ij}$:

Πίνακας 3.14: Πλήρες Τρίγωνο Εξέλιξης Προσομοιωμένων Προσαυξητικών Ζημιών

i/j	1	2	3	4	5	6	7	8	9	10
1	286171	789076	565163	730870	580149	168705	154380	88171	342048	122428
2	177419	901936	1121833	1298265	650639	503863	337663	155342	308809	127937
3	534634	823315	479486	1183465	493465	600876	294535	226920	345358	169970
4	350373	1210588	732670	791483	473382	803209	264185	190170	459914	116408
5	709744	1029747	424867	655367	500231	536942	239836	306700	366382	129220
6	483466	1149642	458918	930423	496475	635239	290596	130085	218504	151761
7	437538	950238	533849	980269	504036	327469	171698	191345	536541	71544
8	138644	1313550	1508500	1495574	536300	997949	480239	315581	336280	206260
9	289944	760684	617285	678951	297851	365003	112619	79391	113776	59435
10	511761	1295901	1202428	1102557	339711	821344	236940	172005	404037	219863

14. Προσθέτουμε τις προσομοιωμένες τιμές στο μελλοντικό τρίγωνο για κάθε έτος προσέλευσης, και γενικά, για να πάρουμε ένα απόθεμα Bootstrap για κάθε έτος ατυχήματος.

Πίνακας 3.15: Πλήρες Τρίγωνο Εξέλιξης Προσομοιωμένων Προσαυξητικών Ζημιών

i/j	1	2	3	4	5	6	7	8	9	10 (τελικές ζημιές)	Απόθεμα
1	286171	789076	565163	730870	580149	168705	154380	88171	342048	122428	0
2	177419	901936	1121833	1298265	650639	503863	337663	155342	308809	127937	127937
3	534634	823315	479486	1183465	493465	600876	294535	226920	345358	169970	515328
4	350373	1210588	732670	791483	473382	803209	264185	190170	459914	116408	766492
5	709744	1029747	424867	655367	500231	536942	239836	306700	366382	129220	1042138
6	483466	1149642	458918	930423	496475	635239	290596	130085	218504	151761	1426185
7	437538	950238	533849	980269	504036	327469	171698	191345	536541	71544	1802633
8	138644	1313550	1508500	1495574	536300	997949	480239	315581	336280	206260	4368183
9	289944	760684	617285	678951	297851	365003	112619	79391	113776	59435	2324311
10	511761	1295901	1202428	1102557	339711	821344	236940	172005	404037	219863	5794786

Οπότε σύμφωνα με την σχέση 2.6 το συνολικό απόθεμα είναι το άθροισμα της τελευταίας στήλης, δηλαδή 18167993.

15. Αποθηκεύουμε τα αποτελέσματα και επαναλαμβάνουμε τα Βήματα “8-15” B , συνήθως 1000 ή 10000, φορές προκειμένου να εξάγουμε την κατανομή πρόβλεψης του αποθέματος κάθε έτους αλλά και του συνολικού.

Η Bootstrap εκτίμηση του αποθέματος κάθε έτους και του συνολικού, είναι στην ουσία ο μέσος της αντίστοιχης κατανομής πρόβλεψης (predictive distribution) με την βοήθεια της οποίας υπολογίζουμε και το Bootstrap σφάλμα πρόβλεψης (Bootstrap Prediction Error). Το Bootstrap σφάλμα πρόβλεψης (PE_{bs}) δίνεται από την σχέση (βλέπε England and Verrall (1999)):

$$PE_{bs}(\hat{R}) = \sqrt{\hat{\phi}_p \hat{R} + \frac{N}{N-p} (SE_{bs}(\hat{R}))^2}, \quad (3.25)$$

όπου \hat{R} το απόθεμα του έτους ατυχήματος ή το συνολικό υπολογισμένο με την μέθοδο ChainLadder και $SE_{bs}(\hat{R})$ είναι το Bootstrap τυπικό σφάλμα της εκτίμησης του αποθέματος (Bootstrap standard error of reserve estimate):

$$SE_{bs}(\hat{R}) = \sqrt{V_{bs}(\hat{R})} = \sqrt{\frac{1}{B} \sum_{b=1}^B (\hat{R}_{(b)}^* - \hat{R})^2}, \quad (3.26)$$

όπου B είναι ο αριθμός των επαναλήψεων της διαδικασίας και $\hat{R}_{(b)}^*$ η εκτίμηση του αποθέματος για κάθε επανάληψη της διαδικασίας.

Η παραπάνω συνθέτη διαδικασία μπορεί να υλοποιηθεί μέσα σε λίγα δευτερόλεπτα με την χρήση του λογισμικού προγράμματος R , όπου ο μέσος και η τυπική απόκλιση (δηλαδή το \hat{R} και PE_{bs} αντίστοιχα) της κατανομής πρόβλεψης των αποθεμάτων υπολογίζονται με την χρήση του πακέτου ChainLadder και της συνάρτησης BootChainLadder υποθέτοντας την υπερδιασκορπισμένη (Over-Dispersed) Poisson κατανομή για τις προσαυξητικές απαιτήσεις (ο κώδικας παρουσιάζεται αναλυτικά στο παράρτημα):

Πίνακας 3.16: Αποτελέσματα Μεθόδου Bootstrap - Μοντέλο ODP

	Τιμή Κύριας Διαγωνίου (Latest Claim)	Μέση Τελική Απαίτηση (Mean Ultimate Claim)	Μέσο IBNR (Mean IBNR)	Τυπική Απόκλιση του IBNR (Standard Deviation of IBNR)	IBNR 75%	IBNR 95%
1	3901463	3901463	0	0	0	0
2	5339085	5432786.62	93701.62	114954.28	139607.75	316998.35
3	4909315	5386187.62	476872.62	220618.69	603088	860503
4	4588268	5309084.66	720816.66	265465.39	868587.50	1194269.35
5	3873311	4876195.02	1002884.02	308282.22	1183298.75	1560636
6	3691712	5126350.11	1434638.11	372641.12	1656566.25	2075143.95
7	3483130	5683007.40	2199877.40	497049.03	2516611.25	3077425.40
8	2864498	6801755.38	3937257.38	800776.17	4441046.50	5447472.40
9	1363294	5724448.03	4361154.03	1091196.28	5044105.25	6294812.65
10	344014	5038220.09	4694206.09	2008043.92	5803729.50	8337219.35
Συνολικά	34358090	53279497.93	18921407.93	2973183.28	20756465.25	24024587.75

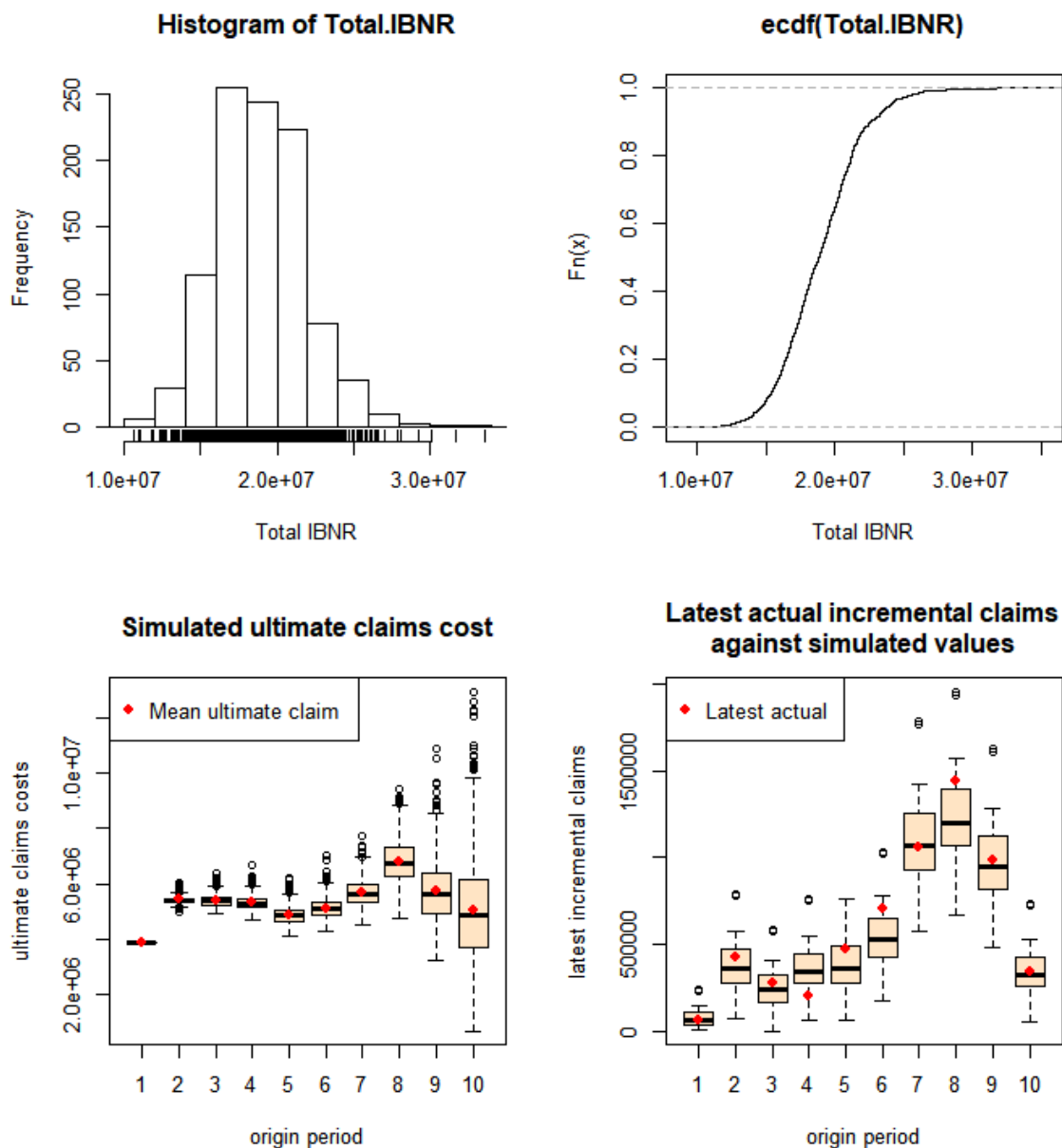
Στον παραπάνω πίνακα έχουμε:

- **Τιμή Κύριας Διαγωνίου (Latest Claim):** Είναι η τελευταία απαίτηση που πληρώθηκε από την ασφαλιστική. Δηλαδή τα στοιχεία της κύριας διαγωνίου του Πίνακα 2.4.
- **Μέση Τελική Απαίτηση (Mean Ultimate Claim):** Είναι οι μέσες τελικές απαιτήσεις για την κατανομή πρόβλεψης (predictive distribution) που εξάγαμε από την διαδικασία Bootstrap για τα έτη εξέλιξης που διαθέτουμε.
- **Μέσο IBNR (Mean IBNR):** Ο μέσος της κατανομής πρόβλεψης του αποθέματος κάθε έτους. Αποτελεί και την εκτίμηση του αποθέματος με την μέθοδο Bootstrap.
- **Τυπική Απόκλιση του IBNR (Standard Deviation of IBNR):** Είναι το Bootstrap τυπικό σφάλμα της εκτίμησης του αποθέματος (Bootstrap standard error of reserve estimate) κάθε έτους αλλά και του συνολικού που δίνεται από την σχέση (3.25).
- **IBNR 75%:** Είναι το 0,75 ποσοστιαίο σημείο της κατανομής πρόβλεψης του αποθέματος κάθε έτους και του συνολικού, δηλαδή το 75% των παρατηρήσεων είναι μικρότερες ή ίσες από αυτή την τιμή.
- **IBNR 95%:** Είναι το 0,95 ποσοστιαίο σημείο της κατανομής πρόβλεψης του αποθέματος κάθε έτους και του συνολικού, δηλαδή το 95% των παρατηρήσεων είναι μικρότερες ή ίσες από αυτή την τιμή.

Η εκτίμηση του συνολικού αποθέματος με την μέθοδο Bootstrap, κάνοντας την υπόθεση ότι οι προσαυξητικές απαιτήσεις ακολουθούν την υπερδιασκορπισμένη Poisson, είναι 18921408 (βλέπε Πίνακα 3.16-Μέσο IBNR).

Όπως σε όλες τις μεθόδους έτσι και στην μέθοδο Bootstrap είναι χρήσιμο να εξετάσουμε διάφορα γραφήματα τα οποία παρουσιάζουμε παρακάτω:

Σχήμα 3.4: Γραφήματα Μεθόδου Bootstrap - Μοντέλο ODP



Στο πρώτο αριστερό γράφημα απεικονίζεται το ιστόγραμμα από τα συνολικά προσομοιωμένα αποθέματα (IBNR's), δηλαδή όλες οι τιμές τις οποίες πήρε το συνολικό απόθεμα σε κάθε επανάληψη της μεθόδου. Στο πρώτο δεξιό γράφημα βλέπουμε την εμπειρική κατανομή (empirical distribution) του συνολικού αποθέματος.

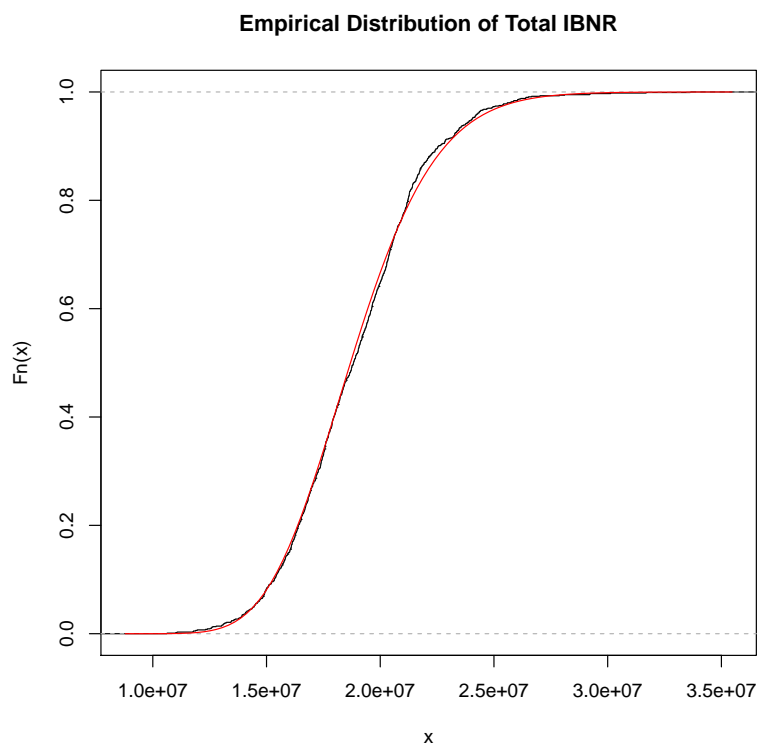
Στην συνέχεια, στη δεύτερη αριστερή εικόνα απεικονίζεται ένα Θηκόγραμμα (box plot) των τελικών προσομοιωμένων ζημιών σε κάθε έτος ατυχήματος. Το τέταρτο γράφημα παριστάνει επίσης ένα Θηκόγραμμα των προσομοιωμένων προσαυξητικών ζημιών για το τελευταίο ημερο-

λογιακό έτος που έχουμε στην διάθεση μας εναντίον των πραγματικών προσαυξητικών ζημιών για την ίδια περίοδο (δηλαδή τα στοιχεία της κύριας διαγωνίου).

Στο δεύτερο δεξιό γράφημα, απεικονίζεται η τάση που ακολουθούν τα προσομοιωμένα δεδομένα. Τα προσομοιωμένα δεδομένα πρέπει να ακολουθούν την ίδια τάση με τα πραγματικά δεδομένα, αλλιώς τα πραγματικά δεδομένα μπορεί να έχουν εγγενείς τάσεις οι οποίες δεν αντιπροσωπεύονται στο μοντέλο που επιλέξαμε.

Απομονώνοντας το πρώτο δεξιό γράφημα, της εμπειρικής κατανομής των 1000 Bootstrap αποθεμάτων, μπορούμε να προσαρμόσουμε μια κατανομή σε αυτά. Στην συγκεκριμένη περίπτωση θα προσαρμόσουμε την λογαριθμοκανονική κατανομή (κόκκινη καμπύλη), με μέσο 16.74 και τυπική απόκλιση 0.15:

Σχήμα 3.5: Εμπειρική Κατανομή των Συνολικών Bootstrap Αποθεμάτων



Όπως είναι φανερό, η κατανομή των συνολικών αποθεμάτων είναι η λογαριθμοκανονική, από την στιγμή που στο Σχήμα 3.5 η εμπειρική κατανομή των 1000 Bootstrap αποθεμάτων (μαύρη καμπύλη) και η συνάρτηση κατανομής της λογαριθμοκανονικής (κόκκινη καμπύλη) για τα ίδια δεδομένα σχεδόν ταυτίζονται (Βλέπε Παράρτημα A-3).

3.6.4. Μοντέλο Γάμμα (Gamma) για Bootstrap

Στο Υποκεφάλαιο 3.4.2 παρουσιάσαμε το μοντέλο Γάμμα (Gamma) σύμφωνα με τις έξης υποθέσεις, για $1 \leq i, j \leq n$:

$$E(X_{ij}) = \mu_{ij} = \alpha_i \beta_j,$$

$$V(X_{ij}) = \phi(E(X_{ij}))^2 = \phi \mu_{ij}^2,$$

$$\eta_{ij} = \ln(\mu_{ij}) = c + a_i + b_j,$$

όπου $a_1 = b_1 = 0$.

Οι προσαρμοσμένες (fitted) τιμές υπολογίζονται πάλι από την σχέση (3.22), ενώ για τα κατάλοιπα Pearson έχουμε:

$$r_{ij}^P = \frac{X_{ij} - \hat{\mu}_{ij}}{\sqrt{\hat{\phi} \hat{\mu}_{ij}^2}}, \quad i + j \leq n + 1. \quad (3.27)$$

Όσον αφορά το Bootstrap σφάλμα πρόβλεψης, ο υπολογισμός του είναι παρόμοιος με το μοντέλο ODP με μια προσαρμογή στη διασπορά της διαδικασίας (process variance). Πιο αναλυτικά, η διασπορά της διαδικασίας δίνεται από (βλέπε Liivika, Meelis and Raunos (2017)):

$$\sum_{i+j>n+1} \hat{\phi}_p \mu_{ij}^2 = \hat{\phi}_p \sum_{i+j>n+1} \mu_{ij}^2, \quad (3.28)$$

όπου στο συγκεκριμένο παράδειγμα το $n = 10$, άρα $i + j > 11$.

Επομένως προκύπτει ο ακόλουθος τύπος:

$$PE_{bs}(\hat{R}) = \sqrt{\hat{\phi}_p \sum_{i+j>11} \mu_{ij}^2 + \frac{N}{N-p} (SE_{bs}(\hat{R}))^2}, \quad (3.29)$$

όπου \hat{R} το απόθεμα του έτους ατυχήματος ή το συνολικό υπολογισμένο με την μέθοδο Chain-Ladder και $SE_{bs}(\hat{R})$ είναι το Bootstrap τυπικό σφάλμα της εκτίμησης του αποθέματος, (3.26).

Ακολουθώντας τα βήματα που παρουσιάσαμε αναλυτικά στο Υποκεφάλαιο 3.6.3, παίρνουμε έναν πίνακα παρόμοιο με τον Πίνακα 3.16, με την διαφορά πως τώρα προσομοιώνουμε μια τιμή για κάθε κελί στο μελλοντικό τρίγωνο (δηλαδή για $i + j > 11$) από την κατανομή Γάμμα με μέση τιμή $E(\hat{X}_{ij}) = \hat{\mu}_{ij}$ και διασπορά $V(\hat{X}_{ij}) = \hat{\phi}_p \hat{\mu}_{ij}^2$ (Βλέπε Παράρτημα Α-4):

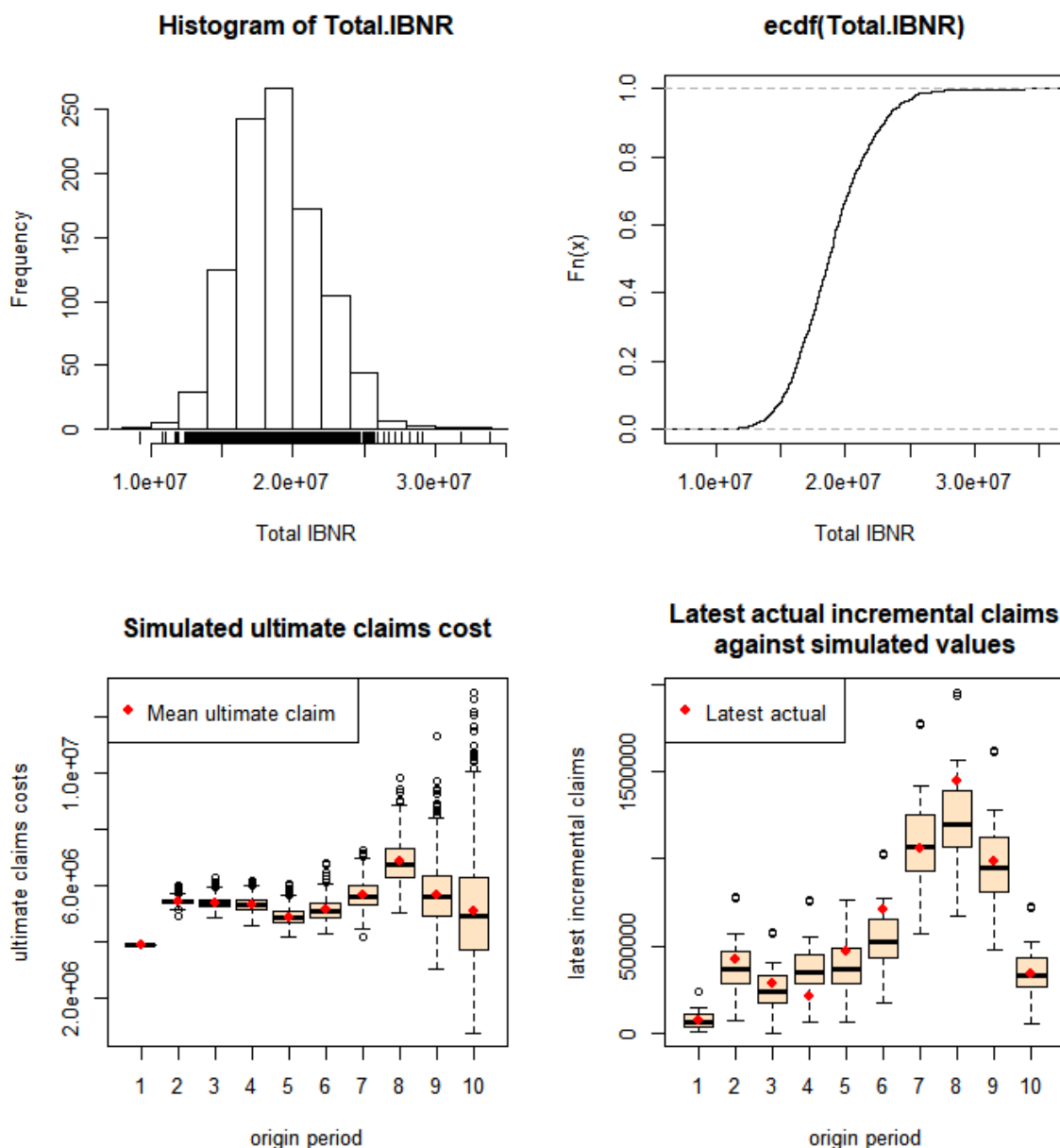
Πίνακας 3.17: Αποτελέσματα Μεθόδου Bootstrap - Μοντέλο Γάμμα

	Τιμή Κύριας Διαγωνίου (Latest Claim)	Μέση Τελική Απαίτηση (Mean Ultimate Claim)	Μέσο IBNR (Mean IBNR)	Τυπική Απόκλιση του IBNR (Standard Deviation of IBNR)	IBNR 75%	IBNR 95%
1	3901463	3901463	0	0	0	0
2	5339085	5434951.87	95866.87	119304.79	147310.62	335493.16
3	4909315	5391371.53	482056.53	223557.49	601515.35	892687.91
4	4588268	5308960.23	720692.23	264600.10	887445.64	1202731.47
5	3873311	4874221.37	1000910.37	304496.20	1188142.72	1547823.72
6	3691712	5126035.25	1434323.25	363833.02	1665254.68	2073083.44
7	3483130	5662203.32	2179073.32	505627.90	2499021.39	3081971.18
8	2864498	6845910.91	3981412.91	782888.41	4473513.67	5382026.76
9	1363294	5672951.60	4309657.60	1097987.49	4960978.98	6312332.95
10	344014	5075322.15	4731308.15	1997393.57	5947129.74	8099999.31
Συνολικά	34358090	53293391	18935301	3037469	20758974	24166153

Η εκτίμηση του συνολικού αποθέματος με την μέθοδο Bootstrap, κάνοντας την υπόθεση ότι οι προσαυξητικές απαιτήσεις ακολουθούν την κατανομή Γάμμα (Gamma), είναι 18935301 (βλέπε Πίνακα 3.16-Μέσο IBNR).

Έπειτα παρουσιάζονται και πάλι τα γραφήματα των συνολικών αποθεμάτων, των τελικών απαιτήσεων άλλα και των ψευδο-προσαυξητικών απαιτήσεων του τελευταίου ημερολογιακού έτους που διαθέτουμε (τα στοιχεία της κύριας διαγωνίου):

Σχήμα 3.6: Γραφήματα Μεθόδου Bootstrap - Μοντέλο Γάμμα



Συνοψίζοντας, παρατηρούμε μια μικρή διαφορά στον υπολογισμό των αποθεμάτων εκκρεμών ζημιών (IBNR) μεταξύ των δυο μοντέλων (13893 μονάδες). Αν είχαμε ως κριτήριο επιλογής το μέγεθος του τυπικού σφάλματος (PE_{bs}), τότε το ιδανικότερο θα ήταν το μοντέλο ODP με 2973183 έναντι 3037469 του Γάμμα. Τέλος, παρατηρούμε πως τα 0.75 και 0.95 ποσοστιαία σημεία της κατανομής των συνολικών αποθεμάτων είναι μεγαλύτερα στο μοντέλο Γάμμα. Όσον αφορά την σταδιακή αύξηση των ποσοστιαίων σημείων από έτος σε έτος, οφείλεται στο γεγονός πως πιο πρόσφατα ατυχήματα (δηλαδή όσο το i μεγαλώνει) περιέχουν μεγαλύτερη αβεβαιότητα και απαιτούνται μεγαλύτερα αποθέματα για την διευθέτησή τους.

4. Μπεϋζιανά MCMC Μοντέλα

4.1. Εισαγωγή

Η χρήση της Μπεϋζιανής συμπερασματολογίας έχει αυξηθεί τα τελευταία χρόνια, γεγονός που φαίνεται και από την δημιουργία της Διεθνούς Ένωση για Μπεϋζιανή Ανάλυση (International Society for Bayesian Analysis, ISBA). Μέχρι τότε, η κλασσική στατιστική κυριαρχούσε στα περισσότερα μοντέλα, στην πορεία όμως πρόεκυψαν αρκετοί περιορισμοί στις δυνατότητές της.

Στην Μπεϋζιανή στατιστική, οι άγνωστες ποσότητες μεταχειρίζονται σαν τυχαίες μεταβλητές. Υποθέτουμε ότι η κατανομή μιας συνεχούς τυχαίας μεταβλητής X εξαρτάται από μια παράμετρο θ , της οποίας την κατανομή θα ονομάσουμε εκ των προτέρων κατανομή (prior distribution) με συνάρτηση πυκνότητας πιθανότητας $\pi(\theta)$. Η δεσμευμένη συνάρτηση πυκνότητας της μεταβλητής X δοθέντος θ συμβολίζεται με $f_{X|\theta}(x | \theta)$ και έχουμε:

$$f_{\theta,X}(\theta, x) = f_{X|\theta}(x | \theta)\pi(\theta), \quad (4.1)$$

όπου $f_{\theta,X}(\theta, x)$ είναι η από κοινού συνάρτηση των τυχαίων μεταβλητών θ και X . Ολοκληρώνοντας ως προς θ παίρνουμε την οριακή κατανομή της X :

$$f_X(x) = \int_{\theta} f_{X|\theta}(x | \theta)\pi(\theta)d\theta. \quad (4.2)$$

Επιπλέον,

$$f_{\theta,X}(\theta, x) = \pi(\theta | X)f_X(x), \quad (4.3)$$

όπου $\pi(\theta | X)$ είναι η εκ των υστέρων κατανομή (posterior distribution) του θ και θα την χρησιμοποιήσουμε για να υπολογίσουμε μια εκτίμηση για την τυχαία μεταβλητή X . Εξισώνοντας την (4.1) και (4.3) καταλήγουμε στην:

$$\pi(\theta | X) = \frac{f_{X|\theta}(x | \theta)\pi(\theta)}{\int f_{X|\theta}(x | \theta)\pi(\theta)d\theta}. \quad (4.4)$$

Στην σχέση (4.4) η οριακή κατανομή της X δεν εξαρτάται από το θ και μπορεί να θεωρηθεί μια σταθερά κανονικοποίησης, έτσι ώστε η εκ των υστέρων κατανομή του θ να είναι καλά ορισμένη (δηλαδή η συνάρτηση πυκνότητας της να αθροίζει στην μονάδα). Για τον λόγο αυτό μπορούμε να γράψουμε:

$$\pi(\theta | X) \propto f_{X|\theta}(x | \theta)\pi(\theta). \quad (4.5)$$

Τις περισσότερες φορές δεν έχουμε καμία πληροφορία εκ των προτέρων για το θ , για τον λόγο αυτό οδηγούμαστε στην χρήση εκ των προτέρων κατανομών που αντιπροσωπεύουν αυτή την έλλειψη γνώσης και θα τις αποκαλούμε μη-πληροφοριακές (non-informative).

Αυτό που θα μας απασχολήσει περισσότερο όμως είναι η κατανομή πρόβλεψης (predictive distribution). Από την στιγμή που έχουμε παρατηρήσει το x μπορούμε να προβλέψουμε την άγνωστη ποσότητα x' , του οποίου η κατανομή εξαρτάται από το θ . Έχοντας υπολογίσει την εκ των υστέρων κατανομή του θ και υποθέτοντας ότι τα x, x' είναι ανεξάρτητα δοθέντος αυτού, θέλουμε την συνάρτηση πυκνότητας $f(x' | x)$:

$$f(x' | x) = \int f(x' | x, \theta)\pi(\theta | x)d\theta = \int f(x' | \theta)\pi(\theta | x)d\theta. \quad (4.6)$$

Το παραπάνω ολοκλήρωμα τις περισσότερες των περιπτώσεων είναι αρκετά δύσκολο στον υπολογισμό του. Για τον λόγο αυτό, χρησιμοποιούμε την μέθοδο *Μόντε Κάρλο με Μαρκοβιανές Αλυσίδες* (*Markov Chain Monte Carlo, MCMC*), προκειμένου να εξάγουμε την κατανομή πρόβλεψης του x' .

Προτού συζητήσουμε τη μέθοδο MCMC γενικά, και ειδικά στον τομέα της αποθεματοποίησης ζιμιών, είναι σημαντικό να αναφέρουμε τα βήματα στην διαδικασία της Μπεϋζιανής μοντελοποίησης (βλέπε Carlin and Louis (2000)):

1. Προσδιορισμός της κατανομής για τα δεδομένα, δοθέντων κάποιων άγνωστων παραμέτρων.
2. Προσδιορισμός των εκ των προτέρων κατανομών για τις παραμέτρους της κατανομής των δεδομένων.

3. Εξάγουμε την συνάρτηση πιθανοφάνειας των παραμέτρων, δοθέντων των δεδομένων.
4. Συνδυάζουμε την συνάρτηση πιθανοφάνειας με τις εκ των προτέρων κατανομές για να εξάγουμε την από κοινού εκ των υστέρων κατανομή των παραμέτρων, δοθέντων των δεδομένων.
5. Παίρνουμε τις παραμέτρους της εκ των υστέρων κατανομής.
6. Παίρνουμε προβλέψεις για νέες παρατηρήσεις χρησιμοποιώντας την κατανομή πρόβλεψης συνδυάζοντας την εκ των υστέρων κατανομή και τις εκ των προτέρων κατανομές.

4.2. Μπεϋζιανή Αποθεματοποίηση Ζημιών

Ας υποθέσουμε ότι τα δεδομένα μας είναι ταξινομημένα πάλι σε ένα τρίγωνο εξέλιξης ζημιών, όπως αυτό που παρουσιάσαμε στον Πινάκα 2.2. Χρησιμοποιώντας το θεώρημα του Bayes και έχοντας ως οδηγό τα βήματα της διαδικασίας που περιγράψαμε προηγουμένως, το πρόβλημα της αποθεματοποίησης μπορεί να προσεγγιστεί με τον παρακάτω τρόπο (βλέπε Piwcewicz (2008)):

- Υποθέτουμε μια κατανομή με συνάρτηση πυκνότητας πιθανότητας $f(C_{ij} | \vec{\theta})$ για όλα τα C_{ij} (για $1 \leq i, j \leq n$), όπου το $\vec{\theta}$ είναι ένα διάνυσμα παραμέτρων με εκ των προτέρων κατανομή $\pi(\theta)$.
- Υπολογίζουμε την συνάρτηση πιθανοφάνειας $L(\vec{\theta} | \mathcal{C})$:

$$L(\vec{\theta} | \mathcal{C}) = \prod_{i+j \leq n+1} f(C_{ij} | \vec{\theta}), \quad (4.7)$$

όπου $\mathcal{C} = C_{ij} : i + j \leq n + 1$.

- Έχοντας πλέον την κατανομή των δεδομένων και την εκ των προτέρων κατανομή, η εκ των υστέρων κατανομή $\pi(\vec{\theta} | \mathcal{C})$ υπολογίζεται με βάση τη σχέση (4.5):

$$\pi(\vec{\theta} | \mathcal{C}) \propto L(\vec{\theta} | \mathcal{C})\pi(\vec{\theta}). \quad (4.8)$$

- Παίρνουμε το διάνυσμα παραμέτρων $\vec{\theta} = (\theta_1, \theta_2, \dots, \theta_n)$ από την εκ των υστέρων κατανομή $\pi(\vec{\theta} | \mathcal{C})$.
- Σκοπός όμως δεν είναι η συμπερασματολογία για το $\vec{\theta}$, αλλά η εκτίμηση των $C_{ij}(i + j > n + 1)$, στο πλαίσιο της στοχαστικής αποθεματοποίησης, με την βοήθεια των μεσών της κατανομής πρόβλεψης (βλέπε de Alba (2002)), σχέση (4.6):

$$f(C_{ij} | \mathcal{C}) = \int f(C_{ij} | \vec{\theta})f(\vec{\theta} | \mathcal{C})d\vec{\theta}, \quad i + j > n + 1. \quad (4.9)$$

Όπως κάθε μέθοδος, έτσι και η Μπεϋζιανή στοχαστική αποθεματοποίηση εμπεριέχει κάποια πλεονεκτήματα αλλά και μειονεκτήματα τα οποία παρουσιάζονται στην συνέχεια:

Πλεονεκτήματα

1. Απαιτεί ένα πλήρως προσδιορισμένο στατιστικό μοντέλο, διασφαλίζοντας ότι οι υποθέσεις του είναι σαφείς.
2. Είναι ευέλικτη όσον αφορά την εφαρμογή της, δηλαδή μπορούμε να χρησιμοποιήσουμε αρκετά μοντέλα και όχι να περιοριστούμε σε ένα.
3. Επιτρέπει την σαφής μοντελοποίηση διαφόρων πηγών αβεβαιότητας.
4. Βοηθά στην εξαγωγή της κατανομής των αποτελεσμάτων.
5. Δεν παράγει ψευδο-δεδομένα όποτε δεν επηρεάζεται από θέματα που επηρεάζουν τη μέθοδο Bootstrap (π.χ. περιορισμένοι συνδυασμοί καταλοίπων, πιθανότητα αρνητικών ψευδο-δεδομένων στο τρίγωνο).
6. Εύκολη εφαρμογή με λογισμικά προγράμματα όπως είναι το WinBUGS και η R.

Μειονεκτήματα

1. Η προσέγγιση της είναι πιο σύνθετη.
2. Ο αλγόριθμος της είναι λιγότερο ανοιχτός σε παραποίηση του σε σχέση με αυτόν της μεθόδου Bootstrap.
3. Η επιλογή της εκ των προτέρων κατανομής μπορεί να δημιουργήσει προβλήματα.
4. Δεν προσελκύει κάποιον στο να μελετήσει την όλη διαδικασία, παρά μόνο τα αποτελέσματα της (black box).

4.3. Μοντέλο Poisson-Γάμμα

Το συγκεκριμένο μοντέλο χρησιμοποιείται συχνά ως παράδειγμα με σκοπό την επεξήγηση της διαδικασίας (βλέπε Carlin and Louis (2000)). Υποθέτουμε ότι τα y_i ($i = 1, \dots, n$) ακολουθούν την κατανομή Poisson με παράμετρο θ και η παράμετρος θ ακολουθεί την κατανομή Γάμμα (Gamma) με τις γνωστές παραμέτρους α, β , δηλαδή:

$$x_i | \theta \sim \text{Poisson}(\theta),$$

$$\theta | \alpha, \beta \sim \text{Gamma}(\alpha, \beta).$$

Σύμφωνα με τα Βήματα “3-5” της διαδικασίας μοντελοποίησης, έχουμε τα εξής:

- $L(\theta | \vec{x}) = \prod_{i=1}^n \frac{\theta^{x_i} e^{-\theta}}{x_i!}$, όπου \vec{x} είναι το διάνυσμα των δεδομένων x_i .
- Υπολογίζουμε την εκ των υστέρων κατανομή, βασιζόμενοι στην σχέση (4.8)

$$f(\theta | \vec{x}, \alpha, \beta) \propto \prod_{i=1}^n \frac{\theta^{x_i} e^{-\theta}}{x_i!} \frac{\beta^\alpha}{\Gamma(\alpha)} \theta^{\alpha-1} e^{-\beta\theta},$$

ή αλλιώς

$$f(\theta | \vec{x}, \alpha, \beta) \propto \theta^{\alpha + \sum_{i=1}^n x_i - 1} e^{-(\beta + n)\theta}.$$

- Η απλοποιημένη μορφή της εκ των υστέρων συνάρτησης πιθανότητας μπορεί να αναγνωρισθεί ως μια Γάμμα κατανομή με γνωστές παραμέτρους $\alpha' = \alpha + \sum_{i=1}^n x_i$ και $\beta' = \beta + n$, δηλαδή,

$$\theta | \vec{x}, \alpha, \beta \sim \text{Gamma}(\alpha + \sum_{i=1}^n x_i, \beta + n).$$

- Τέλος είμαστε πλέον σε θέση να εκτιμήσουμε την άγνωστη ποσότητα \bar{x}' με τη βοήθεια της σχέση (4.9). Η κατανομή πρόβλεψης λοιπόν ορίζεται ως:

$$f(\bar{x}' | \vec{x}) = \frac{\Gamma(\alpha' + \bar{x}')}{\Gamma(\alpha')\Gamma(\bar{x}' + 1)} \left(\frac{\beta'}{1 + \beta'}\right)^{\alpha'} \left(\frac{1}{1 + \beta'}\right)^{\bar{x}'}$$

Αυτή η κατανομή μπορεί να αναγνωρισθεί σαν μια Αρνητική Διωνυμική κατανομή με παραμέτρους α' και $\frac{1}{1 + \beta'}$.

4.4. Μέθοδος MCMC

Η μέθοδος *MCMC* (*Markov Chain Monte Carlo*) είναι ένα χρήσιμο εργαλείο για την κατασκευή της κατανομής πρόβλεψης μέσω προσομοίωσης, όταν το ολοκλήρωμα στην σχέση (4.6) είναι αδύνατον να υπολογιστεί, γεγονός που συμβαίνει αρκετά συχνά στα προβλήματα της Στατιστικής και Αναλογιστικής επιστήμης.

Ο όρος Μαρκοβιανή αλυσίδα (Markov chain) αναφέρεται σε μια στοχαστική διαδικασία, όπου η μεταφορά στην επόμενη κατάσταση εξαρτάται μόνο από την τωρινή κατάσταση, και όχι από τις προηγούμενες (βλέπε Ntzoufras (2009)). Δηλαδή, εάν έχουμε $\{\bar{\theta}^{(1)}, \bar{\theta}^{(2)}, \dots, \bar{\theta}^{(n)}\}$, τότε $\pi(\bar{\theta}^{(t+1)} | \bar{\theta}^{(t)}, \dots, \bar{\theta}^{(1)}) = \pi(\bar{\theta}^{(t+1)} | \bar{\theta}^{(t)})$ και η $\pi(\bar{\theta}^{(t+1)} | \bar{\theta}^{(t)})$ είναι ανεξάρτητη του χρόνου t . Επιπλέον, προκειμένου η κατανομή του $\bar{\theta}^{(t)}$ να συγκλίνει στην κατανομή ισορροπίας, η οποία είναι ανεξάρτητη από την αρχική τιμή $\bar{\theta}^{(0)}$, η Μαρκοβιανή αλυσίδα πρέπει να είναι αμείωτη (irreducible), αperiοδική (aperiodic) και θετικά επαναλαμβανομένη (positive-recurrent).

Για να υπολογίσουμε την εκ των υστέρων κατανομή $\pi(\vec{\theta} | x)$ πρέπει να κατασκευάσουμε μια Μαρκοβιανή αλυσίδα που να έχει ως στάσιμη κατανομή (stationary distribution) την παραπάνω. Για τον λόγο αυτό θα χρησιμοποιήσουμε τον αλγόριθμο των Metropolis-Hastings (βλέπε Wüthrich and Merz (2008)):

Αλγόριθμος Metropolis-Hastings

1. Θέτουμε μια αρχική τιμή $\vec{\theta}^{(0)}$.
2. Προτείνουμε μια καινούρια κατάσταση $\vec{\theta}'$ από την κατανομή με συνάρτηση πιθανότητας $g(\vec{\theta}' | \vec{\theta}^{(0)})$ (προτεινόμενη κατανομή).

3. Υπολογίζουμε την πιθανότητα αποδοχής:

$$\alpha(\vec{\theta}^{(0)}, \vec{\theta}') = \min \left\{ 1, \frac{\pi(\vec{\theta}' | x)g(\vec{\theta}^{(0)} | \vec{\theta}')}{\pi(\vec{\theta}^{(0)} | x)g(\vec{\theta}' | \vec{\theta}^{(0)})} \right\}.$$

4. Παίρνουμε μια τυχαία τιμή u τέτοια ώστε $u \sim U(0, 1)$.

5. Εάν $u \leq \alpha(\vec{\theta}^{(0)}, \vec{\theta}')$ τότε θέτουμε $\vec{\theta}^{(1)} = \vec{\theta}'$, αλλιώς $\vec{\theta}^{(1)} = \vec{\theta}^{(0)}$.

6. Αυξάνουμε τον δείκτη κατά μια μονάδα στην αρχική τιμή του $\vec{\theta}$ και επιστρέφουμε στο Βήμα “2”.

Μια εύλογη ερώτηση που θα μπορούσε να κάνει ο καθένας είναι “πότε η αλυσίδα συγκλίνει;”. Υπάρχουν αρκετά διαγνωστικά τα οποία δείχνουν πως υπάρχει σύγκλιση, όπως είναι το διαγνωστικό του Gelman-Rubin (βλέπε Rubin (1992)) και το διαγνωστικό του Geweke (βλέπε Geweke (1992)). Στα πλαίσια της χρήσης του WinBUGS θα ακολουθήσουμε την μέθοδο που χρησιμοποιείται στο Ntzoufras (2009):

Αλγόριθμος Ntzoufras

- Επιλεγούμε για αρχική τιμή $\vec{\theta}^{(0)}$.

- Επαναλαμβάνουμε την διαδικασία που περιγράψαμε παραπάνω μέχρι να φτάσουμε στην στάσιμη κατανομή (stationary distribution).
- Ελέγχουμε την σύγκλιση του αλγορίθμου μέσω διαφόρων διαγνωστικών.
- Απορρίπτουμε τις πρώτες k παρατηρήσεις καθώς το δείγμα μας (η Μαρκοβιανή αλυσίδα) εξαρτάται από την αρχική τιμή της παραμέτρου. Η φάση αυτή ονομάζεται burn-in και είναι ο χρόνος που απαιτείται έτσι ώστε να σταθεροποιηθεί η αλυσίδα.

Ένας άλλος πιο απλός τρόπος, ο οποίος απαιτεί κάποιο βαθμό εξοικειώσεως με τα γραφήματα, είναι ο έλεγχος των γραφημάτων χρονοσειρών (time-series plots ή trace plots). Τα γραφήματα χρονοσειρών δείχνουν όλες τις τιμές που παίρνει μια παράμετρος για κάθε επανάληψη του αλγορίθμου οι οποίες συνδέονται με μια γραμμή. Εάν διαφορετικά τμήματα αυτής της γραμμής δείχνουν να έχουν διασχίσει διαφορετικά μέρη του δειγματικού χώρου ή υπάρχει κάποιο ξεκάθαρο μοτίβο (π.χ. η αλυσίδα αυξάνεται συνεχώς) τότε δεν έχουμε σύγκλιση.

Ακόμη ένα θέμα που πρέπει να αντιμετωπίσουμε είναι το γεγονός πως υπάρχει μια συσχέτιση μεταξύ των διαφόρων τιμών του $\vec{\theta}$ (autocorrelation) καθώς εξάγονται από μια Μαρκοβιανή διαδικασία. Για τον λόγο αυτό χρησιμοποιούμε μια τιμή κάθε r επαναλήψεις (thinning) και αυξάνουμε το μέγεθος της Μαρκοβιανής αλυσίδας (βλέπε Geyer (1991)).

Στην συνέχεια, πριν αναφέρουμε την μέθοδο MCMC στην αποθεματοποίηση ζημιών, θα περιγράψουμε τον αλγόριθμο που χρησιμοποιεί το λογισμικό πρόγραμμα WinBUGS (Windows Bayesian inference Using Gibbs Sampling), ο οποίος είναι μια παραλλαγή του αλγορίθμου Metropolis-Hastings και ονομάζεται αλγόριθμος του Gibbs:

Αλγόριθμος Gibbs

- Επιλεγούμε μια αρχική τιμή $\vec{\theta}^{(0)} = (\theta_1^{(0)}, \dots, \theta_n^{(0)})$.
- Παίρνουμε μια τυχαία τιμή για το $\theta_1^{(1)}$ από την $\pi(\theta_1^{(1)} | \theta_2^{(0)}, \dots, \theta_n^{(0)}, x)$.
- Παίρνουμε μια τυχαία τιμή για το $\theta_2^{(1)}$ από την $\pi(\theta_2^{(1)} | \theta_1^{(1)}, \theta_3^{(0)}, \dots, \theta_n^{(0)}, x)$.

- Συνεχίζουμε την ίδια διαδικασία μέχρι το $\theta_i^{(1)}$.
- Παίρνουμε μια τυχαία τιμή για το $\theta_i^{(1)}$ από την $\pi(\theta_i^{(1)} | \theta_1^{(1)}, \dots, \theta_{i-1}^{(1)}, \theta_{i+1}^{(0)}, \dots, \theta_n^{(0)}, x)$.
- Συνεχίζουμε την ίδια διαδικασία μέχρι το $\theta_n^{(1)}$.
- Παίρνουμε μια τυχαία τιμή για το $\theta_n^{(1)}$ από την $\pi(\theta_n^{(1)} | \theta_1^{(1)}, \dots, \theta_{n-1}^{(1)}, x)$.
- Αυξάνουμε τον δείκτη κατά μια μονάδα στην αρχική τιμή του $\vec{\theta}$ και επαναλαμβάνουμε τον αλγόριθμο.

Ο αλγόριθμος του Gibbs χρησιμοποιείται για να προσομοιώνουμε τιμές από την εκ των προτέρων κατανομή $\pi(\vec{\theta})$ όταν οι υπόλοιπες μέθοδοι δεν λειτουργούν καλά, γεγονός που οφείλεται κυρίως στην μεγάλη διάσταση του $\vec{\theta}$ (όταν έχουμε δηλαδή πολλές παραμέτρους) και βασίζεται στις πλήρως δεσμευμένες κατανομές (οι κατανομές που χρησιμοποιήθηκαν στον αλγόριθμο).

4.5. Λογαριθμοκανονικό Μοντέλο για MCMC

Το πρώτο μοντέλο που θα παρουσιάσουμε και θα αναλύσουμε εις βάθος είναι το Λογαριθμοκανονικό μοντέλο (βλέπε Verrall (1990)), όπως παρουσιάστηκε στο Υποκεφάλαιο 3.4.3. Οι προσαυξητικές απαιτήσεις X_{ij} , για $1 \leq i, j \leq n$, ακολουθούν την Λογαριθμοκανονική κατανομή με παραμέτρους μ_{ij} και σ^2 (βλέπε Ntzoufras and Dellaportas (2002)):

$$Z_{ij} = \ln(X_{ij}) | \mu_{ij}, \sigma^2 \sim N(\mu_{ij}, \sigma^2),$$

$$\mu_{ij} = c + a_i + b_j,$$

αφού $\eta_{ij} = \mu_{ij}$ και ο γραμμικός εκτιμητής η_{ij} δίνεται από τη σχέση (3.6),

$$c \sim N(0, 1000),$$

$$a_i, b_j \sim N(0, 100),$$

$$\tau = \frac{1}{\sigma^2} \sim G(0.01, 0.01),$$

όπου με $G(a, b)$ συμβολίζουμε την κατανομή Γάμμα με συνάρτηση πυκνότητας πιθανότητας

$$f(x) = \frac{b^a x^{a-1} e^{-bx}}{\Gamma(a)}.$$

Η ανάλυση μας είναι παρόμοια με μια ανάλυση διασποράς (ANOVA) δύο παραμέτρων με τους παραμέτρους a_i και b_j να συμβολίζουν την επίδραση του έτους ατυχήματος i και έτος εξέλιξης j αντίστοιχα. Η παράμετρος μ_{ij} είναι ο μέσος των λογαρίθμων των προσαυξητικών απαιτήσεων και εκφράζεται ως το άθροισμα των c, a_i και b_j , γεγονός που αντικατοπτρίζει τις διαφορές στις αναμενόμενες απαιτήσεις για κάθε έτος ατυχήματος i και έτος εξέλιξης j . Οι μικρές τιμές που χρησιμοποιούνται στην Γάμμα κατανομή και οι τεράστιες διακυμάνσεις στην κανονική κατανομή έχουν σκοπό να παράγουν μη-πληροφοριακές (non-informative) εκ των προτέρων κατανομές. Με λίγα λόγια αυτές οι παράμετροι αντικατοπτρίζουν την αβεβαιότητα μας. Τέλος θα χρησιμοποιούμε τους περιορισμούς $\sum_{i=1}^n a_i = \sum_{j=1}^n b_j = 0$ για να αποφύγουμε τις παραπάνω παραμέτρους από όσες χρειαζόμαστε (over-parametrization).

Συνοψίζοντας κάτω από αυτές τις υποθέσεις και τους περιορισμούς, πρέπει να εκτιμήσουμε τις παραμέτρους $c, a_2, a_3, a_4, a_5, a_6, a_7, a_8, a_9, a_{10}, b_2, b_3, b_4, b_5, b_6, b_7, b_8, b_9, b_{10}$ και σ^2 . Τα a_1 και b_1 δεν χρειάζεται να τα εκτιμήσουμε αφού $a_1 = -\sum_{i=2}^{10} a_i$ και $b_1 = -\sum_{j=2}^{10} b_j$.

4.5.1. Υπολογισμός των εκ των προτέρων κατανομών

Ας υποθέσουμε ότι ο λογάριθμος των προσαυξητικών απαιτήσεων συμβολίζεται με $Z_{ij} = \ln(X_{ij})$ και ως συνέπεια ακολουθούν την κανονική κατανομή με μέσο μ_{ij} και διασπορά σ^2 . Επιπλέον θεωρούμε τις παραμέτρους $\vec{a}, \vec{b}, \sigma^2$ και c , όπου $\vec{a} = (a_1, a_2, \dots, a_{10})$ και $\vec{b} = (b_1, b_2, \dots, b_{10})$, και χωρίζουμε τις προσαυξητικές απαιτήσεις σε λογαριθμική κλίμακα (δηλαδή τα Z_{ij}) σε δύο μέρη, τις γνωστές απαιτήσεις Z και τις άγνωστες προς εκτίμηση απαιτήσεις Z^* .

Με την βοήθεια του θεωρήματος του *Bayes* και έχοντας ως αναφορά τους Ntzoufras and Dellaportas (2002) μπορούμε να υπολογίσουμε τις εκ των προτέρων κατανομές των παραπάνω παραμέτρων. Συμβολίζοντας πλέον με f τις εκ των προτέρων, τις δεσμευμένες και τις περιθώριες συναρτήσεις πυκνότητας πιθανότητας, έχουμε:

$$\begin{aligned} f(c, \vec{a}, \vec{b}, \sigma^2, Z^* | Z) &\propto f(Z | c, \vec{a}, \vec{b}, \sigma^2, Z^*) f(c, \vec{a}, \vec{b}, \sigma^2, Z^*) \\ &\propto f(Z | c, \vec{a}, \vec{b}, \sigma^2) f(c) f(\vec{a}) f(\vec{b}) f(\sigma^2) f(Z^* | c, \vec{a}, \vec{b}, \sigma^2), \end{aligned} \quad (4.10)$$

με την υπόθεση ότι εκ των προτέρων οι παράμετροι είναι ανεξάρτητες. Οι άγνωστες απαιτήσεις μπορούν να εκτιμηθούν από την $f(Z^* | c, \vec{a}, \vec{b}, \sigma^2)$.

Οι δεσμευμένες συναρτήσεις για κάθε μια παράμετρο υπολογίζονται από τις σχέσεις:

$$f(c | \vec{a}, \vec{b}, \sigma^2, Z) \propto f(Z | \vec{a}, \vec{b}, \sigma^2) f(c), \quad (4.11)$$

$$f(\vec{a} | c, \vec{b}, \sigma^2, Z) \propto f(Z | c, \vec{b}, \sigma^2) f(\vec{a}), \quad (4.12)$$

$$f(\vec{b} | c, \vec{a}, \sigma^2, Z) \propto f(Z | c, \vec{a}, \sigma^2) f(\vec{b}), \quad (4.13)$$

$$f(\sigma^2 | c, \vec{a}, \vec{b}, Z) \propto f(Z | c, \vec{a}, \vec{b}) f(\sigma^2). \quad (4.14)$$

Για παράδειγμα θα υπολογίσουμε την $f(c | \vec{a}, \vec{b}, \sigma^2, Z)$. Αρχικά υπολογίζουμε την δεσμευμένη κατανομή της παραμέτρου Z :

$$f(Z | c, \vec{a}, \vec{b}, \sigma^2) = (2\pi\sigma^2)^{\frac{|U|}{2}} \exp\left\{-\frac{1}{2\sigma^2} \left(\sum_{i=1}^n \sum_{j=1}^{n-i+1} (Z_{ij} - c - a_i - b_j)^2\right)\right\}, \quad (4.15)$$

όπου $|U|$ είναι το πλήθος των γνωστών απαιτήσεων (δηλαδή το πάνω τρίγωνο του πίνακα) και ισούται με $\frac{n(n+1)}{2}$. Συνδυάζοντας τώρα τις παραπάνω εξισώσεις με το γεγονός ότι η παράμετρος c ακολουθεί εκ των προτέρων την κανονική κατανομή με μέση τιμή 0 και διασπορά 1000 (για λόγους γενικότητας θα την συμβολίσουμε με σ_c^2) παίρνουμε την δεσμευμένη κατανομή:

$$\begin{aligned}
f(c \mid \vec{a}, \vec{b}, \sigma^2, Z) &\propto \exp\left\{-\frac{c}{2\sigma_c^2} - \frac{1}{2\sigma^2} \left(\sum_{i=1}^n \sum_{j=1}^{n-i+1} (Z_{ij} - c - a_i - b_j)^2\right)\right\} \\
&= \exp\left\{-\frac{1}{2} \left(\frac{c^2}{\sigma_c^2} + \frac{1}{\sigma^2} \sum_{i=1}^n \sum_{j=1}^{n-i+1} (c - (Z_{ij} - a_i - b_j))^2\right)\right\} \\
&= \exp\left\{-\frac{1}{2} \left(\frac{c^2}{\sigma_c^2} + \frac{c^2|U|}{\sigma^2} - 2\frac{c}{\sigma^2} \sum_{i=1}^n \sum_{j=1}^{n-i+1} (Z_{ij} - a_i - b_j)\right)\right\} \quad (4.16) \\
&= \exp\left\{-\frac{1}{2} \left(\frac{1}{\sigma_c^2} + \frac{|U|}{\sigma^2}\right) \left(c^2 - 2c \frac{\sum_{i=1}^n \sum_{j=1}^{n-i+1} (Z_{ij} - a_i - b_j)}{\frac{\sigma^2}{\sigma_c^2 + |U|}}\right)\right\} \\
&= \exp\left\{-\frac{1}{2\sigma'^2} (c - c')^2\right\},
\end{aligned}$$

όπου

$$\sigma'^2 = \frac{\sigma^2}{\frac{\sigma_c^2}{\sigma^2} + |U|}$$

και

$$c' = \frac{\sum_{i=1}^n \sum_{j=1}^{n-i+1} Z_{ij} - a_i - b_j}{\frac{\sigma_c^2}{\sigma^2} + |U|}.$$

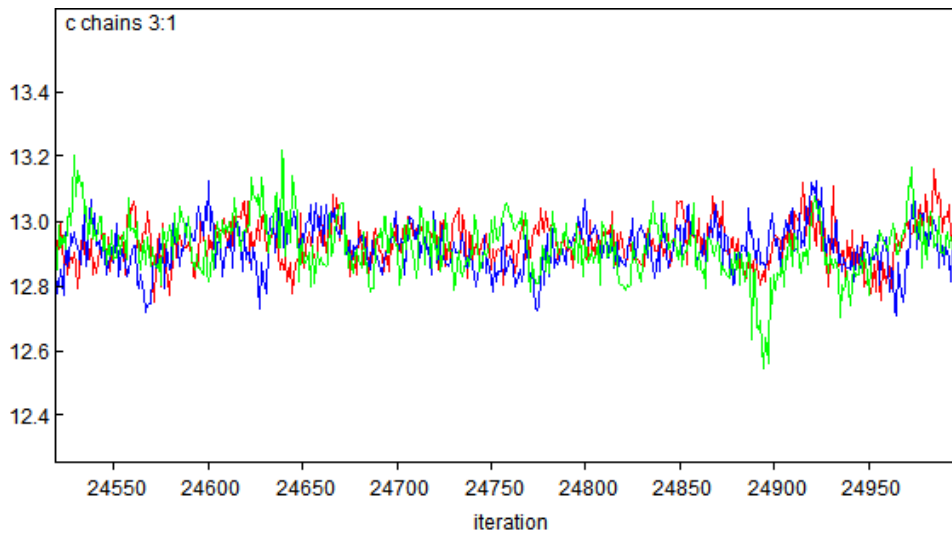
Άρα, η εκ των υστέρων κατανομή του c είναι η κανονική με νέες παραμέτρους c' και σ'^2 , δηλαδή:

$$c \mid \vec{a}, \vec{b}, \sigma^2, Z \sim N(c', \sigma'^2).$$

4.5.2. Υπολογισμός Αποθέματος IBNR

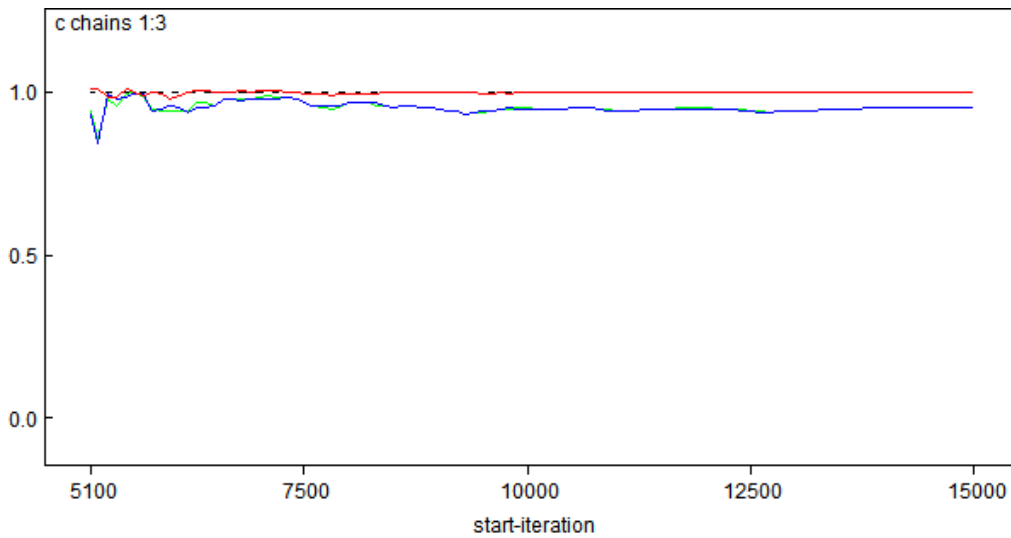
Όπως αναφέραμε και στο Υποκεφάλαιο 4.4, ένα σημαντικό ερώτημα είναι το “πότε η αλυσίδα συγκλίνει;”. Με την χρήση του λογισμικού προγράμματος WinBUGS μπορεί κανείς να ελέγξει την σύγκλιση κοιτώντας τα ίχνη (trace) των παραμέτρων. Για τον σκοπό αυτό, θα παράγουμε 3 αλυσίδες με διαφορετικές αρχικές τιμές, θα έχουμε μια περίοδο burn-in 5000 από τις 25000 συνολικές επαναλήψεις του αλγορίθμου και $thin = 1$. Για παράδειγμα, στο παρακάτω γράφημα απεικονίζεται το ίχνος της παραμέτρου c (Βλέπε Παράρτημα Β-1):

Σχήμα 4.1: Ίχνος της παραμέτρου c



Από το γράφημα φαίνεται πως το ίχνος της παραμέτρου c δεν ακολουθεί κάποιο συγκεκριμένο μοτίβο, οπότε μπορούμε να πούμε πως οι αλυσίδες συγκλίνουν. Ένας άλλος τρόπος να ελέγξουμε την σύγκλιση μέσω του WinBUGS και της επιλογής `bgr diag` είναι το διαγνωστικό Brooks-Gelman-Rubin. Το διαγνωστικό Brooks-Gelman-Rubin βασίζεται στον λόγο της συνολικής (pooled) εκ των υστέρων εκτίμηση της διασποράς και του μέσου των διασπορών σε κάθε δείγμα (βλέπε Brooks and Gelman (1998)). Όταν ο λόγος R είναι περίπου ίσος με 1 τότε έχουμε σύγκλιση. Στο παρακάτω δείγμα βλέπουμε το διαγνωστικό Brooks-Gelman-Rubin για την παράμετρο c . Η συνολική εκ των υστέρων εκτίμηση της διασποράς απεικονίζεται με πράσινο, ο μέσος των διασπορών κάθε δείγματος με μπλε, και ο λόγος R με κόκκινο. Παρατηρούμε ότι ύστερα από 8000 επαναλήψεις ο λόγος σταθεροποιείται και είναι περίπου ίσος με την μονάδα:

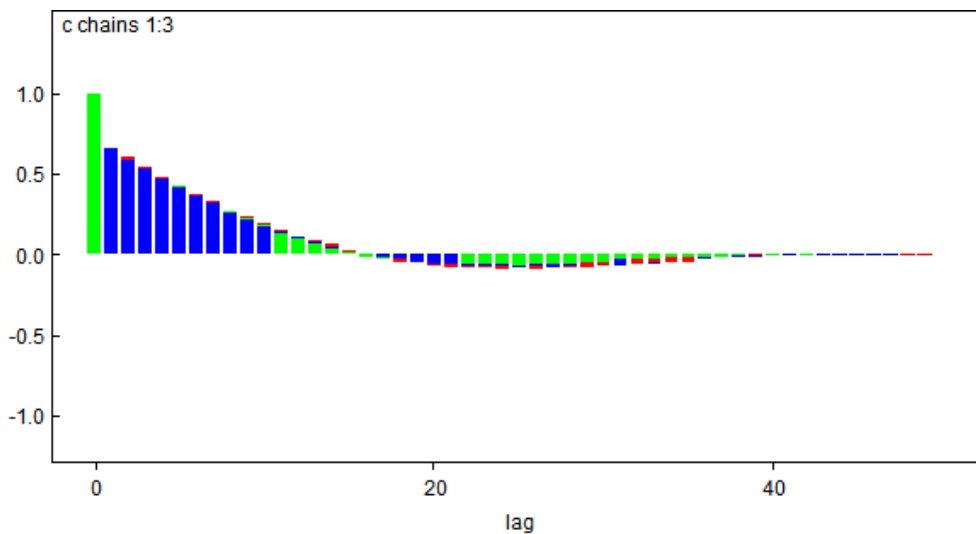
Σχήμα 4.2: Διαγνωστικό Brooks-Gelman-Rubin της παραμέτρου c



Ο κώδικας που χρησιμοποιήθηκε όσο και οι αρχικές τιμές (για την δεύτερη και τρίτη αλυσίδα χρησιμοποιήθηκε η επιλογή "gen inits") αναφέρονται αναλυτικά στο Παράρτημα.

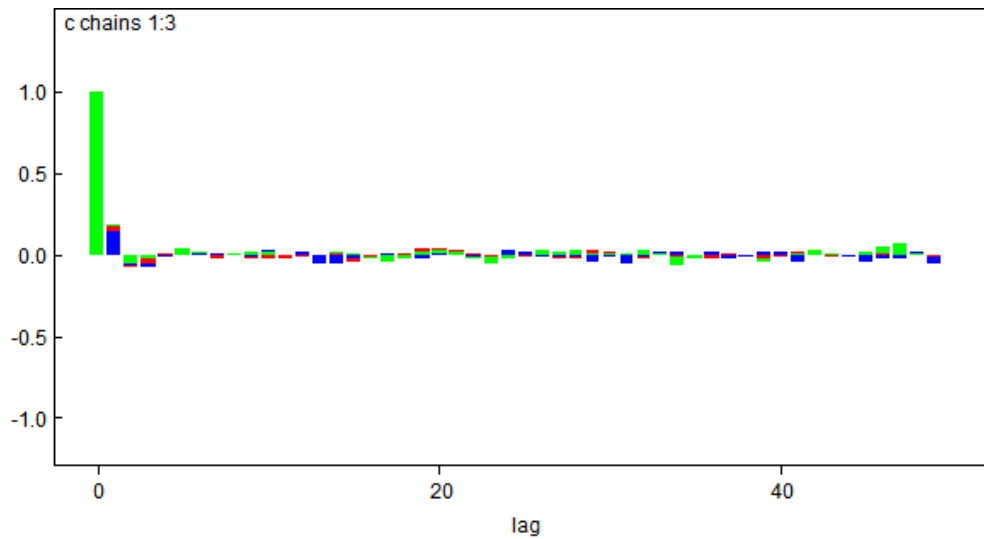
Ένα ακόμη πρόβλημα που αναφέραμε στο Υποκεφάλαιο 4.4 είναι πως ο αλγόριθμος δεν παράγει ένα ανεξάρτητο δείγμα, δηλαδή υπάρχει περίπτωση οι παρατηρήσεις να είναι συσχετισμένες μεταξύ τους. Στο Γράφημα 4.3 παρουσιάζεται η αυτοσυσχέτιση μεταξύ των τιμών της παραμέτρου c , από το οποία είναι φανερό πως υπάρχει συσχέτιση μεταξύ των αρχικών τιμών της:

Σχήμα 4.3: Αυτοσυσχέτιση της παραμέτρου c με $thin = 1$



Προκειμένου να αντιμετωπίσουμε το συγκεκριμένο πρόβλημα μπορούμε να θέσουμε στο WinBUGS το $thin = 10$, δηλαδή να λαμβάνουμε υπόψη μια τιμή κάθε $r = 10$ επαναλήψεις, κάνοντας τις ίδιες υποθέσεις με προηγουμένως όσον αφορά τις συνολικές επαναλήψεις, το burn-in και τις αρχικές τιμές. Στο Γράφημα 4.4 φαίνεται ξεκάθαρα ότι το πρόβλημα της ανεξαρτησίας του δείγματος έχει εξαλειφθεί σε πολύ μεγάλο βαθμό:

Σχήμα 4.4: Αυτοσυσχέτιση της παραμέτρου c με $thin = 10$



Σκοπός του μοντέλου είναι εκτίμηση των παραμέτρων \vec{a} , \vec{b} , c και τ . Έτσι λοιπόν παίρνουμε τον Πίνακα 4.1:

Πίνακας 4.1: Εκτίμηση Παραμέτρων στο Λογαριθμοκανονικό Μοντέλο

	Μέσος (Mean)	Τυπική Απόκλιση (Standard Deviation)	Σφάλμα Monte Carlo (Monte Carlo Error)	2.5%	Διάμεσος (Median)	97.5%
a_1	-0.2861	0.1285	0.0008601	-0.5407	-0.2859	-0.03475
a_2	0.07469	0.1281	0.001078	-0.1775	0.07476	0.329
a_3	-0.002707	0.1324	0.001105	-0.2648	-0.002192	0.2586
a_4	-0.1144	0.1371	0.001105	-0.3837	-0.1146	0.1562
a_5	-0.003014	0.1462	0.001142	-0.2929	-0.003614	0.2867
a_6	0.02635	0.1569	0.001229	-0.2816	0.02618	0.3359
a_7	0.1071	0.1714	0.001426	-0.2291	0.1071	0.4463
a_8	0.195	0.1976	0.001713	-0.1979	0.1953	0.5812
a_9	0.05575	0.2368	0.002344	-0.4103	0.05478	0.5231
a_{10}	-0.05273	0.3348	0.004576	-0.712	-0.0529	0.6036
b_1	-0.1195	0.1288	0.0008858	-0.374	-0.1198	0.1336
b_2	0.7912	0.1283	0.001114	0.5386	0.7914	1.043
b_3	0.8178	0.1317	0.001025	0.5584	0.8183	1.078
b_4	0.8458	0.1365	0.001111	0.5758	0.8456	1.115
b_5	0.2628	0.1438	0.001106	-0.0202	0.2625	0.5464
b_6	-0.1249	0.1549	0.0012	-0.4288	-0.1248	0.1804
b_7	0.2356	0.171	0.001337	-0.5718	-0.2354	0.1022
b_8	-0.5588	0.1955	0.001783	-0.9386	-0.5592	-0.1699
b_9	-0.1735	0.2349	0.00246	-0.6349	-0.1743	0.2938
b_{10}	-1.505	0.3302	0.004385	-2.151	-1.507	-0.8479
c	12.92	0.08125	0.0009001	12.76	12.93	13.08
τ	8.569	2.01	0.01548	5.112	8.405	12.97

Στον παραπάνω Πίνακα 4.1 έχουμε:

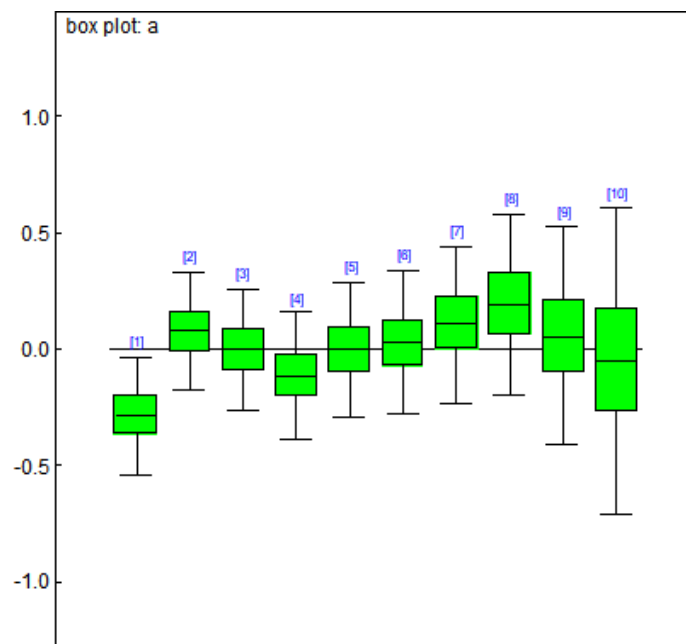
- **Μέσος (Mean):** Είναι η μέση τιμή (εκτίμηση) της αντίστοιχης παραμέτρου.
- **Τυπική Απόκλιση (Standard Deviation):** Η τυπική απόκλιση της εκτίμησης της παραμέτρου.
- **Σφάλμα Monte Carlo (Monte Carlo Error):** Το σφάλμα από την διαδικασία Monte Carlo.

- **2.5%:** Το 0.025 ποσοστιαίο σημείο της κατανομής πρόβλεψης της παραμέτρου.
- **Διάμεσος (Median):** Η διάμεσος της κατανομής πρόβλεψης της παραμέτρου.
- **97.5%:** Το 0.975 ποσοστιαίο σημείο της κατανομής πρόβλεψης της παραμέτρου.

Οι εκτιμήσεις των a_i και b_j δηλώνουν την επίδραση του έτους ατυχήματος i και έτους εξέλιξης j αντίστοιχα. Η εκτίμηση της παραμέτρου c άφορα τον περιορισμό των a_1 και b_1 .

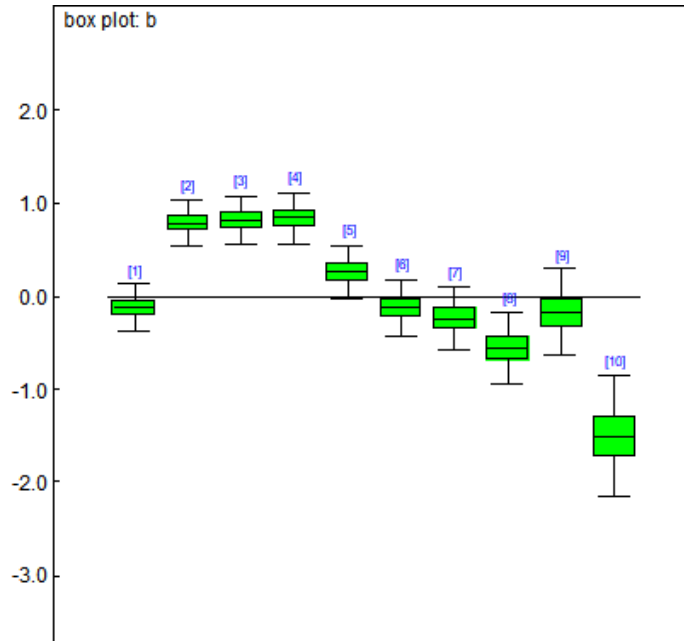
Στο Σχήμα 4.5 παριστάνεται η επίδραση του έτους ατυχήματος. Όπως είναι φανερό δεν υπάρχει κάποιο ξεκάθαρο μοτίβο, μια σταθερή μείωση της παραμέτρου a μπορεί να είναι είτε η μείωση των συμβολαίων, ή απλά ένας χρόνος, όπου το ποσό των απαιτήσεων ελαττώθηκε. Όσον άφορα την σταθερή αύξηση στην εκ των υστέρων διασπορά της παραμέτρου, αυτό οφείλεται στο ότι όσο περνάνε τα χρόνια έχουμε λιγότερα δεδομένα για την εκτίμηση της (για παράδειγμα για το έτος ατυχήματος 2 έχουμε 9 τιμές, ενώ για το έτος ατυχήματος 10 έχουμε μόνο 1 γνωστό ποσό). Αντιπροσωπεύει δηλαδή την αβεβαιότητα για την εκτίμηση των μελλοντικών πληρωμών για τα πρόσφατα έτη ατυχήματος:

Σχήμα 4.5: Επίδραση του έτους ατυχήματος στις απαιτήσεις



Η επίδραση του έτους εξέλιξης παρουσιάζεται στο Σχήμα 4.6. Όπως μπορεί κανείς να δει, εκτός από το πρώτο έτος εξέλιξης όπου οι απαιτήσεις πληρώνονται εκείνη την χρονιά, έχουμε μια σταθερή μείωση της παραμέτρου b . Αυτό οφείλεται στο ότι όσο περνάνε τα χρόνια η ασφαλιστική προσδοκία να πληρώσει λιγότερα για κάθε έτος εξέλιξης. Όπως και προηγουμένως, παρατηρείται μια σταθερή αύξηση στην διασπορά του b για τους ίδιους λόγους που αναφέρθηκαν παραπάνω:

Σχήμα 4.6: Επίδραση του έτους εξέλιξης στις απαιτήσεις



Έχοντας πλέον εκτιμήσει όλες τις παραμέτρους του μοντέλου, είμαστε σε θέση να εκτιμήσουμε και το απόθεμα κάθε έτους αλλά και το συνολικό. Στον Πίνακα 4.2 παρουσιάζονται αναλυτικά τα παραπάνω. Η ερμηνεία κάθε στήλης είναι η ίδια με τον Πίνακα 4.1:

Πίνακας 4.2: Εκτίμηση Αποθεμάτων στο Λογαριθμοκανονικό Μοντέλο

	Μέσος (Mean)	Τυπική Απόκλιση (Standard Deviation)	Σφάλμα Monte Carlo (Monte Carlo Error)	2.5%	Διάμεσος (Median)	97.5%
R_2	112500	63940	610.9	35280	98200	272700
R_3	486700	202300	1515	213300	448700	981700
R_4	665700	221400	1622	343100	629000	1196000
R_5	1099000	323300	2255	611600	1052000	1867000
R_6	1542000	423500	3021	887100	1483000	2532000
R_7	2331000	638700	4674	1344000	2245000	3833000
R_8	3835000	1115000	8369	2123000	3677000	6476000
R_9	4465000	144000	12880	232000	4249000	7919000
R_{10}	5133000	2179000	27270	2151000	4725000	10470000
<i>Total</i>	19670000	341000	37620	14110000	19270000	27460000

Άρα, η μέση τιμή (εκτίμηση) του συνολικού αποθέματος με βάση το Λογαριθμοκανονικό μοντέλο είναι 19670000.

4.6. Μοντέλο Over-Dispersed Poisson για MCMC

Όπως και προηγουμένως με τη μέθοδο Bootstrap έτσι και τώρα θα παρουσιάσουμε το μοντέλο Over-Dispersed Poisson με την μέθοδο MCMC. Οι υποθέσεις που χρησιμοποιούμε είναι παρόμοιες με αυτές που αναφέρθηκαν στα γενικευμένα γραμμικά μοντέλα (GLM's) στο Υποκεφάλαιο 3.4.1 . Πιο αναλυτικά, για $1 \leq i, j \leq 10$ έχουμε:

$$X_{ij} \sim N(\mu_{ij}, \phi\mu_{ij}),$$

$$\eta_{ij} = \ln(\mu_{ij}) = c + a_i + b_j,$$

$$c \sim N(0, 1000),$$

$$a_i, b_j \sim N(0, 100),$$

$$\phi^{-1} \sim \text{Gamma}(1, 0.01).$$

Όπως αναφέρθηκε και προηγουμένως, οι άγνωστοι παράμετροι c, a_i, b_j, ϕ ακολουθούν μια μη-πληροφοριακή εκ των προτέρων κατανομή. Για τον σκοπό μας θα παράγουμε 3 αλυσίδες με διαφορετικές αρχικές τιμές (και για τις 3 αλυσίδες χρησιμοποιήθηκε η επιλογή “gen inits”), 120000 επαναλήψεων η κάθε μία. Τέλος επιλεγούμε μια περίοδο burn-in 20000 επαναλήψεων, $thin = 25$ και εξάγουμε τον παρακάτω πίνακα:

Πίνακας 4.3: Εκτίμηση Παραμέτρων στο Μοντέλο Over-Dispersed Poisson για MCMC

	Μέσος (Mean)	Τυπική Απόκλιση (Standard Deviation)	Σφάλμα Monte Carlo (Monte Carlo Error)	2.5%	Διάμεσος (Median)	97.5%
a_2	0.3029	0.1515	0.003133	0.0004131	0.3023	0.6062
a_3	0.3074	0.1549	0.003171	0.004851	0.3072	0.6162
a_4	0.3193	0.1556	0.003229	0.01518	0.3197	0.6243
a_5	0.1895	0.1653	0.003352	-0.1388	0.1896	0.5128
a_6	0.2469	0.1683	0.003408	-0.09016	0.2475	0.5843
a_7	0.3482	0.1715	0.003342	0.007828	0.3463	0.6931
a_8	0.5511	0.1838	0.003637	0.1864	0.5518	0.915
a_9	0.344	0.2281	0.003884	-0.1083	0.3427	0.7893
a_{10}	0.1984	0.4138	0.006056	-0.6131	0.195	1.008
b_2	0.9337	0.1465	0.00326	0.6509	0.9321	1.227
b_3	0.9827	0.1509	0.003296	0.6837	0.982	1.286
b_4	1.074	0.1539	0.003358	0.7748	1.072	1.384
b_5	0.4778	0.1794	0.003623	0.1196	0.4766	0.8332
b_6	0.174	0.2042	0.00396	-0.2294	0.172	0.5752
b_7	0.04413	0.2299	0.003809	-0.412	0.04647	0.489
b_8	-0.445	0.3025	0.004272	-1.036	-0.4452	0.152
b_9	-0.01085	0.3104	0.004588	-0.6133	-0.0114	0.599
b_{10}	-1.644	0.7725	0.008832	-3.075	-1.662	-0.09413
c	12.48	0.1687	0.004991	12.14	12.48	12.81
ϕ^{-1}	0.0000201	0.000004764	0.00000004755	0.0000119	0.0000197	0.00003061

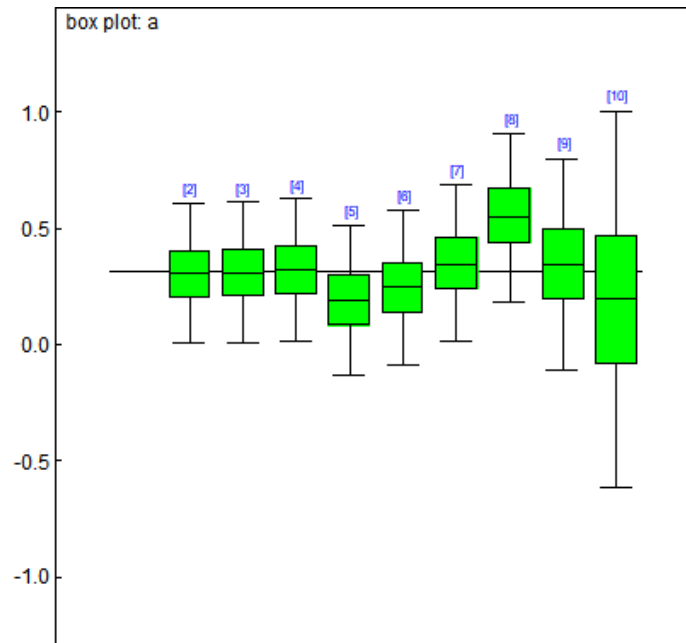
Η ερμηνεία κάθε στήλης είναι ίδια όπως ακριβώς και στον Πίνακα 4.1 στο Λογαριθμοκανονικό μοντέλο.

Οι εκτιμήσεις των a_i και b_j δηλώνουν την επίδραση του έτους ατυχήματος i και έτους εξέλιξης j αντίστοιχα. Η εκτίμηση της παραμέτρου c αφορά τον περιορισμό των a_1 και b_1 και η παράμετρος ϕ είναι η παράμετρος κλίμακας.

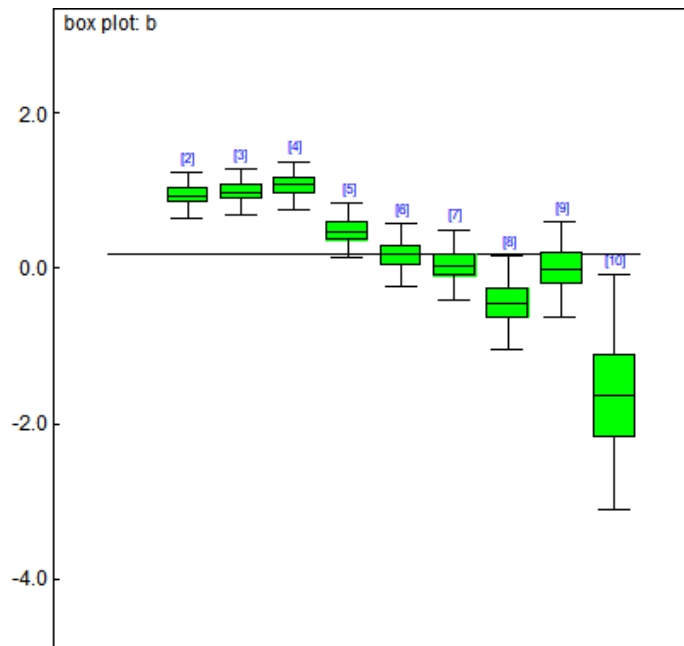
Σε αντιστοιχία με το Λογαριθμοκανονικό μοντέλο, παρουσιάζουμε στην συνέχεια το θηκόγραμμα της επίδρασης του έτους ατυχήματος a_i (Σχήμα 4.7) και το θηκόγραμμα της επίδρα-

σης του έτους εξέλιξης b_j (Σχήμα 4.8) στις απαιτήσεις. Στο πρώτο σχήμα είναι φανερό πως δεν υπάρχει κάποιο ξεκάθαρο μοτίβο, σε αντίθεση με το δεύτερο, όπου παρατηρείται μια σταθερή μείωση στην παράμετρο b_j για τον λόγο ότι η ασφαλιστική εταιρία προσδοκά να πληρώσει λιγότερα χρήματα τα επόμενα έτη εξέλιξης. Τέλος, και στα δυο σχήματα μπορεί κανείς να δει μια σταθερή αύξηση στην διασπορά των παραμέτρων, η οποία αντιπροσωπεύει την αβεβαιότητα για την εκτίμηση των μελλοντικών πληρωμών για τα πρόσφατα έτη ατυχήματος:

Σχήμα 4.7: Επίδραση του έτους ατυχήματος στις απαιτήσεις



Σχήμα 4.8: Επίδραση του έτους εξέλιξης στις απαιτήσεις



Έχοντας πλέον εκτιμήσει όλες τις παραμέτρους του μοντέλου, είμαστε σε θέση να εκτιμήσουμε και το απόθεμα κάθε έτους αλλά και το συνολικό. Στον Πίνακα 4.4 παρουσιάζονται αναλυτικά τα παραπάνω. Η ερμηνεία κάθε στήλης είναι η ίδια με τον Πίνακα 4.3:

Πίνακας 4.4: Εκτίμηση Αποθεμάτων στο Μοντέλο Over-Dispersed Poisson

	Μέσος (Mean)	Τυπική Απόκλιση (Standard Deviation)	Σφάλμα Monte Carlo (Monte Carlo Error)	2.5%	Διάμεσος (Median)	97.5%
R_2	91720	105100	1021	-43580	68940	363500
R_3	463100	214000	2172	108300	440900	940100
R_4	711700	261400	2441	263100	690700	1279000
R_5	966100	299700	3091	430800	948200	1604000
R_6	1429000	374600	3755	762100	1408000	2235000
R_7	2188000	495200	4806	1285000	2163000	3238000
R_8	4050000	800900	8743	2645000	4004000	5778000
R_9	4334000	1037000	11750	2548000	4237000	6636000
R_{10}	4860000	2095000	23320	192000	4489000	997000
<i>Total</i>	19090000	3008000	41080	14090000	18810000	25960000

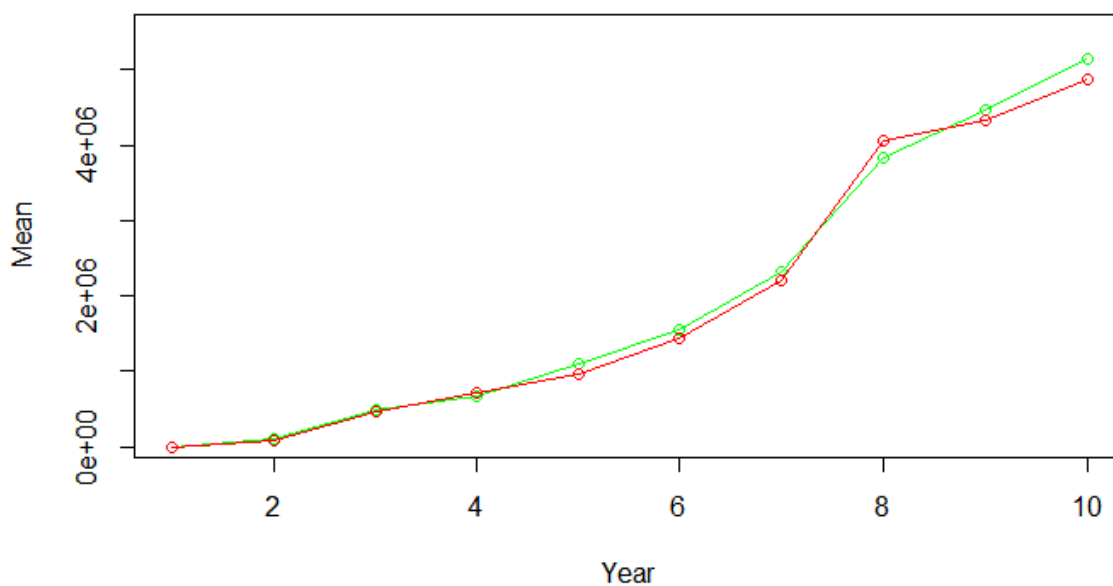
Άρα, η μέση τιμή (εκτίμηση) του συνολικού αποθέματος με βάση το μοντέλο Over-Dispersed Poisson είναι 19090000.

4.7. Σύγκριση των Μοντέλων (DIC)

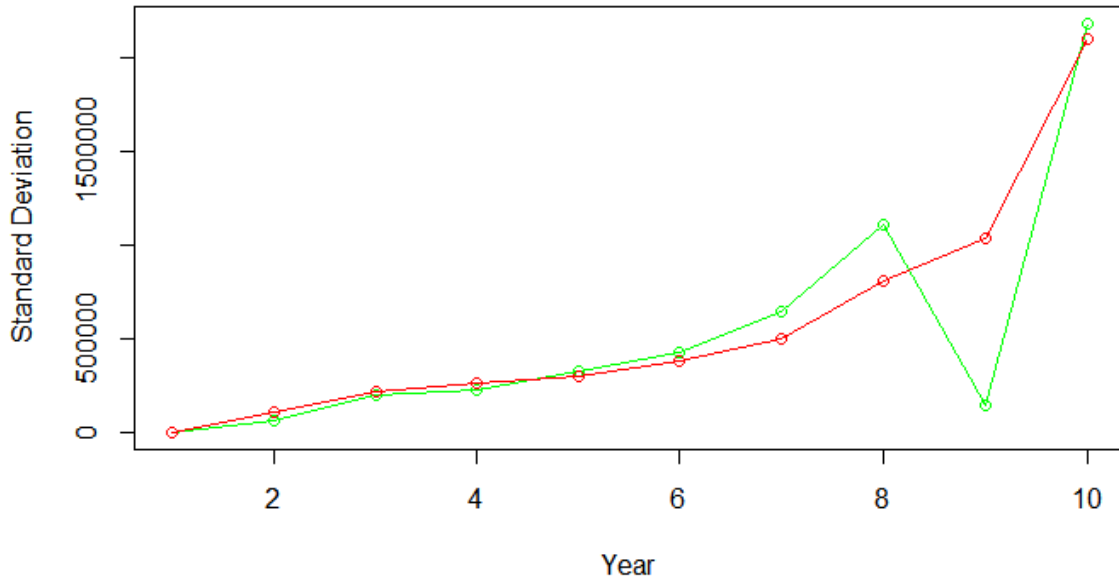
Αφού πλέον παρουσιάσαμε το Λογαριθμοκανονικό (Μοντέλο 1) και το ODP (Μοντέλο 2) μοντέλο, είναι αναγκαίο να συγκρίνουμε τα αποτελέσματα του και να αποφασίσουμε ποιο από τα δύο ταιριάζει καλύτερα στα δεδομένα μας.

Παρακάτω, στο Σχήμα 4.9 παρουσιάζεται η εκ των υστέρων μέση τιμή των εκτιμημένων αποθεμάτων για κάθε έτος ατυχήματος, ενώ στο Σχήμα 4.10 η εκ των υστέρων τυπική απόκλιση τους για κάθε έτος ατυχήματος. Με πράσινο χρώμα απεικονίζεται το Λογαριθμοκανονικό μοντέλο (Μοντέλο 1) και με κόκκινο το μοντέλο ODP (Μοντέλο 2):

Σχήμα 4.9: Σύγκριση των Μέσων των δύο μοντέλων ανά έτος ατυχήματος



Σχήμα 4.10: Σύγκριση των Τυπικών Αποκλίσεων των δύο μοντέλων ανά έτος ατυχήματος



Παρατηρώντας κανείς και τα δύο σχήματα δεν μπορεί να βγάλει ξεκάθαρα συμπεράσματα για το ποιο μοντέλο ταιριάζει καλύτερα στα δεδομένα μας. Για τον λόγο αυτό χρησιμοποιείται το Πληροφοριακό Κριτήριο Απόκλισης (*Deviance Information Criterion, DIC*), μία χαμηλότερη τιμή του κριτηρίου DIC δηλώνει ότι το μοντέλο ταιριάζει καλύτερα (βλέπε Spiegelhalter, Best, Carlin and van der Linde (2002)). Το DIC δίνεται από τον τύπο:

$$DIC(m) = \overline{D(\theta_m, m)} - D(\bar{\theta}_m, m), \quad (4.17)$$

όπου $D(\theta_m, m)$ είναι το μέτρο απόκλισης, $\overline{D(\theta_m, m)}$ ο εκ των υστέρων μέσος του, και $\bar{\theta}_m$ η εκ των υστέρων μέση τιμή των παραμέτρων που χρησιμοποιούμε στο μοντέλο m (στην περίπτωση μας 1 και 2).

Εφαρμόζοντας το παραπάνω και στα δύο μοντέλα παίρνουμε τον ακόλουθο πίνακα:

Πίνακας 4.5: Κριτήριο DIC

Μοντέλο	DIC
Λογαριθμοκανονικό	1508.140
ODP	1499.120

Παρατηρούμε ότι το μοντέλο ODP έχει μικρότερο κριτήριο DIC κατά 9.02 μονάδες, γεγονός που είναι επαρκές για να αποφασίσουμε πως το συγκεκριμένο μοντέλο δίνει καλύτερα αποτελέσματα για τα δεδομένα μας.

4.8. Μοντέλο Chain-Ladder για MCMC

Προηγουμένως, καταλήξαμε ότι το μοντέλο ODP δίνει καλύτερα αποτελέσματα από το Λογαριθμοκανονικό μοντέλο. Τελειώνοντας, θα το συγκρίνουμε με την μέθοδο Chain-Ladder που περιγράψαμε στο Υποκεφάλαιο 2.2 προκειμένου να έχουμε μια ολοκληρωμένη εικόνα για τα δεδομένα μας και να ελέγξουμε τη σταθερότητα των μοντέλων που χρησιμοποιούμε.

Για τον σκοπό μας θα χρησιμοποιήσουμε τον κώδικα που αναπτύχθηκε από τον David P.M. Scollnik (βλέπε Scollnik (2004)). Στο μοντέλο μας, δημιουργήσαμε 3 αλυσίδες με διαφορετικές αρχικές τιμές οι οποίες παράχθηκαν από το πρόγραμμα WinBUGS (οι αρχικές τιμές για τις 3 αλυσίδες παράχθηκαν με την επιλογή “gen inits”), έχουμε μια περίοδο burn-in 5000 επαναλήψεων από τις συνολικές 25000 επαναλήψεις του αλγορίθμου και $thin = 1$. Παρακάτω παρουσιάζεται το απόθεμα κάθε έτους αλλά και του συνολικού με τη μέθοδο Chain-Ladder και το μοντέλο ODP με το πρόγραμμα WinBUGS:

Πίνακας 4.6: Εκτίμηση Αποθεμάτων στο Μοντέλο Over-Dispersed Poisson και με την μέθοδο Chain-Ladder

	IBNR ODP	IBNR Chain-Ladder
R_2	91720	94630
R_3	463100	469500
R_4	711700	709600
R_5	966100	984900
R_6	1429000	1419000
R_7	2188000	2178000
R_8	4050000	3920000
R_9	4334000	4279000
R_{10}	4860000	4626000
<i>Total</i>	19090000	1868000

Συνοψίζοντας στον παραπάνω πίνακα παρατηρούμε ότι το απόθεμα κάθε έτους αλλά και το συνολικό με τη μέθοδο Chain-Ladder μέσω του προγράμματος WinBUGS είναι σχεδόν ίδια με

την ντετερμινιστική μέθοδο που περιγράψαμε στο Υποκεφάλαιο 2.2 και συγκριτικά η διαφορά στην εκτίμηση των δύο αποθεμάτων είναι σχετικά μικρή.

Επίλογος

Μετά το πέρας της παρούσας διπλωματικής εργασίας ο αναγνώστης θα είναι σε θέση να κατανοήσει σε ένα πρώτο επίπεδο την στοχαστική αποθεματοποίηση ζημιών και συγκεκριμένα τις μεθόδους Bootstrap και MCMC, καθώς και την συμβολή τους στον αντίστοιχο τομέα στις ασφαλιστικές εταιρίες.

Αρχικά, παρουσιάστηκαν οι βασικές ορολογίες και οι ορισμοί της αποθεματοποίησης ζημιών, μαζί με την βασική μέθοδο για την εκτίμηση των αποθεμάτων, την Chain-Ladder. Έπειτα αναφέρονται τα δεδομένα τα οποία θα χρησιμοποιήσουμε σε όλη την έκταση της διπλωματικής εργασίας. Στην συνέχεια έγινε μία εισαγωγή όσον αφορά την στοχαστική αποθεματοποίηση ζημιών, με την σύντομη παρουσίαση του μοντέλο του Mack, των Γενικευμένων Γραμμικών Μοντέλων (GLM) και με ιδιαίτερη έμφαση στην μέθοδο Bootstrap και την MCMC, όπου αποτελούν και το θέμα της εργασίας. Προκειμένου να προσαρμόσουμε τις δύο τελευταίες μεθόδους στα πλαίσια της αποθεματοποίησης ζημιών, αναφέρεται αρχικά η γενική θεωρία των μεθόδων και έπειτα αναπτύσσονται αριθμητικά παραδείγματα σχετικά με τα τρίγωνα εξέλιξης ζημιών, με την βοήθεια των λογισμικών προγραμμάτων *R* και WinBUGS.

Στην ερώτηση “ποιο μοντέλο ή ποια μέθοδος είναι καλύτερη;” δεν μπορούμε να απαντήσουμε απόλυτα. Τα δεδομένα χρειάζεται να εξεταστούν με περισσότερη λεπτομέρεια και η επιλογή της καταλληλότερης μεθόδου, ώστε να βρεθεί η καλύτερη εκτίμηση του αποθέματος, είναι στην διακριτική ευχέρεια του αναλογιστή. Είναι βέβαιο πως ακόμα υπάρχει περιθώριο για ανάπτυξη νέων εργαλείων και για βελτίωση των ήδη υπάρχοντων, έτσι ώστε να εφαρμόζονται όλο και περισσότερα σε πραγματικά δεδομένα και καταστάσεις στο κλάδο της στοχαστικής αποθεματοποίησης ζημιών.

Αναφορές

- [1] Adèr H.J., Mellenbergh G.J. and Hand D.J., (2008). Advising on Research Methods: A consultant's companion. Huizen, The Netherlands: Johannes van Kessel Publishing.
- [2] Brooks S.P. and Gelman A., (1998). General Methods for Monitoring Convergence of Iterative Simulations, *Journal of Computational and Graphical Statistics* 7 (4): 434-455.
- [3] Carlin B.P. and Louis T.A., (2000). *Bayes and Empirical Bayes Methods for Data Analysis: texts in statistical science*. 2nd edition, Chapman and Hall/CRC.
- [4] Carrato A., McGuire G. and Scarth R., (2016). *A Practitioner's Introduction to Stochastic Reserving*.
- [5] Christofides S., (1990). Regression Models based on Log-Incremental Payments. *Claims Reserving Manual 2*. Institute of Actuaries, London.
- [6] Davidson R. and MacKinnon J.G., (2000). Bootstrap tests: How many Bootstraps?. *Econometric Reviews* 19 (1): 55-68.
- [7] de Alba E., (2002). Bayesian Estimation of Outstanding Claim Reserves. *North American Actuarial Journal* 6 (4): 1-20.
- [8] Efron B. and Tibshirani R.J., (1993). *An Introduction to the Bootstrap*. Chapman & Hall, New York.
- [9] England P.D. and Verrall R.J., (1999). Analytic and Bootstrap Estimates of Prediction Errors in Claims Reserving. *Insurance: Mathematics and Economics* 25: 281-293.
- [10] England P.D. and Verrall R.J., (2002). Stochastic Claims Reserving in General Insurance (with discussion). *British Actuarial Journal* 8 (3): 443-518.
- [11] Geweke J., (1992). Evaluating the Accuracy of Sampling-Based Approaches to the Calculation of Posterior Moments. In J.M. Bernardo, J.O. Berger, A.P. Dawid and A.F.M. Smith, eds., *Bayesian Statistics 4*, Oxford: Oxford University Press.
- [12] Geyer C.J., (1991). Practical Markov Chain Monte Carlo. *Statistical Science* 7 (4): 473-483.

- [13] Kremer E., (1982). IBNR-Claims and the two-way Model of ANOVA. *Scandinavian Actuarial Journal*: 47–55.
- [14] Kuang D., Nielsen B. and Nielsen J.P., (2009). Chain-Ladder as Maximum Likelihood Revisited. *Annals of Actuarial Science* 4: 105–121.
- [15] Liivika T., Meelis K. and Rauno V., (2017). On Comparison of Stochastic Reserving Methods with Bootstrapping. *Risks* 5 (1).
- [16] Mack T., (1991). A Simple Parametric Model for Rating Automobile Insurance or Estimating IBNR Claims Reserves. *ASTIN Bulletin* 1991 21: 93–109.
- [17] Mack T., (1993). Distribution-free Calculation of the Standard Error of Chain- Ladder Reserve Estimates. *ASTIN Bulletin* 23: 213-225.
- [18] Malik K., (2011). Bayesian Modelling of Outstanding Liabilities in Non-Life Insurance. Heriot-Watt University.
- [19] McCullagh P. and Nelder J.A., (1989). *Generalized Linear Models*. 2nd Edition Chapman & Hall, London.
- [20] Ntzoufras I. and Dellaportas P., (2002). Bayesian Modelling of Outstanding Liabilities Incorporating Claim Count Uncertainty (with discussion). *North American Actuarial Journal* 6: 113-128.
- [21] Ntzoufras I., (2009). *Bayesian Modeling Using WinBUGS*. New York: John Wiley & Sons, Inc.
- [22] Pinheiro P.J., Andrade e Silva J.M. and de Lourdes Centeno, (2003). M. Bootstrap Methodology in Claim Reserving. *Journal of Risk Insurance* 70: 701–714.
- [23] Piwcewicz B., (2008). Stochastic Reserving-Case Study Using a Bayesian Approach. Institute of Actuaries of Australia, 16th General Insurance Seminar.
- [24] Renshaw A.E., (1989). Chain Ladder and Interactive Modelling (Claims Reserving and GLIM). *Journal of the Institute of Actuaries* 116 (3): 559–587.
- [25] Renshaw A.E. and Verrall R.J., (1994). A Stochastic Model Underlying the Chain Ladder Technique. *Proceedings XXV ASTIN Colloquium, Cannes*.

- [26] Rubin D.B., (1992). Inference from Iterative Simulation Using Multiple Sequences Andrew Gelman. *Statistical Science* 7 (4): 457-472.
- [27] Schmidt K.D., (2002). A Note on the Overdispersed Poisson Family. *Insurance: Mathematics and Economics* 30: 21–25.
- [28] Scollnik D.P.M., (2004). Bayesian Reserving Models Inspired by ChainLadder Methods and Implemented Using WinBUGS.
- [29] Spiegelhalter D.J., Best N.G., Carlin B.R. and van der Linde A., (2002). Bayesian Measures of Model Complexity and Fit. *Proceedings of the Royal Society of London Series B - Biological Sciences* 64 (4): 583–639.
- [30] Verrall R.J., (1989). A State Space Representation of the Chain Ladder Linear Model. *Journal of the Institute of Actuaries* 116: 589-609.
- [31] Verrall R.J., (1990). Bayes and Empirical Bayes Estimation for the Chain Ladder Model. *ASTIN Bulletin* 20: 217-243.
- [32] Verrall R.J., (1991a). On the Estimation of Reserves from Loglinear Models. *Insurance: Mathematics and Economics* 10: 75–80.
- [33] Verrall R.J., (1991b). Chain ladder and Maximum Likelihood. *Journal of the Institute of Actuaries* 18 (3): 489–499.
- [34] Verrall R.J., (2000). An Investigation into Stochastic Claims Reserving Models and the Chain-Ladder Technique. *Insurance: Mathematics and Economics* 26: 91-99.
- [35] Taylor G. and Ashe F., (1983). Second Moments of Estimates of Outstanding Claims. *Journal of Econometrics* 23: 37-61.
- [36] Taylor G., (2000). *Loss Reserving: An Actuarial Perspective*. Kluwer Academic Publishers: Dordrecht, The Netherlands.
- [37] Wüthrich M.V. and Merz M., (2008). *Stochastic Claims Reserving Methods in insurance*. Wiley, Chichester.
- [38] Zehnwirth B., (1989). The Chain Ladder Technique - A Stochastic Model. *Claims Reserving Manual 2*, London Institute of Actuaries.

- [39] Zehnwirth B., (1991). Interactive Claims Reserving Forecasting System (ICRFS). Insurware P/L, E. St Kilda, Victoria 3183, Australia.
- [40] Πιτσέλης Γ., (2017). Αποθεματοποίηση Ζημιών. Πανεπιστημιακές Σημειώσεις, Τμήμα Στατιστικής και Ασφαλιστικής Επιστήμης, Πανεπιστήμιο Πειραιώς.

Παραρτήματα

Παράρτημα Α: Κώδικας R

A-1: Μέθοδος Chain-Ladder

```
#PACKAGES-----  
library(apc)  
library(ChainLadder)  
library(xtable)  
library(MASS)  
  
data=data.loss.TA()  
incr.tri=data$response  
incr.tri=round(incr.tri)  
incr.tri  
  
#CUMULATIVE CLAIMS-----  
cum.tri=matrix(NA,10,10)  
cum.tri[,1]=incr.tri[,1]  
for (i in 1:10){  
  for (j in 2:10){  
    cum.tri[i,j]=cum.tri[i,j-1]+incr.tri[i,j]  
  }  
}  
rownames(cum.tri)=c(1:10)  
colnames(cum.tri)=c(1:10)  
cum.tri=round(cum.tri)  
cum.tri  
  
#DEVELOPMENT FACTORS-----  
f=ata(cum.tri)  
dev.factors=attr(f, "vwtd")  
names(dev.factors)=c(2:10)  
dev.factors  
  
#PREDICTED CUMULATIVE CLAIMS-----
```

```

full.cum.tri=cum.tri
for (i in 1:10){
for (j in 2:10){
if (i+j>11) {
full.cum.tri[i,j]=dev.factors[j-1]*full.cum.tri[i,j-1]
}
}
}
full.cum.tri=round(full.cum.tri)
full.cum.tri

#PREDICTED INCREMENTAL CLAIMS-----
full.incr.tri=full.cum.tri
for (i in 1:10){
for (j in 2:10){
full.incr.tri[i,j]=full.cum.tri[i,j]-full.cum.tri[i,j-1]
}
}
full.incr.tri=round(full.incr.tri)
full.incr.tri

#IBNR CHAIN LADDER-----
IBNR.CL=matrix(NA,10,1)
for (i in 2:10){
IBNR.CL[i-1]=full.cum.tri[i,10]-full.cum.tri[i,10-i+1]
}
IBNR.CL[10]=sum(IBNR.CL[1:9])
rownames(IBNR.CL)=c("R2","R3","R4","R5","R6","R7","R8","R9","R10",
"Total")
colnames(IBNR.CL)=c("Reserve_□each_□year")
t(IBNR.CL)

```

A-2: Mack Chain-Ladder

```

MCL=MackChainLadder(cum.tri,est.sigma="Mack")
MCL

```

A-3: Μέθοδος Bootstrap - Μοντέλο Over-Dispersed Poisson

```
#OVER-DISPERSED POISSON DISTRIBUTION-----
rpois.od<-function (n, lambda,d=1) {
if (d==1)
rpois(n, lambda)
else
rnbinom(n, size=(lambda/(d-1)), mu=lambda)}

#FITTED CUMULATIVE CLAIMS-----
fitted.cum.tri=matrix(NA,10,10)
for (i in 1:10){
for (j in 1:10){
if (i+j==11) fitted.cum.tri[i,j]=cum.tri[i,j]
}}
for (i in 1:9){
for (j in rev(1:9)){
if (i+j<11){
fitted.cum.tri[i,j]=fitted.cum.tri[i,j+1]/dev.factors[j]
}
}
}
fitted.cum.tri

#FITTED INCREMENTAL CLAIMS-----
fitted.incr.tri=fitted.cum.tri
for (i in 1:10){
for (j in 2:10){
fitted.incr.tri[i,j]=fitted.cum.tri[i,j]-fitted.cum.tri[i,j-1]
}
}
fitted.incr.tri=round(fitted.incr.tri)
fitted.incr.tri

#PEARSON RESIDUALS-----
pearson=matrix(NA,10,10)
for (i in 1:10){
for (j in 1:10){
pearson[i,j]=(incr.tri[i,j]-fitted.incr.tri[i,j])/+
```

```

sqrt(fitted.incr.tri[i,j])
}
}
pearson=round(pearson)
pearson

#DISPERSION PARAMETER-----
sum.pearson=0
N=55
p=19
for (i in 1:10){
for (j in 1:10){
if (i+j<=11) sum.pearson=sum.pearson+pearson[i,j]^2
}
}
hat.phi=sum.pearson/N-p
hat.phi

#ADJUSTED PEARSON RESIDUALS-----
adj.pearson=round(sqrt(55/(55-19))*pearson)
adj.pearson

#RESAMPLING RESIDUALS-----
#0's CORRECTION
x=c(209,142,-138,-386,210,644,-292,-122,
-108,-48,-68,-59,162,-220,-167,311,31,91,
-166,95,-57,-27,286,-499,
256,73,-115,252,-229,659,-483,
-89,-324,227,-195,151,-215,-26,218,
88,74,-98,-226,266,96,-161,
133,-36,-199,-124,243,-27,17)

set.seed(1993)
adj.pearson.b=matrix(NA,10,10)
for (i in 1:10){
for (j in 1:10){
if (i+j<=11) adj.pearson.b[i,j]=sample(x,1,replace=T)
}}
adj.pearson.b

```

#PSEUDO-INCREMENTAL CLAIMS-----

```
pseudo.incr.tri=matrix(NA,10,10)
for (i in 1:10){
for (j in 1:10){
pseudo.incr.tri[i,j]=adj.pearson.b[i,j]*+
sqrt(fitted.incr.tri[i,j])+fitted.incr.tri[i,j]
}}
pseudo.incr.tri=round(pseudo.incr.tri)
pseudo.incr.tri
```

#PSEUDO-CUMULATIVE CLAIMS-----

```
pseudo.cum.tri=matrix(NA,10,10)
pseudo.cum.tri[,1]=pseudo.incr.tri[,1]
for (i in 1:10){
for (j in 2:10){
pseudo.cum.tri[i,j]=pseudo.cum.tri[i,j-1]+pseudo.incr.tri[i,j]
}
}
rownames(pseudo.cum.tri)=c(1:10)
colnames(pseudo.cum.tri)=c(1:10)
pseudo.cum.tri
```

#PSEUDO-DEVELOPMENT FACTORS-----

```
pseudo.f=ata(pseudo.cum.tri)
pseudo.dev.factors=attr(pseudo.f, "vwtd")
names(pseudo.dev.factors)=c(2:10)
pseudo.dev.factors
```

#PREDICTED PSEUDO-CUMULATIVE CLAIMS-----

```
full.pseudo.cum.tri=pseudo.cum.tri
for (i in 1:10){
for (j in 2:10){
if (i+j>11) {
full.pseudo.cum.tri[i,j]=pseudo.dev.factors[j-1]*+
full.pseudo.cum.tri[i,j-1]
}
}
}
```



```

}
full.pseudo.cum.tri=round(full.pseudo.cum.tri)
full.pseudo.cum.tri

#PREDICTED PSEUDO-INCREMENTAL CLAIMS-----
full.pseudo.incr.tri=full.pseudo.cum.tri
for (i in 1:10){
for (j in 2:10){
full.pseudo.incr.tri[i,j]=full.pseudo.cum.tri[i,j]-+
full.pseudo.cum.tri[i,j-1]
}
}
full.pseudo.incr.tri=round(full.pseudo.incr.tri)
full.pseudo.incr.tri

#SIMULATED INCREMENTAL CLAIMS-----
set.seed(1993)
sim.incr.tri=full.pseudo.incr.tri
for (i in 1:10){
for (j in 1:10){
if (i+j>11) sim.incr.tri[i,j]=rpois.od(1,full.pseudo.incr.tri[i,j],
hat.phi)
}}
sim.incr.tri

#SIMULATED CUMULATIVE CLAIMS-----
sim.cum.tri=matrix(NA,10,10)
sim.cum.tri[,1]=sim.incr.tri[,1]
for (i in 1:10){
for (j in 2:10){
sim.cum.tri[i,j]=sim.cum.tri[i,j-1]+sim.incr.tri[i,j]
}
}
sim.cum.tri

```

```

#IBNR BOOTSTRAP (OVER-DISPERSED POISSON)-----
IBNR.B=matrix(NA,10,1)
for (i in 2:10){
IBNR.B[i-1]=sim.cum.tri[i,10]-sim.cum.tri[i,10-i+1]
}
IBNR.B[10]=sum(IBNR.B[1:9])
rownames(IBNR.B)=c("$R2","R3","R4","R5","R6","R7","R8","R9","R10",
"Total_(IBNR)")
colnames(IBNR.B)=c("Reserve_each_year")
t(IBNR.B)

```

```

#BOOTCHAINLADDER (ODP)-----
set.seed(1993)
boot.ibnr.odp=BootChainLadder(cum.tri,R=1000,process.distr="od.pois")
boot.ibnr.odp
Bodp=summary(boot.ibnr.odp)
plot(boot.ibnr.odp)

```

```

#FIT A LOG-NORMAL DISTRIBUTION -----
fit <- fitdistr(boot.ibnr.odp$IBNR.Totals[boot.ibnr.odp$IBNR.Totals>0],
"lognormal")
fit
plot(ecdf(boot.ibnr.odp$IBNR.Totals))
curve(plnorm(x,fit$estimate["meanlog"], fit$estimate["sdlog"]),
col="red",add=TRUE)

```

A-4: Μέθοδος Bootstrap - Μοντέλο Γάμμα (Gamma)

```

set.seed(1993)
boot.ibnr.gamma=BootChainLadder(cum.tri,R=1000,process.distr="gamma")
boot.ibnr.gamma
Bgamma=summary(boot.ibnr.gamma)
plot(boot.ibnr.gamma)

```

Παράρτημα Β: Κώδικας WinBUGS

B-1: MCMC - Το Λογαριθμοκανονικό Μοντέλο

```
model {
#MODEL 'S LIKELIHOOD-----
for (i in 1:n) {
for (j in 1:n){
x[i,j] ~ dlnorm(m[i,j], t)
m[i,j] <- c + a[i] + b[j]
} }

#CONSTRAINTS-----
#SUM TO ZERO CONSTRAINTS
a[1] <- -sum(a[2:n])
b[1] <- -sum(b[2:n])

#PRIORS-----
c ~ dnorm(0,0.001)
for (i in 2:n) {
a[i] ~ dnorm(0, 0.01)
b[i] ~ dnorm(0,0.01) }
t ~ dgamma(0.01, 0.01)
sigma2 <- 1/t

#ACCIDENT YEAR AND TOTAL RESERVE-----
R[1] <- 0
for (i in 2:n){
R[i] <- sum(x[i,(n+2-i):n])
}
Total <- sum(R[])
}

#DATA-----
list(n=10, x=structure(.Data=c(357848,766940,610542,482940,527326,
574398,146342,139950,227229,67948,352118,884021,
933894,1183289,445745,320996,527804,266172,425046,
NA,290507,1001799,926219,1016654,750816,146923,495992,
```

```

280405,NA,NA,310608,1108250,776189,1562400,272482,
352053,206286,NA,NA,NA,443160,693190,991983,769488,
504851,470639,NA,NA,NA,NA,396132,
937085,847498,805037,705960,NA,NA,NA,NA,NA,
440832,847631,1131398,1063269,NA,NA,NA,NA,NA,NA,
359480,1061648,1443370,NA,NA,NA,NA,NA,NA,NA,376686,
986608,NA,NA,NA,NA,NA,NA,NA,NA,344014,NA,NA,NA,NA,NA,NA,NA,NA),
.Dim=c(10,10) ) )

```

```

#INITIAL VALUES-----
list(a=c(
NA,0.1,0.1,0.1,0.1,0.1,0.1,0.1,0.1,0.1),b=c(
NA,0.5,0.5,0.5,0.5,0.5,0.5,0.5,0.5,0.5),
c=5,
t=1,
x = structure(.Data = c(NA,NA,NA,NA,NA,NA,NA,NA,NA,
NA,NA,NA,NA,NA,NA,NA,NA,NA,
NA,NA,0,NA,NA,NA,NA,NA,NA,NA,NA,
0,0,NA,NA,NA,NA,NA,NA,NA,0,0,0,NA,NA,NA,NA,
NA,NA,0,0,0,0,NA,NA,NA,NA,NA,
0,0,0,0,0,NA,NA,NA,NA,0,0,0,0,0,0,NA,NA,NA,0,0,
0,0,0,0,0,NA,NA,0,0,0,0,0,0,0,0,
NA,0,0,0,0,0,0,0,0,0),.Dim = c(10,10)))

```

B-2: MCMC - Μοντέλο Over-Dispersed Poisson

```

#MODEL-----
model;
{
c ~ dnorm(0 , 0.001)
a[1] <- 0
for( i in 2 : n ){
a[i] ~ dnorm(0 , 0.01)
}
b[1] <- 0
for( i in 2 : n ){
b[i] ~ dnorm(0 , 0.01)
}
for( i in 1 : n ){

```

```

for( j in 1 : n ){
mu[i , j] <- exp( c + a[i] + b[j] )
z[i , j] <- 1*iphi/mu[i , j]  #iphi=1/phi
x[i , j] ~ dnorm( mu[i , j] , z[i , j] )
}
}
iphi ~ dgamma(1 , 0.01)
R[1]<-0
for( i in 2 : n ){
R[i] <- sum( x[i , (n+2 - i) : n] )
}
Total<- sum( R[] )
}

#DATA-----
list(n=10, x=structure(.Data=c(357848,766940,610542,
482940,527326,574398,146342,
139950,227229,67948,352118,884021,933894,
1183289,445745,320996,527804,266172,
425046,NA,290507,1001799,926219,1016654,
750816,146923,495992,280405,NA,NA,
310608,1108250,776189,1562400,272482,352053,
206286,NA,NA,NA,443160,693190,991983,769488,
504851,470639,NA,NA,NA,NA,
396132,937085,847498,805037,705960,
NA,NA,NA,NA,NA,440832,847631,1131398,1063269,
NA,NA,NA,NA,NA,NA,359480,1061648,
1443370,NA,NA,NA,NA,NA,NA,NA,376686,986608,NA,
NA,NA,NA,NA,NA,NA,NA,344014,NA,NA,
NA,NA,NA,NA,NA,NA,NA), .Dim=c(10,10) ) )

#INITIAL VALUES-----
#Initial values for all 3 chains generated from the programm.

```

B-3: MCMC - Μέθοδος Chain-Ladder

```

#MODEL-----
model{
# Define the first column of the cumulative claims triangle.

```

```

for( i in 1 : r ) {
z[ i, 1 ] <- y[ i, 1 ]
}
# Compute the cumulative claim amounts in successive development periods
for( i in 1 : ( r - 1 ) ) {
for( j in 1 : ( r - i ) ) {
z[ i, j + 1 ] <- z[ i, j ] + y[ i, j + 1 ]
}
}

# Calculate the traditional volume weighted development factors.
for( j in 1 : ( r - 1 ) ) {
volwtdevfac[ j ] <- sum( z[ 1 : ( r - j ), j + 1 ] ) /
sum( z[ 1 : ( r - j ), j ] ) + normerror[ j ]
normerror[ j ] ~ dnorm( 0, 1000000000 )
}

# Traditional model for incremental and cumulative claims.
for( i in 2 : r ) {
z.ns[ i, r - i + 1 ] <- z[ i, r - i + 1 ]
for( j in ( r - i + 1 ) : ( r - 1 ) ) {
z.ns[ i, j + 1 ] <- z.ns[ i, j ] * volwtdevfac[ j ]
y.ns[ i, j + 1 ] <- z.ns[ i, j + 1 ] - z.ns[ i, j ]
}
}

for( i in 2 : r ) {
R[ i ] <- sum( y.ns[ i, ( r + 2 - i ) : r ] )
}

Total<- sum( R[])
}

#DATA-----
list(r=10, y=structure(.Data=c(357848,766940,610542,482940,527326,574398,
146342,139950,227229,67948,352118,884021,933894,1183289,445745,320996,527804,
266172,425046,NA,290507,1001799,926219,1016654,750816,146923,495992,280405,
NA,NA,310608,1108250,776189,1562400,272482,352053,206286,NA,NA,NA,443160,
693190,991983,769488,504851,470639,NA,NA,NA,NA,396132,937085,847498,805037,
705960,NA,NA,NA,NA,NA,440832,847631,1131398,1063269,NA,NA,NA,NA,NA,NA,359480,
1061648,1443370,NA,NA,NA,NA,NA,NA,NA,376686,986608,NA,NA,NA,NA,NA,NA,NA,NA,
344014,NA,NA,NA,NA,NA,NA,NA,NA,NA,NA), .Dim=c(10,10) ) )

```