

ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΕΙΡΑΙΩΣ



**ΤΜΗΜΑ ΣΤΑΤΙΣΤΙΚΗΣ
ΚΑΙ ΑΣΦΑΛΙΣΤΙΚΗΣ ΕΠΙΣΤΗΜΗΣ**

**ΜΕΤΑΠΤΥΧΙΑΚΟ ΠΡΟΓΡΑΜΜΑ ΣΠΟΥΔΩΝ
ΣΤΗΝ ΑΝΑΛΟΓΙΣΤΙΚΗ ΕΠΙΣΤΗΜΗ ΚΑΙ
ΔΙΟΙΚΗΤΙΚΗ ΚΙΝΔΥΝΟΥ**

**“ ΣΤΟΧΑΣΤΙΚΗ ΜΟΝΤΕΛΟΠΟΙΗΣΗ
ΑΠΟΘΕΜΑΤΙΚΩΝ ΑΣΦΑΛΙΣΤΙΚΩΝ
ΧΑΡΤΟΦΥΛΑΚΙΩΝ ”**

Νικολέττα Μαρούντα

Διπλωματική Εργασία

που υποβλήθηκε στο Τμήμα Στατιστικής και Ασφαλιστικής Επιστήμης του Πανεπιστημίου Πειραιώς ως μέρος των απαιτήσεων για την απόκτηση Μεταπτυχιακού Διπλώματος Ειδίκευσης στην Αναλογιστική επιστήμη και Διοικητική Κινδύνου

Πειραιάς,
2018

UNIVERSITY OF PIRAEUS



**DEPARTMENT OF STATISTICS
AND INSURANCE SCIENCE**

**POSTGRADUATE PROGRAM IN
ACTUARIAL SCIENCE AND RISK
MANAGEMENT**

**“STOCHASTIC MODELLING OF
RESERVES FOR INSURANCE
PORTFOLIOS”**

BY
Nikoletta Marounta

MSc Dissertation

Submitted to the Department of Statistics and Insurance
Science of the University of Piraeus in partial fulfillment
of the requirements for the degree of Master of Science in
Actuarial Science and Risk Management

Piraeus, Greece

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. 5^η/13.06.2016 συνεδρίασή του σύμφωνα με τον Εσωτερικό Κανονισμό Λειτουργίας του Προγράμματος Μεταπτυχιακών Σπουδών στην Αναλογιστική Επιστήμη και Διοικητική Κινδύνου.

Τα μέλη της Επιτροπής ήταν:

- Ε. Χατζηκωνσταντινίδης, Αναπλ. Καθηγητής (Επιβλέπων)
- Μ. Νεκτάριος, Αναπλ. Καθηγητής
- Γ. Τζαβελάς, Επικ. Καθηγητής

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

Ευχαριστίες

θα ήθελα να ευχαριστήσω τον καθηγητή μου κ. Ευστάθιο Χατζηκωνσταντίδη για τη βοήθεια του, την καθοδήγηση του και την στήριξη του κατά τη διάρκεια συγγραφής της παρούσας διπλωματικής εργασίας. Τέλος, θα ήθελα να ευχαριστήσω την οικογένεια μου για την αμέριστη συμπαράσταση και υποστήριξη τους σε κάθε μου προσπάθεια.

Περίληψη

Η εργασία αυτή εξετάζει τα στοχαστικά μοντέλα τα οποία αναπαράγουν τις εκτιμήσεις της μεθόδου chain-ladder(μέθοδος τριγωνικής εξέλιξης ζημιών) που χρησιμοποιούνται για την εκτίμηση των αποθεματικών των γενικών ασφαλίσεων μιας ασφαλιστικής εταιρίας. Η μέθοδος chain-ladder δεν παρέχει πληροφορίες σχετικά με τη μεταβλητότητα του αποτελέσματος, προσθέτοντας έτσι αβεβαιότητα στις μελλοντικές εκτιμήσεις των απαιτήσεων.Επιπλέον μπορεί να προκύψουν προβλήματα πρόβλεψης χρησιμοποιώντας μια ποικιλία στοχαστικών μοντέλων chain-ladder, όμως αυτά τα διαφορετικά μοντέλα βασίζονται και σε διαφορετικές υποθέσεις. Η σχέση μεταξύ μερικών από αυτά τα μοντέλα εξετάστηκε και αποδείχθηκε πώς τα μοντέλα αυτά ορίζονται για ένα τρίγωνο εξέλιξης των ασφαλιστικών απαιτήσεων. Δύο από αυτά τα μοντέλα, το μοντέλο του Mack και η προσέγγιση του αρνητικού διωνυμικού μοντέλου από την κανονική κατανομή, εφαρμόστηκαν σε ένα σύνολο δεδομένων που αποτελείται από απαιτήσεις μιας ασφαλιστικής εταιρίας για την κάλυψη των γενικών ασφαλίσεων και συγκεκριμένα των αυτοκινήτων. Αυτό έγινε για να βρεθεί το σφάλμα πρόβλεψης των εκτιμήσεων της chain-ladder, καθώς και να επαληθευτεί η ικανότητά τους να διαχειρίζονται αρνητικές τιμές. Στα δύο μοντέλα που χρησιμοποιήθηκαν στην ανάλυση διαπιστώθηκε ότι παράγουν σχεδόν ταυτόσημα σφάλματα πρόβλεψης αλλά και ότι τα δύο ήταν σε θέση να διαχειριστούν τις αρνητικές απαιτήσεις που υπήρχαν στο σύνολο των δεδομένων. Επιπλέον βρέθηκαν ομοιότητες μεταξύ των μοντέλων, σε τέτοιο βαθμό που η προσέγγιση του αρνητικού διωνυμικού μοντέλου από την κανονική κατανομή θα μπορούσε να θεωρηθεί ως το μοντέλο του Mack. Δεδομένου ότι βασίζεται σε ένα γενικευμένο γραμμικό μοντέλο ,η προσέγγιση του αρνητικού διωνυμικού μοντέλου από την κανονική κατανομή προσφέρει μεγαλύτερη ευελιξία στους υπολογισμούς απ'ότι το μοντέλο του Mack.

Abstract

This thesis examines the stochastic models which reproduce chain-ladder estimates used in reserve estimation for nonlife insurance. The chain ladder method provides no information regarding the variability of the outcome, thereby adding uncertainty to future claim estimations. Prediction errors can be found using a variety of stochastic chain ladder models, but the different models are based on different assumptions. The relationship between some of these models was explored, and it was demonstrated how the models are defined for a run-off triangle of insurance claims. Two of these models, Mack's model and the normal approximation to the negative binomial model, were applied to a data set consisting of auto liability insurance claims. This was done in order to find the prediction error of their chain ladder estimates, as well as verify their ability to handle negative values. The two models used in the analysis were found to produce nearly identical prediction errors, and both were able to handle negative insurance claims, which were present in the data set. A number of similarities were found between the models, to the degree that the normal approximation to the negative binomial model should be considered as underlying Mack's model. However, since it is based on a generalized linear model, the normal approximation to the negative binomial model offers greater flexibility in applied calculations than Mack's model.

Περιεχόμενα

ΚΕΦΑΛΑΙΟ 1 : Εισαγωγή	Error! Bookmark not defined.
1.1 Ιστορική αναδρομή	Error! Bookmark not defined.
1.2 Σκοπός διπλωματικής	Error! Bookmark not defined.
1.3 Ορισμοί, συμβολισμοί και περιορισμοί του μοντέλου	Error! Bookmark not defined.
defined.	
1.4 Ο αλγόριθμος της chain-ladder	Error! Bookmark not defined.
1.5 Η χρήση των στοχαστικών μοντέλων στη μέθοδο της chain-ladder	Error! Bookmark not defined.
Bookmark not defined.	
1.6 Δημιουργία ενός στοχαστικού μοντέλου με τη χρήση της chain-ladder μεθόδου	Error! Bookmark not defined.
Error! Bookmark not defined.	
1.7 Στοχαστικά chain-ladder μοντέλα	Error! Bookmark not defined.
1.8 Ντετερμινιστικοί Μέθοδοι	Error! Bookmark not defined.
1.8.1 Η μέθοδος Chain ladder	Error! Bookmark not defined.
1.8.2 Προσαρμοσμένες τιμές	Error! Bookmark not defined.
1.8.3 Η μέθοδος Bornhuetter-Ferguson	Error! Bookmark not defined.
1.8.4 Μέθοδος Διαχωρισμού	Error! Bookmark not defined.
1.8.5 Η μέθοδος Cape-Cod	Error! Bookmark not defined.
ΚΕΦΑΛΑΙΟ 2 : Στοχαστικά μοντέλα	Error! Bookmark not defined.
2.1 Ο αριθμός απαιτήσεων και η διαδικασία μεγέθους απαιτήσεων	Error! Bookmark not defined.
not defined.	
2.2 Το πολλαπλασιαστικό μοντέλο και η chain-ladder μέθοδος	Error! Bookmark not defined.
defined.	
2.3 Το μοντέλο Poisson και η μέθοδος chain-ladder	Error! Bookmark not defined.
2.4 Το μοντέλο poisson και η σχέση του με το αρνητικό διωνυμικό μοντέλο	Error! Bookmark not defined.
Bookmark not defined.	
2.5 Το μοντέλο του Mack	Error! Bookmark not defined.
2.6 Το μοντέλο του Mack και τη σύνδεσή του με την κατανομή Poisson	Error! Bookmark not defined.
not defined.	
2.7 Αρνητικές αυξητικές απαιτήσεις	Error! Bookmark not defined.
2.8 Προβλέψεις και σφάλματα πρόβλεψης	Error! Bookmark not defined.

ΚΕΦΑΛΑΙΟ 3 : Ανάλυση δεδομένων από ασφαλιστικές απαιτήσεις αυτοκινήτων με χρήση μοντέλων στοχαστικής αλυσίδας.....	Error! Bookmark not defined.
3.1 Δείγμα.....	Error! Bookmark not defined.
3.2 Στοχαστικά μοντέλα στην ανάλυση	Error! Bookmark not defined.
3.3 Μια σημαντική τοποθέτηση για τις στοχαστικές υποθέσεις της αλυσίδας-σκάλας	Error! Bookmark not defined.
3.3.1 Η μεροληπτικότητα της chain-ladder μεθόδου.....	Error! Bookmark not defined.
3.3.2 Η διακύμανση των απαιτήσεων.....	Error! Bookmark not defined.
3.4 Η προσέγγιση της αρνητικής διωνυμικής κατανομή από την κανονική κατανομή	Error! Bookmark not defined.
3.4.1 Το μοντέλο.....	Error! Bookmark not defined.
3.4.2 Προβλέψεις αποθεμάτων και σφάλματα πρόβλεψης	Error! Bookmark not defined.
3.4.3 Αποτελέσματα.....	Error! Bookmark not defined.
3.5 Το μοντέλο του Mack	Error! Bookmark not defined.
3.5.1 Το μοντέλο	Error! Bookmark not defined.
3.5.2 Προβλέψεις αποθεμάτων και σφάλματα πρόβλεψης	Error! Bookmark not defined.
3.5.3 Αποτελέσματα.....	Error! Bookmark not defined.
ΚΕΦΑΛΑΙΟ 4 : Σχόλια	Error! Bookmark not defined.
ΚΕΦΑΛΑΙΟ 5 : Συμπεράσματα	Error! Bookmark not defined.
ΚΕΦΑΛΑΙΟ 6 : Βιβλιογραφία	Error! Bookmark not defined.
Παράρτημα 1.....	Error! Bookmark not defined.
Παράρτημα 2.....	Error! Bookmark not defined.
Παράρτημα 3.....	Error! Bookmark not defined.
Παράρτημα 4.....	Error! Bookmark not defined.
Παράρτημα 5.....	Error! Bookmark not defined.
Παράρτημα 6.....	Error! Bookmark not defined.

ΚΕΦΑΛΑΙΟ 1 : Εισαγωγή

1.1 Ιστορική αναδρομή

Μια ασφαλιστική εταιρεία διαθέτει ένα χαρτοφυλάκιο πελατών. Μερικοί από αυτούς δεν θα κάνουν ποτέ κάποια απαίτηση, ενώ άλλοι θα μπορούσαν να κάνουν μία ή περισσότερες απαιτήσεις. Ο αναλογιστής υπολογίζει τα αποθέματα για να είναι σε θέση να καλύψει αυτές τις απαιτήσεις. Στην ασφάλιση ατυχημάτων, η πολιτική περίοδος είναι συνήθως ένα έτος. Μετά από αυτό το έτος, η πολιτική θα μπορούσε είτε να ανανεωθεί είτε να τερματιστεί. Εάν η πολιτική περίοδος τελειώσει, αυτό δεν σημαίνει απαραίτητα ότι η ευθύνη της ασφαλιστικής έχει λήξει. Όλες οι απαιτήσεις που πραγματοποιήθηκαν εντός αυτής της περιόδου (και οι όροι των ασφαλιστηρίων συμβολαίων) είναι ευθύνη του αναλογιστή. Μεταξύ αυτών υπάρχουν απαιτήσεις που έχουν δηλωθεί αλλά δεν έχουν διευθετηθεί (IBNS-Incurred But Not Settled) και απαιτήσεις που έχουν δηλωθεί αλλά δεν έχουν αναφερθεί (IBNR-Incurred But Not Reported).

Ένας διακανονιστής απαιτήσεων σε μια ασφαλιστική εταιρεία θα πρέπει να είναι σε θέση να προσδιορίσει περίπου το ύψος της αποζημίωσης για τις απαιτήσεις-IBNS. Ωστόσο, οι απαιτήσεις-IBNR είναι πολύ πιο δύσκολο να εκτιμηθούν. Σε ορισμένες περιπτώσεις, ακόμη και ο πελάτης μπορεί να μην γνωρίζει ότι έχει κάνει κάποια απαίτηση. Αυτό μπορεί να συμβεί, παραδείγματος χάριν, σε περιπτώσεις όπου ο πελάτης αντιλαμβάνεται τη βαρύτητα του τραυματισμού αρκετές εβδομάδες μετά το αρχικό τραύμα. Ένας άλλος τύπος IBNR-απαίτησης θα μπορούσε να είναι ζημιά από νερό σε ένα σπίτι, όπου η διαρροή δεν ανακαλύφθηκε άμεσα.

Μία κοινή μέθοδος που χρησιμοποιείται για την εκτίμηση IBNR-απαιτήσεων είναι η μέθοδος chain-ladder. Αυτό βασίζεται σε έναν αλγόριθμο που κάνει μια εκτίμηση σημείων για μελλοντικές απαιτήσεις. Η μέθοδος chain-ladder είναι απλή και λογική και χρησιμοποιείται ευρέως στην ασφάλιση ατυχημάτων. Παρά τη δημοτικότητά του, υπάρχουν αδυναμίες στη μέθοδο αυτή. Το πιο σημαντικό είναι ότι δεν παρέχει πληροφορίες σχετικά με τη μεταβλητότητα του αποτελέσματος. Με τη δύναμη επεξεργασίας των σύγχρονων υπολογιστών, η απλότητα της μεθόδου αυτής δεν αποτελεί πλέον σημαντικό εγχείρημα. Εντούτοις, η μέθοδος chain-ladder χρησιμοποιείται συχνά από τους αναλογιστές. Βελτιώσεις στη μέθοδο chain-ladder έχουν γίνει μέσω της ανάπτυξης στοχαστικών μοντέλων που υποστηρίζουν την τεχνική chain-ladder (England & Verral 2002, Hess & Schmidt 2002, Mack 1994a, Mack 1994b, Neuhaus 2006, Renshaw 1998). Σφάλματα πρόβλεψης μπορούν να ληφθούν όταν χρησιμοποιείται ένα στοχαστικό μοντέλο, επιτρέποντας έτσι τη μεγαλύτερη γνώση της εκτίμησης των αποθεμάτων.

1.2 Σκοπός διπλωματικής

Ο κύριος στόχος αυτής της εργασίας είναι να αναδείξει τις μεθόδους που χρησιμοποιούνται για τον προσδιορισμό της μεταβλητότητας του αποτελέσματος (σφάλμα πρόβλεψης) σε έναν υπολογισμό chain-ladder. Αυτό θα επιτευχθεί με την περιγραφή του αλγόριθμου της chain-ladder, την ανασκόπηση των πιο σημαντικών στοχαστικών μοντέλων chain-ladder, την εξέταση της σύνδεσης μεταξύ των στοχαστικών μοντέλων, την τοποθέτηση των μοντέλων στο τρίγωνο εξέλιξης των ασφαλιστικών απαιτήσεων και την χρήση δύο μοντέλων σε ένα σύνολο ασφαλιστικών δεδομένων για απαιτήσεις αυτοκινήτων. Επίσης, θα δοκιμαστούν οι υποθέσεις των μοντέλων στα δύο μοντέλα αυτά. Τέλος, θα χρησιμοποιηθούν τα

αποτελέσματα από την ανάλυση αυτή έτσι ώστε να συζητήσουμε τα δύο στοχαστικά μοντέλα και τη μέθοδο της chain-ladder.

1.3 Ορισμοί, συμβολισμοί και περιορισμοί του μοντέλου

Ένα στοχαστικό μοντέλο chain-ladder ορίζεται ως ένα στοχαστικό μοντέλο το οποίο παράγει τις ίδιες εκτιμήσεις των μελλοντικών απαιτήσεων όπως η μέθοδος chain-ladder.

Η μέθοδος chain-ladder θεωρείται μια ντετερμινιστική μέθοδος όπου οι μεταβλητές είναι γνωστές. Τα στοχαστικά μοντέλα chain-ladder θα χρησιμοποιούν γενικά τα κεφαλαία γράμματα όταν οι μεταβλητές πρόκειται να θεωρηθούν ως στοχαστικές μεταβλητές και οι γνωστές μεταβλητές συμβολίζονται με τη χρήση μικρών γραμμάτων. Οι εκτιμητές θα είναι γενικά γραμμένοι με κεφαλαία γράμματα και θα σημειώνονται με τον συμβολισμό \wedge .

Υπάρχουν πολλά στοχαστικά μοντέλα που μπορούν να χρησιμοποιηθούν για τη στήριξη της μεθόδου chain-ladder. Σε αυτή τη διπλωματική περιλαμβάνονται μόνο τα μοντέλα που παράγουν εκτιμήσεις ισοδύναμες με τη μέθοδο chain-ladder. Αυτά είναι το πολλαπλασιαστικό μοντέλο, το μοντέλο Poisson, το αρνητικό διωνυμικό μοντέλο και το μοντέλο του Mack. Δεδομένου ότι το σύνολο δεδομένων περιείχε αρνητικές απαιτήσεις, μόνο δύο μοντέλα θα χρησιμοποιηθούν στην ανάλυση.

1.4 Ο αλγόριθμος της chain-ladder

Οι προσαυξητικές απαιτήσεις καθορίζονται από το σημείο c_{ij} όπου το i αναφέρεται ως το έτος ατυχήματος και j το έτος εξέλιξης. Ας δηλώσουμε d_{ij} τις σωρευτικές ζημιές. Το έτος ατυχημάτων είναι το έτος όπου συμβαίνει το ατύχημα και το έτος εξέλιξης αντιπροσωπεύει

την καθυστέρηση υποβολής αναφοράς από την ημερομηνία κατά την οποία έλαβε χώρα η ζημιά. Οι σωρευτικές ζημιές d_{ij} είναι:

$$d_{ij} = \sum_{k=1}^j c_{ik} \quad (1.1)$$

Οι παρατηρούμενες απαιτήσεις μπορούν να απεικονιστούν ως τρίγωνο εξέλιξης, όπως απεικονίζεται στο Σχήμα 1.

$$\begin{bmatrix} c_{11} & c_{12} & c_{13} & c_{14} \\ c_{21} & c_{22} & c_{23} & \\ c_{31} & c_{32} & & \\ c_{41} & & & \end{bmatrix} \quad \begin{bmatrix} d_{11} & d_{12} & d_{13} & d_{14} \\ d_{21} & d_{22} & d_{23} & \\ d_{31} & d_{32} & & \\ d_{41} & & & \end{bmatrix}$$

Σχήμα 1: Δύο τρίγωνα εξέλιξης, όπου το αριστερό τρίγωνο εμφανίζει τις παρατηρούμενες προσαυξητικές και το δεξιό τρίγωνο εμφανίζει τις παρατηρούμενες σωρευτικές ζημιές. Οι σειρές εμφανίζουν το έτος ατυχήματος (i) και οι στήλες εμφανίζουν το έτος εξέλιξης (j), όταν $n = 4$. Οι απαιτήσεις στο κάτω αριστερά έως πάνω δεξιά μέρος του τριγώνου είναι γνωστές τιμές. Ο αλγόριθμος της chain-ladder επιδιώκει να εκτιμήσει μελλοντικές απαιτήσεις στο άδειο τμήμα του τριγώνου.

Ο συντελεστής εξέλιξης μπορεί να ορισθεί ως εξής:

$$f_{ij} = \frac{d_{ij}}{d_{i,j-1}} \quad \text{για } 2 \leq j \leq n - i + 1. \quad (1.2)$$

Οι παρατηρούμενες τιμές f_{ij} μπορούν τώρα να παρατηρηθούν στο Σχήμα 2 όπως αυτές παρουσιάζονται για τις προσαυξητικές απαιτήσεις και σωρευτικές ζημιές στο Σχήμα 1. Οι άγνωστες τιμές για f_{ij} θα αφήσουν κενές θέσεις στο κάτω-δεξιά μέρος του τριγώνου. Το Σχήμα 1 έχει διαστάσεις 4×4 , οι οποίες δημιουργούν ένα τρίγωνο f_{ij} με διαστάσεις 3×3 . Θα

πρέπει να σημειωθεί ότι λόγω του ορισμού στο (1.2), η πρώτη στήλη στο τρίγωνο εξέλιξης f_{ij} έχει δείκτη στήλης 2.

$$\begin{bmatrix} f_{12} & f_{13} & f_{14} \\ f_{22} & f_{23} & \\ f_{32} & & \end{bmatrix}$$

Σχήμα 2 : Ένα τρίγωνο εξέλιξης με συντελεστές ανάπτυξης f_{ij} οι οποίοι αντιστοιχούν σε ένα τρίγωνο εξέλιξης ζημιών με διαστάσεις 4×4 .

Η τελική απαίτηση για το έτος ατυχημάτων $i \geq 2$ ορίζεται ως:

$$d_{in} = d_{i,n-i+1} \prod_{j=n-i+2}^n f_{ij} . \quad (1.3)$$

Οι ατομικοί συντελεστές εξέλιξης f_{ij} δεν είναι παρατηρήσιμοι για $j \geq n-i+2$.

Αντιπροσωπεύουν την κάτω αριστερή έως πάνω δεξιά πλευρά του Σχήματος 2. Για να μπορέσουμε να βρούμε την τελική απαίτηση d_{in} πρέπει να εκτιμηθούν οι μη παρατηρήσιμοι μεμονωμένοι παράγοντες. Μια προφανής προσέγγιση θα ήταν να χρησιμοποιηθεί ο μέσος όρος των συντελεστών εξέλιξης που παρατηρήθηκαν στο έτος εξέλιξης j . Αυτό θα παράγει ίδιους ατομικούς συντελεστές εξέλιξης εντός του έτους εξέλιξης j για τα έτη ατυχημάτων στη κάτω αριστερά έως πάνω δεξιά πλευρά του τριγώνου εξέλιξης. Ωστόσο, ο συντελεστής εξέλιξης που χρησιμοποιείται στον αλγόριθμο chain-ladder δεν είναι ένας απλός μέσος όρος των ατομικών συντελεστών εξέλιξης. Είναι μάλλον ένας σταθμισμένος μέσος όρος των παρατηρούμενων ατομικών συντελεστών εξέλιξης

$\hat{f}_j = \sum_{i=1}^{n-j+1} w_{ij} f_{ij}$, όπου w_{ij} δηλώνουν τα βάρη. Επιπλέον, αυτός ο συντελεστής εξέλιξης είναι

συναρτήσεως του έτους εξέλιξης j και επομένως είναι ταυτόσημος με το έτος εξέλιξης j , για όλα τα έτη ατυχημάτων στην κάτω αριστερά έως πάνω δεξιά πλευρά του τριγώνου εξέλιξης. Χρησιμοποιείται ο συμβολισμός \hat{f}_j , όπου το \hat{f}_j θεωρείται εκτιμητής των ατομικών συντελεστών εξέλιξης. Επιλέγοντας την κατάλληλη στάθμιση, καθίσταται σαφές ότι ο συντελεστής εξέλιξης, στη μέθοδο chain-ladder, είναι ένας σταθμισμένος μέσος όρος των ατομικών συντελεστών εξέλιξης. Ο συντελεστής εξέλιξης της chain-ladder είναι:

$$\hat{f}_j = \frac{\sum_{i=1}^{n-j+1} d_{ij}}{\sum_{i=1}^{n-j+1} d_{i,j-1}} = \sum_{i=1}^{n-j+1} \frac{d_{i,j-1}}{\sum_{h=1}^{n-j+1} d_{h,j-1}} \frac{d_{ij}}{d_{i,j-1}} = \sum_{i=1}^{n-j+1} \frac{d_{i,j-1}}{\sum_{h=1}^{n-j+1} d_{h,j-1}} f_{ij} = \sum_{i=1}^{n-j+1} w_{ij} f_{ij} \quad (1.4)$$

Ο συντελεστής εξέλιξης σταθμίζεται με τις αναλογίες των απαιτήσεων κατά το έτος ατυχημάτων i , στο έτος εξέλιξης $j-1$. Οι λόγοι για τη χρήση ενός σταθμισμένου μέσου όρου θα συζητηθούν αργότερα. Δεδομένου ότι ο συντελεστής εξέλιξης της chain-ladder είναι κεντρικός για τα μοντέλα που περιγράφονται σε αυτή τη διπλωματική, επαναλαμβάνεται ότι:

$$\hat{f}_j = \frac{\sum_{i=1}^{n-j+1} d_{ij}}{\sum_{i=1}^{n-j+1} d_{i,j-1}} \quad \text{για } j = 2, \dots, n \quad (1.5)$$

Η τελική απαίτηση είναι η σωρευτική απαίτηση κατά το τελευταίο έτος εξέλιξης. Αυτό φαίνεται στην τελευταία στήλη του τριγώνου εξέλιξης για τις σωρευτικές ζημιές (d_{ij}). Οι τελικές απαιτήσεις μπορούν τώρα να υπολογιστούν στο επόμενο βήμα,

$$\hat{d}_{in} = d_{i,n-i+1} \prod_{j=n-i+2}^n \hat{f}_j \quad \text{για } i = 2, \dots, n \quad (1.6)$$

Η τελευταία απαίτηση $d_{i,n-i+1}$ που έχει παρατηρηθεί χρησιμοποιείται ως βάση για όλες τις μελλοντικές εκτιμήσεις για το έτος ατυχήματος i . Σε κάθε περίπτωση, οι προηγούμενες παρατηρήσεις ισχυρίζονται ότι το έτος ατυχήματος δεν έχει προσθέσει περαιτέρω πληροφορίες για τον υπολογισμό των μελλοντικών απαιτήσεων.

1.5 Η χρήση των στοχαστικών μοντέλων στη μέθοδο της chain-ladder

Η πρωταρχική αδυναμία της μεθόδου chain-ladder είναι ότι είναι ένας ντετερμινιστικός αλγόριθμος, ο οποίος υποδηλώνει ότι τίποτα δεν είναι γνωστό για τη μεταβλητότητα του πραγματικού αποτελέσματος. Για να τροποποιήσουμε αυτό το μειονέκτημα, έχουν αναπτυχθεί στοχαστικά μοντέλα τα οποία παρέχουν τις ίδιες εκτιμήσεις όπως και η μέθοδος chain-ladder. Αυτά τα μοντέλα καθιστούν δυνατή την εύρεση της μεταβλητότητας της εκτίμησης. Ένα στοχαστικό μοντέλο μπορεί επίσης να χρησιμοποιηθεί για να εκτιμήσει κατά πόσο η μέθοδος chain-ladder είναι κατάλληλη για ένα σύνολο δεδομένων. Ωστόσο, είναι σημαντικό να εξεταστεί λεπτομερώς το συγκεκριμένο στοχαστικό μοντέλο που επιλέχθηκε για την ανάλυση, αφού κάθε μοντέλο βασίζεται σε μια σειρά υποθέσεων (Verral 2000).

1.6 Δημιουργία ενός στοχαστικού μοντέλου με τη χρήση της chain-ladder μεθόδου

Αφού η μέθοδος chain-ladder είναι μια ντετερμινιστική μέθοδος, παρουσιάζεται ένα πολύ απλό στοχαστικό μοντέλο που προκύπτει από τη μέθοδο chain-ladder. Ας υποθέσουμε ότι οι απαιτήσεις D_{ij} , για $i=1, \dots, n$ και $j=1, \dots, n$ είναι στοχαστικές μεταβλητές, και επομένως γράφονται με το κεφαλαίο γράμμα D_{ij} . Το αριστερό μέρος του τριγώνου στο σχήμα 1 είναι μια πραγματοποίηση των στοχαστικών μεταβλητών D_{ij} . Η f_j θεωρείται άγνωστη παράμετρος. Θεωρούμε μια γραμμική σχέση μεταξύ των ετών εξέλιξης. Για $2 \leq j \leq n$ η γραμμική σχέση είναι :

$$D_{ij} = D_{i,j-1} f_j \quad (1.7)$$

Με τον υπολογισμό της αναμενόμενης τιμής και για τις δύο πλευρές της εξίσωσης (1.7) προκύπτει η έκφραση

$$E(D_{ij}) = E(D_{i,j-1}) f_j \quad (1.8)$$

Όταν προβλέψουμε την τελική απαίτηση (ή απλώς μια απαίτηση αρκετά χρόνια εξέλιξης μπροστά) μπορεί να χρησιμοποιηθεί ένας τύπος που αντιστοιχεί στη μέθοδο chain-ladder:

$$E(D_{in}) = E(D_{i,n-i+1}) \prod_{j=n-i+2}^n f_j \quad (1.9)$$

Στην εξίσωση (1.9) η αναμενόμενη τιμή μιας προηγούμενης απαίτησης $E(D_{i,n-i+1})$ μπορεί να χρησιμοποιηθεί για να προβλέψει το μέλλον. Η μέθοδος της chain-ladder, ωστόσο, χρησιμοποιεί την τελευταία παρατηρήσιμη απαίτηση $d_{i,n-i+1}$ και όχι την αναμενόμενη τιμή

αυτής. Το μοντέλο της chain-ladder υποθέτει ότι η τελευταία παρατήρηση είναι πιο σχετική από την αναμενόμενη τιμή της και ένα στοχαστικό μοντέλο ισοδύναμο με τη μέθοδο chain-ladder μπορεί να ορισθεί στην τελευταία παρατηρούμενη απαίτηση. Έστω $d_{i,j-1}$ είναι η τελευταία παρατηρήσιμη απαίτηση. Εάν το $d_{i,j-1}$ εφαρμοστεί στην (1.8) προκύπτει η έκφραση:

$$E(D_{ij} | d_{i,j-1}) = d_{i,j-1} f_j \quad (1.10)$$

Δεν έχει γίνει ακόμα καμία παραδοχή σχετικά με την κατανομή του μοντέλου D_{ij} . Το μοντέλο που παρουσιάζεται στον τύπο (1.10) είναι ένα απλό στοχαστικό μοντέλο του αλγόριθμου chain-ladder (Mack 1994b).

Παρουσιάζοντας τις στοχαστικές μεταβλητές D_{ij} , πρέπει να γίνουν κάποιες επισημάνσεις. Αυτό θα συμβεί αργότερα όταν θα γίνει η εισαγωγή των στοχαστικών μοντέλων. Το τρίγωνο εξέλιξης για D_{ij} μπορεί να παρουσιαστεί με τις στοχαστικές μεταβλητές D_{ij} για $1 \leq i, j \leq n$. Στην πραγματικότητα δεν είναι ένα τρίγωνο, καθώς παρουσιάζονται και οι κενές θέσεις στο κάτω μέρος του τριγώνου.

$$\begin{bmatrix} D_{11} & D_{12} & D_{13} & D_{14} \\ D_{21} & D_{22} & D_{23} & D_{24} \\ D_{31} & D_{32} & D_{33} & D_{34} \\ D_{41} & D_{42} & D_{43} & D_{44} \end{bmatrix}$$

Σχήμα 3: Τρίγωνο εξέλιξης των σωρευτικών ζημιών ως στοχαστικές μεταβλητές, όταν $n = 4$.

Οι σειρές εμφανίζουν το έτος ατυχήματος (i) και οι στήλες εμφανίζουν το έτος εξέλιξης (j)

Προκειμένου να διευκολυνθεί η εύρεση των αναμενόμενων τιμών, εισάγονται οι μεταβλητές K_{ij} , K_j και K . Έστω k είναι η πραγματοποίηση της στοχαστικής μεταβλητής K και $k = \{d_{ij}, i=1, \dots, n, j=1, \dots, n-i+1\}$. K είναι η πληροφορία των σωρευτικών ζημιών στη πάνω-αριστερή γωνία του τριγώνου εξέλιξης. Έστω k_{ij} είναι η πραγματοποίηση της στοχαστικής μεταβλητής K_{ij} και $k_{ij} = \{d_{i1}, \dots, d_{ij}\}$ για το έτος ατυχημάτων $i=1, \dots, n$. Έστω k_j είναι η πραγματοποίηση της στοχαστικής μεταβλητής K_j και $k_j = \{d_{i1}, \dots, d_{ij}, i=1, \dots, n\}$.

1.7 Στοχαστικά chain-ladder μοντέλα

Όταν βρίσκουμε ένα στοχαστικό μοντέλο που αναπαράγει εκτιμήσεις της chain-ladder, πρέπει να γίνουν κάποιες υποθέσεις σχετικά με τις ασφαλιστικές απαιτήσεις. Επομένως, είτε πρέπει να διευκρινιστεί η κατανομή των ασφαλιστικών απαιτήσεων, είτε να αναφερθούν απλώς οι δύο πρώτες ροπές (Verrall & England 2002).

Η κατανομή Poisson μπορεί να είναι κατάλληλη όταν τα συμβάντα πρέπει να μετρηθούν κατά τη διάρκεια ενός διαστήματος. Κατά τη διάρκεια μιας περιόδου ασφάλισης συμβαίνουν ατυχήματα και γίνονται ζημιές. Αρκετοί συγγραφείς προτείνουν ένα μοντέλο Poisson σε αυτή την περίπτωση (Hess & Schmidt 2002, Renshaw 1998, Verrall 2000). Άλλες κατανομές συνδέονται στενά με την κατανομή Poisson, την πολλαπλασιαστική κατανομή και το μοντέλο του Mack (Verrall & England 2002). Σε αντίθεση με την Poisson και το αρνητικό διωνυμικό μοντέλο, το πολλαπλασιαστικό μοντέλο και το μοντέλο του Mack καθορίζουν μόνο τις δύο πρώτες ροπές.

1.8 Ντετερμινιστικοί Μέθοδοι

1.8.1 Η μέθοδος Chain- ladder

Η μέθοδος chain-ladder ή μέθοδος τριγωνικής εξέλιξης ζημιών πιθανότατα είναι η πιο δημοφιλής μέθοδος αποθεματοποίησης που χρησιμοποιείται από τους αναλογιστές για την εκτίμηση των αποθεμάτων. Η μέθοδος αυτή ξεκίνησε ως ντετερμινιστική, όμως προκειμένου να εκτιμήσει την μεταβλητότητα του αποτελέσματος που απέδιδε, εξελίχθηκε σε στοχαστική μέθοδο. Ο Taylor (2000) παρουσιάζει μεθόδους που παράγονται από την chain-ladder · η μία είναι ντετερμινιστική, ενώ μια άλλη βασίζεται στην υπόθεση πως οι προσαυξητικές απαιτήσεις (incremental claims) ακολουθούν την κατανομή Poisson. Παράλληλα, ο Verall (2000) παρουσιάζει μοντέλα που δίνουν ίδια αποτελέσματα με αυτά της μεθόδου chain ladder.

Αρχικά, θα ορίσουμε τις προσαυξητικές απαιτήσεις ως C_{ij} όπου είναι οι ζημιές που πραγματικά πλήρωσε η ασφαλιστική εταιρία μέσα στο έτος εξέλιξης j , όπου το i αντιπροσωπεύει το έτος ατυχήματος ή το ασφαλιστικό έτος και το j αντιπροσωπεύει το έτος εξέλιξης, δηλαδή το έτος μετά από το έτος που πραγματοποιήθηκε το ατύχημα. Τα δεδομένα μας εισάγονται σε ένα τρίγωνο εξέλιξης ζημιών, στο οποίο κάθε γραμμή του τριγώνου παριστάνει το έτος ατυχήματος (accident year) και κάθε στήλη το έτος εξέλιξης (development year). Το τρίγωνο εξέλιξης έχει την παρακάτω μορφή:

	Έτος εξέλιξης					
Έτος ατυχήματος	1	2	3	...	n-1	m
1	C_{11}	C_{12}	C_{13}	...	$C_{1,n-1}$	$C_{1,n}$
2	C_{21}	C_{22}	C_{23}	...	$C_{2,n-1}$	
..		
...		
n-1	$C_{n-1,1}$	$C_{n-1,2}$				
n	$C_{n,1}$					

Πίνακας 1: Τρίγωνο εξέλιξης ζημιών-Incremental data

Στη συνέχεια, ορίζουμε τις σωρευτικές ζημιές ή πληρωμές (cumulative claims) για το έτος ατυχήματος i και για το έτος εξέλιξης j , ορίζονται ως D_{ij} , και είναι οι ζημιές που πληρώθηκαν μέχρι το έτος εξέλιξης j , όπου:

$$D_{ij} = \sum_{k=1}^j C_{ik}, \quad i = 1, \dots, n \quad (1.11)$$

Το τρίγωνο εξέλιξης με τις σωρευτικές ζημιές έχει την παρακάτω μορφή:

		Έτος εξέλιξης					
Έτος ατυχήματος	1	2	3	...	n-1	n	
1	D_{11}	D_{12}	D_{13}	...	$D_{1,n-1}$	$D_{1,n}$	
2	D_{21}	D_{22}	D_{23}	...	$D_{2,n-1}$		
..			
				
n-1	$D_{n-1,1}$	$D_{n-1,2}$					
n	$D_{n,1}$						

Πίνακας 2: Τρίγωνο εξέλιξης ζημιών-Cumulative data

Τα στοιχεία της διαγωνίου (κάτω αριστερά έως πάνω δεξιά) αντιπροσωπεύουν τα ποσά που πλήρωσε η ασφαλιστική εταιρία σε αποζημιώσεις κάθε έτος. Στόχος της μεθόδου είναι να εκτιμήσει τις μελλοντικές ζημιές που θα παρουσιαστούν στην ασφαλιστική εταιρία. Τελικά όλες οι ζημιές για την συγκεκριμένη χρονική περίοδο θα πληρωθούν ή θα διακανονιστούν, αλλά είναι άγνωστο πόσα χρόνια θα χρειαστούν για να γίνει αυτό.

Για να γίνει αυτό εκτιμούμε τους ατομικούς συντελεστές εξέλιξης ως το πηλίκο δύο διαδοχικών σωρευτικών απαιτήσεων, δηλαδή παίρνουμε:

$$f_{ij} = \frac{D_{ij}}{D_{i,j-1}}, \quad j = 2, \dots, n \quad (1.12)$$

Οι ατομικοί συντελεστές εξέλιξης παρουσιάζονται στον παρακάτω πίνακα (Ατομικοί συντελεστές εξέλιξης) και δουλεύοντας όπως παραπάνω προσπαθούμε να εκτιμήσουμε τα κενά στοιχεία του τριγώνου εξέλιξης, να υπολογίσουμε δηλαδή τους μελλοντικούς δείκτες εξέλιξης.

Ατομικοί Συντελεστές Εξέλιξης

$$\begin{array}{ccc} f_{12} & f_{13} & f_{14} \\ f_{22} & f_{23} & \\ f_{32} & & \end{array}$$

Έπειτα, εκτιμούμε τον συντελεστή εξέλιξης chain-ladder (chain-ladder development factor)

\hat{f}_j ως εξής:

$$\hat{f}_j = \frac{\sum_{i=1}^{n-j+1} D_{ij}}{\sum_{i=1}^{n-j+1} D_{i,j-1}}, j = 2, \dots, n \quad (1.13)$$

Από την σχέση (1.13) συμπεραίνουμε πως ο συντελεστή εξέλιξης chain-ladder προκύπτει ως το πηλίκο της διαίρεσης δύο διαδοχικών ποσών ζημιών .

Έπειτα, οι συντελεστές της σχέσης (1.13), πολλαπλασιάζονται με τις σωρευτικές ζημίες κάθε γραμμής $D_{i,n-i+1}$ προκειμένου να πάρουμε τις μελλοντικές εκτιμώμενες αποζημιώσεις σύμφωνα με την παρακάτω σχέση :

$$\hat{D}_{i,n} = D_{i,n-i+1} \prod_{j=n-i+2}^n \hat{f}_j, i = 2, \dots, n \quad (1.14)$$

Οι σχέσεις (1.13) και (1.14) αποτελούν τη βάση για το μοντέλο chain-ladder. Η τελευταία σωρευτική ζημία $D_{i,n-i+1}$ χρησιμοποιείται ως βάση για όλες τις μελλοντικές εκτιμήσεις για το έτος ατυχήματος i , από το οποίο άμεσα προκύπτει πως από τα προηγούμενα έτη ατυχήματος δεν αντλούμε πληροφορίες για την εκτίμηση των μελλοντικών αποζημιώσεων.

Τέλος, για να εκτιμήσουμε το απόθεμα που πρέπει να κρατάει η ασφαλιστική εταιρία ανά έτος ατυχήματος, θα πρέπει να αφαιρέσουμε τις πληρωθείσες ζημιές από τις τελικές εκτιμήσεις των ζημιών. Δηλαδή, παίρνουμε :

Η εκτίμηση αποθέματος ζημιών είναι ίση με τη διαφορά των εκτιμώμενων τελικών ζημιών και των πληρωθεισών ζημιών , δηλαδή

$$\hat{R}_i = \hat{D}_{i,n} - D_{i,n-i+1}, i = 2, \dots, n, \hat{R} = \sum_{i=2}^n \hat{R}_i$$

1.8.2 Προσαρμοσμένες τιμές

Εφαρμόζοντας τους ατομικούς συντελεστές εξέλιξης f δουλεύοντας αναδρομικά προς τα πίσω και ξεκινώντας από τα στοιχεία της τελευταίας διαγωνίου, μπορούμε να πάρουμε τις προσαρμοσμένες τιμές (fitted values) των σωρευτικών πληρωθεισών ζημιών των ετών του παρελθόντος, οι οποίες για $i = 2, \dots, n$ δίνονται από:

$$\hat{D}_{i,k-1} = \frac{\hat{D}_{ik}}{f_k}$$

Ακόμα, οι εκτιμώμενες προσαυξητικές πληρωθείσες ζημιές (estimated incremental payments) προκύπτουν από τις διαφορές:

$$\hat{C}_{ij} = \hat{D}_{ij} - \hat{D}_{i,j-1}, i = 2, \dots, n, j = 2, \dots, n$$

$$\hat{C}_{i1} = \hat{D}_{i1}, i = 2, \dots, n$$

1.8.3 Η μέθοδος Bornhuetter-Ferguson

Η μέθοδος Bornhuetter-Ferguson χρησιμοποιεί τους δείκτες εξέλιξης που χρησιμοποιήσαμε και στη μέθοδο Chain Ladder, σε συνδυασμό με τον δείκτη ζημιών. Η μέθοδος αυτή είναι ιδιαίτερα χρήσιμη σε περιπτώσεις όπου οι τελικές πληρωθείσες ζημιές στα αρχικά έτη παρουσιάζουν ανομοιομορφία, γεγονός που καθιστά την μέθοδο chain-ladder αναποτελεσματική στην παραγωγή ασφαλών εκτιμήσεων για τις μελλοντικές πληρωθείσες ζημιές. Ο σκοπός της μεθόδου είναι να σταθεροποιήσει αυτά τα μη ικανοποιητικά αποτελέσματα, χρησιμοποιώντας τις τελικές εκτιμώμενες ζημιές κάθε έτους, (οι οποίες προκύπτουν σαν το γινόμενο του δείκτη ζημιών επί το δεδουλευμένο ασφάλιστρο), σε συνδυασμό με τους συντελεστές εξέλιξης της μεθόδου chain-ladder προκειμένου να εκτιμήσει τα αποθέματα ζημιών. Οι εκτιμώμενες τελικές ζημιές UBF κάθε έτους θα μπορούσαν να θεωρηθούν σαν χρήση εξωτερικής πληροφορίας για την εκτίμηση των αποθεμάτων, γεγονός που κάνει τη μέθοδο Bornhuetter-Ferguson να θεωρείται από αρκετούς πως ακολουθεί τη Bayesian μεθοδολογία.

Συνοπτικά, η μέθοδος Bornhuetter-Ferguson ακολουθεί τα εξής βήματα για την εκτίμηση των αποθεμάτων ζημιών :

- ▶ Αρχικά, υπολογίζουμε τις εκτιμώμενες τελικές ζημιές κάθε έτους ατυχήματος ως το γινόμενο του Αναμενόμενου Δείκτη Ζημιών επί το Δεδουλευμένο Ασφάλιστρο.

Εκτιμώμενες τελικές ζημιές = Αναμενόμενος Δείκτης ζημιών · Δεδουλευμένο Ασφάλιστρο

- ▶ Στη συνέχεια, έχοντας υπολογίσει τους συντελεστές εξέλιξης chain ladder και σε συνδυασμό με τις εκτιμώμενες τελικές ζημιές που υπολογίσαμε παραπάνω, παίρνουμε τα εκτιμώμενα αποθέματα ζημιών.

$$\text{Εκτιμώμενα αποθέματα ζημιών} = \left(U_i^{(BF)} \right) \left(1 - \frac{1}{f_{ult}} \right) \quad (1.15)$$

Όπου $U_i^{(BF)}$ ισούται με τις εκτιμώμενες τελικές ζημιές .

Σύμφωνα με τη μέθοδο chain ladder υπολογίσαμε το εκτιμώμενο απόθεμα ζημιών ως εξής:

$$U_i^{(CL)} - D_{i,n-i+1} = U_i^{(CL)} \frac{1}{f_{n-i+2} f_{n-i+3} \dots f_n} (f_{n-i+2} f_{n-i+3} \dots f_n - 1) = U_i^{(CL)} \left(1 - \frac{1}{f_{n-i+2} f_{n-i+3} \dots f_n} \right)$$

Όπου $U_i^{(CL)}$ είναι οι εκτιμώμενες τελικές ζημιές που υπολογίσαμε με τη μέθοδο την chain-ladder. Συνεπώς:

$$U_i^{(CL)} = D_{i,n-i+1} \prod_{j=n-i+2}^n f_j .$$

Ενώ με την μέθοδο Bornhuetter-Ferguson βρήκαμε πως το εκτιμώμενο απόθεμα ζημιών υπολογίζεται ως εξής:

$$U_i^{(BF)} \left(1 - \frac{1}{\prod_{j=n-i+2}^n f_j} \right) .$$

Οι εκτιμώμενες τελικές ζημιές $U_i^{(BF)}$ προέκυψαν από τη χρήση κάποια εξωτερικής πληροφορίας.

Από τα παραπάνω καταλήγουμε στο συμπέρασμα πως η μόνη διαφορά μεταξύ των δύο μεθόδων είναι πως στη μέθοδο Bornhuetter-Ferguson προκειμένου να εκτιμήσουμε το απόθεμα ζημιών, χρησιμοποιούμε τις εκτιμώμενες τελικές ζημιές $U_i^{(BF)}$, χρησιμοποιούμε δηλαδή μια εξωτερική πληροφορία. (από σχέσεις (1.14) και (1.15)).

1.8.4 Μέθοδος Διαχωρισμού

Η Separation Method ξεκίνησε, όπως και η μέθοδος chain-ladder, χωρίς να κάνει κάποια υπόθεση για την κατανομή των δεδομένων (ζημιών). Σε αυτή τη μέθοδο υποθέτουμε πως οι προσαυξητικές απαιτήσεις (incremental claims) προκύπτουν από συντελεστές οι οποίοι εξαρτώνται από το έτος συμβάντος, το έτος εξέλιξης και το ημερολογιακό έτος. Ορίζουμε λοιπόν ως c το μέσο ποσό ζημίας, ως r_j ορίζουμε το αναμενόμενο ποσοστό ζημιών στο έτος εξέλιξης j (απαιτώντας $\sum_j r_j = 1$), ως n_i ορίζουμε τον αριθμό ζημιών που παρατηρήθηκαν στο έτος συμβάντος i και τέλος ως λ_{i+j} ορίζουμε την επίδραση κάποιου παράγοντα (π.χ. του πληθωρισμού) στο ημερολογιακό έτος.

Συνεπώς, παίρνουμε:

$$E(D_{ij}) = c r_j n_i \lambda_{i+j} \quad (1.16)$$

$$\frac{E(D_{ij})}{n_i} = c r_j \lambda_{i+j} \quad (1.17)$$

όπου τα D_{ij} παριστάνουν τις αθροιστικές ζημιές για το έτος συμβάντος i και για το έτος εξέλιξης j . Η παραπάνω σχέση μπορεί να θεωρηθεί και ως η τιμή του στοιχείου (i, j) ενός τριγωνικού πίνακα.

Έτος εξέλιξης					
Έτος ατυχήματος	0	1	...	k-1	k
0	$c r_0 \lambda_0$	$c r_1 \lambda_1$		$c r_{k-1} \lambda_{k-1}$	$c r_k \lambda_k$
1	$c r_0 \lambda_1$	$c r_1 \lambda_2$		$c r_{k-1} \lambda_k$	
.		
.		
.			
k-1	$c r_0 \lambda_{k-1}$	$c r_1 \lambda_k$			
k	$c r_0 \lambda_k$				

Πίνακας 3: Τρίγωνο Εξέλιξης

Για αυτόν το τριγωνικό πίνακα τα αθροίσματα των διαγωνίων είναι:

$$d_0 = c r_0 \lambda_0$$

$$d_1 = c r_0 \lambda_1 + c r_1 \lambda_1 = c \lambda_1 (r_0 + r_1)$$

$$d_2 = c r_0 \lambda_2 + c r_1 \lambda_2 + c r_2 \lambda_2 = c \lambda_2 (r_0 + r_1 + r_2)$$

....

$$d_k = c \lambda_k (r_0 + r_1 + r_2 + \dots + r_k) = c \lambda_k$$

Στη συνέχεια, ορίζουμε την ποσότητα $B_{ij} = D_{ij} / \hat{n}_{ij}$ όπου D_{ij} είναι οι ζημιές που έχουμε παρατηρήσει (δηλαδή οι αποζημιώσεις που χρειάζεται να πληρώσουμε) και \hat{n}_{ij} ο εκτιμώμενος αριθμός ζημιών. Ξεκινώντας από την τελευταία σχέση και δουλεύοντας αναδρομικά, υπολογίζουμε τα εκτιμώμενα c , $\hat{\lambda}_{i+j}$ και \hat{r}_j .

$$c \hat{\lambda}_k = \hat{d}_k$$

$$\hat{r}_k = B_{1k} / (c \hat{\lambda}_k)$$

$$c \hat{\lambda}_{k-1} = \hat{d}_{k-1} / (1 - \hat{r}_k)$$

$$\hat{r}_{k-1} = (B_{1k-1} + B_{2k-1}) / (c \hat{\lambda}_k + c \hat{\lambda}_{k-1})$$

$$c \hat{\lambda}_{k-2} = \hat{d}_{k-2} / (1 - \hat{r}_k - \hat{r}_{k-1})$$

$$\hat{r}_{k-2} = (B_{1k-2} + B_{2k-2} + B_{3k-2}) / (c \hat{\lambda}_k + c \hat{\lambda}_{k-1} + c \hat{\lambda}_{k-2})$$

Συνεπώς, το εκτιμώμενο απόθεμα είναι το άθροισμα των εκτιμώμενων ζημιών \hat{D}_{ij} .

Ουσιαστικά, με την Separation Method κάνουμε την υπόθεση πως σε κάθε έτος ατυχήματος η ασφαλιστική εταιρία πληρώνει ένα σταθερό ποσοστό των συνολικών ζημιών και αυτή η διαδικασία επαναλαμβάνεται.

1.8.5 Η μέθοδος Cape-Cod

Μια ακόμα μέθοδος αποθεματοποίησης που έχει αναπτυχθεί στη βιβλιογραφία είναι η μέθοδος Cape-Cod. Η μέθοδος αυτή έχει αρκετά κοινά χαρακτηριστικά με την μέθοδο Bornhuetter-Ferguson. Αντί να προϋποθέτει την ύπαρξη ενός συντελεστή loss ratio, εκτιμά έναν χρησιμοποιώντας το ύψος των τελικών ζημιών και ένα μέτρο έκθεσης στον κίνδυνο. Ο συντελεστής αυτός ονομάζεται Cape-Cod συντελεστής ζημιών και αποτελεί έναν εκτιμητή του αναμενόμενου τελευταίου συντελεστή ζημιών, ο οποίος παραμένει σταθερός σε όλα τα χρόνια ατυχήματος. Ο συντελεστής αυτός ερμηνεύεται σαν ένας σταθμισμένος μέσος όρος των τελικών ζημιών για κάθε έτος ατυχήματος

$$\hat{L} = \frac{\sum_{i=1}^n \{(D_{i,n-i+1} / F_{n-i+1}) (P_i / F_{n-i+1})\}}{\sum_{i=1}^n P_i / F_{n-i+1}} \quad (1.18)$$

όπου :

- \hat{L} είναι ο αναμενόμενος συντελεστής ζημιών με την μέθοδο Cape-Cod
- $D_{i,n-i+1}$ είναι οι πληρωθείσες ζημιές
- F_{n-i+1} είναι οι δείκτες εξέλιξης από το έτος ατυχήματος i στο συνολικό (ultimate)

-
- P_i είναι ένα μέτρο έκθεσης στον κίνδυνο που είναι γνωστό (ασφάλιστρα ή κάποια άλλα μέτρα έκθεσης στον κίνδυνο)

Παρατηρούμε ότι τα βάρη που χρησιμοποιούνται για την εκτίμηση του συντελεστή Cap-Cod είναι ανάλογα προς τον αριθμό των ζημιών (exposure) και αντιστρόφως ανάλογα προς τον δείκτη εξέλιξης. Αυτό γίνεται, διότι, δίνεται περισσότερο βάρος στα έτη με μεγαλύτερο αριθμό ζημιών.

Το εκτιμώμενο απόθεμα των IBNR-ζημιών δίνεται από την σχέση (1.19) χρησιμοποιώντας τη μέθοδο Bornhuetter-Ferguson, μόνο που σε αυτή τη μέθοδο ο δείκτης ζημιών προέρχεται από ζημιές που έχουν υποστεί διόρθωση στην τάση τους. Συνεπώς, το εκτιμώμενο απόθεμα των IBNR-ζημιών είναι:

$$\sum_{i=1}^n \hat{R}_i = L \sum_{i=1}^n \left(1 - \frac{1}{F_{m-i+1}} \right) P_i \quad (1.19)$$

ΚΕΦΑΛΑΙΟ 2 : Στοχαστικά μοντέλα

Ο αριθμός των απαιτήσεων είναι πολύ σημαντικός. Μια εισαγωγή στον αριθμό της ζημιάς και στη διαδικασία του μεγέθους της απαίτησης, εισάγεται σε αυτό το κεφάλαιο. Μερικοί τύποι που σχετίζονται με αυτές τις διαδικασίες παρουσιάζονται και θα χρησιμοποιηθούν αργότερα σε αυτή τη διπλωματική.

Όταν παρουσιάσουμε τα στοχαστικά μοντέλα, ο στόχος είναι να δείξουμε ότι παρέχουν πράγματι τις ίδιες εκτιμήσεις με τη μέθοδο της chain-ladder. Υπάρχει επίσης μια στενή σχέση μεταξύ των μοντέλων, την οποία θα αποδείξουμε. Πρώτο παρουσιάζεται το πολλαπλασιαστικό μοντέλο. Μόνο η πρώτη ροπή, η οποία έχει πολλαπλασιαστική δομή, καθορίζεται στο μοντέλο. Μερικά από τα μοντέλα που θα παρουσιαστούν αργότερα μπορούν να θεωρηθούν ως ειδικές περιπτώσεις του πολλαπλασιαστικού μοντέλου. Επίσης, ένας εναλλακτικός τρόπος έκφρασης του συντελεστή εξέλιξης chain-ladder προκύπτει από το πολλαπλασιαστικό μοντέλο και θα παρουσιαστεί.

Το μοντέλο Poisson είναι μια ειδική περίπτωση του πολλαπλασιαστικού μοντέλου. Έχει την ίδια πολλαπλασιαστική δομή στην πρώτη ροπή. Χρησιμοποιώντας τον εκτιμητή μέγιστης πιθανοφάνειας δημιουργείται ο ίδιος συντελεστής εξέλιξης με τον συντελεστή εξέλιξης της chain-ladder και αυτό θα αποδειχθεί.

Το μοντέλο του Mack είναι το τελευταίο μοντέλο που θα παρουσιαστεί. Οι υποθέσεις του Mack δηλώνουν ότι η πρώτη ροπή είναι ισοδύναμη με την εκτίμηση της chain-ladder, οπότε η σχέση μεταξύ του στοχαστικού μοντέλου και της μεθόδου chain-ladder είναι ασήμαντη. Στην προσπάθεια να κατανοήσουμε περαιτέρω το μοντέλο του Mack, διερευνώνται οι λόγοι πίσω από τις υποθέσεις.

2.1 Ο αριθμός απαιτήσεων και η διαδικασία μεγέθους απαιτήσεων

Η προσαυξητική απαίτηση C_{ij} ή η σωρευτική απαίτηση D_{ij} δεν έχουν καθοριστεί περαιτέρω. Μπορεί να αντιπροσωπεύει τον αριθμό των απαιτήσεων που έχει λάβει μία ασφαλιστική εταιρεία ή μπορεί να είναι το συνολικό μέγεθος που χρησιμοποιήθηκε για να παρουσιαστούν οι ασφαλιστικές απαιτήσεις. Το συνολικό μέγεθος των απαιτήσεων είναι επίσης μια συνάρτηση του αριθμού των απαιτήσεων, ο οποίος εισάγει τη σύνθετη κατανομή Poisson. Έστω, $N(t)$, είναι ο αριθμός των απαιτήσεων, που είναι μια κατανεμημένη μεταβλητή Poisson, και είναι μια συνεχής συνάρτηση του χρόνου t . Το $N(t)$ υπολογίζει τον αριθμό των απαιτήσεων στο διάστημα $(0, t]$. Το $N(t)$ αυξάνεται σταδιακά και είναι μια αύξουσα συνάρτηση συναρτήσει του χρόνου t . Έστω, Y_k , το μέγεθος του αριθμού των απαιτήσεων k . Το συνολικό μέγεθος, $X(t)$, των απαιτήσεων $N(t)$ μέχρι το χρόνο t είναι:

$$X(t) = \sum_{k=1}^{N(t)} Y_k \quad (2.1)$$

Αν Y_k είναι ανεξάρτητη και κατανεμημένη με τον ίδιο τρόπο, τότε $X(t)$ ακολουθεί μια κατανομή σύνθετης Poisson. Η αναμενόμενη τιμή και η διακύμανση μπορούν να βρεθούν υπολογίζοντας τη διπλή αναμενόμενη τιμή:

$$E_x(X(t)) = E_N E_x(X(t) | N(t)) = E_N(N(t) E_Y(Y_k)) = E_N(N(t)) E_Y(Y_k) \quad (2.2)$$

Και η διπλή διακύμανση:

$$\begin{aligned} Var_x(X(t)) &= E_N(Var_x(X(t) | N(t))) + Var_N(E_x(X(t) | N(t))) \\ &= E_N(N(t) Var_Y(Y_k)) + Var_N(N(t) E_Y(Y_k)) \end{aligned} \quad (2.3)$$

$$= E_N(N(t))Var_Y(Y_k) + (E_Y(Y_k))^2 Var_N(N(t))$$

Δεδομένου ότι $N(t)$ και $X(t)$ είναι συνεχείς συναρτήσεις του χρόνου t , C_{ij} ή D_{ij} είναι οι μετρήσεις των δύο αυτών μεγεθών σε μια συγκεκριμένη στιγμή. Ο καθορισμός του χρόνου ατυχήματος C_{ij} και D_{ij} μεταβάλλεται κατά τη διάρκεια των ετών εξέλιξης. Μπορούν να είναι είτε ένα μέτρο του αριθμού των απαιτήσεων ή του συνολικού μεγέθους των απαιτήσεων. Για μεταγενέστερους σκοπούς, όταν C_{ij} ή D_{ij} αντιπροσωπεύει το συνολικό μέγεθος των απαιτήσεων, θα αναφέρεται ως το μέγεθος των απαιτήσεων και όχι ως το συνολικό μέγεθος των απαιτήσεων.

2.2 Το πολλαπλασιαστικό μοντέλο και η chain-ladder μέθοδος

Το πολλαπλασιαστικό μοντέλο μπορεί να θεωρηθεί ως υποκείμενο τόσο του μοντέλου του Mack, όσο και του μοντέλου Poisson. Το πολλαπλασιαστικό μοντέλο παρουσιάζεται παρακάτω, όπου διευκρινίζεται η σύνδεση με τη μέθοδο chain-ladder. Σε αυτό το κεφάλαιο θα χρησιμοποιηθούν τα σύμβολα x_i και y_j . Αυτές είναι παράμετροι στο πολλαπλασιαστικό μοντέλο (και όχι υλοποιήσεις $X(t)$ και Y_k , οι οποίες εισήχθησαν στο προηγούμενο κεφάλαιο).

Το πολλαπλασιαστικό μοντέλο ορίζεται από την πρώτη ροπή, και για $1 \leq i, j \leq n$, είναι:

$$E(C_{ij}) = x_i y_j, \quad (2.4)$$

Όπου C_{ij} είναι μια στοχαστική μεταβλητή, x_i και y_j είναι άγνωστες παράμετροι και $y_1 + y_2 + \dots + y_n = 1$.

Με τον ορισμό στο (2.4) και την ιδιότητα ότι το άθροισμα y_i ισούται με ένα, προκύπτει ότι $x_i = E(D_{in})$. Η (2.4) αναφέρει ότι η αναμενόμενη τιμή της προσαυξητικής απαίτησης μπορεί να γραφεί ως προϊόν μιας εξαρτώμενης παραμέτρου από το έτος ατυχήματος x_i και μιας παραμέτρου που εξαρτάται από το έτος εξέλιξης y_i . Δεδομένου ότι, x_i είναι η αναμενόμενη τελική απαίτηση, είναι λογικό ότι το άθροισμα y_i να είναι ίσο με ένα. Εάν C_{ij} αντιπροσωπεύει τον αριθμό των απαιτήσεων, y_i είναι η πιθανότητα ότι μια απαίτηση που προκύπτει κατά το έτος ατυχημάτων i , αναφέρεται στο έτος εξέλιξης j . Η ερμηνεία αυτή παρουσιάζει σιωπηρώς έναν άλλο περιορισμό $y_i, y_i \geq 0$ για $j = 1, \dots, n$.

Ένα πολύ απλό στοχαστικό μοντέλο της chain-ladder προέκυψε στο (1.8). Ο Mack (1994) δήλωσε ότι αυτό ήταν ισοδύναμο με το πολλαπλασιαστικό μοντέλο. Αυτό μπορεί να αποδειχθεί με την εύρεση κατάλληλων υποψηφίων για x_i και y_i .

Με τη χρήση (1.9) μπορεί να γραφτεί η αναμενόμενη τιμή της προσαυξητικής απαίτησης:

$$\begin{aligned}
 E(C_{ij}) &= E(D_{ij}) - E(D_{i,j-1}) & (2.5) \\
 &= (f_{j+1}f_{j+2}\dots f_n)^{-1} E(D_{in}) - (f_j f_{j+1} \dots f_n)^{-1} E(D_{in}) \\
 &= E(D_{in}) \left((f_{j+1}f_{j+2}\dots f_n)^{-1} - (f_j f_{j+1} \dots f_n)^{-1} \right)
 \end{aligned}$$

Το επόμενο βήμα είναι να αναγνωρίσουμε ποιες πρέπει να είναι οι μεταβλητές y_i , ώστε να ισούται $x_i y_j$ (2.5). Η μεταβλητή x_i έχει ήδη αναγνωρισθεί, $x_i = E(D_{in})$ και είναι σαφές

$y_j = (f_{j+1}f_{j+2}\dots f_n)^{-1} - (f_jf_{j+1}\dots f_n)^{-1}$. Για το έτος εξέλιξης, $2 \leq j \leq n$, οι μεταβλητές y_j

είναι:

$$y_1 = (f_1f_2\dots f_n)^{-1} \quad (2.6)$$

$$y_j = (f_{j+1}f_{j+2}\dots f_n)^{-1} - (f_jf_{j+1}\dots f_n)^{-1}$$

$$y_n = 1 - (f_n)^{-1}.$$

Εάν οι πρόσφατα καθορισμένες μεταβλητές y_i πληρούν τον περιορισμό $\sum_{j=1}^n y_j = 1$, μπορούν

να γίνουν αποδεκτές. Συνοψίζοντας τους όρους στο (2.6), αποκαλύπτεται μια τηλεσκοπική

σειρά και χρησιμοποιώντας αυτή την ιδιότητα είναι σαφές ότι $\sum_{j=1}^n y_j = 1$. Επιπροσθέτως,

$y_j \geq 0$, αν $f_j \geq 1$ για $j = 1, \dots, n$. Αυτός ο ορισμός φαίνεται να είναι μια καλή επιλογή. Η

σωρευτική απαίτηση στο έτος i και το έτος εξέλιξης j μπορεί να γραφτεί ως άθροισμα από

τις προσαυξητικές απαιτήσεις, και χρησιμοποιώντας τον περιορισμό y_j , μπορεί να

διαπιστώσει κανείς ότι για το έτος ατυχήματος $i = 2, \dots, n$:

$$E(D_{in}) = x_i (y_1 + \dots + y_n) \quad (2.7)$$

$$= x_i y_1 + x_i y_2 + \dots + x_i y_n$$

$$= E(C_{i1}) + \dots + E(C_{in}).$$

Επιλέγοντας σωστά x_i και y_j , είναι ξεκάθαρο ότι το απλό στοχαστικό μοντέλο από το κεφάλαιο 1.4 είναι ισοδύναμο με το πολλαπλασιαστικό μοντέλο (Mack 1994b).

Ο συντελεστής εξέλιξης μπορεί να ορισθεί γράφοντας την έκφραση (1.8) και χρησιμοποιώντας τις ταυτότητες από το πολλαπλασιαστικό μοντέλο. Για $2 \leq j \leq n$, η έκφραση είναι:

$$\begin{aligned} f_j &= \frac{E(D_{ij})}{E(D_{i,j-1})} & (2.8) \\ &= \frac{x_i (y_1 + y_2 + \dots + y_j)}{x_i (y_1 + y_2 + \dots + y_{j-1})} \\ &= \frac{y_1 + y_2 + \dots + y_j}{y_1 + y_2 + \dots + y_{j-1}}. \end{aligned}$$

Αυτός ο συντελεστής εξέλιξης δεν έχει την ίδια εμφάνιση με τον συντελεστή εξέλιξης της chain-ladder, αλλά είναι ο ίδιος. Αυτό μπορεί να αποδειχθεί με επαγωγή.

2.3 Το μοντέλο Poisson και η μέθοδος chain-ladder

Το μοντέλο Poisson μπορεί να θεωρηθεί ως ειδική περίπτωση του πολλαπλασιαστικού μοντέλου. Έχει την ίδια βασική πολλαπλασιαστική δομή της πρώτης ροπής, αλλά θεωρείται επιπλέον μια κατανομή Poisson των προσαυξητικών απαιτήσεων C_{ij} . Verral (2000) ισχυρίστηκε ότι το μοντέλο poisson θα παράγει ακριβώς τις ίδιες εκτιμήσεις αποθεματικών με τη μέθοδο της chain ladder. Αυτό ισχύει όταν χρησιμοποιούνται οι εκτιμητές μέγιστης πιθανοφάνειας (MLE), οι οποίοι θα αποδειχθούν.

C_{ij} είναι προσαυξητικές απαιτήσεις, και έστω C_{ij} είναι ανεξάρτητη κατανομή Poisson με

$E(C_{ij}) = x_i y_j$ και $\sum_{j=1}^n y_j = 1$. Από το πολλαπλασιαστικό μοντέλο προσδιορίστηκε η

παράμετρος x_i , $x_i = E(D_{in})$. Η x_i είναι η αναμενόμενη τιμή των σωρευτικών απαιτήσεων έως τα τελευταία χρόνια εξέλιξης που παρατηρήθηκαν μέχρι τώρα.

Η πρώτη ροπή μπορεί να παραμετροποιηθεί ως:

$$E(C_{ij}) = x_i y_j = E(D_{in}) y_j = \frac{E(D_{i,n-i+1}) y_j}{\sum_{j=1}^{n-i+1} y_j} = \frac{z_i y_j}{s_{n-i+1}} \quad (2.9)$$

Όπου, $z_i = E(D_{i,n-i+1})$ και $s_k = \sum_{j=1}^k y_j$.

Δεδομένου ότι, y_j , μπορεί να ερμηνευθεί ως η αναλογία της τελικής απαίτησης κατά το έτος εξέλιξης j , είναι λογικό ότι, $E(D_{i,n-i+1})$, θα χωρίζεται από το ποσοστό των απαιτήσεων μέχρι $j = n - i + 1$ να ισούται $E(D_{in})$.

Η εξίσωση (2.9) μπορεί να γραφεί έτσι ώστε να είναι ένας τύπος για την πρόβλεψη της αναμενόμενης τιμής της τελικής απαίτησης, $E(D_{in})$. Η προσέγγιση, $E(D_{in})$, με την εξίσωση

\hat{D}_{in} είναι:

$$\hat{D}_{in} = E(D_{in}) = x_i = \frac{z_i}{\sum_{k=1}^{n-i+1} y_k} = \frac{z_i}{1 - \sum_{k=n-i+2}^n y_k} .$$

(2.10)

Verral (2000) ισχυρίζεται ότι αυτό είναι ισοδύναμο με τον εκτιμητή της chain-ladder:

$$\hat{D}_{n-j+1,n} = d_{n-j+1,j} \hat{f}_{j+1} \hat{f}_{j+2} \dots \hat{f}_n \quad \text{όπου} \quad \hat{f}_j = \frac{\sum_{i=1}^{n-j+1} d_{ij}}{\sum_{i=1}^{n-j+1} d_{i,j-1}}. \quad (2.11)$$

Για να δούμε ότι τα (2.10) και (2.11) είναι στην πραγματικότητα ισοδύναμα, είναι φυσικό να ψάξουμε για τους εκτιμητές των άγνωστων παραμέτρων στο (2.10). Η συνάρτηση μέγιστης πιθανοφάνειας θα χρησιμοποιηθεί για την εύρεση εκτιμητών. Στην περίπτωση αυτή, οι παρατηρήσεις, c_{ij} , θεωρούνται γνωστές και οι παράμετροι θεωρούνται μεταβλητές. Η συνάρτηση μέγιστης πιθανοφάνειας μπορεί να γραφεί ως εξής:

$$L = \prod_{i=1}^n \prod_{j=1}^{n-i+1} \left(\frac{\left(z_i y_j \mid s_{n-i+1} \right)^{c_{ij}} e^{-z_i y_j \mid s_{n-i+1}}}{c_{ij}!} \right). \quad (2.12)$$

είναι η μέγιστη συνάρτηση πιθανότητας μίας κατανεμημένης Poisson μεταβλητής με

πaráμετρο $\frac{z_i y_j}{s_{n-i+1}}$. Οι περαιτέρω υπολογισμοί δείχνουν ότι αυτό μπορεί να γραφεί ως:

$$L = \prod_{i=1}^n \left(\frac{z_i^{d_{i,n-i+1}} e^{-z_i}}{d_{i,n-i+1}!} \left(\frac{d_{i,n-i+1}!}{\prod_{j=1}^{n-i+1} c_{ij}!} \prod_{j=1}^{n-i+1} \left(\frac{y_j}{s_{n-i+1}} \right)^{c_{ij}} \right) \right) = L_c L_d \quad (2.13)$$

όπου

$$L_c = \prod_{i=1}^n \left(\frac{d_{i,n-i+1}!}{\prod_{j=1}^{n-i+1} c_{ij}!} \prod_{j=1}^{n-i+1} \left(\frac{y_j}{s_{n-i+1}} \right)^{c_{ij}} \right) \quad \text{και} \quad L_d = \prod_{i=1}^n \left(\frac{z_i^{d_{i,n-i+1}} e^{-z_i}}{d_{i,n-i}!} \right).$$

(2.14)

L_c , είναι η δεσμευμένη συνάρτηση μέγιστης πιθανοφάνειας, όπου C_{ij} εξαρτάται $d_{i,n-i+1}$

από την πολυωνυμική κατανομή με πιθανότητες $\frac{y_j}{s_{n-i+1}}$ (βλ. Παράρτημα 2). Η πολυωνυμική

κατανομή είναι λογική, λαμβάνοντας υπόψη την πιθανότητα μιας απαίτησης / ή πολλών

απαιτήσεων που αναφέρονται σε αύξηση (i, j) . Η πολυωνυμική κατανομή αντιπροσωπεύει

την πιθανότητα των απαιτήσεων C_{ij} , οι οποίες πραγματοποιήθηκαν κατά το έτος

ατυχημάτων i , αναφέρονται στο έτος εξέλιξης j . L_d , είναι η μέγιστη συνάρτηση

πιθανοτήτων όπου, $D_{i,n-i+1}$, η κατανομημένη Poisson με μέσο όρο z_i , και με αυτή την

έκφραση βρίσκεται ο μέγιστος εκτιμητής πιθανοτήτων (MLE), z_i . Το MLE του z_i είναι

$d_{i,n-i+1}$, από τότε $D_{i,n-i+1}$ που κατανέμεται με Poisson.

Χρησιμοποιώντας το MLE του z_i εκτιμητή της τελικής απαίτησης γίνεται:

$$\hat{D}_{in} = \frac{d_{i,n-i+1}}{1 - \sum_{k=n-i+2}^n y_k} \quad (2.15)$$

Για το έτος ατυχήματος, $n - j + 1$, αυτή η έκφραση είναι

$$\hat{D}_{n-j+1,n} = \frac{d_{in-j+1,j}}{1 - \sum_{k=j+1}^n y_k}. \quad (2.16)$$

Στην έκφραση (2.16) η μοναδική άγνωστη παράμετρος είναι y_k . Αυτό μπορεί να προσδιοριστεί με την εύρεση του MLE με τη χρήση L , αλλά επίσης η L_c μπορεί να χρησιμοποιηθεί. Ο λογάριθμος του L_c βρέθηκε και η προκύπτουσα έκφραση διαφοροποιείται σε σχέση με y_k , για $k = 1, \dots, n$. Αυτό πρέπει να γίνει αναδρομικά, σε μια διαδικασία που περιγράφεται από τον Renshaw (1998). Η παράμετρος \hat{y}_n καθορίζεται πρώτα, τότε \hat{y}_{n-1} και ούτω καθεξής. Οι υπολογισμοί της εύρεσης \hat{y}_n και ο γενικός τύπος \hat{y}_j παρουσιάζονται παρακάτω:

$$\ln(L_c) = l_c \propto \sum_{i=1}^n \sum_{j=1}^{n-i+1} c_{ij} \log \left(\frac{y_i}{\sum_{k=1}^{n-i+1} y_k} \right) = \sum_{i=1}^n \sum_{j=1}^{n-i+1} c_{ij} \left(\log y_i - \log \left(\sum_{k=1}^{n-i+1} y_k \right) \right)$$

$$\frac{\mathcal{G}_c}{\mathcal{G}_{y_n}} = 0 \Rightarrow \frac{c_{1n}}{\hat{y}_n} - \sum_{j=1}^n \frac{c_{1j}}{\sum_{k=1}^n \hat{y}_k} = \frac{c_{1n}}{\hat{y}_n} - \sum_{j=1}^n \frac{c_{1j}}{1} = 0$$

$$\Rightarrow \hat{y}_n = \frac{c_{1n}}{\sum_{j=1}^n c_{1j}} = \frac{c_{1n}}{d_{1n}} \quad (2.17)$$

$$\frac{\mathcal{G}_c}{\mathcal{G}_{y_n}} = 0 \Rightarrow \sum_{i=1}^{n-j+1} \left(\frac{c_{ij}}{\hat{y}_j} - \sum_{j=1}^{n-j+1} \frac{c_{ij}}{\sum_{k=1}^{n-i+1} \hat{y}_k} \right) = \frac{\sum_{k=1}^{n-i+1} c_{ij}}{\hat{y}_j} - \sum_{i=1}^{n-j+1} \left(\frac{d_{i,n-i+1}}{\sum_{k=1}^{n-i+1} \hat{y}_k} \right) = 0$$

$$\Rightarrow \hat{y}_j = \frac{\sum_{i=1}^{n-j+1} c_{ij}}{\sum_{i=1}^{n-y+1} \left(\frac{d_{i,n-i+1}}{\sum_{k=1}^{n-i+1} \hat{y}_k} \right)} = \frac{c_{1j} + \dots + c_{n-j+1,j}}{d_{1n} + \frac{d_{2,n-1}}{1-\hat{y}_n} + \dots + \frac{d_{n-j+1,j}}{1-\hat{y}_{j+1} - \dots - \hat{y}_n}} \quad (2.18)$$

Ένας εκτιμητής μέγιστης πιθανοφάνειας y_j , για $j = 1, \dots, n$ εκφράζεται σε (2.18). Το επόμενο βήμα είναι να βρούμε μια έκφραση για τον συντελεστή εξέλιξης \hat{f}_j , χρησιμοποιώντας το MLE \hat{y}_j . Με την αναδιάταξη της εξίσωσης chain-ladder στο (2.11), προκύπτει μια έκφραση του προϊόντος των συντελεστών εξέλιξης:

$$\hat{f}_{j+1} \hat{f}_{j+2} \dots \hat{f}_n = \frac{\hat{D}_{n-j+1,n}}{d_{n-j+1,j}}$$

Η εισαγωγή της έκφρασης $\hat{D}_{n-j+1,n}$ από την εξίσωση (2.16) και χρησιμοποιώντας τον εκτιμητή \hat{y}_j αντί του y_j , το αποτέλεσμα των συντελεστών εξέλιξης γίνεται:

$$\hat{f}_{j+1} \hat{f}_{j+2} \dots \hat{f}_n = \frac{1}{1 - \hat{y}_{j+1} - \hat{y}_{j+2} - \dots - \hat{y}_n} \quad (2.19)$$

και

$$\hat{f}_j \hat{f}_{j+1} \dots \hat{f}_n = \frac{1}{1 - \hat{y}_j - \hat{y}_{j+1} - \dots - \hat{y}_n} \quad (2.20)$$

Με την αναδιάταξη της (2.19) μια έκφραση $1 - \hat{y}_{j+1} - \hat{y}_{j+2} - \dots - \hat{y}_n$ προκύπτει, και αυτό μπορεί να εισαχθεί στο (2.20). Έτσι

$$\hat{f}_j \hat{f}_{j+1} \dots \hat{f}_n = \frac{1}{\frac{1}{\hat{f}_{j+1} \hat{f}_{j+2} \dots \hat{f}_n} - \hat{y}_j} \quad (2.21)$$

Τέλος, βρέθηκε ένας εκτιμητής του συντελεστή εξέλιξης \hat{f}_j

$$\hat{f}_j = \frac{1}{1 - \hat{y}_j \hat{f}_{j+1} \hat{f}_{j+2} \dots \hat{f}_n} \quad (2.22)$$

Χρησιμοποιώντας το MLE από (2.17) η έκφραση γίνεται:

$$\hat{f}_n = \frac{1}{1 - \hat{y}_n} = \frac{1}{1 - \frac{c_{1n}}{d_{1n}}} = \frac{d_{1n}}{d_{1n} - c_{1n}} = \frac{d_{1n}}{d_{1,n-1}} \quad .$$

(2.23)

Ο εκτιμητής που λαμβάνεται στο (2.23) είναι ο ίδιος με τον εκτιμητή της chain ladder για $j = n$. Για να δείξουμε ότι οι υπόλοιποι συντελεστές εξέλιξης στο μοντέλο Poisson είναι ίδιοι με τους συντελεστές εξέλιξης της chain ladder μπορεί να χρησιμοποιηθεί μια εισαγωγή. Δεδομένου ότι, έχει αποδειχθεί ότι $j = n$, το πρώτο μέρος της επαγωγής ολοκληρώθηκε. Το επόμενο βήμα είναι να βρούμε τον γενικό τύπο για το \hat{f}_j . Για να γίνει αυτό, η έκφραση \hat{y}_j χρειάζεται κάποια απλοποίηση. Η εξίσωση (2.18) δίνει μια έκφραση \hat{y}_j και τα κλάσματα στον παρονομαστή μπορούν να γραφούν χρησιμοποιώντας τις σχέσεις (2.19), (2.20) και είναι ισοδύναμες. Έτσι:

$$\hat{y}_j = \frac{c_{1j} + c_{2,j} + \dots + c_{n-j+1,j}}{d_{1n} + d_{1,n-1} \hat{f}_n \dots + d_{n-j+1,j} \hat{f}_{j+1} \hat{f}_{j+2} \dots \hat{f}_n} \quad (2.24)$$

Με την εξέταση της έκφρασης \hat{y}_j , μπορούμε επίσης να δούμε ότι είναι το ποσοστό της τελικής απαίτησης. Ο αριθμητής μετρά τις προσαυξημένες απαιτήσεις σε όλα τα έτη ατυχημάτων που παρατηρούνται για το έτος εξέλιξης j και ο παρονομαστής μετράει τις εκτιμώμενες τελικές απαιτήσεις σε σχέση με τα ίδια τα ατυχήματα. Η εξίσωση (2.22) είναι μια γενική έκφραση για \hat{f}_j . Η νέα προέκταση έκφρασης, για το \hat{y}_j , εισάγεται στο (2.22).

Έτσι,

$$\hat{f}_j = \frac{1}{1 - \frac{c_{1j} + c_{2,j} + \dots + c_{n-j+1,j}}{d_{1n} + d_{2,n-1}\hat{f}_n \dots + d_{n-j+1,j}\hat{f}_{j+1}\hat{f}_{j+2}\dots\hat{f}_n}} \hat{f}_{j+1}\hat{f}_{j+2}\dots\hat{f}_n \quad (2.25)$$

Αυτός είναι ο γενικός τύπος. Έχει ήδη αποδειχθεί ότι ο εκτιμητής, \hat{f}_n , είναι ο συντελεστής εξέλιξης της chain-ladder.

Ως μέρος της επαγωγής, θεωρείται ότι για $k = j+1, \dots, n$, \hat{f}_k ισούται με τον συντελεστή εξέλιξης της chain-ladder. Το τελευταίο βήμα είναι να αποδειχθεί ότι, \hat{f}_k ισούται με τον συντελεστή εξέλιξης της chain-ladder, για $k = j$

Ο παρονομαστής στο σημείο (2.25) πρέπει να απλουστευθεί, πράγμα που μπορεί να γίνει δείχνοντας ότι:

$$d_{1n} + d_{2,n-1}\hat{f}_n + \dots + d_{n-j+1,j}\hat{f}_{j+1}\hat{f}_{j+2}\dots\hat{f}_n = \hat{f}_{j+1}\hat{f}_{j+2}\dots\hat{f}_n \sum_{i=1}^{n-j+1} d_{ij} \quad (2.26)$$

Αυτό ισχύει για, $j = n - 1$

$$d_{1n} + d_{2,n-1} \hat{f}_n = d_{1n} + d_{2,n-1} \frac{d_{1,n}}{d_{1,n-1}} = \frac{d_{1,n}}{d_{1,n-1}} (d_{1,n-1} + d_{2,n-1}) = \hat{f}_n (d_{1,n-1} + d_{2,n-1})$$

Ομοίως, για $j = n - 2$, υπάρχει η ίδια σχέση,

$$\begin{aligned} d_{1n} + d_{2,n-1} \hat{f}_n + d_{3,n-2} \hat{f}_{n-1} \hat{f}_n &= \hat{f}_n (d_{1,n-1} + d_{2,n-1} + d_{3,n-2} \hat{f}_{n-1}) \\ &= \hat{f}_n (d_{1,n-1} + d_{2,n-1}) \frac{d_{1,n-1} + d_{2,n-1}}{d_{1,n-2} + d_{2,n-2}} + d_{3,n-2} \hat{f}_{n-1} \\ &= \hat{f}_{n-1} \hat{f}_n (d_{1,n-2} + d_{2,n-2} + d_{3,n-2}). \end{aligned}$$

Εκτελώντας αυτό $n - j$ φορές (2.26) αποδεικνύεται, και η εξίσωση για \hat{f}_j στη (2.25) μπορεί να μειωθεί στο

$$\begin{aligned} \hat{f}_j &= \frac{1}{1 - \frac{c_{1j} + c_{2,j} + \dots + c_{n-j+1,j}}{d_{1n} + d_{2,n-1} \hat{f}_n \dots + d_{n-j+1,j} \hat{f}_{j+1} \hat{f}_{j+2} \dots \hat{f}_n} \hat{f}_{j+1} \hat{f}_{j+2} \dots \hat{f}_n} \\ &= \frac{\sum_{i=1}^{n-j+1} d_{ij}}{\sum_{i=1}^{n-j+1} d_{ij} - \sum_{i=1}^{n-j+1} c_{ij}} = \frac{\sum_{i=1}^{n-j+1} d_{ij}}{\sum_{i=1}^{n-j+1} d_{i,j-1}} \end{aligned} \quad (2.27)$$

Η απόδειξη επαγωγής εκπληρώνεται, αφού \hat{f}_j είναι ίση με τον συντελεστή εξέλιξης της chain-ladder. Έχει αποδειχθεί ότι, η χρήση του MLE σε ένα μοντέλο Poisson θα παράγει ακριβώς τις ίδιες εκτιμήσεις με τη μέθοδο chain-ladder.

2.4 Το μοντέλο poisson και η σχέση του με το αρνητικό διωνυμικό μοντέλο

Το προηγούμενο κεφάλαιο ξεκίνησε θεωρώντας C_{ij} ως μια τυχαία μεταβλητή poisson. Αυτό συμβαίνει και εδώ, αλλά στην περίπτωση αυτή η ένταση της κατανομής Poisson θα είναι επίσης στοχαστική. Μέσω των ακόλουθων ορισμών ο Verral (2000) έκανε ένα αναδρομικό μοντέλο που συνέδεε τη Poisson, το μοντέλο και το αρνητικό διωνυμικό μοντέλο.

Η C_{ij} , εξαρτάται από τη $Z_{ij} = z_{ij}$ κατανομή Poisson με μέσο όρο $\frac{Z_{ij}y_j}{s_j}$, όπου $z_{ij} = E(D_{ij})$

$$\text{και } s_j = \sum_{k=1}^j y_k .$$

Η μεταβλητή, Z_{ij} , ορίζεται με τον δείκτη j (εκτός από το δείκτη i), καθώς αυτό είναι ένα μοντέλο υπό δέσμευση, όπου $z_{ij} = E(D_{ij})$, οι αλλαγές με το έτος εξέλιξης j . Πριν από οποιαδήποτε παραδοχή Z_{ij} , η σχέση μεταξύ Z_{ij} και $Z_{i,j-1}$ θα καθοριστεί ως:

$$Z_{ij} = E(D_{ij}) = E(D_{i,j-1}) + E(C_{ij}) = Z_{i,j-1} + \frac{Z_{ij}y_j}{s_j}$$

$$\Rightarrow Z_{ij} = \frac{Z_{i,j-1}s_j}{s_{j-1}} \quad (2.28)$$

Δεδομένης αυτής της σχέσης, η κατανομή, C_{ij} , είναι:

C_{ij} δοθέντος ότι $z_{i,j-1}$ έχει κατανομή Poisson με μέσο όρο $\frac{z_{i,j-1}s_j}{s_{j-1}}$.

Η παράμετρος γ_j μπορεί ακόμα να θεωρηθεί ως η παράμετρος της στήλης και είναι η πιθανότητα να καταγραφεί μια απαίτηση στο έτος εξέλιξης j . Ο παράγοντας $\frac{z_{i,j-1}}{s_{j-1}}$ δίνει την αναμενόμενη τελική απαίτηση.

Ο στόχος είναι να διαπιστωθεί ότι το C_{ij} εξαρτάται από τις προγενέστερες παρατηρηθείσες απαιτήσεις $c_{i1}, \dots, c_{i,j-1}$, είναι αρνητικά δυωνυμικά κατανεμημένο. Για να γίνει αυτό, είναι απαραίτητο να γίνουν κάποιες υποθέσεις για $Z_{i,j-1}$. Θεωρείται ότι, η κατανομή $Z_{i,j-1}$ είναι γνωστή, έτσι ώστε αυτό το μοντέλο παίρνει μια Bayesian προσέγγιση. Στο έτος εξέλιξης j , υπάρχουν παρατηρήσεις των απαιτήσεων μέχρι το έτος εξέλιξης $j-1$.

Θεωρείται ότι:

$Z_{i,j-1}$ δοθέντος ότι $c_{i1}, \dots, c_{i,j-1}$ με κατανομή γάμμα η οποία έχει παραμέτρους α και β .

Με τη χρήση της Bayesian ανάλυσης, μπορεί κανείς να βρει την κατανομή $Z_{i,j-1}$, δοθέντος ότι $c_{i1}, \dots, c_{i,j-1}$. Σε αυτήν την περίπτωση, η προηγούμενη κατανομή

$\pi_{Z_{i,j-1}|c_{i1}, \dots, c_{i,j-1}}(z_{i,j-1} | c_{i1}, \dots, c_{i,j-1})$ είναι η κατανομή γάμμα, η εξαρτημένη κατανομή

$f_{c_{ij}|z_{i,j-1}}(c_{ij} | z_{i,j-1})$ είναι η κατανομή Poisson. Ο Bayesian τύπος που χρησιμοποιείται για την

επίλυση αυτού του προβλήματος είναι:

$$\pi_{Z_{i,j-1}|c_{i1}, \dots, c_{ij}}(z_{i,j-1} | c_{i1}, \dots, c_{ij}) = \frac{f_{c_{ij}|z_{i,j-1}}(c_{ij} | z_{i,j-1}) \pi_{Z_{i,j-1}|c_{i1}, \dots, c_{i,j-1}}(z_{i,j-1} | c_{i1}, \dots, c_{i,j-1})}{\int_0^{\infty} f_{c_{ij}|z_{i,j-1}}(c_{ij} | z_{i,j-1}) \pi_{Z_{i,j-1}|c_{i1}, \dots, c_{i,j-1}}(z_{i,j-1} | c_{i1}, \dots, c_{i,j-1}) dz_{i,j-1}}$$

$$\begin{aligned}
& \frac{\left((z_{i,j-1} / s_{j-1})^{c_{ij}} / c_{ij}! \right) e^{z_{i,j-1} y_j / s_{j-1}} \frac{1}{\Gamma(\alpha) \beta^\alpha} (z_{i,j-1})^{\alpha-1} e^{z_{i,j-1} / \beta}}{\int_0^\infty \left((z_{i,j-1} / s_{j-1})^{c_{ij}} / c_{ij}! \right) e^{z_{i,j-1} y_j / s_{j-1}} \frac{1}{\Gamma(\alpha) \beta^\alpha} (z_{i,j-1})^{\alpha-1} e^{z_{i,j-1} / \beta} dz_{i,j-1}} \\
&= \frac{1}{\Gamma(\alpha + c_{ij})} \left(\frac{y_j}{s_{j-1}} + \beta \right)^{\alpha + c_{ij}} (z_{i,j-1})^{\alpha + c_{ij} - 1} e^{-z_{i,j-1} \left(\frac{y_j}{s_{j-1}} + \beta \right)} \\
&= \Gamma \left(\alpha + c_{ij}, \frac{y_j}{s_{j-1}} + \beta \right) \tag{2.29}
\end{aligned}$$

Βρίσκεται η κατανομή, $Z_{i,j-1}$, δοθέντος ότι $c_{i1}, \dots, c_{i,j-1}$, και η Verral (2000) προχωρά με την εύρεση της κατανομής $Z_{i,j}$, δοθέντος ότι c_{i1}, \dots, c_{ij} . Η σχέση μεταξύ $Z_{i,j-1}$ και Z_j , δίνεται από τη (2.28). Με την απλή αλλαγή που παρατηρείται παρακάτω, μπορεί κανείς να βρει

$$\begin{aligned}
\pi_{Z_{i,j-1} | c_{i1}, \dots, c_{ij}}(z_{ij} | c_{i1}, \dots, c_{ij}) &= \pi_{Z_{i,j-1} | c_{i1}, \dots, c_{ij}} \left(z_{i,j-1} = \frac{s_{j-1}}{s_j} z_j | c_{i1}, \dots, c_{ij} \right) \frac{dz_{i,j-1}}{dz_{ij}} \\
&= \frac{1}{\Gamma(\alpha + c_{ij})} \left(\frac{s_{j-1}}{s_j} \left(\frac{y_j}{s_{j-1}} + \beta \right) \right)^{\alpha + c_{ij}} (z_{ij})^{\alpha + c_{ij} - 1} e^{-z_{ij} \left(\frac{y_j}{s_{j-1}} + \beta \right)} \\
&= \Gamma \left(\alpha + c_{ij}, \frac{s_{j-1}}{s_j} \left(\frac{y_j}{s_{j-1}} + \beta \right) \right) \tag{2.30}
\end{aligned}$$

Οι παραπάνω υπολογισμοί αποδίδουν την κατανομή Z_{ij} , δοθέντος ότι c_{i1}, \dots, c_{ij} και στη συνέχεια είναι ενδιαφέρον να βρούμε την κατανομή για κάθε j , όπου $j = 1, \dots, n$. Είναι

φυσικό να αρχίσουμε με την εύρεση της κατανομής για $j=1$. Για να γίνει αυτό είναι απαραίτητο να υποθέσουμε μια προηγούμενη κατανομή του Z_{i1} . Η Verral (2000) υποθέτει ότι $\pi_{z_{i1}}(z_{i1}) \propto (z_{i1})^{-1}$. Όπως στο (2.29), ο Bayesian τύπος μπορεί να χρησιμοποιηθεί για να βρει την κατανομή z_{i1} δοθέντος ότι c_{i1} :

$$\pi_{z_{i1}|c_{i1}}(z_{i1} | c_{i1}) \propto \frac{\pi_{z_{i1}}(z_{i1}) f_{c_{i1}|z_{i1}}(c_{i1} | z_{i1})}{\int_0^{\infty} \pi_{z_{i1}}(z_{i1}) f_{c_{i1}|z_{i1}}(c_{i1} | z_{i1}) dz_{i1}} = \frac{\frac{1}{z_{i1}} \frac{e^{-z_{i1}}}{c_{i1}!}}{\int_0^{\infty} \frac{1}{z_{i1}} \frac{e^{-z_{i1}}}{c_{i1}!} dz_{i1}} = \Gamma(c_{i1}, 1) \quad (2.31)$$

Ο γενικός τύπος φάνηκε στο (2.30) και ο συγκεκριμένος τύπος βρίσκεται τώρα για την πρώτη περίπτωση, (2.31). Αυτή η κατανομή είναι η περίπτωση όταν $j=2, \alpha=c_{i1}$ και $\beta=1$ όταν (2.30). Δεδομένου ότι, η (2.30) θα παράγει την κατανομή για $j=3$, όταν α και β είναι γνωστό, είναι απαραίτητο να εισαγουμε αυτές τις τιμές έτσι ώστε

$$\pi_{z_{i2}|c_{i1}, c_{i2}}(z_{i2} | c_{i1}, c_{i2}) = \Gamma\left(c_{i1} + c_{i2}, \frac{s_{j-1}}{s_j} \left(\frac{y_j}{s_{j-1}} + 1\right)\right) = \Gamma(d_{i2}, 1) \quad (2.32)$$

Για να αποδειχθεί αυτό για όλα τα j , μπορεί να χρησιμοποιηθεί η επαγωγή. Ο τύπος θεωρείται ότι είναι $k=j-1$, $\pi_{z_{i,j-1}|c_{i1}, \dots, c_{i,j-1}}(z_{i,j-1} | c_{i1}, \dots, c_{i,j-1}) = \Gamma(d_{i,j-1}, 1)$. Όπως έγινε παραπάνω, ο τύπος (2.30) μπορεί να χρησιμοποιηθεί για να το αποδείξει, όταν $k=j$.

$$\pi_{z_{ij}|c_{i1}, \dots, c_{ij}}(z_{ij} | c_{i1}, \dots, c_{ij}) = \Gamma\left(d_{i,j-1} + c_{ij}, \frac{s_{j-1}}{s_j} \left(\frac{y_j}{s_{j-1}} + 1\right)\right) = \Gamma(d_{ij}, 1) \quad (2.33)$$

Το τρίγωνο εξέλιξης έχει μόνο γνωστές τιμές στη πάνω-αριστερή γωνία και για να προβλέψει κάποιος τις υπόλοιπες τιμές C_{ij} , είναι επιθυμητό να βρεθεί η κατανομή του C_{ij}

δοθέντος ότι $C_{i1}, \dots, C_{i,j-1}$. Αυτό μπορεί να βρεθεί με αυτή τη διατύπωση:

$$\begin{aligned}
 f_{c_{ij}|c_{i1}, \dots, c_{i,j-1}}(c_{ij} | c_{i1}, \dots, c_{i,j-1}) &= \int f_{c_{ij}|z_{i,j-1}}(c_{ij} | z_{i,j-1}) f_{z_{i,j-1}|c_{i1}, \dots, c_{i,j-1}}(z_{i,j-1} | c_{i1}, \dots, c_{i,j-1}) dz_{i,j-1} \\
 &= \int_0^{\infty} \frac{(z_{i,j-1}/s_{j-1})^{c_{ij}} e^{-(z_{i,j-1}y_j/s_{j-1})}}{c_{ij}!} \frac{1}{\Gamma(d_{i,j-1})} (z_{i,j-1})^{d_{i,j-1}-1} e^{-z_{i,j-1}} dz_{i,j-1} \\
 &= \frac{(y_j/s_{j-1})^{c_{ij}}}{c_{ij}! \Gamma(d_{i,j-1})} \frac{\Gamma(d_{i,j-1} + c_{ij})}{(y_j/s_{j-1} + 1)^{d_{i,j-1} + c_{ij}}} * \\
 &\int_0^{\infty} \frac{1}{\Gamma(d_{i,j-1} + c_{ij})} (y_j/s_{j-1} + 1)^{d_{i,j-1} + c_{ij}} (z_{i,j-1})^{d_{i,j-1} + c_{ij} - 1} e^{-z_{i,j-1}(y_j/s_{j-1} + 1)} dz_{i,j-1} \\
 &= \frac{\Gamma(d_{ij})}{c_{ij}! \Gamma(d_{i,j-1})} \frac{(y_j/s_{j-1})^{c_{ij}}}{(s_j/s_{j-1})^{d_{i,j-1} + c_{ij}}} \\
 &= \frac{\Gamma(d_{ij})}{c_{ij}! \Gamma(d_{i,j-1})} \left(\frac{s_{j-1}}{s_j}\right)^{d_{i,j-1}} \left(\frac{s_{j-1}}{s_j}\right)^{c_{ij}} \left(\frac{y_j}{s_{j-1}}\right)^{c_{ij}} \\
 &= \frac{\Gamma(d_{ij})}{c_{ij}! \Gamma(d_{i,j-1})} \left(\frac{s_{j-1}}{s_j}\right)^{d_{i,j-1}} \left(\frac{y_j}{s_j}\right)^{c_{ij}} \\
 &= \frac{(c_{i,j-1} + c_{ij} - 1)!}{c_{ij}! (d_{i,j-1} - 1)!} \left(\frac{s_{j-1}}{s_j}\right)^{d_{i,j-1}} \left(1 - \frac{s_{j-1}}{s_j}\right)^{c_{ij}}
 \end{aligned} \tag{2.34}$$

έτσι

C_{ij} δοθέντος ότι $C_{i1}, \dots, C_{i,j-1}$ το οποίο είναι μια αρνητική διωνυμική με μέση τιμή $\frac{d_{i,j-1}y_j}{s_{j-1}}$

και διακύμανση $\frac{d_{i,j-1}s_j}{(s_{j-1})^2}$

Οι συντελεστές εξέλιξης της chain-ladder, θα μπορούσαν επίσης να εκφραστούν ως συνάρτηση των συντελεστών της στήλης y_j , όπου

$$f_j = \frac{\sum_{k=1}^j y_k}{\sum_{k=1}^{j-1} y_k} = \frac{s_j}{s_{j-1}}.$$

Η αποδοχή του ορισμού στο (2.35), αποκαλύπτει ότι η κατανομή C_{ij} , δοθέντος ότι $C_{i1}, \dots, C_{i,j-1}$, μπορεί να γραφεί μόνο ως συνάρτηση των παρατηρούμενων σωρευτικών απαιτήσεων και των συντελεστών εξέλιξης. Η κατανομή της C_{ij} , δοθέντος ότι $C_{i1}, \dots, C_{i,j-1}$, είναι:

$$\frac{(d_{i,j-1} + c_{ij} - 1)! \left(\frac{1}{f_j}\right)^{d_{i,j-1}} \left(1 - \frac{1}{f_j}\right)^{c_{ij}}}{c_{ij}! (d_{i,j-1} - 1)!} \quad (2.36)$$

όπου η μέση τιμή είναι $(f_j - 1)d_{i,j-1}$ και η διακύμανση είναι $f_j(f_j - 1)d_{i,j-1}$.

Δεδομένου ότι, $D_{ij} = D_{i,j-1} + C_{ij}$, η κατανομή D_{ij} , δοθέντος ότι $C_{i1}, \dots, C_{i,j-1}$, είναι επίσης αρνητική διωνυμική κατανομή και η κατανομή είναι:

$$\frac{(d_{i,j-1} + c_{ij} - 1)! \left(\frac{1}{1+f_j}\right)^{d_{i,j-1}} \left(1 - \frac{1}{1+f_j}\right)^{c_{ij}}}{c_{ij}!(d_{i,j-1} - 1)!} \quad (2.37)$$

και μπορεί να γραφεί ως:

$$\frac{(d_{ij} - 1)!}{(d_{ij} - d_{i,j-1})!(d_{i,j-1} - 1)!} \left(\frac{1}{1+f_j}\right)^{d_{i,j-1}} \left(1 - \frac{1}{1+f_j}\right)^{d_{ij} - d_{i,j-1}}$$

όπου η μέση τιμή είναι $f_j d_{i,j-1}$ και η διακύμανση είναι $f_j(f_j - 1)d_{i,j-1}$.

Οι τύποι στις σχέσεις (2.36) και (2.37) δείχνουν ότι δεν είναι απαραίτητο να εξαρτώνται από όλες τις προγενέστερες προσαυξητικές απαιτήσεις $c_{i1}, \dots, c_{i,j-1}$, η κατανομή D_{ij} , δοθέντος ότι $d_{i,j-1}$, είναι ταυτόσημη με (2.37).

2.5 Το μοντέλο του Mack

Το έργο που είχε την μεγαλύτερη επιρροή στα πρώτα στάδια της ανάπτυξης των στοχαστικών μεθόδων αποθεματοποίησης, ήταν το άρθρο του Mack (1994), όπου παρουσιάστηκαν οι πρώτες απόπειρες για τον σχηματισμό στοχαστικών μοντέλων αποθεματοποίησης. Ο Mack παρουσίασε για πρώτη φορά ένα στοχαστικό μοντέλο το οποίο έδινε τα ίδια ακριβώς αποτελέσματα για τις εκτιμήσεις των αποθεμάτων με αυτά που προέκυπταν από τη μέθοδο chain-ladder. Όπως αναφέραμε παραπάνω, συμβολίζουμε με, C_{ij} , την προσαυξανόμενη αλλαγή μεταξύ του έτους εξέλιξης j και, D_{ij} , τις αθροιστικές ζημιές-πληρωμές που έγιναν στο έτος ατυχήματος i και για το έτος εξέλιξης j . Η μεταβλητή K_{ij} καθορίστηκε στο κεφάλαιο 1.6 και ο ορισμός της εξακολουθεί να ισχύει. Υποθέτουμε ότι, το πρώτο έτος ατυχημάτων είναι πλήρως ανεπτυγμένο.

Ο Mack έκανε τρεις παραδοχές για να καθορίσει αυτό το μοντέλο. Οι παραδοχές είναι οι εξής:

1. Υπάρχουν σταθερές f_2, \dots, f_n , τέτοιες ώστε $E(D_{i,j} | K_{i,j-1} = k_{i,j-1}) = f_j d_{i,j-1}$, για $j = 2, \dots, n$
2. Υπάρχουν σταθερές g_2, \dots, g_n , τέτοιες ώστε $Var(D_{i,j} | K_{i,j-1} = k_{i,j-1}) = g_j d_{i,j-1}$, για $j = 2, \dots, n$
3. K_{in} και K_{kn} είναι ανεξάρτητες, για $i \neq k$.

Ο εκτιμητής, \hat{f}_j , στο μοντέλο του Mack είναι ο ίδιος με τον συντελεστή εξέλιξης της chain-ladder. Ο συντελεστής εξέλιξης, \hat{f}_j , είναι ο ίδιος για όλα τα έτη ατυχημάτων στο έτος εξέλιξης j και γι' αυτό γίνεται μια υπόθεση ανεξαρτησίας μεταξύ των ετών ατυχημάτων.

Η παράμετρος, g_j , μπορεί να εκτιμηθεί με:

$$\hat{g}_j = \frac{1}{n-j} \sum_{i=1}^{n-j+1} d_{i,j-1} \left(\frac{d_{i,j}}{d_{i,j-1}} - \hat{f}_j \right)^2$$

Το μοντέλο του Mack καθορίζεται μόνο από τις τρεις παραπάνω παραδοχές. Με την εξέταση των ταυτοτήτων (f_j, g_j) που εισάγονται στις παραδοχές είναι δυνατό να υπάρξει περαιτέρω κατανόηση του μοντέλου. Ο Mack παρουσίασε αυτό το μοντέλο στο άρθρο του (Mack 1994α) και τα ακόλουθα αποτελέσματα προέρχονται από το άρθρο αυτό. Ο συντελεστής εξέλιξης της chain-ladder θα εξεταστεί πρώτα:

Οι συντελεστές εξέλιξης \hat{f}_j είναι αμερόληπτοι εκτιμητές f_j . Χρησιμοποιώντας τον κανόνα του υπολογισμού της διπλής αναμενόμενης τιμής είναι σαφές ότι

$$\begin{aligned}
 &= E \left(\frac{1}{\sum_{i=1}^{n-j+1} D_{i,j-1}} \left(\sum_{i=1}^{n-j+1} E(D_{i,j} | K_{j-1}) \right) \right) \\
 &= E \left(\frac{1}{\sum_{i=1}^{n-j+1} D_{i,j-1}} \left(\sum_{i=1}^{n-j+1} E(D_{i,j} | K_{j-1}) \right) \right).
 \end{aligned}$$

Ο τελευταίος υπολογισμός χρησιμοποίησε την παραδοχή 3. Δεδομένου ότι υπάρχει ανεξαρτησία μεταξύ K_{in} και K_{kn} για $i \neq k$, είναι απαραίτητο να υπολογίσουμε τις άγνωστες τιμές στο σχετικό έτος ατυχήματος. Χρησιμοποιώντας την υπόθεση 1, είναι εύκολο να δούμε ότι το (2.38) είναι ίσο με:

$$E \left(\frac{1}{\sum_{i=1}^{n-j+1} D_{i,j-1}} \left(\sum_{i=1}^{n-j+1} D_{i,j-1} f_j \right) \right) = f_j. \quad (2.39)$$

Αυτό αποδεικνύει ότι, \hat{f}_j , είναι αμερόληπτος (Mack 1994α).

Οι ατομικοί συντελεστές εξέλιξης δεν είναι συσχετισμένοι. Αυτό μπορεί να αποδειχθεί δείχνοντας ότι

$$E \left(\frac{D_{i,k+1}}{D_{ik}} \frac{D_{ik}}{D_{i,k-1}} \right) = E \left(\frac{D_{i,k+1}}{D_{ik}} \right) E \left(\frac{D_{ik}}{D_{i,k-1}} \right). \quad (2.40)$$

Για, $j \leq k$, μπορεί να φανεί χρησιμοποιώντας τον κανόνα του υπολογισμού της διπλής αναμενόμενης τιμής,

$$\begin{aligned} E\left(\frac{D_{i,k+1}}{D_{ij}}\right) &= E\left(E\left(\frac{D_{i,k+1}}{D_{ij}} \mid D_{i1}, \dots, D_{ik}\right)\right) = E\left(\frac{1}{D_{ij}} E(D_{i,k+1} \mid D_{i1}, \dots, D_{ik})\right) \\ &= E\left(\frac{1}{D_{ij}} D_{ik} f_{k+1}\right) = f_{k+1} E\left(\frac{D_{ik}}{D_{ij}}\right). \end{aligned} \quad (2.41)$$

Από το δεύτερο έως το τρίτο βήμα στο (2.41), η δέσμευση κάνει το D_{ij} γνωστό, για $j \leq k$.

Το επόμενο βήμα χρησιμοποιεί την πρώτη υπόθεση του Mack. Όταν, $j = k$, η εξίσωση (2.41) είναι:

$$E\left(\frac{D_{i,k+1}}{D_{ik}}\right) = f_{k+1} E\left(\frac{D_{ik}}{D_{ik}}\right) = f_{k+1}$$

Όταν $j = k - 1$ (2.41), είναι:

$$E\left(\frac{D_{i,k+1}}{D_{i,k}} \frac{D_{i,k}}{D_{i,k-1}}\right) = f_{k+1} E\left(\frac{D_{ik}}{D_{i,k-1}}\right) = E\left(\frac{D_{i,k+1}}{D_{ik}}\right) E\left(\frac{D_{ik}}{D_{i,k-1}}\right) \quad (2.43)$$

Το πρώτο βήμα στο (2.43) χρησιμοποίησε την ταυτότητα που βρέθηκε στο (2.41), και το δεύτερο βήμα χρησιμοποίησε την ταυτότητα που βρέθηκε στο (2.42). Αυτό αποδεικνύει ότι οι μεμονωμένοι συντελεστές εξέλιξης δεν είναι συσχετισμένοι. Αυτό σημαίνει πως εάν μετά από ένα έτος εξέλιξης, όπου παρατηρήσαμε ένα μεγάλο αριθμό ζημιών, στο επόμενο έτος παρατηρήσουμε ένα μικρό αριθμό ζημιών, ο συντελεστής εξέλιξης ζημιών της μεθόδου chain-ladder δεν θα δώσει ασφαλείς προβλέψεις για την εκτίμηση των μελλοντικών ζημιών. Με άλλα λόγια, εάν σε ένα έτος εξέλιξης παρατηρήσουμε ένα υψηλό-μεγάλο συντελεστή

εξέλιξης ζημιών δεν υπάρχει κάποια ένδειξη πως ο ακόλουθος-επόμενος συντελεστής θα είναι ιδιαίτερα χαμηλός .

Ο συντελεστής εξέλιξης της chain-ladder είναι ένας σταθμισμένος μέσος όρος των ατομικών συντελεστών εξέλιξης. Ο παραπάνω δείκτης είναι αμερόληπτος και η πιο χρήσιμη ιδιότητα ενός εκτιμητή είναι να έχει όσο το δυνατόν μικρότερη διακύμανση. Ο μη σταθμισμένος μέσος όρος των επιμέρους συντελεστών εξέλιξης είναι επίσης αμερόληπτος, γεγονός που υποδηλώνει ότι ο λόγος της χρήσης των συντελεστών εξέλιξης της chain-ladder οφείλεται σε μικρότερη διακύμανση. Η δεύτερη υπόθεση της Mack καθορίζει τη δεύτερη ροπή. Τώρα διερευνάται:

Όταν, D_{ij} , θεωρούνται στοχαστικές μεταβλητές, για $i = 1, \dots, n$ και $j = 1, \dots, n - i + 1$ είναι

επίσης, ο συντελεστής εξέλιξης $\hat{f}_j = \frac{\sum_{i=1}^{n-j+1} D_{ij}}{\sum_{i=1}^{n-j+1} D_{i,j-1}}$, για $j = 2, \dots, n$ είναι στοχαστικές

μεταβλητές. Ο ατομικός παράγοντας εξέλιξης γράφεται με κεφαλαίο γράμμα F_{ij} , όταν θεωρείται μια στοχαστική μεταβλητή. Ο συντελεστής εξέλιξης της chain-ladder είναι ένας σταθμισμένος μέσος όρος των επιμέρους συντελεστών εξέλιξης, και γενικά αυτό μπορεί να γραφεί:

$$\hat{f}_j = \sum_{i=1}^{n-j+1} W_{ij} F_{ij} \text{ όπου } \sum_{i=1}^{n-j+1} W_{ij} = 1 \quad (2.44)$$

Οι ατομικοί συντελεστές εξέλιξης, F_{ij} , θεωρούνται ότι δεν είναι συσχετισμένοι και αμερόληπτοι, για $1 \leq i \leq n$. Η διακύμανση της \hat{f}_j , δοθέντος ότι k_{j-1} , είναι:

$$\text{Var}(\hat{f}_j^w | k_{j-1}) = \text{Var}\left(\sum_{i=1}^{n-j+1} W_{ij} F_{ij} | k_{j-1}\right) = \sum_{i=1}^{n-j+1} w_{ij}^2 \text{Var}(F_{ij} | k_{j-1}) \quad (2.45)$$

Η (2.45) ελαχιστοποιείται σε σχέση με W_{ij} όπου $j = 1, \dots, n$ όπου το γράμμα i υποδηλώνει το έτος ατυχήματος, δεδομένου ότι η διακύμανση θεωρείται ίση για όλα τα έτη ατυχημάτων. Η ελαχιστοποίηση πρέπει να γίνει υπό τον περιορισμό W_{ij} (βλέπε (2.44)) και η μέθοδος των παραγόντων Lagrange μπορεί να χρησιμοποιηθεί. Η Lagrangian συνάρτηση ορίζεται ως $L(x, \lambda) = k(x) + \lambda g(x)$, όπου k η συνάρτηση πρέπει να ελαχιστοποιηθεί όσον αφορά το x , g είναι ο περιορισμός και λ είναι ο πολλαπλασιαστής Lagrange. Το ελάχιστο (2.45) είναι:

$$\frac{\partial}{\partial w_{ij}} \left(\sum_{i=1}^{n-j+1} w_{ij}^2 \text{Var}(F_{ij} | k_{j-1}) + \lambda \left(1 - \sum_{i=1}^{n-j+1} w_{ij} \right) \right) = 0$$

Αυτό το ελάχιστο, από αυτές τις λειτουργίες, βρίσκεται όταν τα βάρη είναι αντιστρόφως ανάλογα με την διακύμανση της F_{ij} :

$$w_{ij} = \frac{\lambda}{2 \text{Var}(F_{ij} | k_{j-1})} \quad (2.46)$$

Το βάρος πρέπει να είναι αντιστρόφως ανάλογο της διακύμανσης, εάν η ελάχιστη διακύμανση είναι ο στόχος. Με άλλα λόγια, η διακύμανση των επιμέρους συντελεστών εξέλιξης πρέπει να είναι αντιστρόφως ανάλογη προς το βάρος. Το βάρος του συντελεστή εξέλιξης της chain-ladder είναι:

$$\frac{d_{i,j-1}}{\sum_{i=1}^{n-j+1} d_{i,j-1}}$$

Έτσι, η διακύμανση του ατομικού συντελεστή εξέλιξης είναι αντιστρόφως ανάλογη $d_{i,j-1}$. Ο παρονομαστής στην παραπάνω λειτουργία μπορεί να αντικατασταθεί από μια σταθερά αναλογικότητας. Η τρίτη παραδοχή του Mack μπορεί να ξαναγραφτεί, έτσι ώστε να είναι ξεκάθαρο ότι ο συντελεστής εξέλιξης της chain-ladder είναι στην πραγματικότητα ο εκτιμητής με ελάχιστη διακύμανση:

$$\text{Var}\left(\frac{D_{ij}}{D_{i,j-1}} \mid k_{j-1}\right) = \frac{g_j}{d_{i,j-1}} \quad (2.47)$$

(2.47) είναι αντιστρόφως ανάλογη προς το βάρος $d_{i,j-1}$ και πολλαπλασιάζεται με μια σταθερά αναλογικότητας g_j .

Η παράμετρος g_j πρέπει να εκτιμηθεί. Ο εκτιμητής που προτείνεται είναι για $j = 2, \dots, n$

$$\hat{g}_j = \frac{1}{n-j} \sum_{i=1}^{n-j+1} d_{i,j-1} \left(\frac{d_{ij}}{d_{i,j-1}} - \hat{f}_j \right)^2$$

Αυτός ο εκτιμητής είναι αμερόληπτος και θα αποδειχθεί ότι $E(\hat{g}_j) = g_j$. Πρώτα η ταυτότητα $(n-j)E(\hat{g}_j \mid k_{j-1})$ θα ανακτηθεί και αυτή η ιδιότητα μπορεί να χρησιμοποιηθεί για δούμε ότι $E(\hat{g}_j) = E(E(\hat{g}_j \mid k_{j-1})) = g_j$. Θα χρησιμοποιηθεί το γνωστό τέχνασμα της προσθήκης και αφαίρεσης μιας σταθεράς. Στην περίπτωση αυτή, η σταθερά f_j , θα χρησιμοποιηθεί.

$$(n-j)E(\hat{g}_j \mid k_{j-1}) = \sum_{i=1}^{n-j+1} d_{i,j-1} E\left(\left(\frac{D_{ij}}{D_{i,j-1}} - f_j - \hat{f}_j + f_j\right)^2 \mid k_{j-1}\right) \quad (2.48)$$

$$\begin{aligned}
&= \sum_{i=1}^{n-j+1} d_{i,j-1} E\left((F_{ij} - f_j)^2 \mid k_{j-1}\right) \\
&- 2 \sum_{i=1}^{n-j+1} d_{i,j-1} E\left((F_{ij} - f_j)(\hat{f}_j + f_j) \mid k_{j-1}\right) \\
&+ \sum_{i=1}^{n-j+1} d_{i,j-1} E\left((\hat{f}_j - f_j)^2 \mid k_{j-1}\right) \\
&= \sum_{i=1}^{n-j+1} d_{i,j-1} \text{Var}(F_{ij} \mid k_{j-1}) \\
&- 2 \sum_{i=1}^{n-j+1} d_{i,j-1} \text{Cov}(F_{ij}, \hat{f}_j \mid k_{j-1}) + \sum_{i=1}^{n-j+1} d_{i,j-1} \text{Var}(\hat{f}_j \mid k_{j-1})
\end{aligned}$$

Ο συντελεστής εξέλιξης της chain-ladder είναι ένας σταθμισμένος μέσος όρος των μεμονωμένων παραγόντων εξέλιξης. Χρησιμοποιώντας την τρίτη παραδοχή του Mack, είναι σαφές ότι $\text{Cov}(F_{kj}, F_{lj}) = 0$ και για $k \neq l$, προκύπτει ότι:

$$\text{Cov}(F_{ij}, \hat{f}_j \mid k_{j-1}) = \text{Cov}(F_{ij}, w_{ij} F_{ij} \mid k_{j-1}) = \text{Cov}\left(F_{ij}, \frac{D_{i,j-1}}{\sum_{l=1}^{n-j+1} D_{l,j-1}} F_{ij} \mid k_{j-1}\right) = \frac{d_{i,j-1}}{\sum_{l=1}^{n-j+1} d_{l,j-1}} \text{Var}(F_{ij} \mid k_{j-1})$$

Χρησιμοποιώντας αυτή την ιδιότητα, καθώς και την πρώτη και δεύτερη εξίσωση υποθέσεων του Mack (2.48) προκύπτει ότι:

$$(n-j)E(\hat{g}_j \mid k_{j-1}) = \left(\sum_{i=1}^{n-j+1} g_j\right) - 2g_j + g_j = (n-j)g_j$$

Χρησιμοποιώντας τον κανόνα του υπολογισμού των διπλών αναμενόμενων τιμών, είναι σαφές ότι, \hat{g}_j , είναι ένας αμερόληπτος εκτιμητής για g_j

$$E(\hat{g}_j) = E(E(\hat{g}_j | k_{j-1})) = E(g_j) = g_j.$$

2.6 Το μοντέλο του Mack και τη σύνδεσή του με την κατανομή Poisson

Ο Neuhaus (2006) δηλώνει ότι εάν οι προσαυξητικές απαιτήσεις C_{ij} δεσμευμένες μέχρι το έτος εξέλιξης $j-1$, κατανέμονται ως σύνθετες μεταβλητές Poisson. Έστω, C_{ij} , δεσμευμένη ως προς k_j θα είναι κατανεμημένη Poisson (H_j) (2.49)

και

$$f_j - 1 = \int_{-\infty}^{\infty} u dH_j(u) \quad \text{και} \quad g_j = \int_{-\infty}^{\infty} u^2 dH_j(u). \quad (2.50)$$

U_j είναι η ένταση της απαίτησης και η U_j έχει κατανομή H_j . Τώρα, αποδεικνύεται ότι, η χρήση (2.49) και (2.50) θα οδηγήσει στις υποθέσεις του Mack.

Ο ορισμός των σωρευτικών και προσαυξητικών απαιτήσεων δίνει:

$$E(D_{ij} | k_{i,j-1}) = E(D_{i,j-1} | k_{i,j-1}) + E(C_{ij} | k_{i,j-1}) = d_{i,j-1} + E(C_{ij} | k_{i,j-1}). \quad (2.51)$$

Με το μοντέλο του Mack, έχουμε $E(D_{ij} | k_{i,j-1}) = f_j d_{i,j-1}$ και η υπόθεση αυτή σε συνδυασμό με τον παραπάνω τύπο μας δίνει:

$$E(C_{ij} | k_{i,j-1}) = (f_j - 1) d_{i,j-1}. \quad (2.52)$$

Δεδομένου ότι, C_{ij} , είναι μια μεταβλητή Poisson, έχουμε από (2.2), ότι $E(C_{ij}) = E(N)E(U)$ όπου N είναι ο αριθμός των απαιτήσεων. Μπορούμε να δεσμεύσουμε ως προς, $k_{i,j-1}$, δεδομένου ότι η διαδικασία αριθμού απαιτήσεων έχουν ανεξάρτητες προσαυξητικές και η διαδικασία αριθμού απαιτήσεων είναι ανεξάρτητη από τη διαδικασία μεγέθους απαιτήσεων. Δεδομένου ότι, η U έχει κατανομή H_j , βρήκαμε ότι:

$$\begin{aligned} E(C_{ij} | k_{i,j-1}) &= E(N | k_{i,j-1})E(U | k_{i,j-1}) \\ &= E(N | k_{i,j-1}) \int_{-\infty}^{\infty} U dH_j(U) \\ &= d_{i,j-1}(f_j - 1). \end{aligned} \quad (2.53)$$

Ο τελευταίος υπολογισμός επιτυγχάνεται αφού $\int_{-\infty}^{\infty} U dH_j(U) = f_j - 1$ και αφήνοντας την παράμετρο Poisson αναλογική ή ίση με $d_{i,j-1}$. Με την προσθήκη $D_{i,j-1}$ και στις δύο πλευρές του (2.53), έχουμε επιβεβαιώσει την υπόθεση αριθμού 1.

Ο αριθμός 2 μπορεί να παρουσιαστεί με παρόμοιο τρόπο,

$$Var(D_{ij} | k_{i,j-1}) = Var((D_{i,j-1} + C_{ij}) | k_{i,j-1}) = Var(C_{ij} | k_{i,j-1}). \quad (2.54)$$

Η διακύμανση μιας σύνθετης Poisson βρίσκεται στον τύπο (2.3). Έστω, D_{ij} , θεωρούμε ότι είναι το μέγεθος της απαίτησης, λ είναι η παράμετρος της μεταβλητής Poisson N και U είναι το μέγεθος μιας απαίτησης, έχουμε ότι:

$$Var(D_{ij}) = \lambda Var U + \lambda(EU)^2 = \lambda E(U^2) \quad (2.55)$$

Για να δούμε πώς σχετίζεται η σύνθετη Poisson και το μοντέλο του Mack, οι δύο αυτές ιδιότητες μπορούν να συγκριθούν:

$$\text{Var}(D_{ij} | k_{i,j-1}) = \text{Var}(C_{ij} | k_{i,j-1}) = d_{i,j-1} g_j \quad (2.56)$$

$$\lambda E(U^2) = \lambda \int_{-\infty}^{\infty} U^2 dH_j(U) = \lambda g_j \quad (2.57)$$

Η ένταση του Poisson είναι ανάλογη ή ίση $d_{i,j-1}$ και αυτό δείχνει ότι το μοντέλο σύνθετης Poisson ικανοποιεί επίσης τη δεύτερη παραδοχή του Mack.

2.7 Αρνητικές αυξητικές απαιτήσεις

Οι αρνητικές προσαυξητικές απαιτήσεις είναι συνέπεια των ήδη αναφερθεισών απαιτήσεων που μειώνονται. Με αυτόν τον ορισμό δεν μπορούν να προκύψουν σωρευτικές απαιτήσεις. Οι αρνητικές προσαυξητικές απαιτήσεις μπορούν να εμφανιστούν λόγω της διάσωσης, των συντηρητικών εκτιμήσεων ή της υποκατάστασης (Kunkler 2006). Η υποκατάσταση είναι μια τεχνική που χρησιμοποιούν οι ασφαλιστικές εταιρείες όταν μια απαίτηση έχει καλυφθεί, αλλά υπάρχει μια τρίτη πλευρά που μπορεί να θεωρηθεί υπεύθυνη για την απαίτηση. Η ασφαλιστική εταιρεία υποβάλλει απαίτηση για αποζημίωση από τρίτο μέρος.

Έχουν προταθεί διαφορετικές λύσεις για το πώς να χειριστούν τις αρνητικές προσαυξητικές απαιτήσεις. Μια μέθοδος περιλαμβάνει την προσθήκη μιας θετικής σταθεράς σε όλες τις επιπρόσθετες απαιτήσεις. Μετά την ολοκλήρωση της ανάλυσης, η σταθερά αφαιρείται. Η μέθοδος αυτή παρέχει τα κατάλληλα αποτελέσματα εφόσον δεν υπάρχουν πάρα πολλές αρνητικές απαιτήσεις. Από την άλλη πλευρά, αυτή η διαδικασία καθιστά τη μεταβλητότητα του αποτελέσματος εξαρτώμενη από τη προσθήκη της σταθεράς, η οποία δεν μπορεί να θεωρηθεί εύλογη. (Kunkler 2006). Αν οι αρνητικές απαιτήσεις δεν χρησιμοποιούνται όπως

προτείνεται παραπάνω, το μοντέλο που θα χρησιμοποιηθεί χρειάζεται να διαχειριστεί τις αρνητικές απαιτήσεις. Αν η κατανομή καθορίζεται στο μοντέλο, πρέπει να οριστεί τόσο για αρνητικούς, όσο και για θετικούς αριθμούς. Μια προτεινόμενη ιδέα είναι η κανονική κατανομή, η οποία ορίζεται τόσο για τους θετικούς όσο και για τους αρνητικούς αριθμούς.

2.8 Προβλέψεις και σφάλματα πρόβλεψης

Η κάτω-αριστερή γωνία του τριγώνου εξέλιξης είναι γεμάτη με εκτιμήσεις \hat{D}_{ij} . Το τελευταίο έτος εξέλιξης, αντιπροσωπεύει την τελική απαίτηση, \hat{D}_{in} , για $i = 2, \dots, n$. Το μέσο τετραγωνικό σφάλμα (MSE), μπορεί να είναι ένα κατάλληλο μέτρο. Ο τύπος για το MSE της \hat{D}_{in} , θα βρεθεί παρακάτω. Το ριζικό τετραγωνικό σφάλμα πρόβλεψης (RMSE) θα χρησιμοποιηθεί ως μέτρο του σφάλματος πρόβλεψης. Το MSE θα αναφέρεται, επίσης, ως η διακύμανση πρόβλεψης.

Υπάρχουν ήδη παρατηρηθείσες τιμές στη πάνω-δεξιά γωνία στο τρίγωνο εξέλιξης. Το MSE πρέπει να λάβει υπόψη αυτά τα δεδομένα και εξαιτίας αυτού, προκύπτει ότι το MSE δοθέντος ότι k . Για να απλοποιηθεί ο συμβολισμός, θα αναφέρεται μόνο ως MSE του \hat{D}_{in} .

Το MSE της \hat{D}_{in} είναι:

$$E\left((D_{in} - \hat{D}_{in})^2 \mid k\right) = E\left((D_{in} - E(D_{in}) + E(D_{in}) - \hat{D}_{in})^2 \mid k\right) \quad (2.58)$$

$$= \text{Var}(D_{in} \mid k) + E\left(E(D_{in} \mid k) - \hat{D}_{in}\right)^2 \quad (2.59)$$

$$= \text{Var}(D_{in} \mid k) + \left(E(D_{in} \mid k) - \hat{D}_{in}\right)^2. \quad (2.60)$$

Δεδομένου ότι, $E(D_{in} | k)$ και \hat{D}_{in} , είναι βαθμιδωτές, η εξωτερική απόκλιση αφαιρείται στο δεύτερο όρο της (2.59) και είναι απαραίτητο να εξαρτάται μόνο από το k στην στοχαστική μεταβλητή D_{in} . Ο πρώτος όρος στη (2.60), είναι η μεταβλητή γύρω από την πραγματική τιμή D_{in} και θα είναι πάντοτε παρούσα. Ο δεύτερος όρος στη (2.60), είναι ένα μέτρο για το πόσο ο \hat{D}_{in} χάνει τον στόχο του και αναφέρεται ως η διακύμανση της εκτίμησης (Mack 1994).

ΚΕΦΑΛΑΙΟ 3 : Ανάλυση δεδομένων από ασφαλιστικές απαιτήσεις αυτοκινήτων με χρήση μοντέλων στοχαστικής αλυσίδας

3.1 Δείγμα

Το σύνολο δεδομένων που χρησιμοποιήθηκε για την ανάλυση είναι ένα σύνολο ασφαλιστικών δεδομένων για τις απαιτήσεις αυτοκινήτων από την TrygVesta. Οι ασφαλιστικές απαιτήσεις οργανώθηκαν κατά έτος ατυχήματος και κατά έτος εξέλιξης. Περιέχει τον αριθμό των δηλωθεισών απαιτήσεων, το μέγεθος των απαιτήσεων που είχαν καταβληθεί, τον αριθμό των απαιτήσεων-RBNS, το ποσό των αποθεματικών για τις απαιτήσεις-RBNS συνδυασμένες με τις καταβληθείσες απαιτήσεις και τον αριθμό των διακανονισμένων απαιτήσεων.

Το σύνολο δεδομένων περιέχει τον αριθμό και το μέγεθος των απαιτήσεων που θα χρησιμοποιηθούν κατά την εφαρμογή των στοχαστικών μοντέλων chain-ladder. Δεν έχει γίνει ανάλυση σχετικά με αυτό το σύνολο δεδομένων και δεν έχει προστεθεί κάποιο αποθεματικό και, λόγω αυτού, οι παρατηρήσεις αυτές φαίνονται να είναι οι πλέον κατάλληλες για περαιτέρω ανάλυση. Χρησιμοποιώντας αυτές τις παρατηρήσεις θα προβούμε σε προβλέψεις για τον συνολικό αριθμό και το συνολικό μέγεθος μελλοντικών απαιτήσεων. Τα σύνολα δεδομένων παρουσιάζονται στο προσάρτημα 1.

Όπως αναφέρεται και στην εισαγωγή, οι IBNR-απαιτήσεις είναι οι πιο ενδιαφέρουσες να προβλέψουμε. Το ποσό που καταβλήθηκε σε ένα πρώιμο έτος εξέλιξης χρησιμοποιείται για να προβλέψει τις συνολικές μελλοντικές πληρωμές. Η πρόβλεψη του μεγέθους των απαιτήσεων περιλαμβάνει τόσο τις πληρωμές-RBNS, όσο και τις πληρωμές-IBNR. Η ασφαλιστική εταιρεία ενδέχεται να έχει ειδοποιηθεί για μια απαίτηση και να έχει προβεί σε εκτίμηση αποθεματικού για την απαίτηση αυτή. Η εύρεση των IBNR-απαιτήσεων πρέπει

απλώς να αποσύρει τις RBNS-απαιτήσεις. Δεδομένου ότι, η ασφαλιστική εταιρεία δεν έχει πληροφορίες πριν από την πρώτη κοινοποίηση της απαίτησης, ο αριθμός των απαιτήσεων από το σύνολο δεδομένων είναι ίσος με τον αριθμό των απαιτήσεων-IBNR.

Με δεδομένο το τρίγωνο εξέλιξης, όπως απεικονίζεται στο Σχήμα 1, είναι φυσικό να υπολογίσουμε το υπόλοιπο τρίγωνο, αλλά το πιο σημαντικό είναι η τελική απαίτηση, \hat{D}_{in} . Η ασφαλιστική εταιρεία χρειάζεται αποθεματικό, R_i , για την κάλυψη μελλοντικών απαιτήσεων. Έχουμε ότι, $\hat{R}_i = \hat{D}_{in} - D_{i,n-i+1}$. Οι εκτιμήσεις, R_i , με αντίστοιχα σφάλματα πρόβλεψης θα παρουσιαστούν σε πίνακες.

3.2 Στοχαστικά μοντέλα στην ανάλυση

Και τα δύο σύνολα δεδομένων περιέχουν αρνητικές προσαυξητικές απαιτήσεις. Αυτό περιορίζει ορισμένα από τα στοχαστικά μοντέλα που παρουσιάστηκαν νωρίτερα. Ούτε το μοντέλο Poisson ούτε το αρνητικό διωνυμικό μοντέλο μπορεί να περιέχει αρνητικές προσαυξητικές απαιτήσεις, εάν χρησιμοποιείται ο κανονικός εκτιμητής μέγιστης πιθανοφάνειας των παραμέτρων. Όσο το άθροισμα των προσαυξητικών απαιτήσεων ανήκουν σε ένα έτος εξέλιξης δεν είναι αρνητικό, το πρόβλημα αυτό μπορεί να λυθεί χρησιμοποιώντας μία λογαριθμική πιθανοφάνεια (Renshaw 1998). Στο σύνολο δεδομένων που θα χρησιμοποιηθούν σε αυτή την ανάλυση, εμφανίζονται αρκετές αρνητικές απαιτήσεις, ιδιαίτερα κατά το δεύτερο έτος εξέλιξης. Ακόμη και το άθροισμα του αριθμού των προσαυξητικών απαιτήσεων αποδεικνύεται αρνητικό, ανεξάρτητα από τη δυνατότητα χρήσης της Poisson ή του αρνητικού διωνυμικού μοντέλου. Η στενή σύνδεση μεταξύ της Poisson και του αρνητικού διωνυμικού μοντέλου περιγράφηκε προηγουμένως και μπορεί να χρησιμοποιηθεί μια προσέγγιση του αρνητικού διωνυμικού μοντέλου από την κανονική

κατανομή. Αυτό το μοντέλο μπορεί να χειριστεί τους αρνητικούς αριθμούς και μπορεί επίσης να δημιουργήσει εκτιμήσεις αποθεμάτων και σφάλματα πρόβλεψης.

Το μοντέλο του Mack χρησιμοποιείται επίσης για να βρεθούν εκτιμήσεις αποθεμάτων και σφάλματα πρόβλεψης. Το μοντέλο του Mack κάνει μόνο παραδοχή σχετικά με τις δύο πρώτες ροπές και δεν φαίνεται να υπάρχει προφανής λόγος για τον οποίο αυτό θα ήταν πρόβλημα όταν προκύπτουν αρνητικές απαιτήσεις, εφόσον οι προσαυξητικές απαιτήσεις είναι θετικές.

3.3 Μια σημαντική τοποθέτηση για τις στοχαστικές υποθέσεις της αλυσίδας-σκάλας

Τα δύο μοντέλα που χρησιμοποιούνται στην ανάλυση κάνουν παρόμοιες υποθέσεις σχετικά με τις δύο πρώτες ροπές. Πρώτον, θεωρείται ότι υπάρχει γραμμική σχέση μεταξύ μιας ασφαλιστικής απαίτησης κατά το έτος εξέλιξης $j+1$ και μιας απαίτησης κατά το έτος εξέλιξης j , και ότι ο παράγοντας της γραμμικής σχέσης είναι ο παράγοντας της chain-ladder. Μια αντίστοιχη γραμμική σχέση θεωρείται ότι υπάρχει και για τη δεύτερη ροπή. Πριν από την εφαρμογή των μοντέλων στο σύνολο δεδομένων, θα εξεταστούν αυτές οι υποθέσεις.

3.3.1 Η μεροληπτικότητα της chain-ladder μεθόδου

Το μοντέλο της chain-ladder υποθέτει μια γραμμική σχέση στις απαιτήσεις μεταξύ των ετών εξέλιξης όπου χρησιμοποιείται ο συντελεστής εξέλιξης της chain-ladder. Η μέθοδος της chain-ladder δεν περιέχει σταθερό όρο και σε αυτή την περίπτωση είναι ενδιαφέρον να δούμε αν ένα γραμμικό μοντέλο με σταθερό όρο θα μπορούσε να προβλέψει ακόμα καλύτερα τις απαιτήσεις. Ο Halliwell (2007) πρότεινε ότι η μεροληψία της μεθόδου

αλυσίδας-σκάλας θα μπορούσε να δοκιμαστεί συγκρίνοντας το πιο γενικό γραμμικό μοντέλο, όπου ο σταθερός όρος δεν αναγκάστηκε να περάσει από την αρχή των αξόνων.

Η γραμμική σχέση που λαμβάνεται στα δύο μοντέλα είναι:

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

Ένα πιο γενικό γραμμικό μοντέλο θα μπορούσε να εκφραστεί όπως

$$D_{ij} = \beta_0 + \beta_j d_{i,j-1} + \varepsilon \quad (3.2)$$

Όπου, D_{ij} , είναι μια στοχαστική μεταβλητή, $d_{i,j-1}$ θεωρείται γνωστή, β_0 και β_j είναι παραμέτροι που πρέπει να εκτιμηθούν, ε είναι ο όρος σφάλματος και πρέπει να ακολουθήσει την ίδια κατανομή όπως D_{ij} . Τρία μοντέλα θα εφαρμοστούν στα δεδομένα. Η διαφορά μεταξύ των τριών μοντέλων είναι η αλλαγή των εκτιμητών των β_0 και β_j . Οι εκτιμητές που θα χρησιμοποιηθούν είναι ο συντελεστής εξέλιξης της chain-ladder \hat{f}_j και οι εκτιμητές ελαχίστου τετραγώνου \hat{f}_0^{LS} και \hat{f}_j^{LS} . Τα τρία μοντέλα είναι

1. το μοντέλο chain-ladder, $\hat{\beta}_0 = 0$ και $\hat{\beta}_j = \hat{f}_j$
2. το γενικό μοντέλο, $\hat{\beta}_0 = \hat{f}_0^{LS}$ και $\hat{\beta}_j = \hat{f}_j^{LS}$
3. και το περιορισμένο μοντέλο $\hat{\beta}_0 = 0$ και $\hat{\beta}_j = \hat{f}_j^{LS}$.

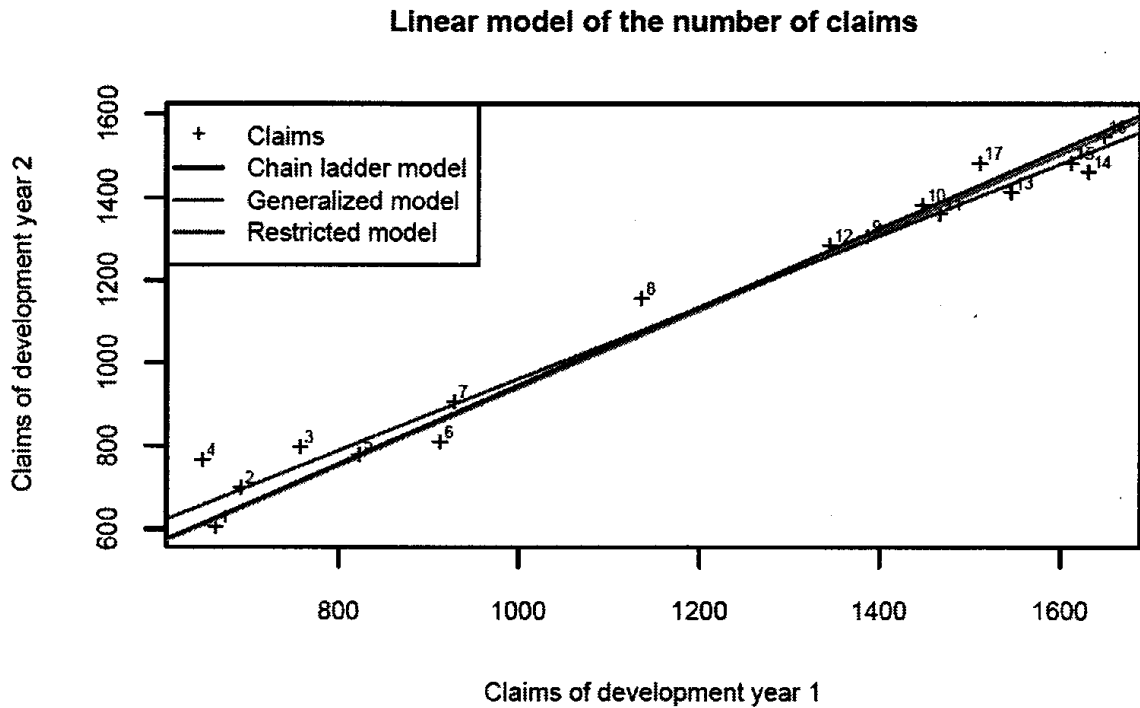
Το γενικό μοντέλο προσφέρει μεγαλύτερη ευελιξία εξαιτίας της πιθανότητας μιας δεύτερης παραμέτρου, ενός σταθερού όρου. Είναι λογικό να υποθέσουμε ότι αυτό το μοντέλο μπορεί να

εφαρμοστεί καλύτερα σε ένα σύνολο δεδομένων παρά τα άλλα δύο μοντέλα. Αυτή η υπόθεση είναι η βάση για την υπόθεση της μεροληπτικότητας της chain-ladder μεθόδου.

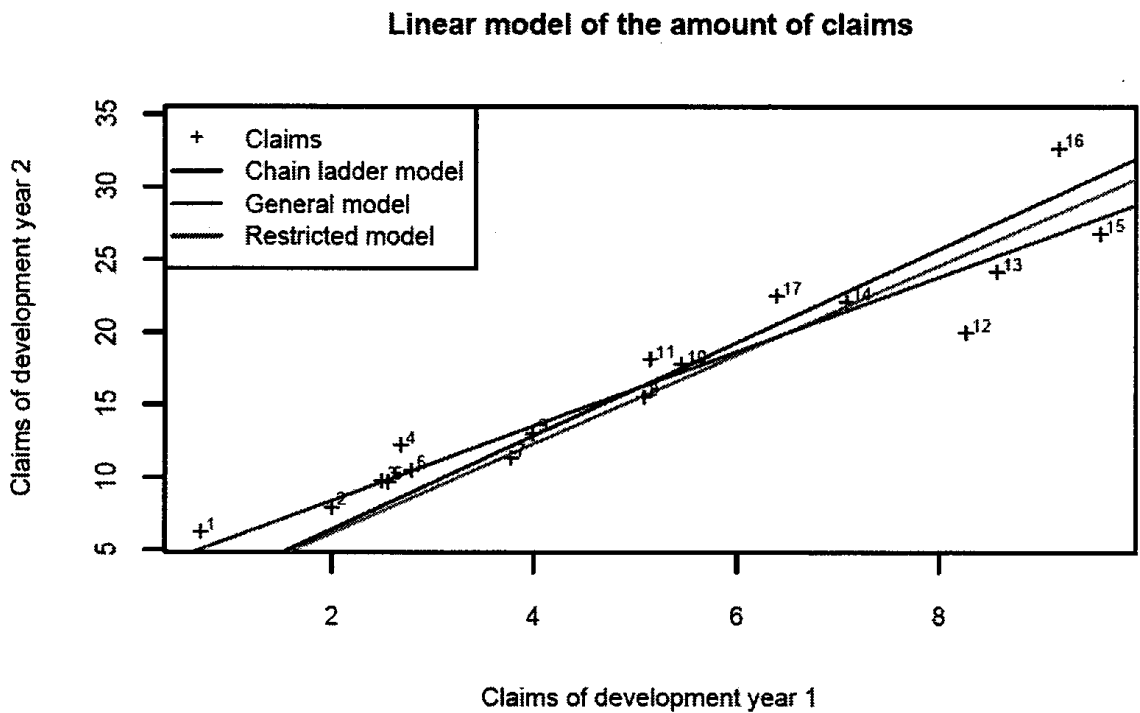
Τα τρία μοντέλα εφαρμόστηκαν στο σύνολο δεδομένων του αριθμού και του ποσού των απαιτήσεων. Τα δύο σύνολα δεδομένων έχουν 18 έτη εξέλιξης τα οποία μπορούν να συγκριθούν με το προηγούμενο έτος εξέλιξης. Τουλάχιστον τρεις παρατηρήσεις χρειάζονται σε κάθε έτος εξέλιξης, οπότε είναι δυνατό να γίνουν 16 γραφήματα, αλλά η ανάλυση με λίγες παρατηρήσεις είναι λιγότερο αξιόπιστη. Το γενικό γραμμικό μοντέλο και το περιορισμένο γραμμικό μοντέλο γίνονται στην R χρησιμοποιώντας τις ακόλουθες εντολές:

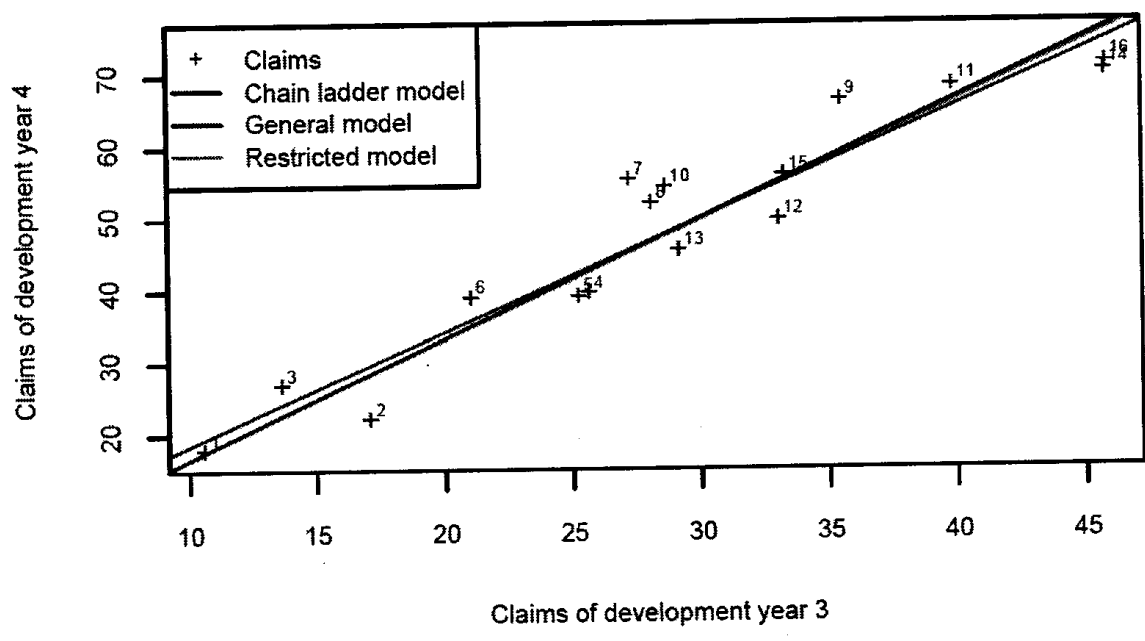
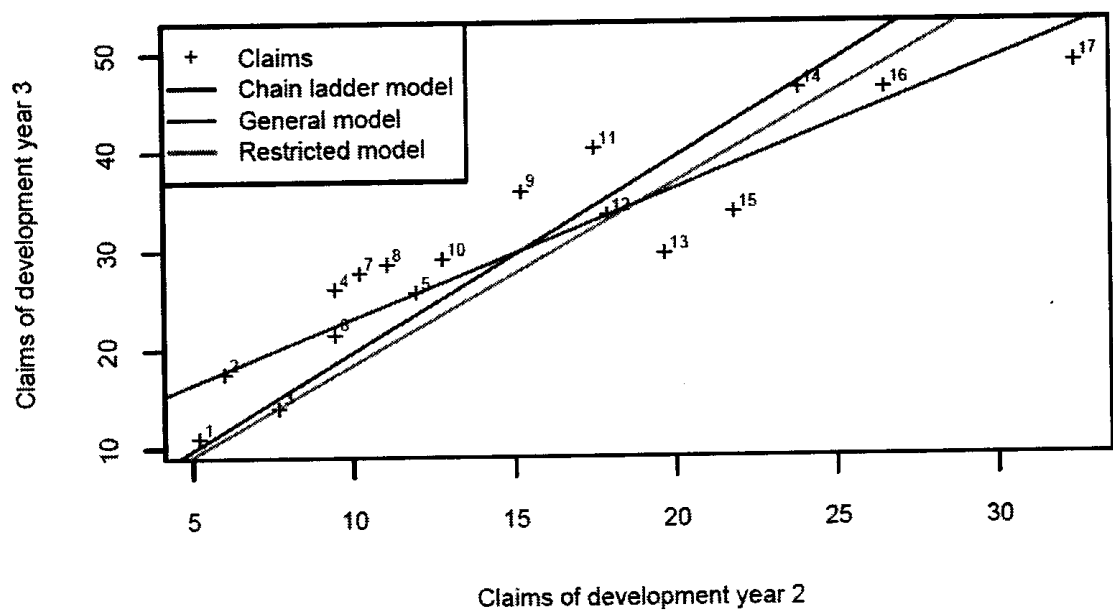
- ❖ $\text{lm}(\text{developmentyear}(j+1) \sim \text{developmentyear}(j) - 1)$
- ❖ $\text{lm}(\text{developmentyear}(j+1) \sim \text{developmentyear}(j))$

Τα γραφικά αποτελέσματα παρουσιάζονται παρακάτω. Παρόλο που τα μοντέλα εφαρμόστηκαν για όλα τα χρόνια εξέλιξης, παρουσιάζονται μόνο λίγα γραφήματα. Τα τρία διαφορετικά μοντέλα καταλήγουν να έχουν σχεδόν ταυτόσημες εκτιμήσεις των παραμέτρων στα γραφήματα των ακόλουθων ετών εξέλιξης.



Σχήμα 4 : Τρία γραμμικά μοντέλα τοποθετημένα στα σωρευτικά δεδομένα του αριθμού των απαιτήσεων. Τα δεδομένα του έτους εξέλιξης 2 σχεδιάζονται ως συνάρτηση των δεδομένων του έτους εξέλιξης 1.





Σχήμα 5: Τρία γραμμικά μοντέλα προσαρμόζονται στα σωρευτικά δεδομένα του ποσού των απαιτήσεων. Τα δεδομένα σχεδιάζονται ως συνάρτηση του προηγούμενου έτους εξέλιξης. Τα τρία διαφορετικά γραφήματα γίνονται για το έτος εξέλιξης 2,3 και 4.

Η υπόθεση σχετικά με τη γραμμικότητα της πρώτης ροπής δεν αμφισβητήθηκε πραγματικά. Ωστόσο, με την εξέταση των γραφικών αποτελεσμάτων, φαίνεται ότι ένα γραμμικό μοντέλο της μορφής (3.2) ταιριάζει καλά στα σύνολα δεδομένων. Εάν ήταν αναγκαία η περαιτέρω διερεύνηση, θα μπορούσε να γίνει ανάλυση της διακύμανσης. Η γραμμική σχέση στην πρώτη ροπή είναι αποδεκτή και η εστίαση αυτής της ανάλυσης είναι η μεροληπτικότητα της chain-ladder.

Τα γραμμικά μοντέλα του αριθμού των απαιτήσεων είναι σχεδόν πανομοιότυπα ήδη στο πρώτο έτος εξέλιξης, αλλά μπορούν να αναφερθούν ορισμένες διαφορές. Τόσο η chain-ladder, όσο και το περιορισμένο μοντέλο αναγκάζονται να περάσουν από την αρχή (το μηδέν) και φαίνεται ότι οδηγούν σε μια υποτίμηση των μικρών απαιτήσεων, δεδομένου ότι οι απαιτήσεις των πρόωρων ετών ατυχήματος φαίνονται να είναι μικρότερες σε σχέση με τα μεταγενέστερα έτη. Το παράγωγο του μοντέλου chain-ladder και το περιορισμένο μοντέλο είναι μεγαλύτερα για να αντισταθμίσουν το θετικό σταθερό όρο στο γενικό μοντέλο και τα δύο μοντέλα μπορεί να υπερεκτιμούν τις απαιτήσεις είτε στα τελευταία έτη ατυχημάτων, είτε στις μεγαλύτερες απαιτήσεις από το γενικό μοντέλο.

Για να εξεταστεί η μεροληπτικότητα της chain-ladder, υποθέτουμε ότι το γενικό μοντέλο είναι καλύτερο από τα άλλα δύο. Αν β_0 είναι σημαντικά διαφορετικό από το μηδέν, θα έδινε ένα λόγο να πιστέψουμε ότι η μέθοδος της chain-ladder είναι μεροληπτική. Για να γίνουν κάποια συμπεράσματα σχετικά με την πρώτη παράμετρο β_0 μπορεί να θεωρηθεί η κατανομή των απαιτήσεων D_{ij} . Υποθέσουμε ότι, D_{ij} , ακολουθεί μια κανονική κατανομή, οι t-τιμές της παραμέτρου ακολουθούν την κατανομή t με n-2 βαθμούς ελευθερίας. Χρησιμοποιώντας ένα επίπεδο σημαντικότητας 0,1 του έτους εξέλιξης 2 και 9 προκύπτουν

σημαντικά αποτελέσματα για την πρώτη παράμετρο β_0 (βλέπε προσάρτημα 3). Αυτό υποδηλώνει ότι ένα γραμμικό μοντέλο που περιέχει ένα σταθερό όρο που είναι διαφορετικός από το μηδέν θα ταιριάζει καλύτερα στα δεδομένα για αυτά τα χρόνια εξέλιξης.

Τα τρία γραφήματα για το ποσό των απαιτήσεων δείχνουν ότι για ένα υψηλότερο έτος εξέλιξης τα μοντέλα γίνονται περισσότερο παρόμοια. Μετά το έτος εξέλιξης 3 είναι σχεδόν πανομοιότυπα. Τα γραμμικά μοντέλα του ποσού των απαιτήσεων παρουσιάζουν πιο αποκλίνουσα συμπεριφορά από τα μοντέλα για τον αριθμό απαιτήσεων. Η ίδια τάση είναι εμφανής και για τον αριθμό των απαιτήσεων. Το γενικό μοντέλο έχει ένα θετικό σταθερό όρο για τα τρία έτη εξέλιξης που εμφανίζονται παραπάνω, γεγονός που αναγκάζει το μοντέλο chain-ladder και το περιορισμένο μοντέλο να αντισταθμιστεί με υψηλότερη κλίση. Η κλίση του μοντέλου chain-ladder υπερβαίνει την κλίση του περιορισμένου μοντέλου.

Εάν οι ισχυρισμοί D_{ij} θεωρούνται ότι ακολουθούν κανονική κατανομή, οι εκτιμήσεις των β_0 είναι σημαντικά διαφορετικές από το μηδέν που ελήφθησαν για το έτος εξέλιξης 2 και 3 (χρησιμοποιώντας ένα επίπεδο σημαντικότητας 0,1). Μόνο 5 από 16 εκτιμήσεις παραμέτρων ήταν αρνητικές για τον αριθμό των απαιτήσεων. Τρεις αρνητικές εκτιμήσεις της παραμέτρου β_0 υπολογίστηκαν για το σύνολο δεδομένων του ποσού των απαιτήσεων. Αυτό σημαίνει ότι ο σταθερός όρος πρέπει να είναι θετικός. Είτε πρόκειται για υποτίμηση των μικρών απαιτήσεων, είτε για υπερεκτίμηση μεγαλύτερων απαιτήσεων, ή για υποτιμήσεις κατά τα πρώτα έτη ατυχημάτων και η υπερεκτίμηση των τελευταίων ετών ατυχημάτων είναι δύσκολο να προσδιοριστεί από αυτές τις αναλύσεις.

3.3.2 Η διακύμανση των απαιτήσεων

Τα δύο μοντέλα που θα χρησιμοποιηθούν στην ανάλυση έχουν την ίδια διατύπωση στην παραδοχή της διακύμανσης:

$$Var(D_{ij} | k_{j-1}) = c_j d_{i,j-1} \text{ όπου } c_j \text{ είναι μια σταθερά} \quad (3.3)$$

Ο συντελεστής εξέλιξης της chain-ladder είναι ένας σταθμισμένος μέσος όρος των επιμέρους συντελεστών εξέλιξης και αποδείχθηκε (2.46) ότι η διακύμανση του συντελεστή εξέλιξης της chain-ladder έπρεπε να είναι αντιστρόφως ανάλογη με τα βάρη. Η ισοδύναμη παραδοχή της διακύμανσης D_{ij} δίνεται στο (3.3). Αν επιλέξουμε έναν άλλο συντελεστή εξέλιξης, η παραδοχή της διακύμανσης ενδέχεται να χρειαστεί να τροποποιηθεί για να επιτευχθεί η ελάχιστη διακύμανση. Δύο άλλοι συντελεστές εξέλιξης προτείνονται ως εναλλακτικοί συντελεστές εξέλιξης και είναι επίσης ένας σταθμισμένος μέσος όρος των επιμέρους συντελεστών εξέλιξης. Αυτά είναι

$$\hat{f}_j^{mean} = \frac{1}{n-j+1} \sum_{i=1}^{n-j+1} w_{ij} f_{ij}, \text{ όπου } w_{ij} = 1 \text{ για } i = 1, \dots, n-j+1$$

$$\hat{f}_j^{ls} = \frac{\sum_{i=1}^{n-j+1} d_{i,j-1} d_{ij}}{\sum_{i=1}^{n-j+1} d_{i,j-1}^2} = \sum_{i=1}^{n-j+1} \frac{d_{i,j-1}^2}{\sum_{i=1}^{n-j+1} d_{i,j-1}^2} \frac{d_{ij}}{d_{i,j-1}}$$

$$= \sum_{i=1}^{n-j+1} \frac{d_{i,j-1}^2}{\sum_{i=1}^{n-j+1} d_{i,j-1}^2} f_{ij} \propto \sum_{i=1}^{n-j+1} w_{ij} f_{ij} \text{ όπου } w_{ij} = d_{i,j-1}^2 \text{ για } i = 1, \dots, n-j+1.$$

\hat{f}_j^{mean} είναι ο μέσος όρος των ατομικών συντελεστών εξέλιξης και \hat{f}_j^{ls} είναι ο ελάχιστος τετραγωνικός εκτιμητής των ατομικών συντελεστών εξέλιξης. Μια ανάλυση μπορεί να πραγματοποιηθεί με τη χρήση των διαφορετικών συντελεστών εξέλιξης και της παραδοχής της διακύμανσης που μπορεί να προκύψει από (2.46).

Τα τρία κατάλοιπα γίνονται:

$$r_{ij}^{mean} = \frac{d_{ij} - \hat{f}_j^{mean} d_{i,j-1}}{1}$$

$$r_{ij} = \frac{d_{ij} - \hat{f}_j d_{i,j-1}}{\sqrt{d_{i,j-1}}}$$

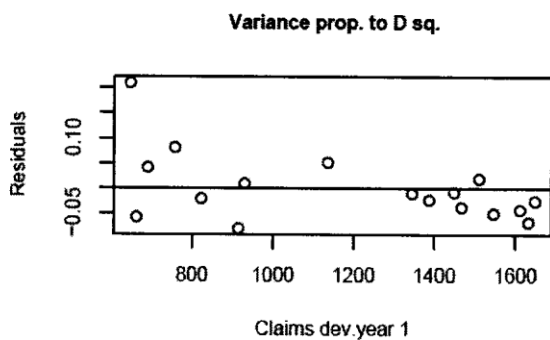
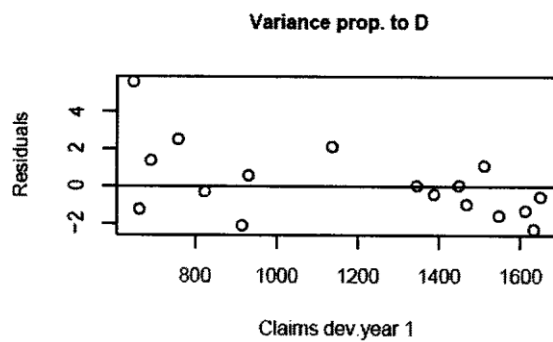
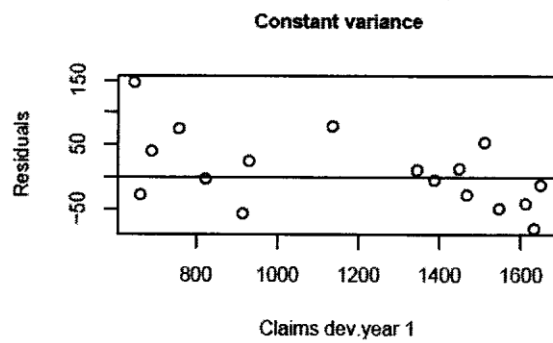
$$r_{ij}^{ls} = \frac{d_{ij} - \hat{f}_j^{ls} d_{i,j-1}}{d_{i,j-1}}.$$

Τα τρία κατάλοιπα γραφημάτων παράγονται για όλα τα j . r_{ij}^{mean} υποθέτει ότι D_{ij} έχει διακυμάνσεις, r_{ij} υποθέτει ότι D_{ij} έχει διακύμανση ανάλογη προς $d_{i,j-1}$ και r_{ij}^{ls} υποθέτει ότι D_{ij} έχει διακύμανση είναι ανάλογη προς $d_{i,j-1}^2$. Αν ένα από τα παραπάνω κατάλοιπα φαίνεται να έχει πιο τυχαία συμπεριφορά, η επιλογή του συντελεστή εξέλιξης θα πρέπει να επανεξεταστεί.

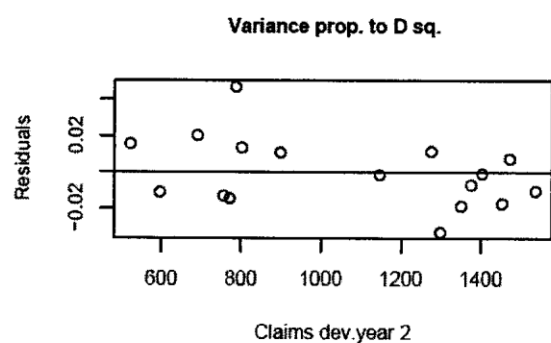
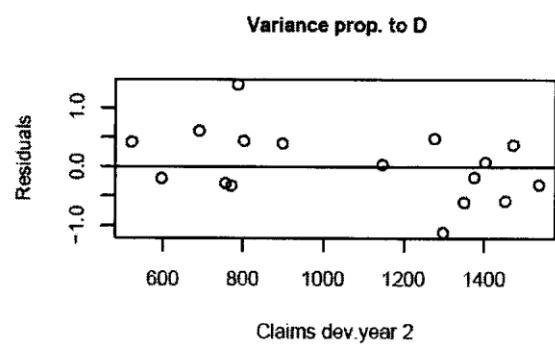
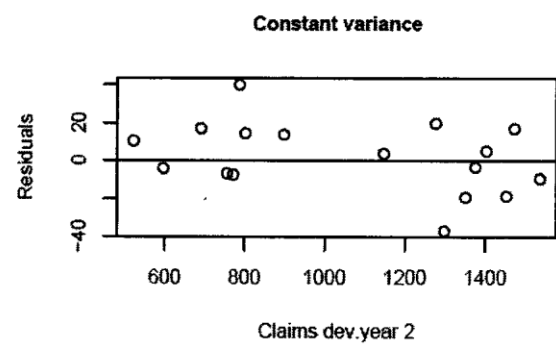
Αυτά τα κατάλοιπα γραφήματα δημιουργήθηκαν για το σύνολο δεδομένων του αριθμού και του ποσού των απαιτήσεων. Από τα 19 έτη εξέλιξης στο σύνολο δεδομένων, 14 έχουν 6 ή περισσότερες παρατηρήσεις και έχουν δημιουργηθεί κατάλοιπα γραφήματα για αυτά. Έχει παρατηρηθεί μια πιθανή εξέλιξη της διακύμανσης, τάση ή διαφορά μεταξύ των συντελεστών εξέλιξης. Τα τρία κατάλοιπα γραφήματα για το έτος εξέλιξης 2 και 3 παρουσιάζονται παρακάτω για τον αριθμό των απαιτήσεων. Τα κατάλοιπα γραφήματα είναι

με οπτική εξέταση όμοια για τους τρεις διαφορετικούς συντελεστές εξέλιξης. Εξαιτίας της ομοιότητας των γραφημάτων, μόνο τα αποτελέσματα που χρησιμοποιήθηκαν r_{ij} περιλαμβάνονται στο Παράρτημα 4 για τα υπόλοιπα έτη εξέλιξης, για τον αριθμό των απαιτήσεων και το ποσό των απαιτήσεων.

Residuals of development year 2



Residuals of development year 3



Σχήμα 6: Τα τρία κατάλοιπα γραφήματα στην αριστερή πλευρά r_{i2}^{mean} , r_{i2} και r_{i2}^{ls} σχεδιάζονται σε σχέση με τις απαιτήσεις από το έτος εξέλιξης 1 για τον αριθμό των απαιτήσεων. Τα τρία κατάλοιπα γραφήματα στη δεξιά πλευρά είναι r_{i3}^{mean} , r_{i3} και r_{i3}^{ls} σχεδιάζονται σε σχέση με τις απαιτήσεις του έτους εξέλιξης 2 για τον αριθμό των απαιτήσεων. Η παραδοχή διακύμανσης που ανήκει στο κατάλοιπο γράφεται πάνω από όλα τα γραφήματα.

Αναλύονται πρώτα τα κατάλοιπα γραφημάτων για τον αριθμό των απαιτήσεων. Οι τρεις διαφορετικοί συντελεστές εξέλιξης δημιουργούν εντυπωσιακά παρόμοια κατάλοιπα γραφημάτων. Φαίνεται ότι δεν υπάρχει διαφορά στην τυχαία συμπεριφορά τριών συντελεστών εξέλιξης. Υπάρχει μια υπερεκπροσώπηση θετικών κατάλοιπων για μικρές ζημιές και αρνητικών κατάλοιπων για μεγάλες ζημιες. Αυτό υποδηλώνει ότι υπάρχει μια τάση, η οποία ήταν επίσης ελάχιστα ορατή στο γραμμικό μοντέλο που δημιουργήθηκε στο προηγούμενο κεφάλαιο. Αυτή η τάση είναι εξίσου παρούσα για όλους τους συντελεστές εξέλιξης. Μετά το έτος εξέλιξης 9, η τάση δεν είναι πλέον ορατή, αλλά στο σημείο αυτό το τρίγωνο εξέλιξης είναι σχεδόν πλήρως ανεπτυγμένο. Κατά το έτος εξέλιξης 6,7 και 8, τα κατάλοιπα είναι μεγαλύτερα για μικρές ζημιές και τα κατάλοιπα είναι μικρότερα για μεγάλες απαιτήσεις και αυτό ισχύει εξίσου για τους τρεις συντελεστές εξέλιξης.

Στο σύνολο δεδομένων του ποσού των απαιτήσεων η οπισθοδρομική τάση είναι ορατή για τα δύο πρώτα έτη εξέλιξης και το έτος εξέλιξης 6. Εκτός από αυτό, το σύνολο δεδομένων δείχνει μια πιο τυχαία συμπεριφορά στα κατάλοιπα γραφημάτων. Η εξαίρεση είναι το έτος εξέλιξης 7, το οποίο φαίνεται να δείχνει μια μη τυχαία συμπεριφορά.

Κανένας από τους συντελεστές εξέλιξης δεν υπολογίζει καθαρά τυχαία κατάλοιπα γραφημάτων. Η τάση θα μπορούσε πιθανότατα να αφαιρεθεί συμπεριλαμβάνοντας μια

δεύτερη παραμέτρο στο μοντέλο, την παρακολούθηση. Δεδομένου ότι δεν υπάρχει σχεδόν καμία διαφορά μεταξύ των συντελεστών εξέλιξης \hat{f}_j , \hat{f}_j^{mean} και \hat{f}_j^{ls} , θα συνεχιστεί η χρήση του συντελεστή εξέλιξης της chain-ladder και η αντίστοιχη παραδοχή της διακύμανσης.

3.4 Η προσέγγιση της αρνητικής διωνυμικής κατανομή από την κανονική κατανομή

3.4.1 Το μοντέλο

Μια προσέγγιση του αρνητικού διωνυμικού μοντέλου από την κανονική κατανομή χρησιμοποιήθηκε για την ανάλυση του συνόλου δεδομένων από την TrygVesta. Αρχικά το μοντέλο εισάγεται γενικά. Ένα γραμμικό και ένα γενικευμένο γραμμικό μοντέλο θα προσαρμοστούν στα δεδομένα για τη διαδικασία πρόβλεψης μελλοντικών απαιτήσεων. Αυτό θα παρουσιαστεί στον γενικό τύπο, αλλά θα δοθεί επίσης ένα παράδειγμα χρησιμοποιώντας τις διαστάσεις του συνόλου δεδομένων από την TrygVesta. Το αρνητικό διωνυμικό μοντέλο παρουσιάστηκε στο (2.37) και ο μέσος όρος και η διακύμανση του D_{ij} δοθέντος ότι $d_{i,j-1}$:

$$E(D_{ij} | d_{i,j-1}) = f_j d_{i,j-1} \quad \text{και} \quad \text{Var}(D_{ij} | d_{i,j-1}) = f_j (f_j - 1) d_{i,j-1} \quad \text{με } j = 2, \dots, n. \quad (3.4)$$

Λόγω των αρνητικών προσυζητικών απαιτήσεων στο σύνολο δεδομένων, ο συντελεστής εξέλιξης γίνεται μικρότερος από ένα και αυτό δημιουργεί αρνητική διακύμανση. Το μοντέλο που θα χρησιμοποιηθεί πρέπει να διαχειρίζεται τις θετικές και τις αρνητικές τιμές των προσαυξητικών απαιτήσεων. Η κανονική κατανομή είναι μια καλή πρόταση. Δεδομένου ότι

το αρνητικό διωνυμικό μοντέλο δεν ταιριάζει στο σύνολο δεδομένων, η δεσμευμένη κατανομή της D_{ij} θεωρείται ότι προσεγγιστικά ακολουθεί μια κανονική κατανομή.

Στην ανάλυση θα δοθεί έμφαση στις ιδιότητες f_{ij} , στους ατομικούς συντελεστές εξέλιξης.

Στο μοντέλο F_{ij} θεωρείται μια στοχαστική μεταβλητή, και από το σύνολο δεδομένων παρατηρούνται τιμές f_{ij} των οποίων οι υλοποιήσεις $F_{ij} \cdot f_{ij}$ παρατηρούνται στην πάνω-δεξιά γωνία του τριγώνου εξέλιξης. Η \hat{f}_{ij} είναι η προβλεπόμενη τιμή f_{ij} της οποίας θα αντικαταστήσει τα κενά σημεία στην κάτω-αριστερή γωνία του τριγώνου εξέλιξης. Αυτά είναι κατά προσέγγιση ανεξάρτητα και κανονικά κατανεμημένα για το έτος εξέλιξης j .

Έστω $F_{ij} = \frac{D_{ij}}{w_{ij}}$ όπου $w_{ij} = d_{i,j-1}$. Η μέση τιμή και η διακύμανση είναι :

$$E\left(\frac{D_{ij}}{w_{ij}} \mid d_{i,j-1}\right) = f_j \quad \text{και} \quad \text{Var}\left(\frac{D_{ij}}{w_{ij}} \mid d_{i,j-1}\right) = \frac{\phi_j}{w_{ij}} \quad (3.5)$$

Η μεταβλητή w_{ij} έχει εισαχθεί επειδή ένα σταθμισμένο γραμμικό μοντέλο θα χρησιμοποιηθεί στην ανάλυση της εύρεσης των άγνωστων ατομικών συντελεστών εξέλιξης.

Τα βάρη στην ανάλυση είναι $W_{ij} = \frac{w_{ij}}{\phi_j}$. Τα βάρη είναι αντιστρόφως ανάλογα με τη διακύμανση, έτσι ώστε τα δεδομένα με μεγαλύτερη διακύμανση να είναι λιγότερο σταθμισμένα. Η συνιστώσα διακύμανσης εξαρτάται από το έτος εξέλιξης και θα πρέπει επίσης να εκτιμηθεί.

Το γραμμικό μοντέλο έχει ως εξής

$$E(F_{ij}) = c + a_{j-1} \quad \text{για } j \geq 2 \text{ και με τον περιορισμό } a_1 = 0 \quad (3.6)$$

Θεωρείται ότι F_{ij} είναι ανεξάρτητη και πρέπει να σημειωθεί ότι το μοντέλο δεν δεσμεύεται από την τελευταία παρατήρηση κατά το έτος ατυχήματος i , $d_{i,n-i+1}$. Δεδομένου ότι F_{ij} είναι κανονικά κατανομημένη, η συνάρτηση σύνδεσης είναι μόνο η συνάρτηση ταυτότητας. Προκειμένου να βρεθούν εκτιμήσεις και των δύο f_{ij} και ϕ_j , μπορεί να χρησιμοποιηθεί το μοντέλο από κοινού. Αυτή η τεχνική περιγράφεται στο (Renshaw 1994, Verrall & England 2002). Αυτή η τεχνική θα περιγραφεί εδώ χρησιμοποιώντας το σύνολο δεδομένων από την TrygVesta.

Το σχήμα 1 απεικόνιζε δύο τρίγωνα εξέλιξης για ζημιές d_{ij} και c_{ij} , και το τρίγωνο εξέλιξης που ανήκε στους συντελεστές εξέλιξης f_{ij} , είχε μικρότερη διάσταση από το ένα. Ένα τρίγωνο εξέλιξης αθροιστικών απαιτήσεων d_{ij} , με διάσταση 19x19, έχει ένα αντίστοιχο τρίγωνο εξέλιξης f_{ij} με διάσταση 18x18. Μόνο οι τιμές στην πάνω-αριστερή γωνία είναι γνωστές τιμές και αυτές θα χρησιμοποιηθούν ως μεταβλητές απόκρισης όπως φαίνεται στο γραμμικό μοντέλο στη σχέση (3.6). Υπάρχουν είναι 171 ($i + j \leq 19$) γνωστές τιμές f_{ij} , και υπάρχουν 153 ($j \geq 2, i + j < 18$) τιμές \hat{f}_{ij} που πρέπει να προβλεφθούν στην κάτω-αριστερή γωνία του τριγώνου εξέλιξης. Για το έτος εξέλιξης j όλες οι προβλεπόμενες τιμές \hat{f}_{ij} , θα είναι ίσες, και ο δείκτης θα μπορούσε να έχει μείνει εκτός.

Τα δύο σύνολα δεδομένων από την TrygVesta που περιέχουν τον αριθμό και το ποσό των απαιτήσεων έχουν και ένα κενό σημείο για το έτος ατυχήματος 1 και το έτος εξέλιξης 1.

Αυτό σημαίνει ότι υπάρχουν μόνο 170 παρατηρήσεις, σε αντίθεση με τις 171 οι οποίες γενικά θα είχαν από ένα τρίγωνο εξέλιξης με διάσταση 19x19. Αυτό έχει ως αποτέλεσμα μια έλλειψη παρατηρήσεων στο διάνυσμα \mathbf{f} και μία λιγότερη σειρά στον πίνακα σχεδίασης. Κατά την χρήση του γραμμικού μοντέλου στο στατιστικό πρόγραμμα R, f_{ij} αναδιατάσσεται ως διάνυσμα διάστασης 170x1. Το γραμμικό μοντέλο στο (3.6) μπορεί να γραφεί

$$E(F) = X\theta \quad (3.7)$$

Όπου \mathbf{F} είναι η μεταβλητή απόκρισης. Το \mathbf{X} είναι ο πίνακας σχεδιασμού και το θ είναι η παράμετρος που πρέπει να εκτιμηθεί. Το διάνυσμα \mathbf{f} και η παράμετρος θ , θα είναι

$$\mathbf{f}^T = [f_{2,2} \dots f_{18,2} \ f_{1,3} \dots f_{17,3} \dots f_{1,19}] \text{ και}$$

$$\theta^T = [c \ a_2 \ a_3 \ a_4 \ a_5 \ \dots \ a_{18}] \quad (3.8)$$

Το διάνυσμα θ έχει διάσταση 18x1. Ο πίνακας σχεδιασμού μπορεί τώρα να οριστεί. Σε αυτή τη μορφή το γραμμικό μοντέλο θα έχει ένα πίνακα σχεδιασμού διάστασης 170x18. Στο διάνυσμα \mathbf{f} υπάρχουν 18 χρόνια εξέλιξης. Ο πίνακας σχεδιασμού μπορεί να παρουσιαστεί για κάθε έτος εξέλιξης και η αντίστοιχη διάσταση γράφεται στη δεξιά πλευρά

$$\text{Έτος ανάπτυξης 1} \quad [\mathbf{1 \ 0 \ 0 \ \dots \ 0}]_{(17 \times 18)} \quad (3.9)$$

$$\text{Έτος ανάπτυξης 2} \quad [\mathbf{1 \ 1 \ 0 \ \dots \ 0}]_{(17 \times 18)}$$

$$\text{Έτος ανάπτυξης 3} \quad [\mathbf{1 \ 0 \ 1 \ 0 \ \dots \ 0}]_{(16 \times 18)}$$

.....

$$\text{Έτος ανάπτυξης 18} \quad [\mathbf{1 \ 0 \ 0 \ \dots \ 0 \ 1}]_{(1 \times 18)}$$

Τα βάρη που χρησιμοποιούνται στο γραμμικό μοντέλο μπορούν να διαμορφωθούν ως διάνυσμα. Θα έχει παρόμοια δομή με το \mathbf{f} . Παρατηρούμε ότι εκτός από τις γνωστές τιμές w_{ij} , οι μεταβλητές ϕ_j είναι άγνωστες, αλλά εξαρτώνται μόνο από το έτος εξέλιξης, τι \mathbf{W} μπορεί να γραφτεί ως:

$$\mathbf{W}^T = \left[\begin{array}{cccccc} \frac{w_{1,2}}{\phi_2} & \dots & \frac{w_{18,2}}{\phi_2} & \frac{w_{1,3}}{\phi_3} & \dots & \frac{w_{17,3}}{\phi_3} & \dots & \frac{w_{1,19}}{\phi_{19}} \end{array} \right]$$

Επιλέγονται αυθαίρετες τιμές για ϕ_j . Για να το κάνουμε απλό, το πρώτο σύνολο ϕ_j είναι ίσο

1. Το γραμμικό μοντέλο στο (3.7) μπορεί να λυθεί στο R από την εντολή:

`lm.wfit (f, W, X)`

Αυτή η εντολή παράγει μια εκτίμηση για την παράμετρο θ και από τον γραμμικό συνδυασμό μπορεί να ληφθεί μια εκτίμηση για το \mathbf{f} . Δεδομένου ότι οι τιμές ϕ_j είναι ακόμη άγνωστες, πρέπει να τοποθετηθεί ένα δεύτερο γραμμικό μοντέλο. Το δεύτερο μοντέλο χρησιμοποιεί το τετράγωνο καταλοίπων ως τις νέες μεταβλητές απόκρισης. Έστω r_{ij}^2 είναι το τετράγωνο καταλοίπων, και ορίζονται ως:

$$r_{ij}^2 = w_{ij} (f_{ij} - \hat{f}_{ij})^2. \quad (3.11)$$

Το γενικευμένο γραμμικό μοντέλο που πρέπει να τοποθετηθεί είναι:

$$g(E(R_{ij}^2)) = c_2 + \gamma_{j-1} \quad \text{για } i = 1, \dots, n-j+1 \quad \text{και } j = 1, \dots, n \quad (3.12)$$

Αυτό είναι ένα γενικευμένο γραμμικό μοντέλο R_{ij}^2 αφού δεν μπορεί να εξηγηθεί άμεσα μέσω ενός γραμμικού μοντέλου. Μια λειτουργία ζεύξης (g), δίνει τη δυνατότητα R_{ij}^2 να

εξηγηθεί μέσω ενός γραμμικού μοντέλου. Αυτή η συνάρτηση σύσδεσης είναι στενά συνδεδεμένη με την κατανομή της μεταβλητής απόκρισης. Εφόσον F_{ij} κατανέμεται

κανονικά με μέση τιμή f_j και διακύμανση $\sqrt{\frac{\phi_j}{w_{ij}}}$, η $R_{ij} = \sqrt{w_{ij}}(f_{ij} - \hat{f}_{ij})$ κατανέμεται

κανονικά με $E(R_{ij})=0$ και $Var(R_{ij})=\phi_j$. Έτσι $\frac{R_{ij}^2}{\phi_j} = w_{ij} \frac{(f_{ij} - \hat{f}_{ij})^2}{\phi_j}$ έχει χ-τετράγωνο

κατανομή με $E\left(\frac{R_{ij}^2}{\phi_j}\right)=1$ και $Var\left(\frac{R_{ij}^2}{\phi_j}\right)=2$. Ο μέσος όρος και η διακύμανση της μεταβλητής

απόκρισης R_{ij}^2 είναι $E(R_{ij}^2)=\phi_j$ και $Var(R_{ij}^2)=2\phi_j^2$. Έστω c είναι χ-τετράγωνο κατανομή της

μεταβλητής $\frac{R_{ij}^2}{\phi_j}$ και έστω η νέα μεταβλητή $Z = R_{ij}^2$ κατανέμεται με τη συνάρτηση f . Η

κατανομή f μπορεί να βρεθεί μέσω ενός γραμμικού μετασχηματισμού του c

$$f_z(z) = c_{R_{ij}^2/\phi_j} \left(\frac{z}{\phi_j} \right) \left| \frac{dr}{dz} \right| = \frac{1}{\Gamma\left(\frac{1}{2}\right) 2^{1/2}} \left(\frac{z}{\phi_j} \right)^{1/2-1} e^{-\left(\frac{z}{2\phi_j}\right)} \frac{1}{\phi_j} \quad (3.13)$$

$$= \frac{1}{\Gamma\left(\frac{1}{2}\right) (2\phi_j)^{1/2}} z^{-1/2} e^{-\left(\frac{z}{2\phi_j}\right)}$$

Είναι σαφές ότι (3.13) είναι η κατανομή γάμμα με παραμέτρους $\frac{1}{2}$ και $2\phi_j$. Η λειτουργία

κανονικής συνάρτησης μιας κατανομής γάμμα είναι η αντίστροφη συνάρτηση (McCullagh &

nelder 1989). Υπάρχουν άλλες πιθανές συναρτήσεις σύσδεσης στην κατανομή γάμμα. Αυτή

είναι η ταυτότητα και η λογαριθμική κατανομή. Ο Verral (2000) πρότεινε τη χρήση της

λογαριθμικής κατανομής. Οι διάφορες συναρτήσεις σύνδεσης ελέγχθηκαν για τα σύνολα δεδομένων από την TrygVesta. Η χρήση της λογαριθμικής συνάρτησης έδειξε μια γραμμική σχέση σε αντίθεση με τις άλλες συναρτήσεις σύνδεσης.

Το γενικευμένο γραμμικό μοντέλο γραμμένο σε μορφή διανύσματος είναι

$$g[E(R^2)] = g(\mu) = X\theta_1 \quad (3.14)$$

όπου g είναι η συνάρτηση σύνδεσης και $g(\mu) = \log(\mu)$. Η μεταβλητή απόκρισης R^2 και το διάνυσμα παραμέτρων θ_1 , έχουν την ίδια μορφή με τη μεταβλητή απόκρισης \mathbf{f} και του διανύσματος παραμέτρων $\boldsymbol{\theta}$ στη (3.7). Ο πίνακας ταυτότητας \mathbf{X} είναι ισοδύναμος με τον πίνακα ταυτότητας στη (3.7)

Στο R αυτό μπορεί να γίνει με τη λειτουργία:

```
glm.fit(R, W, X, Γάμμα (σύνδεσμος = log))
```

Μπορούν να ληφθούν νέες τιμές για $\phi_j, j = 2, \dots, 18$. Η ϕ_{19} δεν μπορεί να επιτευχθεί αφού υπάρχει μόνο ένα κατάλοιπο στο γενικό γραμμικό μοντέλο. Τα αποτελέσματα που παρουσιάζονται αργότερα χρησιμοποιούν δύο διαφορετικές επιλογές $\phi_{19} = \phi_{18}$ και $\phi_{19} = \phi_{17}$,

Το σταθμισμένο \mathbf{W} ενημερώνεται με νέες τιμές του ϕ_j . Οι εκτιμήσεις του c και α_j για $j = 2, \dots, 18$ προέρχονται από το πρώτο γραμμικό μοντέλο για άλλη μια φορά. Οι συντελεστές εξέλιξης μπορούν να υπολογιστούν από αυτές τις εκτιμήσεις. Αυτή είναι η από κοινού διαδικασία μοντελοποίησης.

Οι προβλεπόμενες τιμές μπορούν τώρα να βρεθούν μέσω

$$\hat{\mathbf{f}}^p = \mathbf{X}^p \hat{\boldsymbol{\theta}}$$

$(153 \times 1) \quad (153 \times 18) \quad (18 \times 1)$
(3.15)

όπου \mathbf{f}^p είναι το διάνυσμα των προβλεπόμενων τιμών του \mathbf{f} , \mathbf{X}^p είναι ο πίνακας σχεδιασμού των προβλεπόμενων συντελεστών εξέλιξης και $\hat{\boldsymbol{\theta}}$ είναι το διάνυσμα των εκτιμήσεων των παραμέτρων. Όπως και οι παρατηρήσεις του f_{ij} , οι προβλεπόμενες τιμές του f_{ij} θα παραταχθούν ως διάνυσμα. Το διάνυσμα \mathbf{f}^p και $\boldsymbol{\theta}$ μπορεί να γραφτεί όπως

$$(\mathbf{f}^p)^T = [f_{19,2} \ f_{18,3} \ f_{19,3} \ f_{17,4} \ f_{18,4} \ f_{19,4} \ \dots \ f_{19,19}] \text{ και}$$

$$\hat{\boldsymbol{\theta}}^T = [\hat{c} \ \hat{a}_2 \ \hat{a}_3 \ \hat{a}_4 \ \hat{a}_5 \ \dots \ \hat{a}_{18}]$$

Ο πίνακας σχεδιασμού για τα διαφορετικά έτη εξέλιξης των προβλεπόμενων τιμών παρουσιάζεται παρακάτω με την αντίστοιχη διάσταση που είναι γραμμένη στη δεξιά πλευρά.

Έτος ανάπτυξης 2 $[\mathbf{1 \ 1 \ 0 \ \dots \ 0}]_{(1 \times 18)}$

Έτος ανάπτυξης 3 $[\mathbf{1 \ 0 \ 1 \ 0 \ \dots \ 0}]_{(2 \times 18)}$

.....

Έτος ανάπτυξης 18 $[\mathbf{1 \ 0 \ 0 \ \dots \ 0 \ 1}]_{(17 \times 18)}$

Είναι επιθυμητό να βρεθούν τα τυπικά σφάλματα των παραμέτρων και των συντελεστών εξέλιξης στο γραμμικό μοντέλο. Ο θεωρητικός υπολογισμός παρουσιάζεται παρακάτω.

Έστω Σ^{-1} ένας πίνακας με διάσταση 170×170 . Όλα τα στοιχεία είναι μηδέν εκτός από τη διαγώνιο που είναι το βάρος \mathbf{W} . Η διακύμανση της παραμέτρου $\boldsymbol{\theta}$ μπορεί να βρεθεί από τη

διαγώνιο του πίνακα $Var(\theta) = (X^T \Sigma^{-1} X)^{-1}$. Αυτός είναι ένας πίνακας διαστάσεων 18x18, και η τετραγωνική ρίζα της διαγωνίου παράγει το τυπικό σφάλμα.

Είναι επίσης ενδιαφέρον να βρούμε την διακύμανση της \mathbf{f} . Ο πίνακας συνδιακύμανσης του \mathbf{f} , με τις αντίστοιχες διαστάσεις που είναι γραμμένες κάτω από τους πίνακες, είναι

$$Cov(\hat{\mathbf{f}}) = \mathbf{X}^p \left(\mathbf{X}^T \Sigma^{-1} \mathbf{X} \right)^{-1} \left(\mathbf{X}^p \right)^T \quad (3.16)$$

$(153 \times 153) \quad 153 \times 18 \quad 18 \times 18 \quad 18 \times 153$

Η από κοινού μοντελοποίηση έχει παραγάγει εκτιμήσεις των παραμέτρων στο πρώτο γραμμικό μοντέλο, του συντελεστή εξέλιξης f_j και της συνιστώσας διακύμανσης ϕ_j . Έχει αποδειχθεί πως βρέθηκαν τα τυπικά σφάλματα των παραμέτρων στο πρώτο γραμμικό μοντέλο και οι συντελεστές εξέλιξης. Τα αποτελέσματα παρουσιάζονται παρακάτω στους πίνακες 1-6.

Parameter	Estimate	Standard error
C	0,947	0,014
Alfa2	0,060	0,014
Alfa3	0,080	0,014
Alfa4	0,075	0,014
Alfa5	0,070	0,014
Alfa6	0,064	0,014
Alfa7	0,064	0,014
Alfa8	0,062	0,014
Alfa9	0,061	0,014
Alfa10	0,059	0,014
Alfa11	0,057	0,014
Alfa12	0,057	0,014
Alfa13	0,054	0,014
Alfa14	0,055	0,014
Alfa15	0,056	0,014
Alfa16	0,056	0,014
Alfa17	0,056	0,014
Alfa18	0,057	0,014

Πίνακας 4 : Εκτιμήσεις των παραμέτρων για τον αριθμό των απαιτήσεων στο γραμμικό μοντέλο

Parameter	Estimate
Phi2	3,678
Phi3	0,320
Phi4	0,328
Phi5	0,125
Phi6	0,056
Phi7	0,012
Phi8	0,032
Phi9	0,017
Phi10	0,021
phi11	0,009
phi12	0,002
phi13	0,007
phi14	0,002
phi15	0,000
phi16	0,004
phi17	0,005
phi18	0,001
phi19	-

Πίνακας 5: Εκτιμήσεις phi για τον αριθμό των απαιτήσεων

Parameter	Estimate	Standard error
Dev. Factor 2	0,947	0,006
Dev. Factor 3	1,007	0,004
Dev. Factor 4	1,027	0,004
Dev. Factor 5	1,022	0,003
Dev. Factor 6	1,017	0,002
Dev. Factor 7	1,011	0,001
Dev. Factor 8	1,010	0,002
Dev. Factor 9	1,009	0,001
Dev. Factor 10	1,008	0,001
Dev. Factor 11	1,005	0,001
Dev. Factor 12	1,004	0,001
Dev. Factor 13	1,003	0,001
Dev. Factor 14	1,000	0,001
Dev. Factor 15	1,001	0,0003
Dev. Factor 16	1,002	0,001
Dev. Factor 17	1,003	0,002
Dev. Factor 18	1,002	0,001
Dev. Factor 19	1,003	0,001

Πίνακας 6: Εκτιμήσεις των συντελεστών εξέλιξης και των τυπικών σφαλμάτων για τον αριθμό των απαιτήσεων

Parameter	Estimate	Standard error
C	3,215	0,163
Alfa2	-1,252	0,190
Alfa3	-1,553	0,169
Alfa4	-1,827	0,163
Alfa5	-1,976	0,163
Alfa6	-2,067	0,163
Alfa7	-2,133	0,163
Alfa8	-2,152	0,163
Alfa9	-2,183	0,163
Alfa10	-2,180	0,163
Alfa11	-2,193	0,163
Alfa12	-2,203	0,163
Alfa13	-2,193	0,163
Alfa14	-2,207	0,163
Alfa15	-2,211	0,163
Alfa16	-2,214	0,163
Alfa17	-2,207	0,163
Alfa18	-2,215	0,163

Πίνακας 7: Εκτιμήσεις των παραμέτρων για το ποσό των απαιτήσεων στο γραμμικό μοντέλο

Parameter	Estimate
Phi2	2,266
Phi3	2,498
Phi4	0,934
Phi5	0,199
Phi6	0,208
Phi7	0,219
Phi8	0,099
Phi9	0,08
Phi10	0,021
phi11	0,029
phi12	0,021
phi13	0,006
phi14	0,001
phi15	0,003
phi16	0,010
phi17	0,0002
phi18	0,004
phi19	-

Πίνακας 8: Εκτιμήσεις phi για το μέγεθος των απαιτήσεων

Parameter	Estimate	Standard error
Dev.factor 2	3,215	0,163
Dev.factor 3	1,963	0,098
Dev.factor 4	1,663	0,045
Dev.factor 5	1,388	0,017
Dev.factor 6	1,239	0,015
Dev.factor 7	1,148	0,015
Dev.factor 8	1,083	0,010
Dev.factor 9	1,063	0,009
Dev.factor 10	1,032	0,005
Dev.factor 11	1,036	0,006
Dev.factor 12	1,022	0,005
Dev.factor 13	1,013	0,003
Dev.factor 14	1,023	0,002
Dev.factor 15	1,008	0,003
Dev.factor 16	1,005	0,006
Dev.factor 17	1,002	0,001
Dev.factor 18	1,008	0,007
Dev.factor 19	1,000	0,011

Πίνακας 9: Εκτιμήσεις των συντελεστών εξέλιξης και των τυπικών σφαλμάτων για το μέγεθος των απαιτήσεων.

3.4.2 Προβλέψεις αποθεμάτων και σφάλματα πρόβλεψης

Όταν υπάρχουν εκτιμήσεις για τους συντελεστές εξέλιξης, μπορούν να γίνουν εκτιμήσεις απαιτήσεων. Τα κενά σημεία του τριγώνου εξέλιξης μπορούν να εκτιμηθούν χρησιμοποιώντας το μοντέλο στο (3.4). Η τελική εκτίμηση της απαίτησης \hat{D}_{in} υπολογίζεται χρησιμοποιώντας την εξίσωση της chain-ladder (1.6). Η εκτίμηση των αποθεματικών μπορεί να υπολογιστεί αφού υπάρχει μια απλή σύνδεση μεταξύ του αποθεματικού και της τελικής απαίτησης. Το αποθεματικό είναι $R_i = D_{in} - d_{i,n-i+1}$ και ισοδύναμα η εκτίμηση αποθεματικών είναι $\hat{R}_i = \hat{D}_{in} - d_{i,n-i+1}$ για $i = 2, \dots, n$ όπου \hat{D}_{in} έχει υπολογιστεί και $d_{i,n-i+1}$ παρατηρείται στη διαγώνιο του τριγώνου εξέλιξης.

Οι συντελεστές διακύμανσης ϕ_j έχουν βρεθεί. Το μοντέλο στο (3.4) καθορίζει τη διακύμανση των εκτιμώμενων απαιτήσεων $\hat{D}_{i,n-i+1}$ για $i = 2, \dots, n$. Για να βρούμε το σφάλμα πρόβλεψης και το αποθεματικό της τελικής απαίτησης χρειάζονται μερικοί επιπλέον υπολογισμοί. Η εξίσωση (2.59) έδωσε μια έκφραση για το MSE \hat{D}_{in} :

$$MSE(\hat{D}_{in} | k) = Var(D_{in} | k) + E(E(D_{in} | k) - \hat{D}_{in})^2.$$

Εάν θεωρηθεί ότι υπάρχει ανεξαρτησία μεταξύ των ετών ατυχήματος, δεν είναι απαραίτητο να δεσμεύσουμε όλες τις παρατηρήσεις. Επιπλέον, μόνο η τελευταία παρατήρηση σε κάθε έτος ατυχήματος χρησιμοποιείται στο μοντέλο (3.4). Αρκεί στην περίπτωση αυτή να δεσμεύσουμε ως προς $d_{i,n-i+1}$. Η προσέγγιση του $E(\hat{D}_{in} | d_{i,n-i+1})$ με $E(D_{in} | d_{i,n-i+1})$ κάνει μια νέα έκφραση:

$$MSE(\hat{D}_{in} | d_{i,n-i+1}) \approx Var(D_{in} | d_{i,n-i+1}) + Var(\hat{D}_{in} | d_{i,n-i+1}) \quad (3.17)$$

Το MSE του αποθεματικού είναι το ίδιο με το MSE της τελικής απαίτησης. Αυτό είναι σαφές δεδομένου ότι

$$\begin{aligned} Var(\hat{R}_i | d_{i,n-i+1}) &= Var(D_{in} - D_{i,n-i+1} | d_{i,n-i+1}) = Var(D_{in} | d_{i,n-i+1}) \quad \text{και} \\ Var(\hat{R}_i | d_{i,n-i+1}) &= Var(\hat{D}_{in} - \hat{D}_{i,n-i+1} | d_{i,n-i+1}) = Var(\hat{D}_{in} | d_{i,n-i+1}) \end{aligned}$$

Επομένως,

$$MSE(\hat{R}_i | d_{i,n-i+1}) = MSE(\hat{D}_{in} | d_{i,n-i+1}) \approx Var(D_{in} | d_{i,n-i+1}) + Var(\hat{D}_{in} | d_{i,n-i+1}) \quad (3.18)$$

Verral (2000) υποδηλώνει $Var(D_{in} | d_{i,n-i+1})$ ως τη διακύμανση της διαδικασίας και $Var(\hat{D}_{in} | d_{i,n-i+1})$ ως τη διακύμανση της εκτίμησης, και αυτοί οι όροι θα χρησιμοποιηθούν επίσης σε αυτή την εργασία. Για να ληφθεί το σφάλμα πρόβλεψης θα χρησιμοποιηθούν δύο προσεγγίσεις, οι οποίες παρουσιάζονται στο Verral (2000).

Η διακύμανση της εκτίμησης μπορεί να βρεθεί για το έτος ατυχήματος $2, \dots, n$ και είναι

$$\begin{aligned} \text{Var}(\hat{D}_{in} | d_{i,n-i+1}) &= \text{Var}\left(D_{i,n-i+1} \prod_{j=n-i+2}^n f_j | d_{i,n-i+1}\right) \\ &= d_{i,n-i+1}^2 \text{Var}\left(\prod_{j=n-i+2}^n f_j | d_{i,n-i+1}\right). \end{aligned} \quad (3.19)$$

Το δεύτερο έτος ατυχημάτων μπορεί να βρεθεί άμεσα, δεδομένου ότι (3.19) γίνεται $d_{2,n-1}^2 \text{Var}(\hat{f}_n | d_{2,n-1})$. Η διακύμανση της εκτίμησης του τρίτου έτους ατυχημάτων είναι πιο περίπλοκη δεδομένου ότι είναι απαραίτητο να βρεθεί η διακύμανση δύο συντελεστών εξέλιξης. Το τέταρτο έτος ατυχημάτων απαιτεί μια εκτίμηση της διακύμανσης από τρία έτη εξέλιξης, και ούτω καθεξής. Για να βρεθούν αυτές οι ανεξάρτητες διακυμάνσεις ή τουλάχιστον οι ασύνδετοι συντελεστές εξέλιξης πρέπει να θυπολογισθούν. Υποθέτοντας ανεξαρτησία ή τουλάχιστον καμία σχέση μεταξύ των συντελεστών εξέλιξης η διακύμανση των δύο τελευταίων συντελεστών εξέλιξης είναι:

$$\begin{aligned} \text{Var}(\hat{f}_{n-1}\hat{f}_n) &= (E[\hat{f}_{n-1}])\text{Var}(\hat{f}_n) + (E[\hat{f}_n])\text{Var}(\hat{f}_{n-1}) + \text{Var}(\hat{f}_{n-1})\text{Var}(\hat{f}_n) \\ &\approx (\hat{f}_{n-1})^2 \text{Var}(\hat{f}_n) + (\hat{f}_n)^2 \text{Var}(\hat{f}_{n-1}) + \text{Var}(\hat{f}_{n-1})\text{Var}(\hat{f}_n). \end{aligned} \quad (3.20)$$

Το παράρτημα 5 αποδεικνύει τον τύπο που χρησιμοποιείται στο (3.20). Δεν είναι η δεσμευμένη διακύμανση $\text{Var}(\hat{f}_{n-1}\hat{f}_n | d_{i,n-i+1})$ που έχει ανακτηθεί αλλά η $\text{Var}(\hat{f}_{n-1}\hat{f}_n)$. Στον πραγματικό υπολογισμό του σφάλματος πρόβλεψης είναι το αριθμητικό αποτέλεσμα της (3.16) που θα χρησιμοποιηθεί, η οποία είναι η αδέσμευτη διακύμανση.

Όταν βρεθεί η διακύμανση από τους τρεις συντελεστές εξέλιξης, το προηγούμενο αποτέλεσμα (διακύμανση συντελεστών εξέλιξης) θα χρησιμοποιηθεί. Έτσι

$$\text{Var}(\hat{f}_{n-2}[\hat{f}_{n-1}\hat{f}_n]) \approx (\hat{f}_{n-2})^2 \text{Var}(\hat{f}_{n-1}\hat{f}_n) + (\hat{f}_{n-1}\hat{f}_n)^2 \text{Var}(\hat{f}_{n-2}) + \text{Var}(\hat{f}_{n-2}) \text{Var}(\hat{f}_{n-1}\hat{f}_n).$$

Το τελευταίο βήμα βρίσκεται όταν βρίσκεται το $\text{Var}\left(\prod_{j=2}^n \hat{f}_j\right)$.

Για να βρεθεί η διακύμανση της διαδικασίας, το $\text{Var}(D_{in} | d_{i,n-i+1})$ μπορεί να χρησιμοποιηθεί ως μια αναδρομική διαδικασία. Αυτή η διαδικασία χρησιμοποιεί τον κανόνα της διπλής αναμενόμενης τιμής και της διακύμανσης. Το μοντέλο δίνει ότι $\text{Var}(D_{ij} | d_{i,j-1}) = \phi_j d_{i,j-1}$, έτσι η διακύμανση της διαδικασίας για το επόμενο έτος εξέλιξης έχει ήδη καθοριστεί. Αν αφήσουμε τον δείκτη i , η διακύμανση της διαδικασίας είναι δύο βήματα μπροστά

$$\begin{aligned} \text{Var}(D_{j+1} | d_{j-1}) &= E(\text{Var}[D_{j+1} | D_j] | d_{j-1}) + \text{Var}(E[D_{j+1} | D_j] | d_{j-1}) \\ &= E(\phi_{j+1} D_j | d_{j-1}) + \text{Var}(f_{j+1} D_j | d_{j-1}) \\ &= \phi_{j+1} E(D_j | d_{j-1}) + (f_{j+1})^2 \text{Var}(D_j | d_{j-1}) \\ &= \phi_{j+1} f_j d_{j-1} + (f_{j+1})^2 \phi_j d_{j-1}. \end{aligned}$$

Η διακύμανση της διαδικασίας είναι τρία βήματα μπροστά

$$\begin{aligned} \text{Var}(D_{j+2} | d_{j-1}) &= E(\text{Var}[D_{j+2} | D_{j+1}] | d_{j-1}) + \text{Var}(E[D_{j+2} | D_{j+1}] | d_{j-1}) \\ &= E(\phi_{j+2} D_{j+1} | d_{j-1}) + \text{Var}(f_{j+2} D_{j+1} | d_{j-1}) \\ &= \phi_{j+2} E(D_{j+1} | d_{j-1}) + (f_{j+2})^2 \text{Var}(D_{j+1} | d_{j-1}) \\ &= \phi_{j+2} E(E(D_{j+1} | D_j) | d_{j-1}) + (f_{j+2})^2 \text{Var}(D_{j+1} | d_{j-1}) \\ &= \phi_{j+2} E(f_{j+1} D_j | d_{j-1}) + (f_{j+2})^2 \text{Var}(D_{j+1} | d_{j-1}) \\ &= \phi_{j+2} f_{j+1} f_j d_{j-1} + \phi_{j+1} (f_{j+1})^2 f_j d_{j-1} + \phi_j (f_{j+2})^2 (f_{j+1})^2 d_{j-1}. \end{aligned}$$

Αυτή η διαδικασία μπορεί να εκτελεστεί για τα επόμενα τρία χρόνια και ούτω καθεξής. Η πρόθεση να χρησιμοποιηθεί αυτή η επαναληπτική προσέγγιση είναι για να βρεθεί η διακύμανση της τελικής απαίτησης. Το δεύτερο έτος ατυχημάτων δεν χρειάζεται περισσότερο από ένα βήμα, το τρίτο έτος ατυχήματος χρειάζεται δύο βήματα της προσέγγισης και ούτω καθεξής.

Το συνολικό MSE του αποθεματικού είναι το άθροισμα της διακύμανσης της εκτίμησης και της διαδικασίας, αλλά και ένα στοιχείο συνδιακύμανσης που προστέθηκε λόγω της συνδιακύμανσης μεταξύ των εκτιμώμενων τιμών. Η συνολική εκτίμηση και η διακύμανση της διαδικασίας είναι:

$$MSE(\hat{R} | k) = \sum_{i=2}^n Var(D_{in} | d_{i,n-i+1}) + \sum_{i=2}^n Var(\hat{D}_{in} | d_{i,n-i+1}) + 2 \sum_{\substack{i=2 \\ l>i}}^n Cov(\hat{D}_{in}, D_{ln} | d_{i,n-i+1})$$

(3.21)

Η εκτίμηση και η διακύμανση της προσέγγισης υπολογίστηκε για το σύνολο δεδομένων από την TrygVesta στο R. Δεδομένου ότι ο υπολογισμός της εκτιμητικής μεταβλητότητας απαιτούσε μη συσχετισμένους συντελεστές εξέλιξης, εξετάστηκε ο πίνακας της συνδιακύμανσης της \hat{f} . Η \hat{f} είναι ένα διάνυσμα 153 στοιχείων. Τα 153 στοιχεία μπορούν να τοποθετηθούν στην κάτω-αριστερή γωνία του τριγώνου εξέλιξης και τα στοιχεία που βρίσκονται στο ίδιο έτος εξέλιξης είναι ίσα. Ελέγχθηκε ότι μόνο τα στοιχεία συνδιακύμανσης του πίνακα ήταν διαφορετικά από το μηδέν, και τα υπόλοιπα ήταν μηδέν. Αυτή ήταν η περίπτωση τόσο για τον αριθμό όσο και για το μέγεθος των απαιτήσεων.

Η διακύμανση της εκτίμησης βρέθηκε με τη δημιουργία μιας επανάληψης για κάθε έτος ατυχήματος στο R. Η αναδρομική διαδικασία χρησιμοποίησε την προηγούμενη ανακαλυφθείσα διακύμανση. Το δεύτερο έτος ατυχημάτων απαιτούσε μόνο έναν

υπολογισμό ενώ το έτος ατυχήματος 19 απαιτούσε τους ίδιους υπολογισμούς επαναλαμβάνοντάς τους 18 φορές έως ότου επιτευχθεί η διακύμανση της προβλεπόμενης τελικής απαίτησης. Για να βρεθεί η διακύμανση της διαδικασίας, έγινε επίσης μία επανάληψη για αυτόν τον υπολογισμό. Όπως και στην εκτίμηση της διακύμανση, η επανάληψη έτρεξε μία φορά για το έτος ατυχήματος 2 και 18 φορές για το έτος ατυχήματος 19 να βρεθεί η διακύμανση της απαίτησης στο διηλεκές.

Η συνολική εκτίμηση και η διακύμανση της διαδικασίας υπολογίστηκαν με το άθροισμα της εκτίμησης και της διακύμανσης της διαδικασίας. Μεγάλη προσοχή δόθηκε όταν εντοπίσαμε τον τελευταίο όρο στη σχέση (3.21).

3.4.3 Αποτελέσματα

Οι εκτιμήσεις των αποθεματικών με τα αντίστοιχα σφάλματα πρόβλεψης παρουσιάζονται στις ενότητες 7 και 8.

Accident year	Reserve	Prediction error	Prediction error %
2	2,2696	0,901	39,698
3	5,0462	1,440	28,536
4	8,5419	3,031	35,484
5	9,3872	3,554	37,860
6	10,4740	3,594	34,314
7	12,3098	4,263	34,631
8	15,9718	5,265	32,964
9	24,7007	6,426	26,015
10	33,7194	7,738	22,948
11	47,9351	10,177	21,231
12	59,6699	11,453	19,194
13	73,0450	13,474	18,446
14	93,8263	14,769	15,741
15	121,8697	18,137	14,882
16	162,3809	24,390	15,020
17	208,6802	35,895	17,201
18	211,5965	43,583	20,597
19	118,2655	93,339	78,923
Overall	1 219,6896	78,571	6,442

Πίνακας 10: Απόθεμα και σφάλμα πρόβλεψης για τον αριθμό απαιτήσεων

Accident year	Reserve	Prediction error	Prediction error %
2	0,000	0,756	315539,2
3	0,596	1,222	205,0647
4	0,980	1,605	163,8432
5	1,725	2,304	133,5831
6	2,302	2,163	93,96275
7	6,109	2,7573	45,13254
8	8,155	3,058	37,4985
9	12,553	3,954	31,49889
10	14,976	4,137	27,62489
11	22,616	5,126	22,66539
12	26,523	5,918	22,31235
13	31,296	6,754	21,58113
14	64,976	11,289	17,37403
15	67,443	11,523	17,08554
16	115,679	15,368	13,28507
17	163,470	24,899	15,23153
18	168,715	41,325	24,49396
19	171,178	52,306	30,55655
Overall	879,291	56,541	6,430291

Πίνακας 11: Απόθεμα και σφάλμα πρόβλεψης για το ποσό των απαιτήσεων

3.5 Το μοντέλο του Mack

3.5.1 Το μοντέλο

Το μοντέλο του Mack αποτελείται από τρεις υποθέσεις. Οι δύο πρώτες παραδοχές αφορούσαν τις δύο πρώτες ροπές και είναι για $j = 2, \dots, n$:

$$E(D_{ij} | K_{i,j-1} = k_{i,j-1}) = f_j d_{i,j-1}$$

$$\text{Var}(D_{ij} | K_{i,j-1} = k_{i,j-1}) = g_j d_{i,j-1}.$$

Οι παράμετροι έχουν εκτιμητές:

$$\hat{f}_1 = \frac{\sum_{i=1}^{n-j+1} d_{ij}}{\sum_{i=1}^{n-j+1} d_{i,j-1}} \quad \text{και} \quad \hat{g}_j = \frac{1}{n-j} \sum_{i=1}^{n-j+1} d_{i,j-1} \left(\frac{d_{ij}}{d_{i,j-1}} - \hat{f}_j \right)^2$$

Οι εκτιμητές αυτοί βρέθηκαν για το σύνολο δεδομένων από την TrygVesta. Οι υπολογισμοί αυτοί έγιναν σε ένα excel και τα αποτελέσματα παρουσιάζονται στον πίνακα 9, όπου \hat{f}_j σημειώνεται ως συντελεστής εξέλιξης και ως \hat{g}_j συντελεστής διακύμανσης. Δεν υπάρχουν αρκετές πληροφορίες για υπολογισθεί το \hat{g}_{19} και για τους επόμενους σκοπούς το \hat{g}_{19} θα είναι ίσο με είτε \hat{g}_{18} είτε \hat{g}_{17} .

Accident year	Number of claims		Amount of claims	
	Dev.factor	Variance factor	Dev. Factor	Variance factor
1	-	-	-	-
2	0,947	3,678	3,215	2,266
3	1,007	0,340	1,963	2,654
4	1,027	0,328	1,663	0,996
5	1,022	0,119	1,388	0,179
6	1,017	0,059	1,239	0,221
7	1,011	0,012	1,148	0,232
8	1,010	0,032	1,083	0,099
9	1,009	0,019	1,063	0,080
10	1,008	0,023	1,032	0,023
11	1,005	0,008	1,036	0,033
12	1,004	0,003	1,022	0,024
13	1,003	0,006	1,013	0,007
14	1,000	0,003	1,023	0,001
15	1,001	0,000	1,008	0,004
16	1,002	0,002	1,005	0,002
17	1,003	0,007	1,002	0,0002
18	1,002	0,001	1,008	0,004
19	1,003	-	1,000	-

Πίνακας 12: εκτιμήσεις του συντελεστή εξέλιξης \hat{f}_j και του συντελεστή διακύμανσης \hat{g}_j για τον αριθμό και το ποσό των απαιτήσεων.

3.5.2 Προβλέψεις αποθεμάτων και σφάλματα πρόβλεψης

Ο συντελεστής εξέλιξης είναι ισοδύναμος με τον συντελεστή εξέλιξης της chain-ladder. Οι μελλοντικές εκτιμήσεις απαιτήσεων μπορούν να υπολογισθούν χρησιμοποιώντας την πρώτη παραδοχή του Mack και τα κενά σημεία στο τρίγωνο εξέλιξης των ζημιών μπορούν να καλυφθούν με εκτιμώμενες τιμές. Οι εκτιμήσεις των αποθεματικών βρίσκονται ότι χρησιμοποιούμε το μοντέλο της κανονικής προσέγγισης, δηλαδή $\hat{R}_i = \hat{D}_{in} - d_{i,n-i+1}$. Δεδομένου ότι ο συντελεστής εξέλιξης είναι ισοδύναμος με το προηγούμενο μοντέλο, τα αποτελέσματα είναι προφανώς πανομοιότυπα.

Η δεύτερη παραδοχή του Mack καθορίζει τη διακύμανση των εκτιμώμενων απαιτήσεων. Για να βρεθεί το σφάλμα πρόβλεψης της τελικής ζημιάς απαιτούνται περισσότεροι υπολογισμοί. Από το κεφάλαιο 3.4.2 προκύπτει ότι το σφάλμα πρόβλεψης του αποθεματικού είναι το ίδιο με το σφάλμα πρόβλεψης της εκτιμώμενης τελικής απαίτησης. Το MSE \hat{R}_i είναι

$$MSE(\hat{R}_i | k) \approx \hat{D}_{in}^2 \sum_{j=n-i+2}^n \frac{\hat{g}_i}{\hat{f}_j^2} \left(\frac{1}{\hat{D}_{i,j-1}} + \frac{1}{\sum_{i=1}^{n-j+1} d_{i,j-1}} \right) \quad (3.22)$$

Το συνολικό σφάλμα πρόβλεψης του αποθεματικού είναι

$$MSE(\hat{R} | k) \approx \sum_{i=2}^n \left\{ MSE(\hat{R}_i | k) + \hat{D}_{in} \left(\sum_{l=i+1}^n \hat{D}_{ln} \right) \left(\sum_{k=n-i+2}^n \frac{2\hat{g}_k}{\hat{f}_j^2 \sum_{l=1}^{n-k+1} d_{lk}} \right) \right\}. \quad (3.23)$$

Είναι πολύ εκτεταμένο έργο να βρούμε τους εκτιμητές στις (3.22) και (3.23), και πρόσφατα έγινε σε μια άλλη μεταπτυχιακή εργασία (Gangsoy 2008). Οι υπολογισμοί περιλαμβάνονται στο Παράρτημα 6.

Οι υπολογισμοί της εύρεσης $MSE(\hat{R}_i | k)$ και $MSE(\hat{R} | k)$ του συνόλου δεδομένων από την TrygVesta έγιναν σε φύλλο Excel. Τα εκτιμώμενα αποθεματικά με τα αντίστοιχα σφάλματα πρόβλεψης παρουσιάζονται παρακάτω στους πίνακες 10, 11, 12 και 13. Περιλαμβάνονται επίσης το συνολικό αποθεματικό, το οποίο απλά είναι το άθροισμα των αποθεματικών και το συνολικό σφάλμα πρόβλεψης.

Επίσης, υπάρχουν δυο πίνακες για τον αριθμό των απαιτήσεων και για το ποσό των απαιτήσεων. Δεδομένου ότι ο συντελεστής μεταβλητότητας \hat{g}_{19} δεν μπορεί να υπολογιστεί, έχει οριστεί ίσος με \hat{g}_{18} και \hat{g}_{17} και επίσης υπάρχει ένας πίνακας για κάθε προσέγγιση.

3.5.3 Αποτελέσματα

Οι εκτιμήσεις αποθεμάτων με τα αντίστοιχα σφάλματα πρόβλεψης παρουσιάζονται στους πίνακες 10-13.

Accident year	Reserve	Prediction error	Prediction error %
1	0,000	-	-
2	2,270	0,870	38,333
3	5,046	1,390	27,545
4	8,542	3,440	40,272
5	9,387	3,680	39,202
6	10,474	3,690	35,230
7	12,310	4,400	35,744
8	15,972	5,250	32,870
9	24,701	6,430	26,032
10	33,719	7,660	22,717
11	47,935	10,320	21,529
12	59,670	11,700	19,608
13	73,045	13,720	18,783
14	93,826	14,980	15,966
15	121,870	18,470	15,156
16	162,381	24,440	15,051
17	208,680	35,900	17,203
18	211,596	44,060	20,823
19	118,265	95,550	80,793
Overall	1219,690	139,140	11,408

Πίνακας 13: Εκτιμήσεις αποθεματικών για τον αριθμό των απαιτήσεων και τα σφάλματα πρόβλεψής

τους. Για το έτος ατυχήματος 19 $g_{19} = g_{18}$.

Accident year	Reserve	Prediction error	Prediction error %
1	0,000	-	-
2	2,270	3,030	133,503
3	5,046	3,840	76,096
4	8,542	5,250	61,462
5	9,387	5,240	55,821
6	10,474	5,240	50,029
7	12,310	6,030	48,985
8	15,972	6,750	42,262
9	24,701	8,230	33,319
10	33,719	9,370	27,788
11	47,935	11,840	24,700
12	59,670	13,020	21,820
13	73,045	14,830	20,303
14	93,826	16,120	17,181
15	121,870	19,440	15,951
16	162,381	25,250	15,550
17	208,680	36,500	17,491
18	211,596	44,510	21,035
19	118,265	95,730	80,945
Overall	1219,690	158,930	13,030

Πίνακας 14: Εκτιμήσεις αποθεματικών για τον αριθμό των απαιτήσεων και τα σφάλματα πρόβλεψής τους. Για το έτος ατυχήματος 19 $g_{19} = g_{17}$.

Accident year	Reserve	Prediction error	Prediction error %
1	0,000	-	-
2	0,000	0,680	283818,339
3	0,596	1,110	186,270
4	0,980	1,460	149,041
5	1,725	1,800	104,362
6	2,302	1,760	76,456
7	6,109	2,280	37,320
8	8,155	2,620	32,128
9	12,553	3,640	28,997
10	14,976	3,980	26,577
11	22,616	5,000	22,108
12	26,523	5,830	21,981
13	31,296	6,700	21,409
14	64,976	11,350	17,468
15	67,443	11,690	17,333
16	115,679	15,340	13,261
17	163,470	25,330	15,495
18	168,715	42,420	25,143
19	171,178	53,110	31,026
Overall	879,291	89,860	10,220

Πίνακας 15: Εκτιμήσεις αποθεματικών για το μέγεθος των απαιτήσεων και τα σφάλματα πρόβλεψής τους. Για το έτος ατυχήματος 19 $g_{19} = g_{18}$.

Accident year	Reserve	Prediction error	Prediction error %
1	0,000	-	-
2	0,000	0,160	66780,786
3	0,596	0,710	119,146
4	0,980	0,930	94,937
5	1,725	1,230	71,314
6	2,302	1,330	57,776
7	6,109	1,720	28,154
8	8,155	2,100	25,751
9	12,553	3,190	25,413
10	14,976	3,680	24,573
11	22,616	4,670	20,649
12	26,523	5,630	21,227
13	31,296	6,560	20,961
14	64,976	11,180	17,206
15	67,443	11,580	17,170
16	115,679	15,210	13,148
17	163,470	25,230	15,434
18	168,715	42,370	25,113
19	171,178	53,070	31,003
Overall	879,291	86,340	9,819

Πίνακας 16: Εκτιμήσεις αποθεματικών για τον μέγεθος των απαιτήσεων και τα σφάλματα πρόβλεψής τους. Για το έτος ατυχήματος 19 $g_{19} = g_{17}$.

ΚΕΦΑΛΑΙΟ 4 : Σχόλια

Το μοντέλο poisson είναι μια ειδική περίπτωση του πολλαπλασιαστικού μοντέλου και χρησιμοποιώντας τον εκτιμητή μέγιστης πιθανότητας στο μοντέλο poisson αποδείχθηκε ότι είναι ίσος με τη μέθοδο της chain-ladder (Verral 2000). Το αρνητικό διωνυμικό μοντέλο μπορεί να προκύψει από το μοντέλο poisson αφήνοντας την ένταση και την απαίτηση να είναι στοχαστικές μεταβλητές (Verral 2000). Όταν η προσέγγιση του αρνητικού διωνυμικού μοντέλου και του μοντέλου του Mack από την κανονική κατανομή προσαρμόστηκε στα δεδομένα, και τα δύο μοντέλα παρήγαγαν ίδιους συντελεστές εξέλιξης. Επίσης, και οι δύο ήταν ταυτόσημες με τους συντελεστές εξέλιξης της chain-ladder.

Μεταξύ των μοντέλων που εισήχθησαν μόνο ένα από αυτά μπορούσε να διαχειριστεί τις αρνητικές προσαυξητικές απαιτήσεις, το μοντέλο του Mack. Ως εναλλακτική λύση για το αρνητικό διωνυμικό μοντέλο, χρησιμοποιήθηκε μια κανονική προσέγγιση, καθώς αυτή θα έλυne το πρόβλημα με τις αρνητικές προσαυξητικές απαιτήσεις. Είναι λιγότερο ελκυστικό να γίνεται χρήση αυτής της προσέγγισης, καθώς χρειάζεται να εκτιμηθούν περισσότερες παράμετροι (οι συντελεστές διακύμανσης). Το μοντέλο του Mack ορίζει μόνο τις δύο πρώτες ροπές, ενώ η προσέγγιση του αρνητικού διωνυμικού μοντέλου από την κανονική κατανομή ορίζει επίσης τους ατομικούς συντελεστές εξέλιξης οι οποίοι κατανέμονται κανονικά. Είναι πιθανό να δημιουργηθούν διαστήματα εμπιστοσύνης χρησιμοποιώντας την κανονική προσέγγιση, ενώ το μοντέλο του Mack απαιτεί περαιτέρω παραδοχές για να συμβεί αυτό.

Τα δύο μοντέλα που χρησιμοποιούνται στην ανάλυση προϋποθέτουν μια συμμετρική κατανομή γύρω από τον μέσο όρο. Τα δύο μοντέλα επιλέχθηκαν λόγω της ικανότητάς τους να διαχειρίζονται τις αρνητικές προσαυξητικές απαιτήσεις και όχι επειδή οι απαιτήσεις

κατανέμονται συμμετρικά γύρω από τον μέσο όρο. Αυτό δεν έχει μελετηθεί σε αυτή τη διπλωματική εργασία.

Η μέθοδος chain-ladder είναι ένα γραμμικό μοντέλο και μπορεί να θεωρηθεί ως γραμμική που διέρχεται από την αρχή των αξόνων. Παρατηρείται ότι στα πρώτα χρόνια εξέλιξης, είναι ένα μοντέλο που περιλαμβάνει έναν, επίσης, διαφορετικό από το μηδέν όρο, θα έδινε καλύτερη προσαρμογή του μοντέλου. Το αποτέλεσμα του εξαναγκασμού της γραμμικής παλινδρόμησης μέσω της αρχής (του μηδενός) είναι ότι οι απαιτήσεις στα πρώτα έτη ανάπτυξης των ζημιών είναι υποτιμώνται, και στα τελευταία έτη ανάπτυξης είναι υπερεκτιμημένες. Αυτή η τάση ήταν επίσης εμφανής κατά την εξέταση των κατάλοιπων και ήταν περισσότερο δεσπίζουσα για τον αριθμό των απαιτήσεων (Mack 1994a). Αυτή η υπόθεση δοκιμάστηκε με τη σύγκρισή της με άλλους συντελεστών εξέλιξης που χρειάζονται άλλες υποθέσεις διακύμανσης για να επιτευχθεί αμερόληπτη εκτίμηση (Mack 1994a) με ελάχιστη διακύμανση. Ούτε από τον ελάχιστο τετραγωνικό εκτιμητή, ούτε από τον απλό μέσο δείκτη αποδείχθηκε τυχαία συμπεριφορά στα κατάλοιπα γραφήματα.

Οι εκτιμήσεις του συντελεστή εξέλιξης, των εκτιμήσεων των αποθεμάτων και του σφάλματος πρόβλεψης υπολογίστηκαν χρησιμοποιώντας την κανονική προσέγγιση και το μοντέλο του Mack. Η πρώτη ροπή των δύο μοντέλων ήταν ίδια. Η δεύτερη ροπή είχε διαφορετικό γράμμα δίνοντας τη συνιστώσα της διακύμανσης. Τα αποτελέσματα δείχνουν πολύ παρόμοιες εκτιμήσεις για τα δύο συστατικά διακύμανσης, αλλά δεν είναι πανομοιότυπα. Η συνιστώσα της διακύμανσης g_j στο μοντέλο του Mack αποδείχθηκε ότι είναι αμερόληπτη και μπορεί να θεωρηθεί ως ένας αμερόληπτος, σταθμισμένος μέσος όρος των κατάλοιπων (Mack 1994a). Η συνιστώσα της διακύμανσης ϕ_j στην κανονική προσέγγιση βρέθηκε με την χρήση ενός γενικευμένου γραμμικού μοντέλου στα

τετραγωνικά κατάλοιπα. Είναι απλώς ένας μέσος όρος των κατάλοιπων, αλλά δεν είναι αμερόληπτος. Ο συντελεστής διακύμανσης, g_j , θα ήταν επίσης μερόληπτικός, εάν ο συντελεστής $\frac{1}{n-j}$ δεν είχε αφαιρέσει τις εκτιμώμενες παραμέτρους στον παρονομαστή.

Το απόθεμα και το σφάλμα πρόβλεψης στα δύο μοντέλα βρέθηκαν αλγεβρικά. Το αποθεματικό και οι εκτιμήσεις των σφαλμάτων πρόβλεψης έχουν επίσης ληφθεί εμπειρικά χρησιμοποιώντας και τα δύο μοντέλα. Οι εκτιμήσεις των αποθεμάτων αυξάνονται για μεγαλύτερα έτη ατυχημάτων, καθώς αυξάνεται ο αριθμός των ετών εξέλιξης. Φυσικά, τα σφάλματα πρόβλεψης αυξάνονται επίσης και για τα έτη με υψηλότερο έτος ατυχήματος. Τα εμπειρικά αποτελέσματα είναι σχεδόν πανομοιότυπα για τα δύο μοντέλα και η διαφορά τους μπορεί να θεωρηθεί λόγω των δύο διαφορετικών παραγόντων διακύμανσης. Για να επιτευχθεί ακριβώς το ίδιο αποτέλεσμα, πρέπει να χρησιμοποιηθεί μια αμερόληπτη μορφή της ϕ_j . Λόγω των ομοιοτήτων μεταξύ των δύο μοντέλων, η προσέγγιση του αρνητικού δυωνυμικού μοντέλου από την κανονική κατανομή μπορεί να θεωρηθεί ότι βασίζεται στο μοντέλο του Mack. Η προσέγγιση του αρνητικού δυωνυμικού μοντέλου από την κανονική κατανομή χρησιμοποιεί ένα γενικευμένο γραμμικό μοντέλο στην εκτίμηση και αυτή η προσέγγιση προσφέρει περισσότερη ευλυγισία στην ανάλυση από το μοντέλο του Mack. Ένα γενικευμένο γραμμικό μοντέλο θα μπορούσε να εφαρμοστεί στη Poisson και στο αρνητικό μοντέλο αν δεν υπήρχε η παρουσία αρνητικών προσαυξητικών απαιτήσεων.

Ο συντελεστής διακύμανσης του τελευταίου έτους εξέλιξης έπρεπε να προσεγγιστεί. Το μοντέλο Mack χρησιμοποιήθηκε δύο φορές χρησιμοποιώντας δύο διαφορετικές προσεγγίσεις. Η διαφορά ήταν μικρή, αλλά αυτό θα μπορούσε να εξεταστεί περαιτέρω

χρησιμοποιώντας άλλες προσεγγίσεις και πιθανόν μια μεγαλύτερη διαφορά να εμφανιζόταν.

Υπάρχουν μερικές αδυναμίες στη μέθοδο chain-ladder. Οι εκτιμητές των τελευταίων συντελεστών εξέλιξης υπολογίζονται χρησιμοποιώντας μόνο ένα μικρό αριθμό παρατηρήσεων. Επιπλέον, τα τελευταία έτη ατυχημάτων απαιτούν προβλέψεις πολλών μεταγενέστερων ετών εξέλιξης. Αυτό καθιστά την τελική πρόβλεψη απαίτησης αβέβαιη, και αυτό είναι εμφανές στο σφάλμα πρόβλεψης.

Το σύνολο δεδομένων από την Trygvesta έδειξε μεγάλο αριθμό αρνητικών προσαυξητικών απαιτήσεων κατά το δεύτερο έτος εξέλιξης. Αυτό θα μπορούσε να είναι συνέπεια ενός μεγάλου αριθμού δηλωμένων απαιτήσεων κατά το επόμενο έτος εξέλιξης και θα έδειχνε ότι οι ατομικοί συντελεστές εξέλιξης είναι συσχετισμένοι. Το μοντέλου του Mack υποδηλώνει ασυσχέτιστους ατομικούς συντελεστές εξέλιξης και η κανονική προσέγγιση υποδηλώνει ανεξάρτητους ατομικούς συντελεστές εξέλιξης. Αν αυτό δεν συμβαίνει, τα μοντέλα δεν είναι κατάλληλα για το σύνολο δεδομένων.

Φαίνεται ότι τα μοντέλα έχουν εντοπίσει ότι οι αρνητικές προσαυξητικές απαιτήσεις θα εμφανιστούν μετά το πρώτο έτος εξέλιξης για το έτος ατυχήματος 19, δεδομένου ότι το αποθεματικό είναι μικρότερο από ότι για το έτος ατυχήματος 18. Περισσότερο εμπειρική έρευνα πρέπει να γίνει για να αποκαλυφθεί αν οι ατομικοί συντελεστές εξέλιξης μεταξύ των ετών είναι ασυσχέτιστοι όταν υπάρχει μεγάλη συχνότητα αρνητικών προσαυξητικών απαιτήσεων.

ΚΕΦΑΛΑΙΟ 5 : Συμπεράσματα

Αυτή η εργασία έδειξε θεωρητικά ότι το πολλαπλασιαστικό μοντέλο, το μοντέλο Poisson, το αρνητικό διωνυμικό μοντέλο και το μοντέλο του Mack παράγουν ισοδύναμα αποτελέσματα με τη μέθοδο chain-ladder. Η προσέγγιση του αρνητικού μοντέλου και του μοντέλου του Mack από την κανονική κατανομή είναι δύο πιθανά μοντέλα για την περίπτωση που υπάρχουν αρνητικές προσαυξητικές απαιτήσεις σε ένα σύνολο δεδομένων. Τα δύο αυτά μοντέλα δημιουργούν σχεδόν ταυτόσημα αποτελέσματα και η κανονική προσέγγιση μπορεί να θεωρηθεί ως ένα αδιαμφισβήτητο μοντέλο του μοντέλου του Mack.

ΚΕΦΑΛΑΙΟ 6 : Βιβλιογραφία

- [1] England, P.D. & Verrall, R.J stochastic Claims Reserving in General Insurance. Presented to the Institute of Actuaries 28.january 2002
- [2] Hess, K. Th. & Schmidt, K. D. 2002, “A comparison of models for the chain-ladder method” , Insurance : Mathematics and Economics, vol. 31 , pp.351-364.
- [3] Mack, T. 1994a, Measuring the variability of chain-ladder reserve estimates, Spring Forum.
- [4] Mack, T. 1994b, “Which stochastic model is underlying the chain-ladder method?”, Insurance: Mathematics and Economics, vol. 15, pp. 133-138.
- [5] Neuhaus, W. 2006, Estimating Outstanding Claims in General Insurance. (Unpublished Work)
- [6] Renshaw, A.E. & Verrall, R. J. 1998, “A stochastic model underlying the chain-ladder technique”, British Actuarial Journal, vol. 4, no. 4, pp. 903-923.
- [7] Verrall, R. J. An investigation into stochastic claims reserving models and the chain-ladder technique. Insurance: Mathematics and Economics 26, 91-99.2000.
- [8] Verrall, R. J. & England, P. D. Stochastic Claims Reserving in General Insurance. Presented to the Institute of Actuaries. 2002.

Παράρτημα 1

AY/DY	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
1	-	522	535	536	538	544	547	549	554	558	564	567	566	567	567	570	571	573	575
2	663	596	595	592	607	610	618	625	633	634	641	643	646	647	648	649	649	650	
3	691	690	710	749	765	775	788	804	818	826	829	833	840	838	839	841	846		
4	757	787	831	855	880	899	908	929	939	945	953	958	959	960	962	961			
5	648	756	753	795	815	833	848	859	864	881	881	884	888	887	888				
6	823	771	767	803	825	852	855	859	871	871	874	878	881	883					
7	913	801	819	871	920	953	961	974	986	994	1001	1001	1002						
8	929	897	915	954	987	1000	1009	1012	1019	1027	1031	1034							
9	1136	1147	1156	1207	1231	1248	1264	1272	1282	1291	1297								
10	1387	1298	1267	1289	1301	1322	1338	1353	1359	1369									
11	1448	1374	1377	1403	1425	1447	1460	1473	1478										
12	1467	1351	1338	1352	1386	1409	1426	1438											
13	1345	1275	1301	1337	1360	1382	1396												
14	1545	1402	1414	1427	1457	1472													
15	1631	1452	1440	1466	1486														
16	1612	1473	1497	1532															
17	1648	1537	1535																
18	1511	1474																	
19	1431																		

Πίνακας 17: Τρίγωνο εξέλιξης του αριθμού των απαιτήσεων αυτοκινήτων από την TrygVesta. Οι σειρές εμφανίζουν τα έτη ατυχημάτων (AY) και οι στήλες εμφανίζουν τα έτη εξέλιξης (DY).

AY/DY	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
1	0,000	5,200	10,600	17,300	21,133	27,125	29,825	30,604	30,944	31,896	33,383	33,967	34,067	35,008	35,359	36,644	36,642	37,392	37,393
2	0,700	6,000	17,100	21,557	27,980	30,980	34,236	38,795	43,551	47,320	48,807	50,881	51,496	52,805	52,646	52,695	52,695	52,699	
3	2,000	7,700	13,654	26,334	38,734	48,492	52,685	57,677	61,664	62,646	64,204	66,754	68,493	69,570	70,418	70,217	70,462		
4	2,500	9,458	25,658	39,001	50,045	63,106	71,721	79,218	85,239	89,252	90,710	92,559	93,815	96,383	97,981	98,052			
5	2,688	11,958	25,241	38,562	53,836	70,607	83,326	92,246	100,784	104,161	107,791	111,542	114,025	116,686	117,068				
6	2,560	9,448	21,037	38,454	49,875	67,396	76,950	81,942	87,534	90,925	96,038	96,957	98,121	100,063					
7	2,795	10,220	27,252	54,768	81,070	100,423	115,359	118,617	121,727	126,110	132,068	131,688	131,801						
8	3,778	11,072	28,103	51,360	73,364	95,403	108,446	117,982	121,342	124,977	132,322	136,538							
9	5,083	15,237	35,488	65,717	92,923	106,712	116,284	132,801	146,637	149,505	151,035								
10	3,996	12,764	28,641	53,651	75,726	90,144	109,356	115,232	120,466	122,910									
11	5,451	17,508	39,835	67,750	89,010	112,132	123,015	134,509	142,955										
12	5,146	17,896	33,035	49,087	68,793	84,154	107,485	114,596											
13	8,264	19,681	29,160	44,865	63,715	80,831	93,745												
14	8,569	23,860	45,769	69,704	101,389	122,228													
15	7,088	21,831	33,254	55,304	75,132														
16	9,580	26,534	45,800	70,765															
17	9,172	32,395	48,358																
18	6,393	22,201																	
19	6,423																		

Πίνακας 18: Τρίγωνο εξέλιξης για το ποσό των ζημιών αυτοκινήτων από την TrygVesta. Οι σειρές εμφανίζουν τα έτη ατυχημάτων (AY) και οι στήλες εμφανίζουν τα έτη εξέλιξης (DY).

Παράρτημα 2

Δείχνουμε ότι, L_C , η μέγιστη συνάρτηση πιθανότητας του C_{ij} , δοθέντος ότι $d_{i,n-i+1}$, είναι

πολλαπλάσια κατανομημένη με την παράμετρο $\frac{y_j}{s_{n-i+1}}$.

Είχαμε από αυτόν τον τύπο (2.14) ότι

$$L_C = \prod_{i=1}^n \left(\frac{d_{i,n-i+1}!}{\prod_{j=1}^{n-i+1} c_{ij}!} \prod_{j=1}^{n-i+1} \left(\frac{y_j}{s_{n-i+1}} \right)^{c_{ij}} \right) = \prod_{i=1}^n \left(\frac{d_{i,n-i+1}!}{\prod_{j=1}^{n-i+1} c_{ij}!} \prod_{j=1}^{n-i+1} (y_{(i)j})^{c_{ij}} \right)$$

Όπου, $y_{(i)j}$, είναι η πιθανότητα για μια ζημιά που πραγματοποιήθηκε κατά το έτος i και καταγράφηκε στο έτος j .

Έστω, C_{ij} , για $i = 1, \dots, n$ και $j = 1, \dots, n - i + 1$ είναι ανεξάρτητες τυχαίες μεταβλητές Poisson,

με αναμενόμενη τιμή $y_{i(j)} = \frac{y_j}{\sum_{k=1}^{n-i+1} y_k}$. Δεδομένου ότι είναι μία παράμετρος, δίνουμε σε

αυτόν τον όρο νέο συμβολισμό, ώστε να δείχνει πιο οικείο $y_{i(j)} = p_{i(j)} = \frac{p_j}{\sum_{k=1}^{n-i+1} p_k}$.

Έχουμε τότε ότι, D_{ij} , για $i = 1, \dots, n$ και $j = 1, \dots, n - i + 1$, είναι ανεξάρτητες τυχαίες Poisson μεταβλητές, με αναμενόμενη τιμή $(p_{i(1)} + p_{i(2)} + \dots + p_{i(n-i+1)})$. Αυτό είναι αποτέλεσμα της

D_{ij} που είναι ένα άθροισμα των Poisson τυχαίων μεταβλητών.

Η δεσμευμένη κατανομή είναι ως εξής

$$f_{c_{ij}|D_{i,n-i+1}}(c_{ij} | d_{i,n-i+1}) = \left(\prod_{j=1}^{n-i+1} \frac{p_j^{c_{ij}} e^{-p_j}}{c_{ij}!} \right) / \left(\frac{(p_1 + \dots + p_{n-i+1})^{d_{i,n-i+1}} e^{-(p_1 + \dots + p_{n-i+1})}}{p_{i,n-i+1}!} \right)$$

$$= \frac{d_{i,n-i+1}!}{\prod_{j=1}^{n-i+1} c_{ij}!} \left(\frac{p_1}{\sum_{j=1}^{n-i+1} p_j} \right)^{c_{i1}} \cdots \left(\frac{p_{n-i+1}}{\sum_{j=1}^{n-i+1} p_j} \right)^{c_{i,n-i+1}}$$

$$= \frac{d_{i,n-i+1}!}{\prod_{j=1}^{n-i+1} c_{ij}!} p_{i(1)} p_{i(2)} \cdots p_{i(n-i+1)}$$

Η τελευταία έκφραση που τώρα αναγνωρίζεται ως μια πολυώνυμη κατανομή για C_{ij}

δοθέντος ότι $d_{i,n-i+1}$.

Παράρτημα 3

Dev. Year	Estimate	St. Error	t-value
1	0,939	0,011	82,24
2	1,005	0,004	250,6
3	1,025	0,004	254,2
4	1,021	0,003	384,1
5	1,017	0,002	556,8
6	1,011	0,001	1154
7	1,010	0,002	655,6
8	1,008	0,001	773,7
9	1,008	0,001	700,4
10	1,005	0,001	978,7
11	1,003	0,001	1499
12	1,003	0,001	881,6
13	1,000	0,001	1306
14	1,001	0	3309
15	1,001	0,001	803,6
16	1,003	0,002	532,9

Πίνακας 19: Εκτιμήσεις, τυπικά σφάλματα και t-τιμές της παραμέτρου για το περιορισμένο γραμμικό μοντέλο για το συνολικό μέγεθος των απαιτήσεων.

Dev. Year	Parameter	Estimate	St. Error	t-value
1	Beta 0	97,727	42,424	2,304
1	Beta 1	0,864	0,034	25,293
2	Beta 0	22,412	14,179	,1581
2	Beta 1	0,986	0,013	77,776
3	Beta 0	23,791	14,077	1,69
3	Beta 1	1,004	0,013	78,02
4	Beta 0	13,346	23,791	1,365
4	Beta 1	1,009	0,013	112,328
5	Beta0	5,894	9,779	0,835
5	Beta 1	1,011	0,009	154,749
6	Beta0	-0,115	7,056	-0,033
6	Beta 1	1,011	0,007	308,659
7	Beta 0	4,167	5,868	0,71
7	Beta 1	1,006	0,005	178,95
8	Beta 0	10,442	3,702	2,821
8	Beta 1	0,998	0,004	272,847
9	Beta 0	-0,571	5,921	-0,096
9	Beta 1	1,008	0,006	163,894
10	Beta 0	4,888	4,112	1,189
10	Beta 1	1,000	0,005	222,491
11	Beta 0	3,154	3,193	0,988
11	Beta 1	1,000	0,004	169,611
12	Beta 0	-0,111	5,865	-0,019
12	Beta 1	1,003	0,007	143,126
13	Beta 0	1,866	3,856	0,484
13	Beta 1	0,998	0,005	209,414
14	Beta 0	-1,773	1,070	-1,657
14	Beta 1	1,003	0,001	744,479
15	Beta 0	6,736	3,330	20,023
15	Beta 1	0,993	0,004	229,461
16	Beta 0	-9,541	6,050	-1,577
16	Beta 1	1,007	0,009	116,967

Πίνακας 20: Εκτιμήσεις, τυπικά σφάλματα και t-τιμές της παραμέτρου για το γενικό μοντέλο για τον αριθμό των απαιτήσεων.

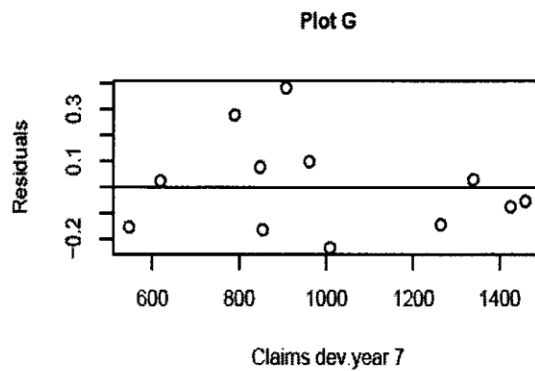
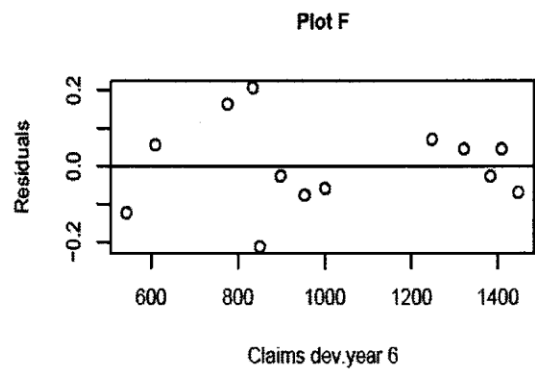
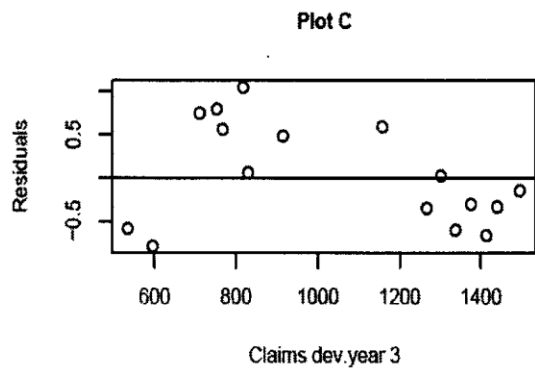
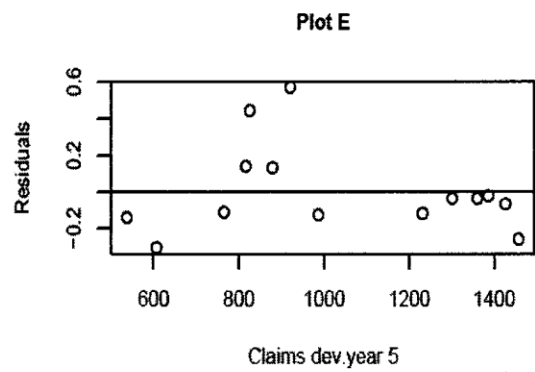
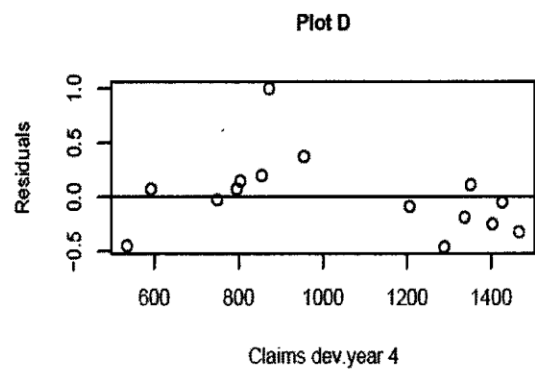
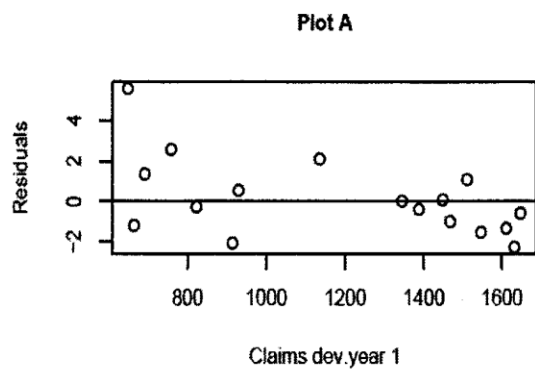
Dev. Year	Estimate	St. Error	t-value
1	3,074	0,115	26,828
2	1,837	0,090	20,322
3	1,652	0,043	38,511
4	1,395	0,016	86,281
5	1,233	0,015	82,725
6	1,148	0,016	73,027
7	1,083	0,010	104,990
8	1,062	0,009	116,790
9	1,030	0,004	258,571
10	1,035	0,007	153,678
11	1,020	0,006	170,537
12	1,012	0,004	277,756
13	1,023	0,002	613,585
14	1,008	0,003	302,115
15	1,002	0,005	186,429
16	1,002	0,001	810,478

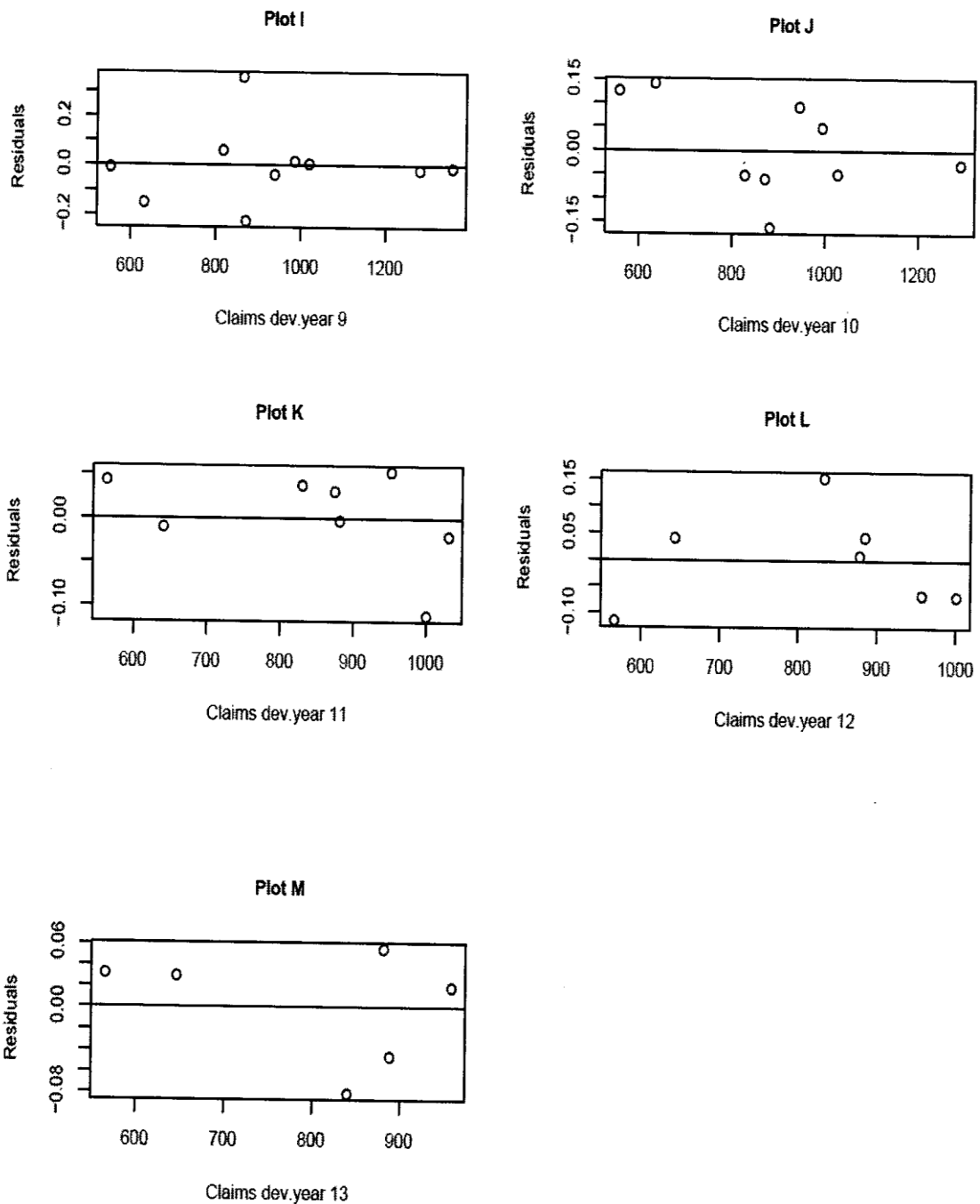
Πίνακας 21: Εκτιμήσεις, τυπικά σφάλματα και t-τιμές της παραμέτρου για το περιορισμένο γραμμικό μοντέλο για το συνολικό μέγεθος των απαιτήσεων.

Dev. Year	Parameter	Estimate	St. Error	t-value
1	Beta 0	3,268	1,174	2,783
1	Beta 1	2,568	0,206	12,477
2	Beta 0	9,949	2,525	3,940
2	Beta 1	1,310	0,149	8,786
3	Beta 0	2,849	4,074	0,699
3	Beta 1	1,563	0,134	11,658
4	Beta 0	-2,822	2,436	-1,158
4	Beta 1	1,449	0,050	29,031
5	Beta0	3,189	2,884	1,106
5	Beta 1	1,189	0,043	27,850
6	Beta0	0,299	3,968	0,075
6	Beta 1	1,144	0,050	22,964
7	Beta 0	0,450	2,861	0,157
7	Beta 1	1,078	0,031	34,409
8	Beta 0	0,706	2,591	0,273
8	Beta 1	1,055	0,027	39,665
9	Beta 0	1,673	0,992	1,687
9	Beta 1	1,014	0,010	100,891
10	Beta 0	0,491	1,907	0,258
10	Beta 1	1,030	0,019	53,463
11	Beta 0	1,196	1,611	0,743
11	Beta 1	1,008	0,017	59,245
12	Beta 0	0,507	0,971	0,522
12	Beta 1	1,007	0,011	92,721
13	Beta 0	0,052	0,444	0,117
13	Beta 1	1,022	0,005	187,784
14	Beta 0	-0,045	0,833	-0,053
14	Beta 1	1,009	0,010	96,506
15	Beta 0	1,389	0,883	1,572
15	Beta 1	0,983	0,013	75,856
16	Beta 0	-0,316	0,219	-1,446
16	Beta 1	1,007	0,004	252,931

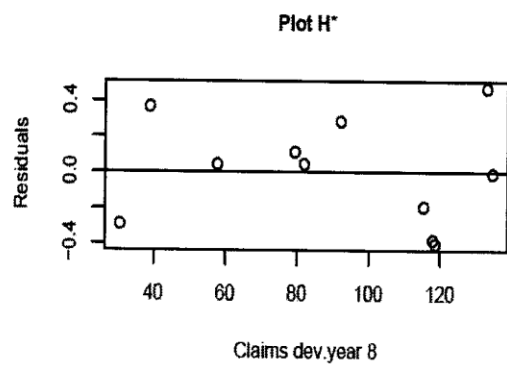
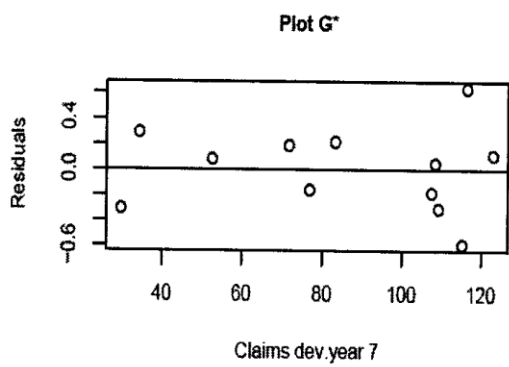
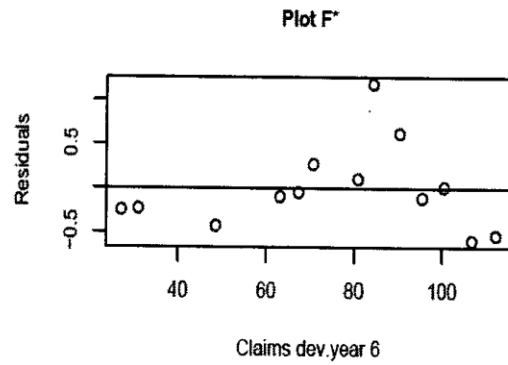
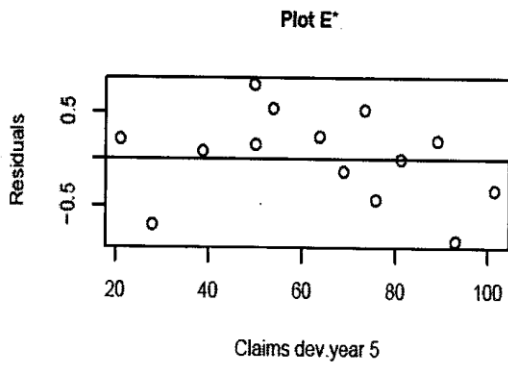
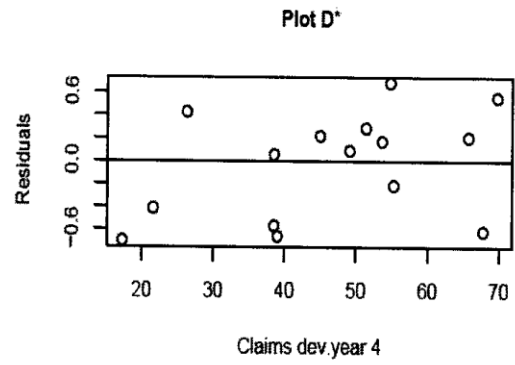
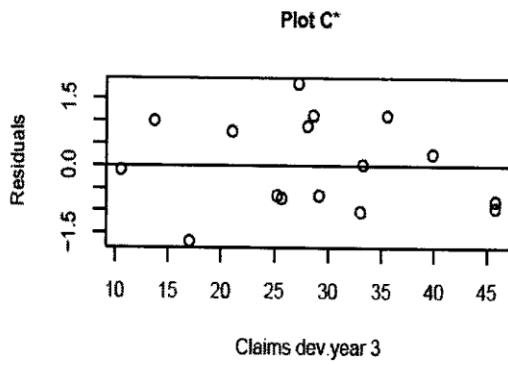
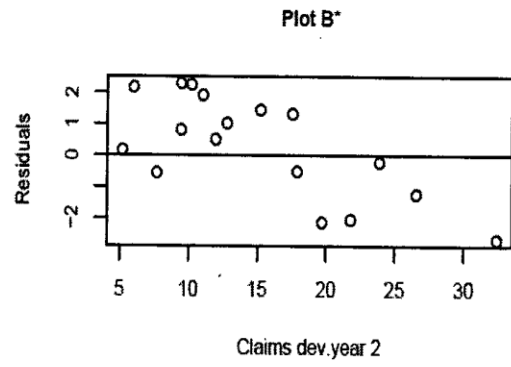
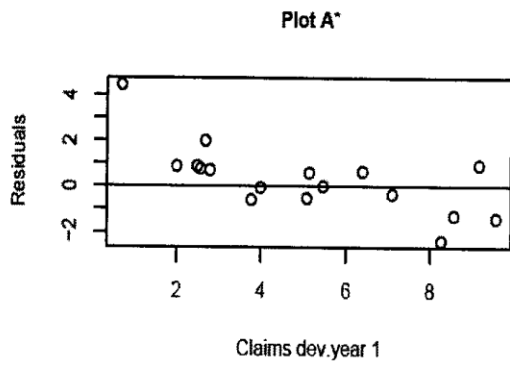
Πίνακας 22: Εκτιμήσεις, τυπικά σφάλματα και t-τιμές της παραμέτρου για το περιορισμένο γραμμικό μοντέλο για τον αριθμό των απαιτήσεων.

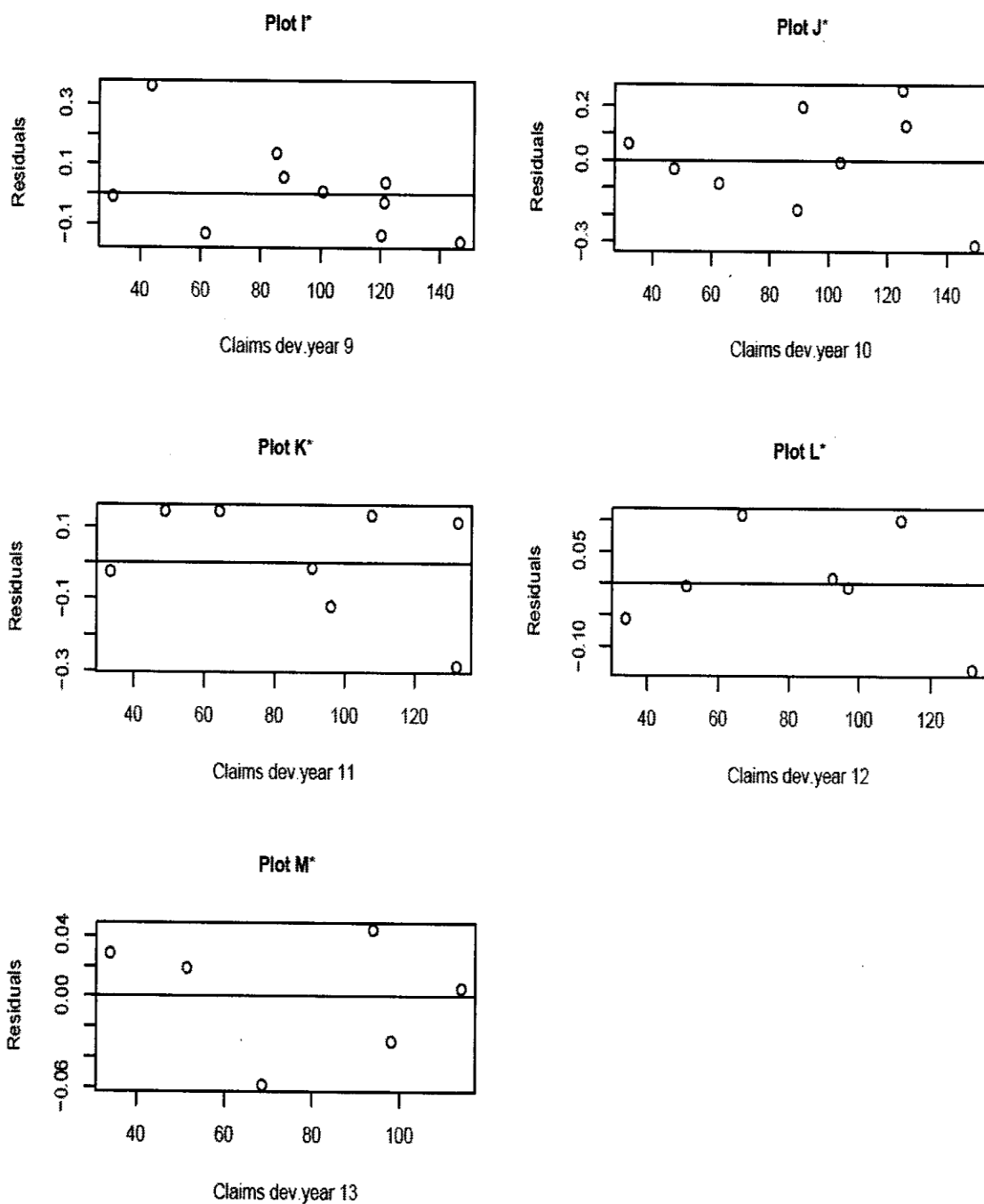
Παράρτημα 4





Γράφημα Α έως Μ. Κατάλοιπα, r_{ij} που καταρτίστηκαν για τις απαιτήσεις του έτους εξέλιξης $j-1$, $d_{i,j-1}$ για τον αριθμό των απαιτήσεων. Στο Γράφημα Α έως Μ παρουσιάζονται τα κατάλοιπα γραφήματα για το έτος εξέλιξης 2 έως το έτος εξέλιξης 14, ως συνάρτηση των απαιτήσεων κατά το προηγούμενο έτος εξέλιξης.





Γράφημα A* έως M*. Κατάλοιπα, r_{ij} που καταρτίστηκαν για τις απαιτήσεις του έτους εξέλιξης $j-1$, $d_{i,j-1}$ για το μέγεθος των απαιτήσεων. Στο Γράφημα A* έως M* παρουσιάζονται τα κατάλοιπα γραφήματα για το έτος εξέλιξης 2 έως το έτος εξέλιξης 14, ως συνάρτηση των απαιτήσεων κατά το προηγούμενο έτος εξέλιξης.

Παράρτημα 5

Έστω X και Y να είναι ανεξάρτητες τυχαίες μεταβλητές . Ο τύπος που πρέπει να αποδειχθεί είναι:

$$\text{Var}(XY) = (E[X]^2)\text{Var}(Y) + E([Y]^2)\text{Var}(X) + \text{Var}(Y)\text{Var}(X)$$

$$\text{Έστω } \delta X = \frac{X - EX}{EX} \text{ και } \delta Y = \frac{Y - EY}{EY}$$

Τότε μπορεί να προκύψει ότι:

$$XY - EXEY = EXEY([\delta X + 1][\delta Y + 1] - 1) = EXEY(\delta X + \delta Y + \delta X\delta Y)$$

η πρώτη συσχέτιση αποδεικνύεται εύκολα με τη χρήση του ορισμού για δX και δY . Για να βρεθεί το $\text{Var}(XY)$ είναι δυνατόν να χρησιμοποιηθεί η γνωστή ταυτότητα, η οποία για τη μεταβλητή X είναι $\text{Var}(X) = E(X - EX)^2$. Έτσι

$$\begin{aligned} \text{Var}(XY) &= E[(XY - EXEY)^2] = E([EXEY]^2[\delta X + \delta Y + \delta X\delta Y]^2) \\ &= [EXEY]^2 E\left[\frac{X - EX}{EX} + \frac{Y - EY}{EY} + \frac{(X - EX)(Y - EY)}{EXEY}\right]^2 \\ &= [EXEY]^2 E\left[\left(\frac{X - EX}{EX}\right)^2 + \left(\frac{Y - EY}{EY}\right)^2 + \left(\frac{(X - EX)(Y - EY)}{EXEY}\right)\right. \\ &\quad \left.+ 2\left(\frac{X - EX}{EX}\right)\left(\frac{Y - EY}{EY}\right) + 2\left(\frac{X - EX}{EX}\right)^2\left(\frac{Y - EY}{EY}\right) + 2\left(\frac{Y - EY}{EY}\right)^2\left(\frac{X - EX}{EX}\right)\right] \end{aligned}$$

Οι τρεις τελευταίοι παράγοντες εξαλείφθηκαν εξαιτίας της ανεξαρτησίας μεταξύ των X και Y , και η έκφραση γίνεται:

$$\text{Var}(XY) = E[(XY - EXEY)^2] = (EX)^2\text{Var}(Y) + (EY)^2\text{Var}(X) + \text{Var}(X)\text{Var}(Y)$$

(Goodman 1960)

Παράρτημα 6

Αυτή η απόδειξη αναπαράγεται από τον Mack (1994a).

Δλινονται οι τρεις υποθέσεις του Mack

1. Υπάρχουν σταθερές f_2, \dots, f_n τέτοιες ώστε $E(D_{i,j} | K_{i,j-1} = k_{i,j-1}) = f_j d_{i,j-1}$ για $j = 2, \dots, n$
2. Υπάρχουν σταθερές g_2, \dots, g_n τέτοιες ώστε $Var(D_{i,j} | K_{i,j-1} = k_{i,j-1}) = g_j d_{i,j-1}$ για $j = 2, \dots, n$
3. K_{in} και K_{kn} είναι ανεξάρτητες για $i \neq k$.

Το MSE του αποθεματικού είναι

$$MSE(\hat{R}_i | k) = \hat{D}_{in}^2 \sum_{j=n-i+2}^n \frac{\hat{g}_j^2}{\hat{f}_j^2} \left(\frac{1}{\hat{D}_{i,j-1}} + \frac{1}{\sum_{i=1}^{n-j+1} d_{i,j-1}} \right) \quad (A.1)$$

Απόδειξη:

Το MSE του αποθέματος είναι το ίδιο με το MSE της τελικής απαίτησης \hat{D}_{in} , το αποτέλεσμα αυτό αποδεικνύεται στις (3.17) και (3.18).

$$MSE(\hat{R}_i | k) = MSE(\hat{D}_{in} | k) \quad (A.2)$$

Ο στόχος είναι να βρεθεί το MSE του αποθεματικού. Είναι απαραίτητο να βρούμε το MSE της τελικής αξίωσης \hat{D}_{in} που θα γίνει εδώ. Η διακύμανση της πρόβλεψης της τελικής απαίτησης ήταν στο (2.60) που βρέθηκε να είναι

$$MSE(\hat{D}_{in} | k) = Var(D_{in} | k) + (E(D_{in} | k) - \hat{D}_{in})^2 \quad (A.3)$$

Με την τρίτη παραδοχή του Mack αρκεί να δεσμεύσουμε τις παρατηρήσεις μέσα στο έτος ατυχήματος i . Είναι απαραίτητο μόνο να βρεθεί $Var(D_{in} | k_{i,n-i+1})$ και $(E(D_{in} | k_{i,n-i+1}) - \hat{D}_{in})^2$

Ο όρος $Var(D_{in} | k_{i,n-i+1})$ θα παρουσιαστεί πρώτος. Ο όρος μπορεί να γραφτεί ως:

$$Var(D_{in} | k_{i,n-i+1}) = E((D_{in})^2 | k_{i,n-i+1}) - (E(D_{in} | k_{i,n-i+1}))^2 \quad (A.4)$$

Θα προσδιοριστεί πρώτα ο δεύτερος όρος της (A.4). Με βάση τον κανόνα της αναμενόμενης τιμής προσδιορίζονται οι αναμενόμενες τιμές των απαιτήσεων στη κάτω-αριστερή γωνία του τριγώνου εξέλιξης ($j \geq n - i + 1$). Οι αναμενόμενες απαιτήσεις των δύο τελευταίων ετών εξέλιξης σε περίπτωση ατυχήματος i είναι

$$\begin{aligned} E(D_{in} | k_{i,n-i+1}) &= E(E(D_{in} | D_{i,n-1}) | k_{i,n-i+1}) = f_n E((D_{i,n-1}) | k_{i,n-i+1}) \text{ και} \\ E(D_{i,n-1} | k_{i,n-i+1}) &= E(E(D_{i,n-1} | D_{i,n-2}) | k_{i,n-i+1}) = f_{n-1} E((D_{i,n-2}) | k_{i,n-i+1}) \end{aligned} \quad (A.5)$$

Η εισαγωγή της πρώτης έκφρασης που προκύπτει στο (A.5) στην επόμενη γίνεται μια νέα έκφραση

$$E(D_{in} | k_{i,n-i+1}) = f_{n-1} f_n E((D_{i,n-2}) | k_{i,n-i+1})$$

Η πραγματοποίηση αυτού του βήματος αρκετές φορές καθιστά δυνατή την εύρεση $E(D_{in} | k_{i,n-i+1})$ ή την αναμενόμενη τιμή οποιασδήποτε μελλοντικής απαίτησης στη κάτω-αριστερή γωνία του τριγώνου εξέλιξης. Γενικά, για το έτος ατυχήματος $i = 2, \dots, n$ και το έτος εξέλιξης $j \geq n - i + 2$, ο τύπος είναι:

$$E(D_{ij} | k_{i,n-i+1}) = f_{n-i+2} \dots f_{j-1} f_j E(D_{i,n-i+1} | k_{i,n-i+1}) = f_{n-i+2} \dots f_{j-1} f_j d_{i,n-i+1} \quad (A.6)$$

Ο γενικός τύπος $E(D_{ij} | k_{i,n-i+1})$ ανακτάται και είναι ασήμαντο να υπολογιστούν οι όροι $E(D_{in} | k_{i,n-i+1})$ και $(E(D_{in} | k_{i,n-i+1}))^2$. Το δεύτερο μέρος του (A.4) προσδιορίζεται.

Το επόμενο βήμα είναι να βρούμε $E\left((D_{in})^2 | k_{i,n-i+1}\right)$. Η ανάκτηση $E\left((D_{in})^2 | k_{i,n-i+1}\right)$ της ταυτότητας $E\left((D_{ij})^2 | k_{i,n-i+1}\right)$ θα υπολογιστεί πρώτα χρησιμοποιώντας τον κανόνα της διπλής αναμενόμενης τιμής και τις δύο πρώτες υποθέσεις του Mack.

$$\begin{aligned}
E\left((D_{in})^2 | k_{i,n-i+1}\right) &= E\left((D_{in})^2 | D_{i,j-1} | k_{i,n-i+1}\right) & (A.7) \\
&= E\left\{\text{Var}(D_{ij} | D_{i,j-1}) + (E(D_{ij} | D_{i,j-1}))^2 | k_{i,n-i+1}\right\} \\
&= E\left\{g_j D_{i,j-1} + (f_j D_{i,j-1})^2 | k_{i,n-i+1}\right\} \\
&= g_j E(D_{i,j-1} | k_{i,n-i+1}) + f_j^2 E(D_{i,j-1}^2 | k_{i,n-i+1}).
\end{aligned}$$

$E\left((D_{ij})^2 | k_{i,n-i+1}\right)$ μπορεί να προσδιοριστεί χρησιμοποιώντας τις (A.7) και (A.5). Οι υπολογισμοί μπορούν να φανούν παρακάτω και οι τύποι που χρησιμοποιούνται είναι γραμμένοι στη δεξιά πλευρά.

$$E\left((D_{in})^2 | k_{i,n-i+1}\right) = g_n E(D_{i,n-1} | k_{i,n-i+1}) + f_n^2 E(D_{i,n-1}^2 | k_{i,n-i+1}) \quad (A.7)$$

$$= g_n (f_{n-i+2} \dots f_{n-1} d_{i,n-i+1}) + \quad (A.5)$$

$$= g_{n-1} E(D_{i,n-2} | k_{i,n-i+1}) f_n^2 + E(D_{i,n-2}^2 | k_{i,n-i+1}) f_{n-1}^2 f_n^2 \quad (A.7)$$

$$= g_n (f_{n-i+2} \dots f_{n-1} d_{i,n-i+1}) + g_{n-1} d_{i,n-i+2} f_{n-i+2} \dots f_{n-2} f_n^2 \quad (A.5)$$

$$+ g_{n-2} E(D_{i,n-3} | k_{i,n-i+1}) f_{n-1}^2 f_n^2 + E(D_{i,n-3}^2 | k_{i,n-i+1}) f_{n-2}^2 f_{n-1}^2 f_n^2 \quad (A.7)$$

$$= g_n (f_{n-i+2} \dots f_{n-1} d_{i,n-i+1}) + g_{n-1} d_{i,n-i+1} f_{n-i+2} \dots f_{n-2} f_n^2$$

$$+ g_{n-2} d_{i,n-i+1} f_{n-i+2} \dots f_{n-3} f_{n-1}^2 f_n^2 \quad (A.5)$$

$$+ g_{n-3} E(D_{i,n-4} | k_{i,n-i+1}) f_{n-2}^2 f_{n-1}^2 f_n^2$$

(A.5)

$$+ E(D_{i,n-4}^2 | k_{i,n-i+1}) f_{n-3}^2 f_{n-2}^2 f_{n-1}^2 f_n^2$$

κτλ.

Τα αποτελέσματα από τις (A.5) και (A.7) χρησιμοποιούνται μέχρι το τελευταίο βήμα όταν είναι σαφές ότι, $E(D_{i,n-i+1}^2 | k_{i,n-i+1}) = d_{i,n-i+1}^2$, δοθέντος $k_{i,n-i+1}$, είναι γνωστό. Αυτό μπορεί να γραφεί ως:

$$E(D_{in}^2 | k_{i,n-i+1}) = d_{i,n-i+1}^2 \sum_{j=n-i+2}^n (f_{n-i+2} \dots f_{j-1} g_j f_{j+1}^2 \dots f_n^2) + d_{i,n-i+1}^2 f_{n-i+2}^2 \dots f_n^2 \quad (\text{A.8})$$

Έχουμε δημιουργήσει εκτιμητές για $E(D_{in}^2 | k_{i,n-i+1})$ και $E(D_{in} | k_{i,n-i+1})$ και (A.4) μπορεί να γραφεί όπως:

$$\begin{aligned} \text{Var}(D_{in} | k_{i,n-i+1}) &= E(D_{in}^2 | k_{i,n-i+1}) - (E(D_{in} | k_{i,n-i+1}))^2 \\ &= d_{i,n-i+1}^2 \sum_{j=n-i+2}^n (f_{n-i+2} \dots f_{j-1} g_j f_{j+1}^2 \dots f_n^2) \\ &\quad + d_{i,n-i+1}^2 f_{n-i+2}^2 \dots f_n^2 - (d_{i,n-i+1} f_{n-i+2} \dots f_n)^2 \\ &= d_{i,n-i+1}^2 \sum_{j=n-i+2}^n f_{n-i+2} \dots f_{j-1} g_j f_{j+1}^2 \dots f_n^2 \end{aligned} \quad (\text{A.9})$$

Γράφοντας πρώτα (A.9) και στη συνέχεια αντικαθιστώντας τις παραμέτρους g_j και f_j με \hat{g}_j και \hat{f}_j η διακύμανση της διαδικασίας είναι:

$$\begin{aligned} \text{Var}(D_{in} | k_{i,n-i+1}) &= d_{i,n-i+1}^2 \hat{f}_{n-i+2}^2 \dots \hat{f}_n^2 \sum_{j=n-i+2}^n \frac{f_{n-i+2} \dots f_n}{d_{i,n-i+1} \hat{f}_{n-i+2}^2 \dots \hat{f}_n^2} \\ &= d_{i,n-i+1}^2 \hat{f}_{n-i+2}^2 \dots \hat{f}_n^2 \sum_{j=n-i+2}^n \frac{\hat{g}_j}{d_{i,n-i+1} \hat{f}_{n-i+2}^2 \dots \hat{f}_j^2} \\ &= \hat{D}_{in}^2 \sum_{j=n-i+2}^n \frac{\left(\frac{\hat{g}_j}{\hat{f}_j^2} \right)}{\hat{D}_{i,j-1}}. \end{aligned} \quad (\text{A.10})$$

Για να φτάσουμε στην τελευταία έκφραση (A.10), χρησιμοποιήσαμε ότι $\hat{D}_{in} = D_{i,j-1} \hat{f}_j \dots \hat{f}_n$

όταν $D_{i,j-1}$ εκτιμάται και $\hat{D}_{in} = d_{i,n-i+1} \hat{f}_{n-i+2} \dots \hat{f}_n$, όπου $d_{i,n-i+1}$, θεωρείται γνωστό.

Το πρώτο μέρος του (A.2) έχει βρεθεί, και το επόμενο που μας ενδιαφέρει να βρούμε

$(E(D_{in} | k_{i,n-i+1}) - \hat{D}_{in})^2$. Χρησιμοποιώντας τον τύπο που βρίσκεται στο (A.6) έχουμε ότι

$$\begin{aligned} (E(D_{in} | k_{i,n-i+1}) - \hat{D}_{in})^2 &= \left((d_{i,n-i+1} f_{n-i+2} \dots f_n) - (d_{i,n-i+1} \hat{f}_{n-i+2} \dots \hat{f}_n) \right)^2 \\ &= d_{i,n-i+1}^2 \left(f_{n-i+2} \dots f_n - \hat{f}_{n-i+2} \dots \hat{f}_n \right)^2 \end{aligned} \quad (\text{A.11})$$

Σε αντίθεση με αυτό που κάναμε από το (A.9) έως το (A.10), δεν είναι καλή ιδέα να

αντικαταστήσουμε την παράμετρο f_j με \hat{f}_j . Εάν γίνει αυτό, θεωρείται ότι ο εκτιμητής \hat{f}_j

είναι στην πραγματικότητα ίδιο με την τμή f_j , αλλά είναι πιο ρεαλιστικό να υπάρχει μια

διαφορά μεταξύ του εκτιμητή και της παραμέτρου. Για να λυθεί αυτό το πρόβλημα, ο Mack

(1994α) εισήγαγε μια νέα ταυτότητα F . Η F δεν έχει καμία σχέση με τον συντελεστή

εξέλιξης. Ορίζεται ως:

$$F = f_{n-i+2} \dots f_n - \hat{f}_{n-i+2} \dots \hat{f}_n = S_{n-i+2} + \dots + S_n \quad (\text{A.12})$$

όπου

$$\begin{aligned} S_j &= \hat{f}_{n-i+2} \dots \hat{f}_{j-1} f_j f_{j+1} \dots f_n - \hat{f}_{n-i+2} \dots \hat{f}_{j-1} \hat{f}_j \hat{f}_{j+1} \dots \hat{f}_n \\ &= \hat{f}_{n-i+2} \dots \hat{f}_{j-1} (f_j - \hat{f}_j) f_{j+1} \dots f_n. \end{aligned} \quad (\text{A.13})$$

Το νέο αναγνωρισμένο F τετράγωνο μπορεί να γραφεί όπως:

$$F^2 = (S_{n-i+2} + \dots + S_n)^2 = \sum_{j=n-i+2}^n S_j^2 + 2 \sum_{\substack{j,k=n-i+2 \\ k < j}} \sum_{k=n-i+2}^n S_k S_j \quad (\text{A.14})$$

S^2_j και $S_k S_j$, μπορούν να προσεγγιστούν με $E(S^2_j | k_{j-1})$ και $E(S_k S_j | k_{j-1})$. Με τη χρήση αυτής της προσέγγισης οι παρατηρήσεις λαμβάνονται υπόψη, και αυτό δε θα συνέβαινε όταν προσεγγίζουμε $E(\hat{f}_j)$ με f_j . Δεδομένου ότι \hat{f}_j είναι ένας αμερόληπτος εκτιμητής (βλ. κεφάλαιο 2.6), $E\left((f_j - \hat{f}_j)^2 | k_{j-1}\right) = 0$ και επίσης $E(S_j S_l | k_{l-1}) = 0$. Για να το δούμε αυτό, οι υπολογισμοί γίνονται από κάτω από την υπόθεση, όπου $j < l$:

$$\begin{aligned} E(S_j S_l | k_{l-1}) &= E\left(\hat{f}_{n-i+2} \dots \hat{f}_{j-1} (f_j - \hat{f}_j) f_{j+1} \dots f_n \right) \left(\hat{f}_{n-i+2} \dots \hat{f}_{l-1} (f_l - \hat{f}_l) f_{l+1} \dots f_n \right) | k_{l-1} \\ &= \left(\hat{f}_{n-i+2} \dots \hat{f}_{j-1} (f_j - \hat{f}_j) f_{j+1} \dots f_n \right) E\left(\hat{f}_{n-i+2} \dots \hat{f}_{l-1} (f_l - \hat{f}_l) f_{l+1} \dots f_n \right) | k_{l-1} \\ &= 0. \end{aligned}$$

Η ταυτότητα της $E(S^2_j | k_{j-1})$ μπορεί να εξετασθεί $E\left((f_j - \hat{f}_j)^2 | k_{j-1}\right)$:

$$E\left((f_j - \hat{f}_j)^2 | k_{j-1}\right) = \text{Var}\left(\hat{f}_j | k_{j-1}\right) = \text{Var}\left(\frac{\sum_{i=1}^{n-j+1} D_{ij}}{\sum_{i=1}^{n-j+1} D_{i,j-1}} | k_{j-1}\right) \quad (\text{A.15})$$

$$= \frac{1}{\left(\sum_{i=1}^{n-j+1} d_{i,j-1}\right)^2} \text{Var}\left(\sum_{i=1}^{n-j+1} D_{ij} | k_{j-1}\right)$$

$$= \frac{1}{\left(\sum_{i=1}^{n-j+1} d_{i,j-1}\right)^2} \sum_{i=1}^{n-j+1} d_{i,j-1} g_j$$

$$= \frac{g_j}{\left(\sum_{i=1}^{n-j+1} d_{i,j-1}\right)} \quad (\text{A.16})$$

Η έκφραση στην (A.16) ήταν μόνο ένα μέρος από αυτό που χρειάζεται για να βρεθεί το $E(S^2_j | k_{j-1})$, αλλά χρησιμοποιώντας την (A.16) είναι σαφές ότι $E(S^2_j | k_{j-1})$ μπορεί να εκφραστεί ως:

$$E(S^2_j | k_{j-1}) = \hat{f}^2_{n-i+2} \dots \hat{f}^2_{j-1} E\left((f_j - \hat{f}_j)^2 | k_{j-1}\right) f^2_{j+1} \dots f^2_n = \hat{f}^2_{n-i+2} \dots \hat{f}^2_{j-1} \frac{g_j}{\sum_{i=1}^{n-j+1} d_{i,j-1}} f^2_{j+1} \dots f^2_n$$

\hat{f}_l για $l < j$ μια κλιμακωτή, γιατί δοθέντος ότι k_{j-1} .

Με την αντικατάσταση των παραμέτρων f_j και g_j με \hat{f}_j και \hat{g}_j στην παραπάνω έκφραση, ένας εκτιμητής για $(E(D_{in} | k_{i,n-i+1}) - \hat{D}_{in})^2$ προκύπτει:

$$\begin{aligned} (E(D_{in} | k_{i,n-i+1}) - \hat{D}_{in})^2 &= d^2_{i,n-i+1} F^2 = d^2_{i,n-i+1} \sum_{i=n-i+2}^n E(S^2_j | k_{j-1}) \\ &= d^2_{i,n-i+1} \hat{f}^2_{n-i+2} \dots \hat{f}^2_n \sum_{j=n-i+2}^n \frac{\left(\frac{\hat{g}_j}{\hat{f}_j^2}\right)}{\sum_{i=1}^{n-j+1} d_{i,j-1}} \\ &= \hat{D}_{in}^2 \sum_{j=n-i+2}^n \frac{\left(\frac{\hat{g}_j}{\hat{f}_j^2}\right)}{\sum_{i=1}^{n-j+1} d_{i,j-1}}. \end{aligned} \quad (A.17)$$

Τέλος, ο εκτιμητής της διακύμανσης πρόβλεψης είναι διαθέσιμος. Χρησιμοποιώντας τους τύπους από (A.2), (A.10) και (A.17) μπορούμε να δούμε ότι η διακύμανση πρόβλεψης \hat{D}_{in} είναι:

$$MSE(\hat{R}_i) = Var(D_{in} | K) + (E(D_{in} | K) - \hat{D}_{in})^2$$

$$\begin{aligned}
&= \hat{D}_{in}^2 \sum_{j=n-i+2}^n \frac{\left(\frac{\hat{g}_j}{\hat{f}_j^2} \right)}{\hat{D}_{i,j-1}} + \hat{D}_{in}^2 \sum_{j=n-i+2}^n \frac{\left(\frac{\hat{g}_j}{\hat{f}_j^2} \right)}{\sum_{i=1}^{n-j+1} d_{i,j-1}} \\
&= \hat{D}_{in}^2 \sum_{j=n-i+2}^n \frac{\hat{g}_j}{\hat{f}_j^2} \left(\frac{1}{\hat{D}_{i,j-1}} + \frac{1}{\sum_{i=1}^{n-j+1} d_{i,j-1}} \right). \tag{A.18}
\end{aligned}$$

Εκτός από τη διακύμανση πρόβλεψης \hat{D}_{in} , είναι σημαντικό να βρεθεί η διακύμανση πρόβλεψης της συνολικής εκτίμησης των αποθεματικών.

Οι εξηγήσεις είναι μικρότερες όταν αποδεικνύεται αυτός ο τύπος δεδομένου ότι έχουν γίνει οι ίδιοι υπολογισμοί κατά την εύρεση της διακύμανσης πρόβλεψης για κάθε έτος ατυχήματος. Αντίθετα, οι ήδη υπάρχουσες ταυτότητες των τύπων που χρησιμοποιούνται θα γραφτούν στη δεξιά πλευρά του υπολογισμού. Η ταυτότητα που πρέπει να αποδειχθεί είναι

$$MSE(\hat{R} | k) = \sum_{i=1}^n \left\{ MSE(\hat{R}_i | k) + \hat{D}_{in} \left(\sum_{l=i+1}^n \hat{D}_{ln} \right) \left(\sum_{k=n-i+2}^n \frac{2\hat{g}_k}{\hat{f}_j^2 \sum_{l=1}^{n-k+1} d_{lk}} \right) \right\}$$

Απόδειξη:

$$\begin{aligned}
MSE(\hat{R} | k) &= MSE\left(\sum_{i=2}^n \hat{R}_i | k \right) \\
&= E\left(\left(\sum_{i=2}^n \hat{R}_i - \sum_{i=2}^n R_i \right)^2 | k \right) \tag{A.19}
\end{aligned}$$

$$= E\left(\left(\sum_{i=2}^n \hat{D}_{in} - \sum_{i=2}^n D_{in}\right)^2 \mid k\right) \quad (3.18)$$

$$= \text{Var}\left(\sum_{i=2}^n D_{in} \mid k\right) + \left(E\left(\sum_{i=2}^n D_{in} \mid k\right) - \sum_{i=2}^n \hat{D}_{in}\right)^2 \quad (2.60)$$

$$= (1) + (2).$$

οι δύο εκφράσεις (1) και (2) θα προσδιοριστούν χωριστά.

$$(1) = \text{Var}\left(\sum_{i=2}^n D_{in} \mid k\right) \underline{\underline{\text{ανεξαρτησ } \acute{\alpha} \sum_{i=2}^n \text{Var}(D_{in} \mid k)}}} \quad (A.10)$$

$$(2) = \left(E\left(\sum_{i=2}^n D_{in} \mid k\right) - \sum_{i=2}^n \hat{D}_{in}\right)^2 = \left(\sum_{i=1}^n (E(D_{in} \mid k) - \hat{D}_{in})\right)^2$$

$$= \left(\sum_{i=2}^n (d_{i,n-i+1} f_{n-i+2} \dots f_n - d^2_{i,n-i+1} \hat{f}^2_{n-i+2} \dots \hat{f}_n^2)\right)^2 \quad (A.11)$$

$$= \left(\sum_{i=2}^n d_{i,n-i+1} (f_{n-i+2} \dots f_n - \hat{f}^2_{n-i+2} \dots \hat{f}_n^2)\right)^2$$

$$= \left(\sum_{i=2}^n d_{i,n-i+1} F_i\right)^2$$

$$= \sum_{i=2}^n d^2_{i,n-i+1} F_i^2 + 2 \sum_{i=2}^n \sum_{\substack{l=2 \\ i < l}}^n d_{i,n-i+1} d_{l,n-i+1} F_i F_l .$$

Η μεταβλητή F εισήχθη στο (A.12) και ορίστηκε ως $F = (f_{n-i+2} \dots f_n - \hat{f}_{n-i+2} \dots \hat{f}_n)$. Στην περίπτωση αυτή η μεταβλητή είναι απαραίτητη $i = 2, \dots, n$ για να συμπεριληφθεί ένας δείκτης i . Επομένως η μεταβλητή γίνεται

$$F_i = (f_{n-i+2} \dots f_n - \hat{f}_{n-i+2} \dots \hat{f}_n).$$

Για να βρεθεί μια απλούστερη έκφραση για (2) πρέπει να προσδιοριστεί ένας εκτιμητής $F_i F_l$. Αυτό γίνεται χρησιμοποιώντας την ίδια διαδικασία όπως στο (A.12-A.16). Ο εκτιμητής

$F_i F_l$ είναι

$$\sum_{q=n-i+1}^n \hat{f}_{n-l+2} \cdot \hat{f}_{n-l+1} \hat{f}_{n-l+2}^2 \dots \hat{f}_{q-1}^2 \frac{\hat{g}_q \hat{f}_{q+1}^2 \dots \hat{f}_n^2}{\sum_{i=1}^{n-q+1} d_{i,q-1}}. \quad (\text{A.12-A.16})$$

Οι δύο ταυτότητες (1) και (2) προστίθενται και χρησιμοποιείται παραπάνω.

Απομνημονεύοντας την έκφραση του $MSE(\hat{R}_i | k)$ είναι σαφές ότι:

$$\begin{aligned} MSE(\hat{R}_i | k) &= (1) + (2) = \sum_{i=2}^n Var(D_{in} | k) + \sum_{i=2}^n d^2_{i,n-i+1} F^2_i + 2 \sum_{i=2}^n \sum_{\substack{l=2 \\ i < l}}^n d_{i,n-i+1} d_{l,n-i+1} F_i F_l \\ &= \sum_{i=1}^n \left\{ MSE(\hat{R}_i | k) + \hat{D}_{in} \left(\sum_{l=i+1}^n \hat{D}_{ln} \right) \left(\sum_{k=n-i+2}^n \frac{2 \hat{g}_k}{\hat{f}_j^2 \sum_{l=1}^{n-k+1} d_{lk}} \right) \right\}. \end{aligned}$$