

ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΕΙΡΑΙΩΣ



ΤΜΗΜΑ ΣΤΑΤΙΣΤΙΚΗΣ ΚΑΙ ΑΣΦΑΛΙΣΤΙΚΗΣ ΕΠΙΣΤΗΜΗΣ

ΠΡΟΓΡΑΜΜΑ ΜΕΤΑΠΤΥΧΙΑΚΩΝ ΣΠΟΥΔΩΝ ΣΤΗΝ
ΑΝΑΛΟΓΙΣΤΙΚΗ ΕΠΙΣΤΗΜΗ ΚΑΙ ΔΙΟΙΚΗΤΙΚΗ ΚΙΝΔΥΝΟΥ

Μέθοδοι αποθεματοποίησης με χρήση των
γενικευμένων γραμμικών μοντέλων

Ελευθέριος Ν. Παπαϊωάννου

Πειραιάς 2019

UNIVERSITY OF PIRAEUS



DEPARTMENT OF STATISTICS AND INSURANCE
SCIENCE

POSTGRADUATE PROGRAM IN ACTUARIAL SCIENCE
AND RISK MANAGEMENT

Generalized linear models in reserving risk

Eleftherios N. Papaioannou

Piraeus 2019

Αφιερώσεις

Στην μητέρα μου Μαρία στον πατέρα μου Νικόλαο και στην αδερφή μου Ελένη.

Ευχαριστίες

Κατ'αρχάς θα ήθελα να ευχαριστήσω ιδιαίτερα τον επιβλέποντα για την παρούσα διπλωματική εργασία κύριο Πιτσέλη Γεώργιο, Επίκουρο Καθηγητή του τμήματος Στατιστικής και Ασφαλιστικής Επιστήμης του Πανεπιστημίου Πειραιώς για την πολύτιμη καθοδήγηση που μου προσέφερε κατά τη διάρκεια εκπόνησης της εργασίας. Επίσης θα ήθελα να ευχαριστήσω και τα άλλα δύο μέλη της Τριμελούς Εξεταστικής Επιτροπής, κύριο Πολίτη Κωνσταντίνο και κύριο Ψαρράκο Γεώργιο για την επίβλεψή τους. Ακόμη θα ήθελα να ευχαριστήσω την οικογένεια μου για την αμέριστη συμπαράσταση και την ψυχολογική στήριξη διότι χωρίς αυτές η ολοκλήρωση της παρούσας εργασίας θα ήταν αδύνατη.

Περίληψη

Βασικός στόχος της παρούσας διπλωματικής εργασίας είναι να αναφέρει τις βασικές μεθόδους αποθεματοποιήσεις με την χρήση των γενικευμένων γραμμικών μοντέλων. Παρουσιάζονται οι βασικές μέθοδοι αποθεματοποίησης και οι βασικοί ορισμοί των γενικευμένων γραμμικών μοντέλων. Τέλος γίνεται η εφαρμογή κάποιων μοντέλων στην R για την εκτίμηση των αποθεμάτων.

Abstract

The purpose of this thesis is to present the basic methods of generalized linear models in reserving risk. The basic methods of reserving and the basic rules of the generalized linear models are analyzed. Last but not least, examples of estimations for the reserves with generalized linear models using distributions from the Tweedie family are performed with the use of R.

Περιεχόμενα

1	Εισαγωγή	3
2	Αποθεματοποίηση	5
2.1	Ορισμός ασφαλιστικής επιχείρησης	5
2.1.1	Ο ρόλος της ασφαλιστικής επιχείρησης	6
2.2	Τι είναι αποθεματοποίηση	7
2.3	Βασική ορολογία	8
2.4	Κλασικές μέθοδοι αποθεματοποίησης	10
2.4.1	Εισαγωγή	10
2.4.2	Μέθοδος αναμενόμενου δείκτη ζημιών	12
2.4.3	Chain Ladder (distribution free model)	12
2.4.4	Bornhuetter-Ferguson (B-F)	17
2.4.5	Μέτρα αβεβαιότητας των αποθεμάτων	20
3	Γενικευμένα γραμμικά μοντέλα	22
3.1	Οι βασικές υποθέσεις και η ορολογία	22
3.1.1	Το κλασικό μοντέλο παλινδρόμησης	22
3.1.2	Το γραμμικό μοντέλο	22
3.1.3	Γενικευμένα γραμμικά μοντέλα	23
3.2	Δομή του γενικευμένου γραμμικού μοντέλου	25
3.2.1	Συνάρτηση σύνδεσης	25
3.2.2	Εκθετική οικογένεια κατανομών	26
3.3	Εκτίμηση παραμέτρων	29
3.3.1	Μέθοδος μεγίστης πιθανοφάνειας	29
3.3.2	Συνάρτηση Wald	33
3.3.3	Score test	33
3.4	Επάρκεια μοντέλου	34
3.4.1	Έλεγχος υποθέσεων	36

3.4.2	Δείκτες καλής προσαρμογής AIC και BIC	37
3.5	Τα βήματα της μεθόδου	38
4	Αποθεματοποίηση με χρήση των γενικευμένων γραμμικών μοντέλων	43
4.1	Υπερσχεδαστικό μοντέλο Poisson	43
4.1.1	Εφαρμογή της διαδικασίας	45
4.1.2	Εφαρμογή του ODP μοντέλου.	51
4.1.3	Γενίκευση του ODP μοντέλου	56
4.2	Υπερσχεδαστική αρνητική διωνυμική κατανομή	58
5	Εφαρμογή σε πραγματικά δεδομένα	60
5.1	Δεδομένα	60
5.2	Εφαρμογή της συνάρτησης βάση των μοντέλων μας	61
5.2.1	Μοντέλο με την χρήση της ODP	61
5.2.2	Μοντέλο με την χρήση της κανονικής κατανομής	65
5.2.3	Μοντέλο με την χρήση της Gamma	67
5.2.4	Μοντέλο με την χρήση της Inverse Gaussian	69
5.3	Σχολιασμός μοντέλων	71

Κεφάλαιο 1

Εισαγωγή

Στην διπλωματική εργασία που ακολουθεί θα δούμε την εφαρμογή των γενικευμένων γραμμικών μοντέλων στην εκτίμηση των αποθεμάτων. Στο Κεφάλαιο 2 παρουσιάζεται ο ορισμός της ασφαλιστικής επιχείρησης καθώς και ο ρόλος των εταιρειών αυτών στην κοινωνία. Στην συνέχεια αναφέρονται οι βασικοί ορισμοί της αποθεματοποίησης και της βασικής της ορολογίας. Στο Κεφάλαιο αυτό παρουσιάζονται οι βασικές ντετερμινιστικές μέθοδοι αποθεματοποίησης. Η πρώτη μέθοδος είναι η μέθοδος αναμενόμενου δείκτη ζημιών. Στην συνέχεια ακολουθεί η Chain Ladder (distribution free model). Στο σημείο αυτό πραγματοποιείται και μια εφαρμογή της μεθόδου αυτής στα δεδομένα με τα οποία ασχολήθηκαν και οι Taylor and Ashe (1983)[13] όπως και ο Verall (1990 [14], 1991 [15]). Τα δεδομένα αυτά θα χρησιμοποιηθούν και στις εφαρμογές στο Κεφάλαιο 5. Ακολουθεί η αναφορά στην μέθοδο που παρουσιάστηκε από τους Bornhuetter and Ferguson το 1972 [1] και θα δούμε και ένα θεωρητικό παράδειγμα της μεθόδου που συνδυάζει την μέθοδο Chain Ladder και του αναμενόμενου δείκτη ζημιών. Τέλος γίνεται μια σύντομη αναφορά στα μέτρα αβεβαιότητας των αποθεμάτων.

Στο Κεφάλαιο 3 γίνεται μια παρουσίαση των βασικών υποθέσεων και της ορολογίας των γενικευμένων γραμμικών μοντέλων. Ξεκινώντας από την δομή του κλασικού μοντέλου περνάμε στο γραμμικό και μετέπειτα στην μορφή ενός γενικευμένου γραμμικού μοντέλου. Αρχικά ορίζεται η δομή του μοντέλου αυτού και παρουσιάζεται η μέθοδος που ακολουθείται προκειμένου να εκτιμήσουμε τις παραμέτρους του. Τέλος γίνεται αναφορά στον τρόπο ελέγχου της επάρκειας ενός γενικευμένου γραμμικού μοντέλου.

Στο Κεφάλαιο 4 πραγματοποιείται η σύνδεση των προηγούμενων δύο κεφαλαίων το οποίο αποτελεί το βασικό θέμα της παρούσας διπλωματικής εργασίας. Θα δούμε μοντέλα με κατανομές που ανήκουν στην εκθετική οικογένεια με απώτερο σκοπό την εκτίμηση των μελλοντικών απαιτήσεων ενός χαρτοφυλακίου ζημιών μιας ασφαλιστικής εταιρείας. Αρχικά ορίζεται το υπερσχεδαστικό μοντέλο Poisson και γίνεται μια πλήρης παρουσίαση της διαδικασίας εκτίμησης του αποθέματος σε ένα απλό τρίγωνο 3 ετών ζημιών σε 3 έτη εξέλιξης. Με σκοπό να γίνει πιο εύκολα κατανοητή η διαδικασία αυτή θα δούμε και μια εφαρμογή με αριθμητικά δεδομένα.

Ακολούθως βρίσκουμε τις βασικές υποθέσεις των μοντέλων με την χρήση της κανονικής, της Gamma και της Inverse Gaussian αντίστοιχα. Τέλος γίνεται αναφορά στο μοντέλο της υπερσχεδαστικής αρνητικής διωνυμικής κατανομής καθώς και της κανονικής προσέγγισής της.

Στο τελευταίο κεφάλαιο ακολουθεί η εφαρμογή των μοντέλων αυτών στα δεδομένα που παρουσιάσαμε και στο κεφάλαιο της αποθεματοποίησης. Η εφαρμογή γίνεται με την βοήθεια της εντολής `glmReserve` του πακέτου `package(ChainLadder)` της R. Τα μοντέλα αυτά γίνονται με την χρήση των κατανομών της Tweedie οικογένειας η οποία ανήκει στην εκθετική οικογένεια κατανομών. Στην οικογένεια αυτή ανήκουν η Poisson, η κανονική, η Gamma και η Inverse Gaussian κατανομές, τις βασικές υποθέσεις των οποίων εντοπίζουμε στο προηγούμενο κεφάλαιο. Ο σχετικός κώδικας παρουσιάζεται στο παράρτημα της παρούσας εργασίας.

Κεφάλαιο 2

Αποθεματοποίηση

2.1 Ορισμός ασφαλιστικής επιχείρησης

Ασφάλιση, με βάση τη νομική έννοια, είναι το σύνολο ομοίων κινδύνων που παρέχει στα μέλη της, με αντάλλαγμα ασφάλιστρο ή εισφορά, αυτόνομη αξίωση για κάλυψη οικονομικής ανάγκης. Ο ορισμός αυτός αφορά όλες τις μορφές ασφάλισης.

Η ασφάλιση υπάρχει σχεδόν από την αρχή της ανθρώπινης κοινωνίας. Στην Ελλάδα ορίστηκε ως έννοια στην ασφαλιστική νομοθεσία από τον ισχύοντα Νόμο 2496/1997. Η βάση της έννοιας της ασφάλισης είναι ο νόμος των μεγάλων αριθμών, έχοντας ως στόχο την ελαχιστοποίηση των υποθετικών παραγόντων και τη μείωση της αστάθειας που δημιουργείτε από τις διακυμάνσεις στις επερχόμενες ζημιές από το ένα έτος στο άλλο. Με άλλα λόγια, βασίζεται στο ότι όσο μεγαλύτερος είναι ο αριθμός των ασφαλιζόμενων κινδύνων τόσο μικρότερη θα είναι η απόκλιση μεταξύ αναμενόμενων και πραγματικών ζημιών. Η ασφάλιση παίρνει τη μορφή σύμβασης μεταξύ αυτού που επιδιώκει την προστασία, ασφαλιζόμενος, και αυτού που την παρέχει, ασφαλιστική εταιρεία.

Οι ασφαλιστικές επιχειρήσεις, βάσει της κείμενης νομοθεσίας λειτουργούν είτε με καθεστώς εγκατάστασης, έχουν δηλαδή την έδρα τους την Ελλάδα ή έχοντας υποκαταστήματα σε αυτή, είτε με καθεστώς Ελεύθερης Παροχής Υπηρεσιών (ΕΠΥ), δηλαδή παρέχουν υπηρεσίες στην Ελλάδα από την έδρα τους ή από υποκατάστημα, εφόσον είναι εγκαταστημένα σε κάποια χώρα-μέλος της Ευρωπαϊκής Ένωσης. Η Ένωση Ασφαλιστικών Εταιριών Ελλάδος (ΕΑΕΕ) είναι το κοινό επαγγελματικό όργανο εκπροσώπησης όλων των ασφαλιστικών εταιρειών που δραστηριοποιούνται στην Ελλάδα.

Μια ασφαλιστική επιχείρηση προκειμένου να μπορεί να ασκήσει εργασίες στην Ελλάδα, θα πρέπει να λάβει πρώτα άδεια λειτουργίας από την εποπτεύουσα αρχή. Στην Ελλάδα την άδεια αυτή παρέχει η Διεύθυνση Εποπτείας Ιδιωτικής Ασφάλισης - Τράπεζα της Ελλάδος (Δ.Ε.Ι.Α.), με κοινοποίηση στην Ευρωπαϊκή Εποπτική αρχή (ΕΙΟΡΑ). Η άδεια λειτουργίας που χορηγείται σε μια ασφαλιστική εταιρεία σε ένα κράτος-μέλος ισχύει σε όλα τα κράτη-μέλη και χορηγείται συνήθως κατά κλάδο

ασφάλισης. Η εποπτική αρχή για να χορηγήσει άδεια λειτουργίας θα πρέπει η έδρα της εταιρείας ή η κεντρική διοίκηση να βρίσκεται στην Ελλάδα, να έχει τη μορφή ανώνυμης εταιρείας ή ευρωπαϊκής εταιρείας, όπως αυτή ορίζεται από τον Κανονισμό Νο 2157/2001 και έχει σαν αποκλειστικό σκοπό δραστηριότητες ασφάλισης και εργασίες που προκύπτουν άμεσα από αυτές, αποκλειόμενης οποιαδήποτε άλλης εμπορικής δραστηριότητας.

2.1.1 Ο ρόλος της ασφαλιστικής επιχείρησης

Σκοπός της ασφάλισης είναι η αποκατάσταση της πραγματικής οικονομικής ζημιάς, ώστε να μην μεταβληθεί η οικονομική κατάσταση του ασφαλισμένου, όσο αυτό είναι εφικτό, κατά την επέλευση του κινδύνου, μέσω της χρηματικής αποζημίωσης.

Η ιδιωτική ασφάλιση επιδρά θετικά και στην οικονομία, συγκεκριμένα

1. Αποκατάσταση ζημιών. Με την αποκατάσταση ζημιών περιορίζεται η σοβαρή διατάραξη των οικονομικών, επιχειρηματικών, επαγγελματικών και κοινωνικών δραστηριοτήτων. Με την μετάθεση των κινδύνων στην ασφάλιση ενδυναμώνεται η ανανεωτική ικανότητα της οικονομίας και ενθαρρύνονται οι επιχειρηματικές πρωτοβουλίες.
2. Ενθάρρυνση της αποταμίευσης. Η ιδιωτική ασφάλιση δημιουργεί αποταμιεύσεις. Ο όγκος των χρημάτων και άλλης περιουσίας που συσσωρεύεται μέσω των ασφαλιστρών δημιουργεί αποθέματα που συγκεντρώνουν οι ασφαλιστικοί φορείς προκειμένου να αντιμετωπίσουν τους κινδύνους που αναλαμβάνουν. Τα αποταμιευμένα κεφάλαια που συσσωρεύονται δημιουργούν μια δεξαμενή πόρων στη διάθεση της οικονομίας ενισχύοντας την εθνική κεφαλαιαγορά. Ο μηχανισμός της ιδιωτικής ασφάλισης με την συγκέντρωση αποθεμάτων και κεφαλαίων, συμβάλλει στη σταθερή και πιο αξιόπιστη πηγή πόρων στη διάθεση της οικονομίας.
3. Χρηματοδότηση επενδύσεων. Με τη δημιουργία αποταμιευτικών κεφαλαίων δημιουργούνται συνθήκες για αξιόπιστη ενίσχυση των παραγωγικών επενδύσεων που οδηγούν σε ανάπτυξη της Οικονομίας και αύξηση του ΑΕΠ. Οι ασφαλιστικοί φορείς αναλαμβάνουν να επενδύσουν τα διαθέσιμα είτε με μορφή δανεισμού σε επιχειρήσεις είτε σε εναλλακτικού τύπου επενδύσεις. Δανεισμός, για τους ασφαλιστικούς φορείς αποτελεί η χρησιμοποίηση των ασφαλιστρών που έχουν κατατεθεί από τους κατόχους των ασφαλιστηρίων συμβολαίων με στόχο να παραχωρούν μακροχρόνια δάνεια και άλλες εναλλακτικές μορφές επένδυσης. Συνεπώς η οικονομία ωφελείται άμεσα καθώς με αυτό τον τρόπο δημιουργούνται πηγές χρηματοδότησης προς νέες επιχειρήσεις, νέα ακίνητα κλπ.
4. Ενθάρρυνση της επιχειρηματικής πρωτοβουλίας. Χωρίς το προστατευτικό δίκτυο των ασφαλίσεων, οι επιχειρηματικές πρωτοβουλίες θα ήταν πολύ λιγότερες, ιδιαίτερα οι πιο καινοτόμες και ριψοκίνδυνες.

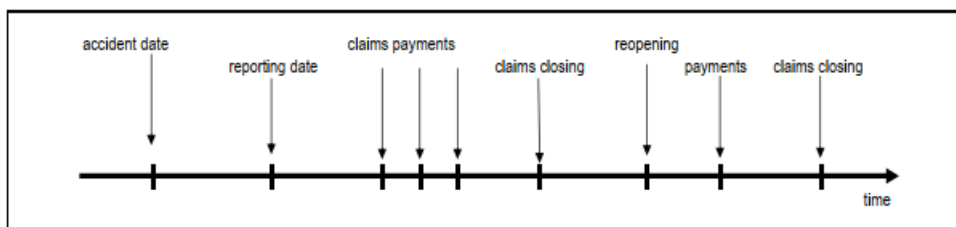
2.2 Τι είναι αποθεματοποίηση

Το χαρτοφυλάκιο που θα μελετήσουμε αφορά απαιτήσεις γενικών ασφαλίσεων, δηλαδή ασφαλίσεις που δεν αφορούν τον κλάδο ζωής, ασφάλιση περιουσίας και ασφάλεια ατυχήματος (Non-life Insurance or Property and Casualty Insurance). Αφορά ουσιαστικά όλα τα ασφαλιστικά προϊόντα εκτός των ασφαλίσεων ζωής. Στις γενικές ασφαλίσεις ανήκουν

1. Ασφαλίσεις οχημάτων (αστική ευθύνης προς τρίτους, ίδιες ζημιές)
2. Ασφαλίσεις περιουσίας (πυρός, πλημμύρας, κλπ.)
3. Ασφαλίσεις ευθύνης προς τρίτους (προσωπική και επαγγελματική ευθύνη)
4. Ασφαλίσεις ατυχημάτων (προσωπική, εργοδοτική ευθύνη, κλπ.)
5. Ασφαλίσεις υγείας (ιδιωτική και ομαδική ασφάλιση)
6. Ασφαλίσεις σκαφών και μεταφορών
7. Λοιπές ασφαλίσεις, όπως προϊόντων, ταξιδιών, ταξιδιωτικών πρακτόρων, νομική ευθύνη, κλπ.

Ο ασφαλισμένος κίνδυνος μετά την πληρωμή ενός ασφαλιστρού μεταφέρεται στην ασφαλιστική εταιρεία η οποία με την σειρά της πρέπει να αποπληρώσει τις ζημιές που συνέβησαν κατά την διάρκεια της περιόδου ασφάλισης.

Η χρονική εξέλιξη μιας τυπικής απαίτησης σε ένα χαρτοφυλάκιο ζημιών περιγράφεται πλήρως στο σχήμα 2.1 [17].



Σχήμα 2.1: Χρονική εξέλιξη απαίτησης

Απο το παραπάνω σχήμα φαίνεται ότι ο η απαίτηση που θα δημιουργηθεί απο την επέλευση του ασφαλιζόμενου κινδύνου συνήθως δεν τακτοποιείται άμεσα. Αυτό οφείλεται σε:

1. Καθηστέρηση στην αναφορά της ζημίας. Για παράδειγμα, στη περίπτωση της αστικής ευθύνης, η χρήση αμύαντου σε κατασκευές ή συμβάν περιβαλλοντικής μόλυνσης όπου τα αποτελέσματα θα φανούν αρκετά χρόνια μετά.

2. Μετά την αναγγελία του συμβάντος, ο χρόνος που απαιτείται για τον διακανονισμό των αποζημιώσεων μπορεί να είναι μεγάλος. Για παράδειγμα σε συμβάντα που αφορούν ευθύνη προς τρίτους ή και σωματικές βλάβες ο χρόνος που απαιτείται για τον διακανονισμό είναι αρκετά μεγάλος.
3. Πολλές φορές φάκελοι απαιτήσεων που έχουν τακτοποιηθεί ανοίγονται ξανά σε περίπτωση που εμφανιστούν νέα στοιχεία.

Ο υπολογισμός των αποθεμάτων, είναι η αναλογιστική διαδικασία εκτίμησης του ποσού που πρέπει να διαθέτει μια ασφαλιστική εταιρεία για την αποπληρωμή ζημιών και για τα έξοδα διακανονισμού ζημιών. Η εκτίμηση των αποθεμάτων για την αποπληρωμή των αποζημιώσεων περιλαμβάνει την εκτίμηση ενδεχόμενων μελλοντικών χρηματοοικονομικών γεγονότων. Τα αποθέματα αυτά μπορεί να είναι αποθέματα που προορίζονται για την αποπληρωμή συγκεκριμένων απαιτήσεων είτε προβλέψεις για μελλοντικές εξελίξεις γνωστών απαιτήσεων είτε προβλέψεις για απαιτήσεις που επαναδραστηριοποιούνται μετά το κλείσιμό τους.

Στο κεφάλαιο αυτό θα γίνει μια σύντομη παρουσίαση των τεχνικών για την εκτίμηση αυτών των αποθεμάτων. Μπορούμε να πούμε πως καμία από αυτές τις τεχνικές δεν είναι απόλυτα σωστή και καταλυτικό παράγοντα στην τελική απόφαση ασχούν η εμπειρία και η κρίση του αναλογιστή.

2.3 Βασική ορολογία

Υπάρχουν δύο κατηγορίες αποθεμάτων ενός χαρτοφυλακίου απαιτήσεων (ζημιών)

1. **Αποθέματα γνωστών απαιτήσεων (ζημιών)**, τα οποία αποτελούνται από το ποσό που χρειάζεται για την κάλυψη ζημιών που είτε έχουν ήδη συμβεί είτε αφορούν μελλοντικές ζημιές που έχουν ήδη γνωστοποιηθεί στην εταιρεία. Πιο συγκεκριμένα είναι αποθέματα που αναφέρονται σε *συγκεκριμένες απαιτήσεις (ζημιές)*, σε αποθέματα για *πρόβλεψη και μελλοντική εξέλιξη γνωστών απαιτήσεων (ζημιών)* και *πρόβλεψη απαιτήσεων (ζημιών)* που έχουν κλείσει και ο φάκελός τους έχει ανοίξει ξανά.
2. **Αποθέματα άγνωστων απαιτήσεων (ζημιών)**, αντιπροσωπεύει το ποσό που απαιτείται για ζημιές που έχουν γίνει και δεν έχουν ακόμη αναγγελθεί ακόμη στην εταιρεία (**IBNR**) και για ζημιές που έχουν αναγγελθεί στην εταιρεία αλλά δεν έχουν υπολογιστεί.

Προκειμένου να γίνει η ανάλυση της διαδικασίας εκτίμησης του αποθέματος θα πρέπει πρώτα να εξηγήσουμε κάποιους βασικούς όρους. Ως ημερομηνία ατύχηματος ορίζεται η ημέρα που πραγματοποιήθηκε το ατύχημα και ως ημερομηνία αναγγελίας η ημερομηνία κατά την οποία δηλώθηκε το ατύχημα. Επίσης, η ημερομηνία κατάρτισης των λογιστικών καταστάσεων μέχρι και την οποία υπολογίζονται οι οφειλόμενες απαιτήσεις για το απόθεμα ζημιών αποτελεί την λογιστική ημερομηνία,

ενώ ημερομηνία αποτίμησης ονομάζεται η ημέρα κατά την οποία έγινε η εκτίμηση του οφειλόμενου αποθέματος (συνήθως στο τέλος μιας οικονομικής περιόδου). Λόγω της μεταβολής του αποθέματος ζημιών εξαιτίας των διαδοχικών ημερομηνιών αποτίμησης θα πρέπει να ξεκαθαρίσουμε τα διάφορα είδη αποθέματος. Το τελικό ποσό που προορίζεται για την διεκπαιρέωση όλων των οφειλούμενων απαιτήσεων καλείται απαιτούμενο απόθεμα και μόνο με την τακτοποίηση όλων των απαιτήσεων γίνεται γνωστό. Το απαιτούμενο απόθεμα είναι σταθερός αριθμός και δεν μεταβάλλεται κατά τις ημερομηνίες αποτίμησης. Ως ενδεικτικό απόθεμα ορίζεται η εκτίμηση του αναλυτή για το σύνολο του απαιτούμενου αποθέματος ζημιών σε μια συγκεκριμένη ημερομηνία αποτίμησης.

Το ποσό της υποχρέωσης για μη πληρωθείσες απαιτήσεις που εμφανίζονται στις εσωτερικές ή εξωτερικές οικονομικές καταστάσεις ονομάζεται μεταφερόμενο απόθεμα. Η διαφορά ανάμεσα στο μεταφερόμενο και το απαιτούμενο απόθεμα ονομάζεται περιθώριο αποθέμα και δεν μπορεί να είναι μηδενικό. Αυτό προκύπτει από το γεγονός ότι το μεταφερόμενο απόθεμα δεν μπορεί να είναι ίσο με το ενδεικτικό απόθεμα και επειδή το μεταφερόμενο απόθεμα αυτό είναι μια άγνωστη ποσότητα μπορούμε να ξέρουμε μόνο το ενδεικτικό περιθώριο. Το απαιτούμενο σύνολο για να καλύψουμε τα μελλοντικά έξοδα που απαιτούνται για τη διερεύνηση και εξόφληση των απαιτήσεων σε δεδομένη χρονική στιγμή, ονομάζονται έξοδα προσαρμογής αποθεμάτων και διακρίνονται σε κατανεμημένα και μη λόγω της πιθανής επιβάρυνσης με αποθέματα συγκεκριμένων ή γενικών απαιτήσεων. Κατανεμημένα έξοδα είναι αυτά που προορίζονται για συγκεκριμένες απαιτήσεις (λογαριασμοί, νόμιμα έξοδα) και μη κατανεμημένα είναι οι μισθοί, το ρεύμα κλπ [19].

Διαδικασία εκτίμησης του αποθέματος

Η στρατηγική προκειμένου να εκτιμήσουμε το απόθεμα που απαιτείται για την ικανοποίηση των απαιτήσεων χωρίζεται σε τέσσερις φάσεις. Αρχικά θα πρέπει να επαληθεύσουμε τα δεδομένα μας και μέσω άλλων έγκυρων πηγών. Επίσης οφείλουμε να εξετάσουμε αν υπάρχουν τυχόν ανωμαλίες στα δεδομένα μας και να αναγνωρίσουμε τα βασικά χαρακτηριστικά τους. Εφόσον πραγματοποιηθεί ο έλεγχος των δεδομένων εφαρμόζουμε την κατάλληλη τεχνική εκτίμησης αποθέματος. Τα αποτελέσματα που δεν συμφωνούν με τη χρήση των διαφόρων μεθόδων εκτίμησης οφείλουμε να τα εξηγήσουμε ή να τα συμβιβάσουμε. Δεδομένου ότι τα παραπάνω βήματα έχουν πραγματοποιηθεί μπορούμε πλέον να παρουσιάσουμε και να προετοιμάσουμε την εξέλιξη του αποθέματος για τα μεταγενέστερα ημερολογιακά έτη.

Η αξιολόγηση της βάσης δεδομένων μας παίζει καθοριστικό ρόλο στην όλη διαδικασία εκτίμησης του αποθέματος. Πρίν εφαρμόσουμε κάποια μέθοδο είναι απαραίτητο να κατανοήσουμε τις τάσεις και τις αλλαγές που επηρεάζουν το χαρτοφυλάκιο το οποίο θα μελετήσουμε. Καθώς τα στοιχεία που μελετάμε αποτελούν απαιτήσεις ασφαλιστικών προϊόντων, πρέπει να γνωρίζουμε τις κύριες

παροχές που προσφέρουν καθώς και τις διάφορες μεταβολές που ενδεχομένως θα υπάρξουν. Αλλαγές στην λειτουργία της επιχείρησης καθώς και αλλαγές στην νομοθεσία ή προηγούμενες δικαστικές αποφάσεις έχουν ιδιαίτερη σημασία.

Το πλήθος των παρατηρήσεών μας θα πρέπει να επιτρέπει τον καθορισμό προτύπων, καθώς για μια καλή αξιοπιστία τα δεδομένα συνήθως εξαρτώνται από το μέσο όρο και από την μεταβολή των μεγεθών των ζημιών. Ιδιαίτερα σημαντικό είναι να υπάρχει η μέγιστη δυνατή ομοιογένεια μεταξύ των δεδομένων και σε περίπτωση που αυτό δεν ικανοποιείται θα πρέπει να προβούμε σε συνδιασμό ή υποδιαίρεση αυτών. Αν τα όρια της πολιτικής ή τα επίπεδα των ζημιών είναι αρκετά όμοια τότε η αξιοπιστία και η ομοιογένεια των δεδομένων είναι υψηλή [19].

2.4 Κλασικές μέθοδοι αποθμεματοποίησης

2.4.1 Εισαγωγή

Στις περισσότερες περιπτώσεις οι οφειλόμενες απαιτήσεις ζημιών παρουσιάζονται με την βοήθεια των τριγώνων εξέλιξης (development triangles). Στα τρίγωνα αυτά οι απαιτήσεις καταχωρούνται σε δύο άξονες όπου στον κατακόρυφο άξονα έχουμε το έτος του ατυχήματος i , δηλαδή το έτος επέλευσης του ασφαλιζόμενου κινδύνου, και στον οριζόντιο άξονα ορίζουμε το έτος εξέλιξης j , δηλαδή το πώς εξελίσσεται η ζημιά με την πάροδο του χρόνου.

Για λόγους παρουσίασης ορίζουμε

- $X_{i,j}$ είναι οι προσαυξητικές απαιτήσεις της ζημιάς που συνέβη το έτος i για το έτος εξέλιξης j .
- $C_{i,j}$ είναι οι αθροιστικές απαιτήσεις για το έτος ατυχήματος i και έτος εξέλιξης j , δηλαδή

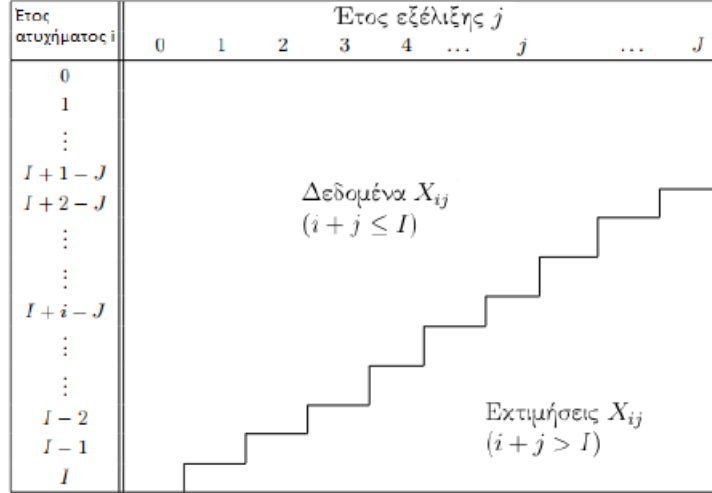
$$C_{i,j} = \sum_{k=0}^j X_{i,k}. \quad (2.1)$$

- τα έτη εξέλιξης είναι τόσα όσα τα έτη ατυχήματος, έστω $I = J$.

Επομένως καταλήγουμε στην κάτωθι μορφή (Σχήμα 2.2)

Μια εναλλακτική παραδοχή είναι

Σχήμα 2.2: Τρίγωνο εξέλιξης



- $X_{i,j}$ ορίζουμε το πλήθος των απαιτήσεων του έτους ατυχήματος i όπου αναφέρθηκαν με μια καθυστέρηση j ετών.
- $C_{i,j} = \sum_{k=0}^j X_{i,k}$ ορίζουμε τον συνολικό αριθμό απαιτήσεων του έτους ατυχήματος i που αναφέρθηκαν μέχρι και $i+j$ χρόνια μετά τη επέλευση του ασφαλισμένου κινδύνου.

Οι παρατηρήσεις $X_{i,j}$, $i+j \leq I$ που βρίσκονται στο πάνω τρίγωνο είναι γνωστές και αποτελούν τα δεδομένα μας. Συνηθίζουμε να τις συμβολίζουμε ως $D_I^U = \{X_{i,j}, i+j \leq I\}$. Τα στοιχεία $X_{i,j}$, $i+j > I$ που βρίσκονται στο κάτω τρίγωνο πρέπει να εκτιμηθούν. Συνήθως συμβολίζονται $D_I^L = \{X_{i,j}, i+j > I\}$. Οι αποζημιώσεις της ασφαλιστικής εταιρείας για το έτος αποτίμησης k είναι το άθροισμα των στοιχείων της $k+1$ διαγωνίου, έστω

$$X_k = \sum_{i+j=k} X_{i,j}.$$

Αυτό που μας ενδιαφέρει είναι να υπολογίσουμε το ποσό της απαίτησης της ζημίας με έτος ατυχήματος i την στιγμή $i+j$, που υπολογίζεται με την βοήθεια του κάτωθι τύπου

$$R_{i,j} = \sum_{k=j+1}^J X_{i,k} = C_{i,J} - C_{i,j}, \quad (2.2)$$

όπου τα $X_{i,j}$ είναι προσαυξητικές απαιτήσεις.

Το $R_{i,j}$ το ονομάζουμε και απόθεμα απαιτήσεων. Ουσιαστικά είναι αυτό που προσπαθούμε να εκτιμήσουμε ούτως ώστε μαζί με τις αποζημιώσεις των

προηγούμενων ετών $C_{i,j}$, θα μας βοηθήσουν να υπολογίσουμε το συνολικό απόθεμα που θα πρέπει να κατέχουμε για την ζημία του έτους ατυχήματος i . Το συνολικό απόθεμα του χαρτοφυλακίου υπολογίζεται ανθροίζοντας τα αποθέματα κάθε έτους ατυχήματος [17].

Στη συνέχεια θα γίνει μια σύντομη παρουσίαση μερικών από τις βασικές μεθόδους αποθεματοποίησης ζημιών.

2.4.2 Μέθοδος αναμενόμενου δείκτη ζημιών

Η μέθοδος του αναμενόμενου δείκτη ζημιών μπορεί να εφαρμοστεί όταν δεν υπάρχει επαρκής εμπειρία ζημιών από τον κλάδο. Η μέθοδος αυτή, αν και είναι σχετικά εύκολη στον υπολογισμό των αποθεμάτων, έχει ένα βασικό μειονέκτημα. Το μειονέκτημά της είναι ότι δεν μπορεί να χρησιμοποιηθεί για ολόκληρο το χαρτοφυλάκιο της εταιρείας επειδή ο αναμενόμενος δείκτης ζημιών μπορεί οποιαδήποτε στιγμή να τροποποιηθεί από την διαχείριση και να οδηγήσει σε ανεπαρκή αποθέματα.

Το απόθεμα των ζημιών που αναζητούμε να βρούμε με αυτή την μέθοδο θα προκύψει αν από τις εκτιμώμενες τελικές ζημίες αφαιρέσουμε τις ζημίες που έχουν εξοφληθεί μέχρι σήμερα.

Οι εκτιμώμενες τελικές ζημίες υπολογίζονται πολλαπλασιάζοντας τον αναμενόμενο δείκτη ζημιών με το κατάλληλο δεδουλευμένο ασφάλιστρο.

Το δεδουλευμένο ασφάλιστρο υπολογίζεται αν από το άθροισμα του εγγεγραμμένου ασφαλιστρού με το αρχικό απόθεμα μη δεδουλευμένου ασφαλιστρού αφαιρέσουμε το τελικό απόθεμα δεδουλευμένου ασφαλιστρού [19]. Επομένως έχουμε:

$$(\text{Εκτιμώμενες τελικές ζημίες})_{i,j} = (\text{Αναμενόμενος δείκτης ζημιών})_{i,j} \times (\text{Δεδουλευμένο Ασφάλιστρο})_{i,j},$$

όπου το i δηλώνει τον κλάδο ασφάλισης και j δηλώνει το την περίοδο ασφάλισης. Αυτό συνεπάγεται ότι

$$(\text{Εκτιμώμενα αποθέματα ζημιών})_{i,j} = (\text{Εκτιμώμενες τελικές ζημίες})_{i,j} - (\text{Ζημίες που έχουν πληρωθεί})_{i,j},$$

από το οποίο προκύπτει ότι

$$\text{Συνολικά Εκτιμώμενα Αποθέματα Ζημιών} = \sum_{i,j} (\text{Εκτιμώμενα Αποθέματα Ζημιών})_{i,j}.$$

2.4.3 Chain Ladder (distribution free model)

Για την αποπληρωμή μελλοντικών καθώς και εκκρεμών ζημιών απαιτείται η δημιουργία αποθεμάτων. Η πιό δημοφιλής μέθοδος αποθεματοποίησης ζημιών είναι

η μέθοδος Chain Ladder. Είναι μία απο τις παλαιότερες και πιο διαδεδομένες αναλογιστικές τεχνικές που εφαρμόζεται ευρέως για την εκτίμηση των αποθεμάτων δαπανών διακανονισμών ζημιών. Θεωρείτε ένας αλγόριθμος ικανός για λύση χωρίς κρίση. Αρχικά η μέθοδος βασιζόταν σε έναν καθαρά υπολογιστικό αλγόριθμο και πολύ αργότερα αναπτύχθηκαν στοχαστικά μοντέλα αλγορίθμων για την μέθοδο αυτή. Ο πρώτος που όρισε ένα πλήρως στοχαστικό μοντέλο ήταν το Mack όπου το 1993 δημοσίευσε το πιο δημοφιλές άρθρο για τον υπολογισμό του σφάλματος στο μοντέλο Chain Ladder [8].

Βασικές υποθέσεις του μοντέλου

Στους πίνακες εξέλιξης ζημιών εμφανίζονται οι ατομικές ζημίες ή οι αθροιστικές ζημίες. Όπως παραπάνω, ορίζουμε

- $X_{i,j}$ τις ατομικές ζημίες για το έτος ατυχήματος i και το έτος εξέλιξης j .
- $C_{i,j}$ τις αθροιστικές ζημίες για το έτος ατυχήματος i και το έτος εξέλιξης j .
- τα έτη εξέλιξης είναι τόσα όσα τα έτη ατυχήματος, έστω $I = J$.

Σε αυτή την μέθοδο μας θα δουλέψουμε με τις αθροιστικές ζημίες $C_{i,j}$, $0 \leq i \leq I$, $0 \leq j \leq J$, όπου θεωρούνται τυχαίες μεταβλητές για $i + j \leq I$.

Το ντετερμινιστικό αυτό μοντέλο έχει δύο βασικές υποθέσεις:

Υπόθεση 1

Επειδή ο αλγόριθμος της Chain Ladder υποθέτει ότι υπάρχει ανεξαρτησία μεταξύ των ετών ατυχήματος οδηγούμαστε στην υπόθεση ότι οι αθροιστικές ζημίες $C_{i,j}$ διαφορετικών ετών ατυχήματος είναι ανεξάρτητες,

$$\{C_{i,0}, \dots, C_{i,I}\}, \{C_{k,0}, \dots, C_{k,I}\}, \forall i \neq k \text{ είναι ανεξάρτητες.} \quad (2.3)$$

Υπόθεση 2

Ορίζουμε τους συντελεστές εξέλιξης $f_0, \dots, f_{J-1} > 0$ όπου

$$E[C_{i,j}|C_{i,0}, \dots, C_{i,j-1}] = E[C_{i,j}|C_{i,j-1}] = f_{j-1}C_{i,j-1}, \forall 0 \leq i \leq I, 1 \leq j \leq J. \quad (2.4)$$

Οι συντελεστές εξέλιξης ή συντελεστές Chain Ladder, f_k είναι τα βασικά στοιχεία της μεθόδου αυτής. Η εκτίμηση αυτών των συντελεστών γίνεται με

$$\hat{f}_j = \frac{\sum_{k=0}^{I-j} C_{k,j}}{\sum_{k=0}^{I-j} C_{k,j-1}}, j = 1, \dots, J. \quad (2.5)$$

Η εκτίμηση των αθροιστικών ζημιών $\hat{C}_{i,j}$, $i + j > I$ για κάθε έτος ατυχήματος i και έτος εξέλιξης j δίνεται απο τον κάτωθι τύπο

$$\hat{C}_{i,j} = C_{i,I-i} \prod_{k=I-i+1}^{j-1} \hat{f}_k. \quad (2.6)$$

Μπορούμε να πούμε ότι η εκτίμηση των συνολικών ανθροιστικών ζημιών $\hat{C}_{i,I}^{CL}$ για κάθε έτος ζημίας υπολογίζεται βάση του κάτωθι τύπου

$$C_{i,I}^{CL} = C_{i,I-i} \hat{f}_{I-i} \dots \hat{f}_{I-1}, \quad (2.7)$$

και η εκτίμηση για το απόθεμα για κάθε έτος ατυχήματος γίνεται με

$$\hat{R}_i^{CL} = C_{i,I-i} \left(\hat{f}_{I-i} \dots \hat{f}_{I-1} - 1 \right). \quad (2.8)$$

Τέλος, το συνολικό απόθεμα είναι

$$\hat{R} = \sum_{i=1}^I \hat{R}_i^{CL}. \quad (2.9)$$

Προκειμένου να αποκατασταθεί ή να διατηρηθεί η σταθερότητα της μεθόδου Chain Ladder επεμβαίνουμε στο μοντέλο εξέλιξης των ζημιών και παραμερίζουμε τους υψηλότερους ή χαμηλότερους συντελεστές εξέλιξης ζημιών. Ένας άλλος τρόπος είναι να θέσουμε και βάρη στους συντελεστές των τελευταίων δεδομένων ή να συμπληρώσουμε το μισό του ορθογωνίου ως γραμμική συνάρτηση του άλλου μισού ή να κάνουμε χρήση του αρμονικού ή γεωμετρικού μέσου [8].

Εφαρμογή

Με την βοήθεια όλων όσων ορίσαμε στην ενότητα αυτή, είμαστε σε θέση να δούμε ένα παράδειγμα εφαρμογής της μεθόδου αυτής σε ένα χαρτοφυλάκιο ζημιών. Τα δεδομένα που θα χρησιμοποιήσουμε είναι αυτά που χρησιμοποίησαν οι Taylor and Ashe το 1983[13] και ο Verall το 1990 [14] και το 1991 [15]. Με την βοήθεια του πακέτου R θα γίνουν οι υπολογισμοί των πράξεων (αναλυτικά ο κώδικας βρίσκεται στο παράρτημα).

Στον Πίνακα (2.1) παρουσιάζονται οι προσαυξητικές απαιτήσεις X_{ij} σε ένα τρίγωνο εξέλιξης ζημιών 10 ετών ατυχημάτων και 10 ετών εξέλιξης. Οι τιμές που βρίσκονται στο πάνω αριστερά τρίγωνο είναι τα δεδομένα μας, είναι οι ζημίες που έχει πληρώσει η εταιρεία μέχρι σήμερα κάθε έτος. Οι τιμές της διαγωνίου εκφράζουν τις απαιτήσεις του τρέχοντος έτους. Στόχος των μεθόδων αποθεματοποίησης είναι να εκτιμηθούν οι τιμές των στοιχείων του κάτω δεξιά τριγώνου.

Πίνακας 2.1: Προσαυξητικές απαιτήσεις X_{ij}

	0	1	2	3	4	5	6	7	8	9
0	357848	766940	610542	482940	527326	574398	146342	139950	227229	67948
1	352118	884021	933894	1183289	445745	320996	527804	266172	425046	
2	290507	1001799	926219	1016654	750816	146923	495992	280405		
3	310608	1108250	776189	1562400	272482	352053	206286			
4	443160	693190	991983	769488	504851	470639				
5	396132	937085	847498	805037	705960					
6	440832	847631	1131398	1063269						
7	359480	1061648	1443370							
8	376686	986608								
9	344014									

Με την χρήση του τύπου (2.1) υπολογίζουμε τις αθροιστικές απαιτήσεις $C_{i,j}$ όπως παρουσιάζονται κάτωθι στον Πίνακα 2.2

Πίνακας 2.2: Αθροιστικές απαιτήσεις $C_{i,j}$

	0	1	2	3	4	5	6	7	8	9
0	357848	1124788	1735330	2217820	2745596	3319994	3466336	3606286	3833515	3901463
1	352118	1236139	2170033	3353322	3799067	4120063	4647867	4914039	5339085	
2	290507	1292306	2218525	3235179	3985995	4132918	4628910	4909315		
3	310608	1418858	2195047	3757447	4029929	4381982	4588268			
4	443160	1136350	2128333	2897821	3402672	3873311				
5	396132	1333217	2180715	2985752	3691712					
6	440832	1288463	2419861	3483130						
7	359480	1421128	2864498							
8	376686	1363294								
9	344014									

Το άθροισμα των τιμών του κάτω δεξιά τριγώνου ισούται με το απόθεμα για τις ζημιές που δεν έχουν αναγγελθεί ακόμα στην εταιρεία (IBNR). Για να εκτιμήσουμε τις απαιτήσεις που δεν έχουν αναγγελθεί ακόμα στην εταιρεία θα εφαρμόσουμε την μέθοδο Chain Ladder.

Αρχικά θα υπολογίσουμε τις εκτιμήσεις των συντελεστών εξέλιξης ζημιών \hat{f}_j για $0 \leq j \leq J - 1$ με την βοήθεια του τύπου (2.5).

Για παράδειγμα,

$$\begin{aligned}
 \hat{f}_6 &= \frac{\sum_{k=0}^{9-6} C_{k,6}}{\sum_{k=0}^{9-6} C_{k,6-1}} \\
 &= \frac{C_{0,6} + C_{1,6} + C_{2,6} + C_{3,6}}{C_{0,5} + C_{1,5} + C_{2,5} + C_{3,5}} \\
 &= \frac{3466336 + 4647867 + 4628910 + 4588268}{3319994 + 4120063 + 4132918 + 4381982} = 1.086269.
 \end{aligned}$$

Χρησιμοποιώντας το πακέτο R υπολογίζουμε κατά τον ίδιο τρόπο τις εκτιμήσεις των συντελεστών εξέλιξης κάθε έτους που παρουσιάζονται στον Πίνακα 2.3.

Πίνακας 2.3: Εκτιμήσεις συντελεστών εξέλιξης ζημιών \hat{f}_j

j	1	2	3	4	5	6	7	8	9
\hat{f}_j	3.490607	1.747333	1.457413	1.173852	1.103824	1.086269	1.053874	1.076555	1.017725

Για να υπολογίσουμε τις τιμές του κάτω δεξιά τριγώνου, δηλαδή τις εκτιμήσεις των απαιτήσεων για κάθε έτος εξέλιξης, \hat{C}_{ij} θα χρησιμοποιήσουμε τον τύπο (2.6).

Παραδείγματος χάριν

$$\begin{aligned}\hat{C}_{7,4} &= C_{7,9-7} \prod_{j=9-7+1}^4 \hat{f}_j = C_{7,2} \prod_{j=3}^4 \hat{f}_j \\ &= (2864498) \times (1.457413) \times (1.173852) = 4900545.\end{aligned}$$

Με την βοήθεια της R , για τον υπολογισμό των πράξεων, παρουσιάζεται ο κάτωθι συμπληρωμένος Πίνακας 2.4.

Πίνακας 2.4: Συμπληρωμένο τρίγωνο εξέλιξης αθροιστικών ζημιών

	0	1	2	3	4	5	6	7	8	9
0	357848	1124788	1735330	2218270	2745596	3319994	3466336	3606286	3833515	3901463
1	352118	1236139	2170033	3353322	3799067	4120063	4647867	4914039	5339085	5433719
2	290507	1292306	2218525	3235179	3985995	4132918	4628910	4909315	5285148	5378826
3	310608	1418858	2195047	3757447	4029929	4381982	4588268	4835458	5205637	5297906
4	443160	1136350	2128333	2897821	3402672	3873311	4207459	4434133	4773589	5297906
5	396132	1333217	2180715	2985752	3691712	4074999	4426546	4665023	5022155	4858200
6	440832	1288463	2419861	3483130	4088678	4513179	4902528	5166649	5562182	5111171
7	359480	1421128	2864498	4174756	4900545	5409337	5875997	6192562	6666635	5660771
8	376686	1363294	2382128	3471744	4075313	4498426	4886502	5149760	5544000	6784799
9	344014	1200818	2098228	3057984	3589620	3962307	4304132	4536015	4883270	4969825

Προκειμένου να εκτιμήσουμε το απόθεμα που απαιτείται για κάθε έτος, που το συμβολίζουμε ως \hat{R}_i , αρκεί να υπολογίσουμε την διαφορά της εκτιμημένης αθροιστικής απαίτησης για την ζημιά του έτους i το τελευταίο έτος εξέλιξης $j = I$ \hat{C}_{iI} με την τελευταία απαίτηση που πλήρωσε η εταιρεία για το έτος i , δηλαδή την $C_{i,j}$, όπου $i + j = I$.

Ενδεικτικά έχουμε

$$\hat{R}_3^{CL} = \hat{C}_{3,9} - C_{3,6} = 5297906 - 4588268 = 709638.$$

Το ίδιο αποτέλεσμα θα έχουμε και αν εφαρμόσουμε τον τύπο (2.8)

$$\begin{aligned}\hat{R}_3^{CL} &= C_{3,9-3} \times \left[\left(\hat{f}_{9-3} \times \hat{f}_{9-2} \times \hat{f}_{9-1} \right) - 1 \right] = C_{3,6} \times \left[\left(\hat{f}_6 \times \hat{f}_7 \times \hat{f}_8 \right) - 1 \right] \\ &= 4588268 \times \left[\left((1.053874) \times (1.076555) \times (1.017725) \right) - 1 \right] = 709638.\end{aligned}$$

Κατά τον ίδιο τρόπο υπολογίζουμε τις εκτιμήσεις των αποθεμάτων κάθε έτους. Τέλος για τον υπολογισμό του συνολικού αποθέματος, σύμφωνα με τον τύπο (2.9), αρκεί να προσθέσουμε τις εκτιμήσεις των αποθεμάτων κάθε έτους ζημιάς. Τα αποτελέσματα παρουσιάζονται στον Πίνακα 2.5.

Πίνακας 2.5: Εκτειμήσεις των αποθεμάτων κάθε έτους \hat{R}_i^{CL}

i	1	2	3	4	5	6	7	8	9	$Total\hat{R}$
\hat{R}_i^{CL}	94634	469511	709638	984889	1419459	2177641	3920301	4278972	4625811	18680856

2.4.4 Bornhuetter-Ferguson (B-F)

Μια άλλη μέθοδος για την αποθεματοποίηση των ζημιών (απαιτήσεων) παρουσιάστηκε από τους Bornhuetter and Ferguson το 1972 [1]. Για την μέθοδο αυτή ορίζουμε το κάτωθι μοντέλο υποθέσεων

Υπόθεση 1

1. Όπως και στη μέθοδο *Chain Ladder*, υποθέτουμε ότι τα διαφορετικά έτη ζημίας είναι ανεξάρτητα. Κατ'επαίχταση οι αθροιστικές ζημιές $C_{i,j}$ διαφορετικών ετών ατυχήματος είναι ανεξάρτητες.
2. Ορίζουμε τις παραμέτρους $\mu_0, \dots, \mu_I > 0$ και $\beta_0, \dots, \beta_I > 0$ με $\beta_I = 1$, έτσι ώστε $\forall i \in \{0, 1, \dots, I\}$, $\forall j \in \{0, 1, \dots, I-1\}$ και $\forall k \in \{0, 1, \dots, I-j\}$ να ισχύει

$$E[C_{i,0}] = \mu_i \beta_0, \quad (2.10)$$

$$E[C_{i,j+k} | C_{i,0}, \dots, C_{i,j}] = C_{i,j} + \mu_i (\beta_{j+k} - \beta_j). \quad (2.11)$$

Ουσιαστικά, λαμβάνοντας τις παραπάνω υποθέσεις, οδηγούμαστε στο συμπέρασμα ότι $E[C_{i,j}] = \mu_i \beta_j$ και $E[C_{i,j}] = \mu_i$. Η ακολουθία $(\beta_j)_j$ ορίζει το μοτίβο εξέλιξης των ζημιών. Αν τα $C_{i,j}$ είναι αθροιστικές απαιτήσεις, τότε τα β_j είναι το συσσωρευμένο πρότυπο ταμειακών ροών.

Το μοντέλο υποθέσεων που ορίσαμε μας οδηγεί σε ένα ασθενέςθερο μοντέλο υποθέσεων όπου μαζί με την υπόθεση της ανεξαρτησίας των αθροιστικών απαιτήσεων $C_{i,j}$ διαφορετικών ετών i είναι:

Υπόθεση 2

1. Τα έτη ατυχήματος i είναι ανεξάρτητα
2. Υπάρχουν οι παράμετροι $\mu_0, \dots, \mu_I > 0$ και τα $\beta_0, \dots, \beta_I > 0$, όπου $\beta_I = 1$ έτσι ώστε

$$E[C_{i,j}] = \mu_i \beta_j, \quad \forall i \in \{0, 1, \dots, I\} \text{ και } \forall j \in \{0, 1, \dots, I-1\}. \quad (2.12)$$

Συνήθως χρησιμοποιούμε το δεύτερο μοντέλο υποθέσεων, ωστόσο αντιμετωπίζουμε κάποιες δυσκολίες.

$$\begin{aligned} E[C_{i,j} | D_I] &= E[C_{i,j} | C_{i,0}, \dots, C_{i,I-i}] \\ &= C_{i,I-i} + E[C_{i,j} - C_{i,I-i} | C_{i,0}, \dots, C_{i,I-i}]. \end{aligned} \quad (2.13)$$

Αν δεν έχουμε κάποιες επιπλέον υποθέσεις, δεν γνωρίζουμε ακριβώς τί πρέπει να κάνουμε με τον τελευταίο όρο. Για παράδειγμα αν γνωρίζαμε ότι οι προσαυξητικές απαιτήσεις $C_{i,J} - C_{i,I-i}$ είναι ανεξάρτητες από τις $C_{i,0}, \dots, C_{i,I-i}$ τότε

$$\begin{aligned} E[C_{i,J}|D_I^U] &= E[C_{i,j}|C_{i,0}, \dots, C_{i,I-i}] \\ &= C_{i,I-i} + (1 - \beta_{I-i})\mu_i. \end{aligned} \quad (2.14)$$

Ουσιαστικά θα καταλήγαμε στο πρώτο μοντέλο υποθέσεων. Αυτό που επιθυμούμε, ανεξάρτητα από ποιο μοντέλο υποθέσεων θα χρησιμοποιήσουμε, είναι να εκτιμήσουμε την ποσότητα $E[C_{i,J}|D_I^U]$ [17]

Η εκτίμηση $B - F$ είναι

$$\hat{C}_{i,J}^{BF} = \hat{E}[C_{i,J}|D_I^U] = C_{i,I-i} + (1 - \hat{\beta}_{I-i})\hat{\mu}_i, \quad (2.15)$$

για $1 \leq i \leq I$, όπου $\hat{\beta}_{I-i}$ είναι ο εκτιμητής του β_{I-i} και $\hat{\mu}_i$ είναι μια αρχική εκτίμηση για το $E[C_{i,J}]$.

Από την υπόθεση της Chain Ladder έχουμε

$$E[C_{i,j}] = E[E[C_{i,j}|C_{i,j-1}]] = f_{j-1}E[C_{i,j-1}] = E[C_{i,0}] \prod_{k=0}^{j-1} f_k, \quad (2.16)$$

$$E[C_{i,J}] = E[C_{i,0}] \prod_{k=0}^{J-1} f_k, \quad (2.17)$$

και λύνοντας το σύστημα

$$E[C_{i,j}] = \prod_{k=j}^{J-1} f_k^{-1} E[C_{i,J}]. \quad (2.18)$$

Σε αυτή την περίπτωση μπορούμε να πούμε ότι ο όρος $\prod_{k=j}^{J-1} f_k^{-1}$ παίζει τον ρόλο του

β_j . Επομένως, λαμβάνοντας υπόψιν την παραπάνω παρατήρηση, η σχέση (2.15), με την βοήθεια της οποίας υπολογίζουμε την εκτίμηση των συνολικών αθροιστικών ζημιών, γίνεται

$$\hat{C}_{i,J}^{BF} = C_{i,I-i} + \left[1 - \left(\prod_{j=I-i}^{J-1} \hat{f}_j \right)^{-1} \right] \hat{\mu}_i. \quad (2.19)$$

Η εκτίμηση της μεθόδου Chain Ladder (2.7), γίνεται

$$\hat{C}_{i,J}^{CL} = C_{i,I-i} \prod_{j=I-i}^{J-1} \hat{f}_j \quad (2.20)$$

$$= C_{i,I-i} + C_{i,I-i} \left(\prod_{j=I-i}^{J-1} \hat{f}_j - 1 \right) \quad (2.21)$$

$$= C_{i,I-i} + \frac{\hat{C}_{i,J}^{CL}}{\prod_{j=I-i}^{J-1} \hat{f}_j} \left(\prod_{j=I-i}^{J-1} \hat{f}_j - 1 \right) \quad (2.22)$$

$$= C_{i,I-i} + \left[1 - \left(\prod_{j=I-i}^{J-1} \hat{f}_j \right)^{-1} \right] \hat{C}_{i,J}^{CL}. \quad (2.23)$$

Όπως παρατηρούμε η διαφορά της Chain Ladder και της Bornhuetter & Ferguson έγκυται στον τρόπο που θα ορίσουμε την εκτίμηση της παραμέτρου $\hat{\mu}_i$. Στην μέθοδο Chain Ladder εκτιμάται μέσα απο τα δεδομένα μας, ενώ στην Bornhuetter & Ferguson την ορίζουμε εμείς βασιζόμενοι σε εξωτερικούς παράγοντες [17].

Παράδειγμα

Για παράδειγμα ας δούμε την μέθοδο $B - F$ που συνδυάζει τη μέθοδο του αναμενόμενου δείκτη ζημιών και τη μέθοδο Chain Ladder [19].

Από την μέθοδο του αναμενόμενου δείκτη ζημιών, δεδομένου ότι για κάθε έτος ατυχήματος έχουμε υπολογίσει τον Αναμενόμενο Δείκτη Ζημιών (ΑΔΖ) λαμβάνοντας υπόψιν όλες τις πληροφορίες που αφορούν τα δεδομένα μας, για να υπολογίσουμε τις Εκτιμώμενες Τελικές Ζημιές (ETZ) έχουμε

$$ETZ^{LR} = A\Delta Z \times (\Delta\epsilon\delta\upsilon\lambda\epsilon\upsilon\mu\epsilon\nu\alpha \text{ Α}\sigma\phi\acute{\alpha}\lambda\iota\sigma\tau\upsilon\alpha)$$

και τα Εκτιμώμενα Αποθέματα Ζημιών (EAZ)

$$EAZ^{LR} = ETZ^{LR} - (\text{Ζημιές που έχουν πληρωθεί μέχρι σήμερα}). \quad (2.24)$$

Επίσης, βάση της Chain Ladder έχουμε

$$ETZ^{CL} = (\text{Ζημιές που έχουν πληρωθεί μέχρι σήμερα}) \times \prod_j f_j, \quad (2.25)$$

όπου f_j είναι ο συντελεστής ανάπτυξης ζημιών του τριγώνου της μεθόδου Chain Ladder και

$$EAZ^{CL} = ETZ^{CL} - (\text{Ζημιές που έχουν πληρωθεί μέχρι σήμερα}) \quad (2.26)$$

$$= (\text{Ζημιές που έχουν πληρωθεί μέχρι σήμερα}) \times \left(\prod_j f_j - 1 \right). \quad (2.27)$$

Σύμφωνα με την μέθοδο B-F οι εκτημώμενες τελικές ζημιές είναι

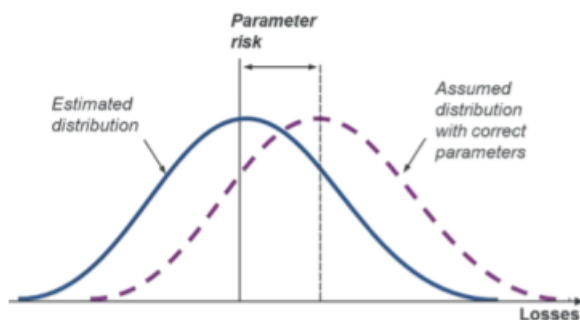
$$ETZ = \prod_j \frac{1}{f_j} EAZ^{CL} + \left(1 - \prod_j \frac{1}{f_j}\right) ETZ^{LR}. \quad (2.28)$$

2.4.5 Μέτρα αβεβαιότητας των αποθεμάτων

Πρίν προχωρήσουμε στην παρουσίαση των γενικευμένων γραμμικών μοντέλων και τον τρόπο εφαρμογής αυτών στην διαδικασία υπολογισμού αποθεμάτων, είναι σημαντικό να γίνει μια σύντομη αναφορά στα είδη σφαλμάτων που περιέχει η κάθε διαδικασία [2].

Μπορούμε να πούμε ότι τα είδη των σφαλμάτων είναι:

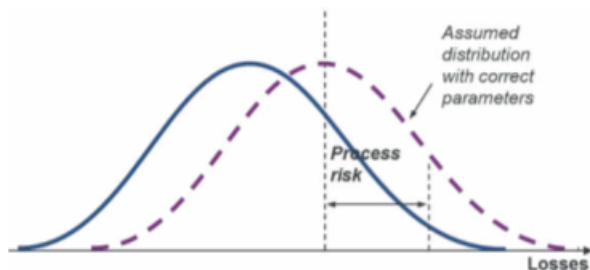
1. **Σφάλμα παραμέτρου (Parameter error):** Έχοντας υποθέσει ότι το μοντέλο έχει προσδιοριστεί σωστά, το συγκεκριμένο σφάλμα, αντιπροσωπεύει την αβεβαιότητα στην εκτίμηση των παραμέτρων.



Σχήμα 2.3: Parameter error

Όπως παρατηρούμε και από το Σχήμα 2.2, το σφάλμα της παραμέτρου είναι η διαφορά της μέσης τιμής της εκτιμημένης κατανομής με την μέση τιμή της θεωρητικής κατανομής με τις σωστές παραμέτρους.

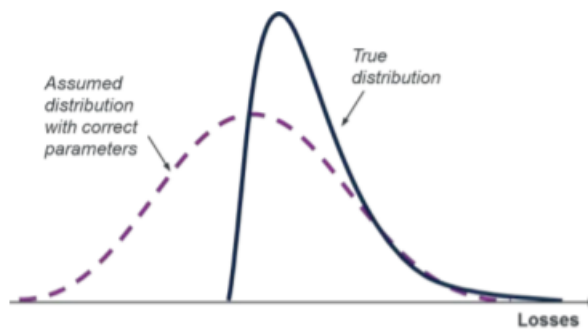
2. **Σφάλμα διαδικασίας (Process error):** Αντιπροσωπεύει την τυχαιότητα των μελλοντικών αποτελεσμάτων. Ακόμη και αν το μοντέλο έχει οριστεί σωστά και οι παράμετροι έχουν εκτιμηθεί σωστά, η μεταβλητότητα των διαδικασιών της ασφαλιστικής εταιρείας μπορεί να οδηγήσει σε διαφορετικά αποτελέσματα σε σχέση με αυτά που αναμέναμε.



Σχήμα 2.4: Process error

Είναι ουσιαστικά το σφάλμα που οφείλεται στην διαφορά της πραγματικής μελλοντικής τιμής από την μέση τιμή της θεωρητικής κατανομής με τις πραγματικές παραμέτρους.

3. **Σφάλμα μοντέλου (Model error):** Υπολογίζει την διαφορά μεταξύ της παραμετρικής μορφής του μοντέλου που έχουμε υποθέσει με την πραγματική αλλά άγνωστη παραμετρική του μορφή. Είναι ουσιαστικά το σφάλμα που παρουσιάζεται εξαιτίας της επιλογής του μοντέλου.



Σχήμα 2.5: Model error

Το σύνηθες εργαλείο για να υπολογίσουμε το σφάλμα εκτίμησης είναι το μέσο τετραγωνικό σφάλμα πρόβλεψης (MSEP). Στην περίπτωση της εκτίμησης αποθεμάτων ζημιών, το μέσο τετραγωνικό σφάλμα είναι η αναμενόμενη τιμή του συνολικού σφάλματος του αποθέματος,

$$MSEP(\hat{R}) = E \left[(R - \hat{R})^2 \right]. \quad (2.29)$$

Σε περιπτώσεις όπου το σφάλμα της παραμέτρου και το σφάλμα της διαδικασίας είναι στοχαστικά ανεξάρτητα τότε το MSEP μπορεί να υπολογιστεί αθροίζοντας τα τετράγωνα και των δύο. Εκτενέστερη αναφορά θα γίνει στο κεφάλαιο της αποθεματοποίησης με την χρήση των γενικευμένων γραμμικών μοντέλων.

Προκειμένου να εκφράσουμε την αβεβαιότητα των αποθεμάτων είναι σύνηθες να χρησιμοποιείται ένας συντελεστής διασποράς (CoV, coefficient of variation). Ο συντελεστής αυτός ορίζεται ως το κλάσμα όπου αριθμητής είναι η τετραγωνική ρίζα του μέσου τετραγωνικού σφάλματος (RMSEP) και παρονομαστής είναι η εκτίμηση του συνολικού αποθέματος [2].

$$\frac{\sqrt{MSEP}}{\hat{R}}. \quad (2.30)$$

Κεφάλαιο 3

Γενικευμένα γραμμικά μοντέλα

3.1 Οι βασικές υποθέσεις και η ορολογία

3.1.1 Το κλασικό μοντέλο παλινδρόμησης

Το απλό γραμμικό μοντέλο έχει την μορφή

$$Y = b_0 + \beta_1 X + \beta_2 X_2 + \dots + \beta_n X_n + \varepsilon, \quad (3.1)$$

όπου

1. Y είναι η εξαρτημένη μεταβλητή,
2. x_1, x_2, \dots, x_n είναι οι n ανεξάρτητες μεταβλητές,
3. $\beta_0, \beta_1, \dots, \beta_n$ είναι οι συντελεστές παλινδρόμησης που επιθυμούμε να εκτιμήσουμε,
4. ε είναι το σφάλμα.

3.1.2 Το γραμμικό μοντέλο

Πολλές φορές για να βρεθεί το μοντέλο εκτίμησης χρησιμοποιούνται δεδομένα τα οποία έχουν προκύψει από μια σειρά N παρατηρήσεων και συχνά δίνονται με τη μορφή του παρακάτω πίνακα:

$$\begin{bmatrix} x_{11} & x_{12} & \dots & x_{1n} \\ x_{21} & x_{22} & \dots & x_{2n} \\ \dots & \dots & \dots & \dots \\ x_{N1} & x_{N2} & \dots & x_{Nn} \end{bmatrix}. \quad (3.2)$$

Το γραμμικό μοντέλο με την μορφή πινάκων έχει την μορφή

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_N \end{bmatrix} = \begin{bmatrix} x_{11} & x_{12} & \dots & x_{1n} \\ x_{21} & x_{22} & \dots & x_{2n} \\ \dots & \dots & \dots & \dots \\ x_{N1} & x_{N2} & \dots & x_{Nn} \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_n \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_N \end{bmatrix}, \quad (3.3)$$

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon},$$

όπου

1. \mathbf{Y} είναι ένας $N \times 1$ πίνακας εξαρτημένων τυχαίων μεταβλητών (τ.μ.),
2. \mathbf{X} είναι ένας $N \times n$ πίνακας ανεξάρτητων μεταβλητών,
3. $\boldsymbol{\beta}$ είναι ένας $n \times 1$ πίνακας παραμέτρων,
4. $\boldsymbol{\varepsilon}$ είναι ένας $N \times 1$ πίνακας όπου τα $\varepsilon = (\varepsilon_1, \varepsilon_2, \dots, \varepsilon_N)$ ακολουθούν την Κανονική κατανομή με $E(\boldsymbol{\varepsilon}) = 0$ και $Var(\boldsymbol{\varepsilon}) = \sigma^2 \mathbf{I}_N$ όπου \mathbf{I}_N είναι ο μοναδιαίος πίνακας. Επιπλέον τα $\varepsilon = (\varepsilon_1, \varepsilon_2, \dots, \varepsilon_N)$ είναι ασυσχέτιστα, πράγμα το οποίο σημαίνει $Cov(\varepsilon_i, \varepsilon_j) = 0, \forall i \neq j$.

3.1.3 Γενικευμένα γραμμικά μοντέλα

Μορφή

Οι Nelder and Wedderburn το 1972 [10] χρησιμοποίησαν την ιδέα των γενικευμένων γραμμικών μοντέλων. Τα γενικευμένα γραμμικά μοντέλα εξόρισμού βρίσκονται σε σχέση με ένα σύνολο ανεξάρτητων τυχαίων μεταβλητών Y_1, Y_2, \dots, Y_N όπου $Y_i, i = 1, 2, \dots, N$ ακολουθεί κατανομή η οποία ανάγεται στην εκθετική οικογένεια κατανομών. Η κατανομή του εκάστοτε $Y_i, i = 1, 2, \dots, N$ έχει τις κάτωθι ιδιότητες:

1. Είναι κανονικής μορφής και εξαρτάται από μία μόνο παράμετρο θ_i όπου δεν είναι απαραίτητο να είναι όλα ίδια, έτσι

$$f(y_i) = \exp [b_i(\theta_i)y_i + c_i(\theta_i) + d_i(y_i)].$$

2. Οι κατανομές των Y_i είναι της ίδιας μορφής και κατ' επέκταση δεν χρειαζόμαστε δείκτες στις συναρτήσεις $b(\cdot), c(\cdot), d(\cdot)$.

Έτσι η από κοινού συνάρτηση πυκνότητας πιθανότητας των Y_1, Y_2, \dots, Y_N είναι:

$$\begin{aligned} f(y_1, y_2, \dots, y_N; \theta_1, \theta_2, \dots, \theta_N) &= \prod_{i=1}^N \exp [y_i b(\theta_i) + c(\theta_i) + d(y_i)] \\ &= \exp \left[\sum_{i=1}^N y_i b(\theta_i) + \sum_{i=1}^N c(\theta_i) + \sum_{i=1}^N d(y_i) \right] \end{aligned} \quad (3.4)$$

Οι παράμετροι θ_i τυπικά δεν είναι στο άμεσο ενδιαφέρον μας, αφού μπορεί να υπάρχει διαφορετικό θ για κάθε παρατήρηση. Ουσιαστικά μας ενδιαφέρει να μελετήσουμε τους συντελεστές $\boldsymbol{\beta}^T = (\beta_1, \beta_2, \dots, \beta_n)$, όπου $n < N$. Υποθέτοντας ότι $E(Y_i) = \mu_i$, όπου μ_i είναι μια συνάρτηση του θ_i . Σε ένα γενικευμένο γραμμικό μοντέλο ορίζουμε

$$g(\mu_i) = \mathbf{x}_i^T \boldsymbol{\beta}.$$

Σημειώνεται ότι η συνάρτηση $g(\cdot)$ είναι μια μονότονη και διαφορίσιμη συνάρτηση η οποία καλείται συνάρτηση σύνδεσης (link function), ο \mathbf{x}_i είναι ένας πίνακας-στήλη $n \times 1$ επεξηγηματικών μεταβλητών

$$\mathbf{x}_i^T = [x_{i1}, \dots, x_{in}]$$

και $\boldsymbol{\beta}$ είναι ένας $n \times 1$ πίνακας-στήλη με

$$\boldsymbol{\beta} = \begin{bmatrix} \beta_1 \\ \vdots \\ \beta_n \end{bmatrix}.$$

Το διάνυσμα \mathbf{x}_i είναι η i -στήλη του πίνακα σχεδιασμού \mathbf{X} .

Ουσιαστικά ένα γενικευμένο γραμμικό μοντέλο αποτελείται από 3 συστατικά:

1. Τις μεταβλητές απόκρισης Y_1, \dots, Y_N όπου υποθέτουμε ότι ακολουθούν κατανομή η οποία ανήκει στην εκθετική οικογένεια κατανομών.
2. Τις παραμέτρους του μοντέλου μας $\boldsymbol{\beta}$ και τις επεξηγηματικές μεταβλητές

$$\mathbf{X} = \begin{bmatrix} \mathbf{x}_1^T \\ \vdots \\ \mathbf{x}_N^T \end{bmatrix} = \begin{bmatrix} x_{11} & \dots & x_{1n} \\ \vdots & & \vdots \\ x_{N1} & & x_{Nn} \end{bmatrix}.$$

3. Την μονότονη συνάρτηση σύνδεσης $g(\cdot)$, τέτοια ώστε

$$g(\mu_i) = \mathbf{x}_i^T \boldsymbol{\beta},$$

όπου

$$\mu_i = E(Y_i).$$

Ο λόγος που χρησιμοποιούμε κατανομές που ανήκουν στην εκθετική οικογένεια κατανομών είναι η ομαλότητά τους με συνέπεια οι εκτιμητές μέγιστης πιθανοφάνειας των παραμέτρων να έχουν καλές ασυμπτωτικές ιδιότητες [7].

3.2 Δομή του γενικευμένου γραμμικού μοντέλου

3.2.1 Συνάρτηση σύνδεσης

Στο κλασικό γραμμικό μοντέλο θεωρούμε ότι ισχύει η σχέση

$$\mu_i = E(Y_i) = \mathbf{x}_i^T \boldsymbol{\beta}, \text{ όπου } \mathbf{x}_i = (x_{i1}, x_{i2}, \dots, x_{iN}).$$

Σε ένα γενικευμένο γραμμικό μοντέλο ορίζουμε τρεις συνιστώσες

1. Τυχαία συνιστώσα (random component) : Αποτελείται από την απόκριση Y για την οποία έχουμε ανεξάρτητες παρατηρήσεις $\mathbf{Y}^T = (Y_1, \dots, Y_N)$ και την κατανομή της που υποθέσαμε ότι ανήκει στην εκθετική οικογένεια κατανομών.
2. Γραμμική συνάρτηση πρόβλεψης (linear predictor) : Για κάποιες παραμέτρους $\boldsymbol{\beta} = (\beta_1, \dots, \beta_n)^T$ και κάποιες επεξηγηματικές μεταβλητές $\mathbf{x}_i^T = (x_{i1}, \dots, x_{in})$ η γραμμική συνάρτηση πρόβλεψης εκφράζει παραμέτρους η_i που σχετίζονται με την μέση τιμή $E(Y_i)$ ως γραμμικούς συνδιασμούς των επεξηγηματικών μεταβλητών

$$\eta_i = \eta_i(\boldsymbol{\beta}) = \mathbf{x}_i^T \boldsymbol{\beta} = \sum_{j=1}^n \beta_j x_{ij}.$$

3. Συνάρτηση σύνδεσης (link function) : Είναι μια συνάρτηση που συνδέει την τυχαία συνιστώσα με την γραμμική συνάρτηση πρόβλεψης.

Στα γενικευμένα γραμμικά μοντέλα θεωρούμε ότι υπάρχει μια μονότονη και διαφορίσιμη συνάρτηση, έστω $g(\cdot)$ και ένα σύνολο παραμέτρων $\boldsymbol{\beta}^T = (\beta_1, \beta_2, \dots, \beta_n)$, $n < N$ τέτοια ώστε ένας γραμμικός συνδιασμός των $\boldsymbol{\beta}$ να είναι ίσος με τη συνάρτηση της αναμενόμενης τιμής μ_i των Y_i , δηλαδή

$$g(\mu_i) = \eta_i = \mathbf{x}_i^T \boldsymbol{\beta}.$$

Η συνάρτηση αυτή ονομάζεται συνάρτηση σύνδεσης (link function). Η επιλογή μιας συνάρτησης σύνδεσης εξαρτάται από το χαρακτήρα των δεδομένων και είναι κατά κάποιο τρόπο αυθαίρετη. Ένα μοντέλο όπου το πλήθος των παραμέτρων είναι το ίδιο με το πλήθος των παρατηρήσεων ονομάζεται κορεσμένο ή πλήρες μοντέλο και ένα μοντέλο με μία παράμετρο, όπου όλες οι αποκλίσεις μεταξύ των απαντήσεων διαμορφώνονται από τυχαίο στοιχείο, ονομάζεται μηδενικό μοντέλο. Το κορεσμένο μοντέλο είναι στην πράξη μη ενημερωτικό, αλλά χρήσιμο κατά τη μέτρηση της επάρκειας ενός προσαρμοσμένου μοντέλου, κάτι που θα αναλυθεί αργότερα [18].

Διωνυμικά δεδομένα

Όταν οι ξαρτημένες μεταβλητές ακολουθούν την διωνυμική κατανομή υπάρχουν τρεις βασικές συναρτήσεις σύνδεσης

1. logit : $\eta_i = g(\mu_i) = \log\left(\frac{\mu_i}{1 - \mu_i}\right)$,
2. probit : $\eta_i = g(\mu_i) = \Phi^{-1}(\mu_i)$, όπου Φ είναι η συνάρτηση κατανομής της κανονικής κατανομής $N(0, 1)$,
3. complementary log-log : $\eta_i = g(\mu_i) = \log(-\log(1 - \mu_i))$.

Αν και οι μετασχηματισμοί logit και probit είναι πανομοιότυποι, προτιμάται ο logit καθώς είναι πιο άμεση και ευκολότερη η ερμηνεία όντας ο λογάριθμος των σχετικών (ή συμπληρωματικών) πιθανοτήτων (odds) [9].

Κανονική συνάρτηση σύνδεσης

Σε κάθε εκθετική οικογένεια κατανομών υπάρχει μια παράμετρος που αναφέρεται ως φυσική παράμετρος. Η συνάρτηση σύνδεσης που μετασχηματίζει τη μέση τιμή μ_i στη φυσική παράμετρο καλείται κανονική συνάρτηση σύνδεσης και είναι αυτή που συνηθίζεται στις περισσότερες περιπτώσεις. Η κανονική συνάρτηση σύνδεσης μας εξασφαλίζει ότι ο λογάριθμος της πιθανοφάνειας είναι κοίλη συνάρτηση (υπάρχει μοναδικό μέγιστο το οποίο βρίσκουν γρήγορα οι αριθμητικές μέθοδοι) και η επαρκής συνάρτηση, στην οποία βασίζεται όλη η συμπερασματολογία, έχει απλή μορφή. Ενδεικτικά για τις παρακάτω κατανομές όταν $\eta_i = \theta_i$, όπου θ_i είναι η κανονική παράμετρος όπως αυτή έχει οριστεί στην (3.7), ορίζουμε τις κάτωθι συναρτήσεις πρόβλεψης [9]

1. Κανονική : $\eta_i = g(\mu_i) = \mu_i$,
2. Poisson : $\eta_i = g(\mu_i) = \log(\mu_i)$,
3. Διωνυμική : $\eta_i = g(\mu_i) = \log\left(\frac{p_i}{1 - \mu_i}\right)$,
4. Γαμμα : $\eta_i = g(\mu_i) = \mu_i^{-1}$.

3.2.2 Εκθετική οικογένεια κατανομών

Για να ανήκει στην εκθετική οικογένεια κατανομών μια κατανομή πιθανότητας θα πρέπει η συνάρτηση πιθανότητάς της να μπορεί να γραφεί με την κάτωθι μορφή

$$f_Y(y; \theta; \phi) = \exp\left[\frac{y\theta - b(\theta)}{a(\phi)} + c(y, \phi)\right], \quad (3.5)$$

όπου

1. $a(\cdot)$, $b(\cdot)$, $c(\cdot)$ γνωστές συναρτήσεις,
2. θ , ϕ παράμετροι.

Σημειώνεται ότι:

1. Αν το ϕ είναι γνωστό τότε έχουμε εκθετική οικογένεια κατανομών με μία παράμετρο και το θ καλείται κανονική παράμετρος της κατανομής.
2. Αν το ϕ είναι άγνωστο τότε μπορούμε να το θεωρήσουμε σαν μια παράμετρο κλίμακας για την κατανομή και καλείται παράγοντας όχλησης της κατανομής [3].

Παρατήρηση 3.2.1. Στην προαναφερθείσα περίπτωση, όπου το ϕ είναι άγνωστο, η συνάρτηση $a(\phi)$ του τύπου της συνάρτησης κατανομής της εκθετικής οικογένειας κατανομών (3.7) έχει την μορφή $a(\phi) = \frac{\phi}{w}$, όπου w είναι μια σταθερά στάθμισης, η οποία είναι διαφορετική για κάθε παρατήρηση. Δηλαδή αν το δείγμα μας αποτελείται από y_1, y_2, \dots, y_N παρατηρήσεις τότε

$$a_i(\phi) = \frac{\phi}{w_i}.$$

Σε αυτή την περίπτωση η συνάρτηση πιθανοφάνειας της εκθετικής οικογένειας κατανομών έχει την μορφή

$$\mathcal{L}_n(\theta) = \prod_{i=1}^N f_Y(y_i; \theta, \phi) = \exp \left[\frac{\sum_{i=1}^N w_i y_i \theta - b(\theta)}{\phi} + c_1(y_i, \phi) \right].$$

Σημειώνεται ότι η συνάρτηση $c_1(y_i, \phi)$ είναι ανεξάρτητη της παραμέτρου θ .

Μέση τιμή εκθετικής οικογένειας κατανομών

Έστω μια κατανομή η οποία ανήκει στην εκθετική οικογένεια κατανομών, δηλαδή μέσω τροποποιήσεων μπορεί να αναχθεί στην μορφή (3.7).

Υπολογίζουμε τον λογάριθμο της συνάρτησης πιθανοφάνειας έστω

$$l(\theta; y) = \log(y; \theta, \phi)$$

και συγκεκριμένα, υπολογίζοντας τον λογάριθμο της (3.7) έχουμε:

$$l(\theta; y) = \frac{y\theta - b(\theta)}{a(\phi)} + c(y, \phi).$$

Επίσης έχουμε

$$\frac{\partial l}{\partial \theta} = \frac{y - b'(\theta)}{a(\phi)},$$

$$\frac{\partial^2 l}{\partial \theta^2} = -\frac{b''(\theta)}{a(\phi)}.$$

Είναι γνωστό ότι για την συνάρτηση l ισχύει

$$\begin{aligned} E\left(\frac{\partial l}{\partial \theta}\right) &= 0 \\ \Rightarrow \frac{\mu - b'(\theta)}{a(\phi)} &= 0 \\ \Rightarrow \mu - b'(\theta) &= 0 \\ \Rightarrow \mu &= b'(\theta). \end{aligned}$$

Συνεπώς καταλήγουμε ότι η μέση τιμή της Y ισούται με την παράγωγο της συνάρτησης b

$$E(Y) = \mu = b'(\theta). \quad (3.6)$$

Διακύμανση εκθετικής οικογένειας κατανομών

Όμοια με προηγουμένως, έστω μια κατανομή η οποία ανήκει στην εκθετική οικογένεια. Δηλαδή η συνάρτηση πιθανότητάς της μπορεί να γραφεί με την κάτωθι μορφή

$$f_y(y; \theta; \phi) = \exp\left[\frac{y\theta - b(\theta)}{a(\phi)} + c(y, \phi)\right], \quad (3.7)$$

όπου

1. $a(\cdot)$, $b(\cdot)$, $c(\cdot)$ άγνωστες συναρτήσεις,
2. θ , ϕ παράμετροι.

Στη συνέχεια δίνουμε τον ορισμό της ροπογεννήτριας ο οποίος είναι χρήσιμος για την εύρεση της διακύμανσης.

Ορισμός 3.2.1. Έστω X μια τυχαία μεταβλητή, $t > 0$, η συνάρτηση $M_X(t)$ ονομάζεται ροπογεννήτρια της X και ορίζεται ως:

$$M_X(t) = \begin{cases} \int_{-\infty}^{+\infty} e^{tx} f_x(x) dx, & \text{όταν η τ.μ. } X \text{ είναι συνεχής} \\ \sum_{-\infty}^{+\infty} e^{tx} f(x), & \text{όταν η τ.μ. } X \text{ είναι διακριτή} \end{cases}.$$

Η $f_x(x)$ στην περίπτωση που η τ.μ. X είναι συνεχής είναι συνάρτηση πυκνότητας πιθανότητας της τ.μ. X ενώ στην περίπτωση που η τ.μ. X είναι διακριτή η $f(x)$ είναι η συνάρτηση πιθανότητας.

Παρατήρηση: Η πρώτη παράγωγος της ροπογεννήτριας στο σημείο $t = 0$ ισούται με την μέση τιμή της τυχαίας μεταβλητής. Η δεύτερη παράγωγος της ροπογεννήτριας στο σημείο $t = 0$ ισούται με την δεύτερη ροπή με την οποία θα υπολογίσουμε την διακύμανση ως κάτωθι

$$\text{Var}(Y) = \left. \frac{d^2 M_x(t)}{dt^2} \right|_{t=0} - \left[\left. \frac{dM_x(t)}{dt} \right|_{t=0} \right]^2.$$

Με την βοήθεια του ορισμού έχουμε [9]

$$\begin{aligned} E \left(\frac{\partial^2 l}{\partial \theta^2} \right) + E \left(\frac{\partial l}{\partial \theta} \right)^2 &= 0 \\ \Rightarrow -\frac{b''(\theta)}{a(\phi)} + \frac{\text{Var}(Y)}{a^2(\phi)} &= 0. \end{aligned}$$

Επομένως καταλήγουμε ότι η διακύμανση της Y ισούται με

$$\text{Var}(Y) = b''(\theta)a(\phi). \quad (3.8)$$

3.3 Εκτίμηση παραμέτρων

Αφού επιλέξουμε ένα συγκεκριμένο μοντέλο στη συνέχεια θα πρέπει να εκτιμήσουμε τις παραμέτρους του μοντέλου βάση των παρατηρήσεων και να καταλήξουμε σε μια εκτίμηση για την αναμενόμενη τιμή της εξαρτημένης μεταβλητής Y . Οι μέθοδοι που χρησιμοποιούμε στα γενικευμένα γραμμικά μοντέλα για να εκτιμήσουμε (κατά σημείο ή κατά διάστημα) τις παραμέτρους της κατανομής της μεταβλητής απόκρισης βασίζονται στην μέθοδο μεγίστης πιθανοφάνειας, όπου η εκτιμητής μεγίστης πιθανοφάνειας (EMΠ) υπάρχει και μπορούμε να βρούμε εύκολα μια μαθηματική του έκφραση. Σε πολλές περιπτώσεις, κατά την διαδικασία εύρεσης των εκτιμητριών, καταλήγουμε σε ένα μη γραμμικό σύστημα εξισώσεων που για την επίλυσή του είναι απαραίτητη η εφαρμογή αριθμητικών μεθόδων, όπως η μέθοδος *Newton – Raphson*.

3.3.1 Μέθοδος μεγίστης πιθανοφάνειας

Έστω ανεξάρτητες τυχαίες μεταβλητές Y_1, Y_2, \dots, Y_n τέτοιες ώστε να ικανοποιούν τις προϋποθέσεις ενός γενικευμένου γραμμικού μοντέλου με $E(Y_i) = \mu_i$ και συνάρτηση σύνδεσης $\eta_i = g(\mu_i) = \mathbf{x}_i^T \boldsymbol{\beta}$, όπου $\mathbf{x}_i^T = (x_{i1}, x_{i2}, \dots, x_{im})$ είναι οι παρατηρήσεις μας και $\boldsymbol{\beta}^T = (\beta_1, \beta_2, \dots, \beta_m)$. Έχοντας θεωρήσει ότι οι ανεξάρτητες τυχαίες μεταβλητές

$$Y_i \sim f(y_i; \theta_i, \phi) = \exp \left[\frac{y_i \theta_i - b(\theta_i)}{a(\phi)} + c(y_i, \phi) \right], \quad i = 1, \dots, n,$$

προκύπτει ότι για $\forall Y_i$

$$\begin{aligned} E(Y_i) &= \mu_i = b'(\theta_i), \\ \text{Var}(Y_i) &= \alpha(\phi)b''(\theta_i). \end{aligned}$$

Η από κοινού πυκνότητα των δεδομένων είναι το γινόμενο

$$\prod_{i=1}^n f(y_i; \theta_i, \phi)$$

και η συνάρτηση πιθανοφάνειας δίνεται απο την σχέση

$$\mathcal{L} = \prod_{i=1}^n f(y_i; \theta_i, \phi) = \prod_{i=1}^n \exp \left[\frac{y_i \theta_i - b(\theta_i)}{a(\phi)} + c(y_i, \phi) \right]. \quad (3.9)$$

Παίρνοντας τον λογάριθμο της (3.9) έχουμε

$$l = \log \mathcal{L} = \sum_{i=1}^n l_i = \sum_{i=1}^n \exp \left[\frac{y_i \theta_i - b(\theta_i)}{a(\phi)} + c(y_i, \phi) \right]. \quad (3.10)$$

Παραγωγίζοντας την (3.10) ως προς τις παραμέτρους του μοντέλου μας και εφαινόζοντας τον κανόνα της αλυσίδας, παίρνουμε την κάτωθι σχέση

$$U_j = \frac{\partial l}{\partial \beta_j} = \sum_{i=1}^n \left[\frac{\partial l_i}{\partial \beta_j} \right] = \sum_{i=1}^n \left[\frac{\partial l_i}{\partial \theta_i} \frac{\partial \theta_i}{\partial \mu_i} \frac{\partial \mu_i}{\partial \beta_j} \right], \quad j = 1, 2, \dots, m. \quad (3.11)$$

Γνωρίζουμε ότι

$$\begin{aligned} \theta_i &= \theta_i(\mu_i), \\ \mu_i &= \mu_i(\eta_i) \equiv g^{-1}(\eta_i), \\ \eta_i &= \eta_i(\boldsymbol{\beta}) = \mathbf{x}_i^T \boldsymbol{\beta} = \sum_{k=1}^m \beta_k x_{ik} \end{aligned}$$

και όπως παρατηρούμε η παράμετρος $\boldsymbol{\beta}^T = (\beta_1, \beta_2, \dots, \beta_m)$ είναι 'κρυμμένη' στις παραμέτρους θ_i , $\theta_i = \theta_i(\mu_i(\eta_i(\boldsymbol{\beta})))$.

Κάνοντας τους υπολογισμούς του δεύτερου μέρους της (3.11) λαμβάνουμε

$$\begin{aligned} \frac{\partial l_i}{\partial \theta_i} &= \frac{y_i - \mu_i}{\alpha(\phi)}, \\ \frac{\partial \theta_i}{\partial \mu_i} &= \left(\frac{\partial \mu_i}{\partial \theta_i} \right)^{-1} = \frac{1}{b''(\theta_i)} = \frac{\alpha(\phi)}{\text{Var}(Y_i)}, \\ \frac{\partial \mu_i}{\partial \beta_j} &= \frac{\partial \mu_i}{\partial \eta_i} \frac{\partial \eta_i}{\partial \beta_j} = \frac{\partial \mu_i}{\partial \eta_i} x_{ij}. \end{aligned}$$

Έτσι οι εξισώσεις πιθανοφάνειας ενός γενικευμένου γραμμικού μοντέλου είναι

$$\mathbf{U}_j = \frac{\partial l}{\partial \beta_j} = \frac{1}{\alpha(\phi)} \sum_{i=1}^n \left[\frac{(y_i - \mu_i)}{\text{Var}(Y_i)} \left(\frac{\partial \mu_i}{\partial \eta_i} \right) x_{ij} \right] = 0, \quad j = 1, 2, \dots, m,$$

όπου $\eta_i = g(\mu_i) = \mathbf{x}_i^T \boldsymbol{\beta} = \sum_{k=1}^m \beta_k x_{ik}$. Η συνάρτηση $U = [U_1, \dots, U_n]$ ονομάζεται συνάρτηση βαθμολόγησης (score).

Έστω ο πίνακας διασποράς-συνδιασποράς \mathbf{I} των \mathbf{U}_i με όρους $\mathcal{I}_{jk} = E[U_j U_k]$, που ονομάζεται πίνακας πληροφορίας.

$$\begin{aligned} \mathcal{I}_{jk} &= E[U_j U_k] \\ &= E \left(\sum_{i=1}^n \left[\frac{Y_i - \mu_i}{\text{Var}(Y_i)} x_{ij} \left(\frac{\partial \mu_i}{\partial \eta_i} \right) \right] \sum_{l=1}^n \left[\frac{Y_l - \mu_l}{\text{Var}(Y_l)} x_{lk} \left(\frac{\partial \mu_l}{\partial \eta_l} \right) \right] \right) \\ &= \sum_{i=1}^n \frac{E[(Y_i - \mu_i)^2]}{[\text{Var}(Y_i)]^2} x_{ij} x_{ik} \left(\frac{\partial \mu_i}{\partial \eta_i} \right)^2 \end{aligned} \quad (3.12)$$

και γνωρίζοντας ότι $E[(Y_i - \mu_i)^2] = 1$ έχουμε

$$\mathcal{I}_{jk} = \sum_{i=1}^n \frac{E[(Y_i - \mu_i)^2]}{[\text{Var}(Y_i)]^2} x_{ij} x_{ik} \left(\frac{\partial \mu_i}{\partial \eta_i} \right)^2. \quad (3.13)$$

Ορίζοντας έναν διαγώνιο $n \times n$ πίνακα, έστω \mathbf{W} , με στοιχεία $w_{ii} = \frac{1}{\text{Var}(Y_i)} \left(\frac{\partial \mu_i}{\partial \eta_i} \right)^2$, ο παραπάνω τύπος μπορεί να γραφεί και ως

$$\mathcal{I}_{jk} = \mathbf{X}^T \mathbf{W} \mathbf{X}. \quad (3.14)$$

Τέλος, ο πίνακας πληροφορίας $\mathbf{I} = E[\mathbf{U} \mathbf{U}^T]$ έχει τα στοιχεία

$$\mathcal{I}_{jk} = E \left[\frac{\partial l}{\partial \beta_j} \frac{\partial l}{\partial \beta_k} \right] = -E \left[\frac{\partial^2 l}{\partial \beta_j \partial \beta_k} \right]. \quad (3.15)$$

Για να εκτιμήσουμε τα β_j θα εφαρμόσουμε την μέθοδο Newton-Raphson μέσω του επαναληπτικού σχήματος

$$\begin{aligned} \hat{\boldsymbol{\beta}}^{(t+1)} &= \hat{\boldsymbol{\beta}}^{(t)} - \frac{\mathbf{U}^{(t)}}{\mathbf{U}'^{(t)}} \\ &= \hat{\boldsymbol{\beta}}^{(t)} - \left[\frac{\partial^2 l}{\partial \hat{\boldsymbol{\beta}}_j \partial \hat{\boldsymbol{\beta}}_k} \right]_{\hat{\boldsymbol{\beta}} = \hat{\boldsymbol{\beta}}^t}^{-1} \mathbf{U}^{(t)} \end{aligned} \quad (3.16)$$

και λόγω της (3.18)

$$\hat{\boldsymbol{\beta}}^{(t+1)} = \hat{\boldsymbol{\beta}}^{(t)} + [\mathbf{I}^{(t)}]^{-1} \mathbf{U}^{(t)}. \quad (3.17)$$

Πολλαπλασιάζοντας και τα δύο μέλη με την $\mathbf{I}^{(t-1)}$ καταλήγουμε στην εξίσωση της score

$$\begin{aligned}\mathbf{X}^T \mathbf{W}^{(t)} \mathbf{X} \hat{\boldsymbol{\beta}}^{(t+1)} &= \mathbf{X}^T \mathbf{W}^{(t)} \mathbf{z}^{(t)} \\ \hat{\boldsymbol{\beta}}^{(t+1)} &= (\mathbf{X}^T \mathbf{W}^{(t)} \mathbf{X})^{-1} \mathbf{X}^T \mathbf{W}^{(t)} \mathbf{z}^{(t)},\end{aligned}\quad (3.18)$$

όπου \mathbf{z} είναι η στήλη με στοιχεία

$$z_i = \mathbf{x}_i^T \hat{\boldsymbol{\beta}} + (y_i - \mu_i) \frac{\partial \mu_i}{\partial \eta_i}, \quad i = 1, \dots, n. \quad (3.19)$$

Επειδή τα μ_i, η_i τα υπολογίσαμε στο $\hat{\boldsymbol{\beta}} = \hat{\boldsymbol{\beta}}^{(t)}$, ο επαναληπτικός τύπος (3.18) μπορεί να γραφεί και

$$\hat{\boldsymbol{\beta}}^{(t+1)} = (\mathbf{X}^T \hat{\mathbf{W}}^{(t)} \mathbf{X})^{-1} \mathbf{X}^T \hat{\mathbf{W}}^{(t)} \hat{\mathbf{z}}^{(t)}, \quad (3.20)$$

όπου $\hat{\mathbf{W}}$ συμβολίζουμε τον αντίστοιχο πίνακα μετά την ακτικατάσταση των $\boldsymbol{\beta}$ από τις εκτιμήτριες $\hat{\boldsymbol{\beta}}$. Η τελευταία μορφή είναι των σταθμισμένων ελαχίστων τετραγώνων.

Παρατηρούμε ότι η σχέση (3.20) μας θυμίζει την μορφή των κανονικών εξισώσεων των γενικευμένων γραμμικών μοντέλων που προκύπτουν από τα σταθμισμένα ελάχιστα τετράγωνα. Η διαφοροποίηση έγκυται στο γεγονός ότι στην περίπτωση των γενικευμένων γραμμικών μοντέλων θα πρέπει να λυθούν με μια επαναληπτική μέθοδο επειδή, όπως είδαμε, τα \mathbf{z} και \mathbf{W} εξαρτώνται σε γεννικές γραμμές από τις εκτιμήτριες $\hat{\boldsymbol{\beta}}$.

Ο αλγόριθμος της διαδικασίας έχει επι της ουσίας τέσσερα βήματα:

1. Ξεκίνημα με μια αρχική εκτίμηση $\mu^{(0)}$.
2. Υπολογισμός ενός προσαρμοσμένου διανύσματος απόκρισης $\mathbf{z}^{(0)}$ και ένα πλέγμα βαρών $\mathbf{W}^{(0)}$.
3. Υπολογισμός $\hat{\boldsymbol{\beta}}^{(1)}$ με τη μέθοδο των σταθμισμένων ελαχίστων τετραγώνων.
4. Επανάληψη των βημάτων (2) και (3) μέχρι τη σύγκλιση.

Πολλά στατιστικά πακέτα για την εκτίμηση των παραμέτρων χρησιμοποιούν τον παραπάνω αλγόριθμο ο οποίος τερματίζει για το διάνυσμα $\hat{\boldsymbol{\beta}}^{(t)}$ για το οποίο ισχύει

$$\left| \frac{\hat{\beta}_i^{(t-1)} - \hat{\beta}_i^{(t)}}{\hat{\beta}_i^{(t)}} \right| < \delta, \quad i = 1, 2, \dots, m, \quad (3.21)$$

όπου $\delta = \frac{1}{\epsilon}$ δηλαδή μια πολύ μικρή ποσότητα [7].

3.3.2 Συνάρτηση Wald

Αφού εκτιμήσουμε τις παραμέτρους του μοντέλου θα πρέπει να εκτιμήσουμε ποιές μεταβλητές είναι στατιστικά σημαντικές για το μοντέλο μας. Εφαρμόζουμε την στατιστική συνάρτηση Wald με σκοπό την συμπερασματολογία για τους εκτιμητές μεγίστης πιθανοφάνειας. Η εκτιμήτρια μεγίστης πιθανοφάνειας $\hat{\beta}$ για την παράμετρο β ακολουθεί ασυμπτωτικά την κανονική κατανομή

$$\hat{\beta} \sim N_p(\beta, \hat{V}(\hat{\beta})),$$

όπου $\hat{V}(\hat{\beta}) = \mathbf{I}^{-1}(\hat{\beta})$ η εκτιμήτρια του πίνακα διασποράς-συνδιασποράς της ασυμπτωτικής κατανομής της $\hat{\beta}$ και $\mathbf{I}(\hat{\beta})$ ο παρατηρούμενος πίνακας πληροφορίας όπως έχει οριστεί στην (3.13). Επομένως γίνεται αντιληπτό ότι οι εκτιμήτριες μεγίστης πιθανοφάνειας $\hat{\beta}$ ακολουθούν συμπτωτικά την κανονική κατανομή με μέση τιμή β και πίνακα διασποράς-συνδιασποράς $\mathbf{X}^T \hat{\mathbf{W}} \mathbf{X}$.

Για τον έλεγχο *Wald* μιας μηδενικής υπόθεσης, έστω $H_0 : \beta_j = 0$ με εναλλακτική $H_1 : \beta_j \neq 0$ χρησιμοποιούμε την συνάρτηση ελέγχου *Wald* η οποία ακολουθεί προσεγγιστικά την κανονική κατανομή

$$Z = \frac{\hat{\beta}_j - \beta_j}{\left(\mathbf{I}^{-1}(\hat{\beta})_{jj}\right)^{\frac{1}{2}}} \sim N(0, 1),$$

όπου $\mathbf{I}^{-1}(\hat{\beta})_{jj}$ είναι το j -οστό διαγώνιο στοιχείο του αντιστρόφου του παρατηρούμενου πίνακα πληροφορίας $\mathbf{I}(\hat{\beta})$. Η συνάρτηση Wald είναι

$$W_n = \frac{1}{a(\phi)} \left(\mathbf{C}\hat{\beta} - \mathbf{r}\right)^T \left[\mathbf{C}(\mathbf{X}^T \hat{\mathbf{W}} \mathbf{X})^{-1} \mathbf{C}^T\right]^{-1} \left(\mathbf{C}\hat{\beta} - \mathbf{r}\right),$$

όπου \mathbf{C} είναι ο πίνακας υπόθεσης που έχει την κάτωθι μορφή

$$\mathbf{C} = \begin{bmatrix} 0 & \dots & 0 & 1 & 0 & \dots & 0 \\ 0 & \dots & 0 & 0 & 1 & 0 & \dots & 0 \\ \vdots & & & & & & & \\ 0 & \dots & 0 & 0 & 0 & \dots & & 1 \end{bmatrix}$$

και r γνωστή παράμετρος [3].

3.3.3 Score test

Σε αντίθεση με τη δοκιμή Wald, στη δοκιμή βαθμολογίας απαιτείται η εκτίμηση β . Η στατιστική δοκιμής βαθμολογίας χρησιμοποιεί το παράγωγο της συνάρτησης πιθανοτήτων *log*, τη στατιστική βαθμολογίας $\mathbf{U}(\beta)$ και τη διακύμανσή της και δίνεται από :

$$R_n = \alpha(\phi) \mathbf{U}^T(\tilde{\beta}) (\mathbf{X}^T \hat{\mathbf{W}} \mathbf{X})^{-1} \mathbf{U} \tilde{\beta}. \quad (3.22)$$

Το στατιστικό αποτέλεσμα της δοκιμής έχει περίπου την κατανομή X_k^2 κάτω από την μηδενική υπόθεση και είναι λιγότερο ακριβές όταν χρησιμοποιούνται οι εκτιμητές για W και ϕ . Η μηδενική υπόθεση απορρίπτεται, όταν ο R_n παίρνει τις τιμές από την άνω ουρά της X_k^2 κατανομής. Σημειώστε ότι η X^2 κατανομή των προηγούμενων στατιστικών δοκιμών ισχύει μόνο ασυμπτωτικά και επομένως οι δοκιμές ισχύουν μόνο σε μεγάλα δείγματα [3].

3.4 Επάρκεια μοντέλου

Αφού έχουμε δημιουργήσει το μοντέλο μας και έχουμε εκτιμήσει ποιές μεταβλητές X_i είναι στατιστικά σημαντικές, ουσιαστικά ποιές μεταβλητές συνεισφέρουν στο μοντέλο μας, θα πρέπει να ερευνήσουμε την επάρκεια προσαρμογής του μοντέλου μας στο σύνολο δεδομένων. Αυτό επιτυγχάνεται μέσω της σύγκρισης της συνάρτησης πιθανοφάνειας αυτού του μοντέλου με τη συνάρτηση πιθανοφάνειας του μέγιστου μοντέλου.

Μέγιστο μοντέλο

Το μέγιστο μοντέλο είναι ένα γενικευμένο γραμμικό μοντέλο με την ίδια κατανομή και ίδια συνάρτηση σύνδεσης όπως το μοντέλο που έχουμε δημιουργήσει. Ο αριθμός των παραμέτρων του πλήρους μοντέλου ισούται με τον αριθμό των παρατηρήσεων, έστω N . Λόγω αυτής της ιδιότητας το μοντέλο αυτό μπορεί να θεωρηθεί ότι περιγράφει πλήρως τα δεδομένα.

Στατιστική συνάρτηση απόκλισης Deviance

Οι συναρτήσεις πιθανοφάνειας υπολογίζονται στον εκτιμητή μεγίστης πιθανοφάνειας $\hat{\beta}_{max}$, όπου είναι οι παράμετροι του κορεσμένου μοντέλου, και $\hat{\beta}$, οι παράμετροι του μοντέλου μας, αντίστοιχα και λαμβάνουμε $\mathcal{L}(\hat{\beta}_{max}; \mathbf{y})$ και $\mathcal{L}(\hat{\beta}; \mathbf{y})$ αντίστοιχα. Σημειώνουμε ότι αν το $\mathcal{L}(\hat{\beta}; \mathbf{y})$ είναι κοντά στο $\mathcal{L}(\hat{\beta}_{max}; \mathbf{y})$ τότε το μοντέλο που έχουμε δημιουργήσει περιγράφει ικανοποιητικά τα δεδομένα μας και αν το $\mathcal{L}(\hat{\beta}; \mathbf{y})$ είναι μικρότερο από το $\mathcal{L}(\hat{\beta}_{max}; \mathbf{y})$ τότε το μοντέλο δεν είναι ικανοποιητικό. Σε αυτή την περίπτωση οδηγούμαστε στην χρήση του γενικευμένου λόγου πιθανοφάνειας

$$\lambda = \frac{\mathcal{L}(\hat{\beta}_{max}; \mathbf{y})}{\mathcal{L}(\hat{\beta}; \mathbf{y})}$$

$$\Rightarrow \log \lambda = l(\hat{\beta}_{max}; \mathbf{y}) - l(\hat{\beta}; \mathbf{y}),$$

σαν μέτρο καλής προσαρμογής του μοντέλου. Ένδειξη καλής προσαρμογής του μοντέλου μας είναι όταν το μέτρο αυτό λαμβάνει μεγάλες τιμές.

Με την βοήθεια του παραπάνου λογαρίθμου του γενικευμένου λόγου πιθανοφάνειας ορίζουμε την στατιστική συνάρτηση λογαριθμικής πιθανοφάνειας Deviance

$$D = 2 \left[l(\hat{\boldsymbol{\beta}}_{max}; \mathbf{y}) - l(\hat{\boldsymbol{\beta}}; \mathbf{y}) \right] \quad (3.23)$$

και ο μετασχηματισμός αυτής

$$D = 2 \left[l(\hat{\boldsymbol{\beta}}_{max}; \mathbf{y}) - l(\boldsymbol{\beta}_{max}; \mathbf{y}) \right] - 2 \left[l(\hat{\boldsymbol{\beta}}; \mathbf{y}) - l(\boldsymbol{\beta}; \mathbf{y}) \right] + 2 \left[l(\boldsymbol{\beta}_{max}; \mathbf{y}) - l(\boldsymbol{\beta}; \mathbf{y}) \right],$$

όπου

- $l(\hat{\boldsymbol{\beta}}_{max}; \mathbf{y}) - l(\boldsymbol{\beta}_{max}; \mathbf{y})$: ακολουθεί την κατανομή χ_m^2 , όπου m ο αριθμός των παραμέτρων του πλήρους μοντέλου,
- $l(\hat{\boldsymbol{\beta}}; \mathbf{y}) - l(\boldsymbol{\beta}; \mathbf{y})$: ακολουθεί την κατανομή χ_p^2 , όπου p ο αριθμός των παραμέτρων στο μοντέλο που μας ενδιαφέρει,
- $l(\boldsymbol{\beta}_{max}; \mathbf{y}) - l(\boldsymbol{\beta}; \mathbf{y})$: είναι μια θετική σταθερά η οποία θα είναι κοντά στο μηδέν αν το μοντέλο που δημιουργήσαμε περιγράφει τόσο καλά τα δεδομένα μας όσο το πλήρες μοντέλο.

Όταν το μοντέλο μας προσαρμόζεται καλά στα δεδομένα μας ισχύει

$$D \sim \chi_{n-p}^2,$$

ενώ όταν το μοντέλο δεν είναι ικανοποιητικό τότε η Deviance ακολουθεί προσεγγιστικά τη μη κεντρική χ^2 κατανομή. Η τιμή της D θα πρέπει να είναι κοντά στο μέσο της κατανομής αν το μοντέλο είναι κατάλληλο. Δηλαδή αν ένα μοντέλο p παραμέτρων περιγράφει καλά ένα σύνολο από N παρατηρήσεις, ώστε $D \sim \chi_{N-p}^2$ τότε θα πρέπει $D \simeq N - p$.

Η στατιστική συνάρτηση Deviance, όπως την ορίσαμε, χρησιμοποιείται και στον έλεγχο υποθέσεων [7].

Residuals

Οι διαφορές των παρατηρήσεων y_i με τις τιμές των εκτιμημένων τιμών \hat{y}_i ονομάζονται υπόλοιπα (residuals). Έστω

$$\hat{e}_i = y_i - \hat{y}_i.$$

Τα υπόλοιπα μας δίνουν πληροφορίες για τα χαρακτηριστικά των μοντέλων. Για να μετρήσουμε την απόκλιση του μοντέλου μας χρησιμοποιούμε την στατιστική συνάρτηση Pearson X^2 που ορίζεται ως

$$X^2 = \sum_{i=1}^n \frac{(y_i - \hat{\mu}_i)^2}{Var(\hat{\mu}_i)}.$$

Ωστόσο στα γενικευμένα γραμμικά μοντέλα δεν μπορούμε να δουλέψουμε με αυτόν τον τρόπο, επομένως ορίζουμε [5]

- Υπόλοιπα deviance

$$r_i^{(D)} = \text{sign}(y_i - \hat{\mu}_i) \sqrt{d_i},$$

όπου $d_i = 2 \left\{ y_i \left(\hat{\theta}_i^{max} - \hat{\theta}_i \right) - \left[b(\hat{\theta}_i^{max}) - b(\hat{\theta}_i) \right] \right\}$ αν ορίσουμε $\hat{\theta}^{max}$ την εκτίμηση του θ_{max} του πλήρους μοντέλου.

- Υπόλοιπα Pearson

$$r_i^{(P)} = \frac{y_i - \hat{\mu}_i}{\sqrt{\text{Var}(\hat{\mu}_i)}}, \quad i = 1, \dots, N.$$

Με την βοήθεια των υπολοίπων που έχουμε ορίσει μπορούμε να εκτιμήσουμε τα κάτωθι [5]:

- Εκτίμηση της ϕ

$$\hat{\phi} = \frac{\sum_{i=1}^I \left(r_i^{(P)} \right)^2}{n - p},$$

όπου n είναι το πλήθος των παρατηρήσεων μας και p το πλήθος των παραμέτρων του μοντέλου μας.

- Υπολογισμός του Deviance.

$$D = \frac{\sum_{i=1}^n \left(r_i^{(D)} \right)^2}{a(\hat{\phi})}.$$

3.4.1 Έλεγχος υποθέσεων

Ορίζουμε ένα μοντέλο για κάθε υπόθεση και συγκρίνουμε τις συναρτήσεις της καλής προσαρμογής για τα συγκεκριμένα μοντέλα τα οποία ακολουθούν την ίδια κατανομή και έχουν την ίδια συνάρτηση σύνδεσης. Έστω η μηδενική υπόθεση H_0 και η εναλλακτική H_1 , όπου

$$H_0 : \beta = \beta_0 = \begin{bmatrix} \beta_1 \\ \vdots \\ \beta_q \end{bmatrix} \text{ και } H_1 : \beta = \beta_1 = \begin{bmatrix} \beta_1 \\ \vdots \\ \beta_p \end{bmatrix}, \text{ όπου } q < p < N, \quad (3.24)$$

και οι συναρτήσεις καλής προσαρμογής τους

$$D_0 = 2 \left[l(\hat{\beta}_{max}; \mathbf{y}) - l(\hat{\beta}_0; \mathbf{y}) \right]$$

$$D_1 = 2 \left[l(\hat{\beta}_{max}; \mathbf{y}) - l(\hat{\beta}_1; \mathbf{y}) \right].$$

Παίρνοντας την διαφορά των στατιστικών συναρτήσεων Deviance έχουμε

$$\begin{aligned}\Delta D &= D_0 - D_1 = 2 \left[l(\hat{\beta}_{max}; \mathbf{y}) - l(\hat{\beta}_0; \mathbf{y}) \right] - 2 \left[l(\hat{\beta}_{max}; \mathbf{y}) - l(\hat{\beta}_1; \mathbf{y}) \right] \\ \Delta D &= 2 \left[l(\hat{\beta}_1; \mathbf{y}) - l(\hat{\beta}_0; \mathbf{y}) \right].\end{aligned}$$

Αν και τα δύο μοντέλα περιγράφουν καλά τα δεδομένα μας, δηλαδή $D_0 \sim \chi_{N-q}^2$ και $D_1 \sim \chi_{N-p}^2$, έτσι ώστε κάτω από κάποιες συνθήκες ανεξαρτησίας $\Delta D \sim \chi_{p-q}^2$, δηλαδή η ΔD είναι συνεπής με την χ_{p-q}^2 , επιλέγουμε το μοντέλο H_0 γιατί είναι πιο απλό (με τις λιγότερες μεταβλητές) [7].

3.4.2 Δείκτες καλής προσαρμογής AIC και BIC

Akaike's information criterion(AIC)

Ορίζουμε την σχέση

$$AIC = 2d - 2\log\mathcal{L},$$

όπου

1. \mathcal{L} η μεγιστοποιημένη τιμή της συνάρτησης πιθανοφάνειας για το εκτιμώμενο μοντέλο,
2. d ο αριθμός των παραμέτρων του μοντέλου.

Το AIC αποτελεί ένα κριτήριο επιλογής του βέλτιστου μοντέλου με όσο το δυνατόν μικρότερο αριθμό παραμέτρων. Βέλτιστο μοντέλο βάση αυτού του κριτηρίου θεωρείται αυτό όπου έχει την μικρότερη τιμή AIC.

Σημειώνουμε ότι όσο αυξάνεται το πλήθος των παραμέτρων d , ανεξάρτητα αν αυτές θεωρούνται στατιστικά σημαντικές, αυξάνεται και ο όρος $\log\mathcal{L}$ και μειώνεται η τιμή του AIC στην περίπτωση που η προσαρμογή του μοντέλου βελτιώνεται. Η ποσότητα $2d$ ονομάζεται ποινή [5].

Bayesian information creterion (BIC)

Ορίζουμε την σχέση

$$BIC = d \log m - 2 \log\mathcal{L},$$

όπου

1. \mathcal{L} η μεγιστοποιημένη τιμή της συνάρτησης πιθανοφάνειας για το εκτιμώμενο μοντέλο,
2. d ο αριθμός των παραμέτρων του μοντέλου,
3. m ο αριθμός των παρατηρήσεων.

Το BIC αποτελεί και αυτό ένα κριτήριο για την επιλογή του βέλτιστου μοντέλου ανάμεσα σε μοντέλα με διαφορετικό πλήθος παραμέτρων. Η λογική επιλογής είναι όμοια με αυτή του AIC όμως η διαφορά τους έγκηται στο γεγονός ότι η εισαγωγή επιπρόσθετων παραμέτρων αποθαρρύνεται σε μεγαλύτερο βαθμό απο το AIC. Σύμφωνα με την κλίμακα του Raftery, η οποία βασίζεται στην απόλυτη τιμή της διαφοράς των τιμών του κριτηρίου BIC δύο μοντέλων μπορούμε να κρίνουμε κατά πόσο ένα μοντέλο έχει καλύτερη προσαρμογή σε σχέση με ένα άλλο [5].

Σχήμα 3.1: Κλίμακα του Raftery

Διαφορά των τιμών BIC	Ένδειξη
0 – 2	Ασθενής (Weak)
2 – 8	Θετική (Positive)
6 – 10	Ισχυρή (Strong)
> 10	Πολύ Ισχυρή (Very Strong)

3.5 Τα βήματα της μεθόδου

Έχοντας ορίσει όσα χρειαζόμαστε για την μέθοδο ας δούμε ποιά είναι τα βασικά βήματα μοντελοποίησης με την χρήση των γενικευμένων γραμμικών μοντέλων [3]. Έστω y η μεταβλητή που επιθυμούμε να μελετήσουμε

- Επιλέγουμε την κατάλληλη $b(\theta)$ καθώς αυτή θα επηρεάσει την κατανομή των δεδομένων μας $f(y)$.
- Επιλέγουμε την συνάρτηση σύνδεσης $g(\mu)$. Συνήθως επιλέγουμε την κανονική συνάρτηση σύνδεσης της κατανομής που επιθυμούμε να δουλέψουμε.
- Επιλέγουμε τις επεξηγηματικές μεταβλητές x με την βοήθεια των οποίων θα μοντελοποιήσουμε την συνάρτηση σύνδεσης $g(\mu)$.
- Συλλέγουμε τις παρατηρήσεις y_1, y_2, \dots, y_n για τις αντίστοιχες επεξηγηματικές μεταβλητές x_1, x_2, \dots, x_n . Επιτυχημένες θεωρούνται οι παρατηρήσεις που είναι ανεξάρτητες μεταξύ τους καθώς θα μας οδηγήσουν στο να δημιουργήσουμε ένα τυχαίο δείγμα.
- Προκειμένου να φτιάξουμε το μοντέλο μας θα πρέπει να εκτιμήσουμε τις παραμέτρους β και ϕ , δεδομένου ότι η παράμετρος ϕ είναι άγνωστη. Αυτή η διαδικασία γίνεται επί το πλείστον με την χρήση προγραμματιστικών λογισμικών, όπως το SAS.
- Τέλος, αφού έχουμε εκτιμήσει τις παραμέτρους β του μοντέλου μας, παράγουμε εκτιμήσεις για την y και ελέγχουμε αν το μοντέλο μας εφαρμόζεται καλά στα δεδομένα. Αυτό επιτυγχάνεται τόσο μέσω των

διαγνωστικών ελέγχων για το μοντέλο όσο και εξετάζοντας κατά πόσο αυτές οι εκτιμήσεις διαφέρουν από τις πραγματικές παρατηρήσεις.

Παραδείγματα κατανομών

Παρακάτω παρουσιάζονται κάποια παραδείγματα γνωστών κατανομών που ανήκουν στην εκθετική οικογένεια κατανομών.

Κανονική κατανομή

Έστω ότι $Y \sim N(\mu, \sigma^2)$, με $\sigma > 0$ γνωστό, τότε η πυκνότητα της Y γράφεται

$$f(y; \mu) = (2\pi\sigma^2)^{-\frac{1}{2}} e^{-\frac{1}{2\sigma^2}(y-\mu)^2}, \quad y, \mu \in \mathfrak{R}, \quad (3.25)$$

η οποία γράφεται ως

$$\begin{aligned} f_Y(y; \mu) &= \frac{1}{\sqrt{2\pi\sigma^2}} \exp \left\{ -\frac{(y-\mu)^2}{2\sigma^2} \right\} \\ &= \exp \left\{ -\log \sqrt{2\pi\sigma^2} - \frac{y^2 - 2y\mu + \mu^2}{2\sigma^2} \right\} \\ &= \exp \left\{ \left[-\frac{1}{2} \log(2\pi\sigma^2) - \frac{y^2}{2\sigma^2} \right] + \frac{y\mu - \frac{\mu^2}{2}}{\sigma^2} \right\} \\ &= \exp \left\{ c(y, \phi) + \frac{y\theta - b(\theta)}{\alpha(\phi)} \right\}, \end{aligned} \quad (3.26)$$

καταλήγουμε στην επιθυμητή μορφή (3.7), όπου

$$\theta = \mu, \quad b(\theta) = \frac{\theta^2}{2}, \quad \phi = \sigma^2, \quad a(\phi) = \phi, \quad c(y, \phi) = -\frac{1}{2} \log(2\pi\phi) - \frac{y^2}{2\phi}, \quad \text{και}$$

$$D = \frac{1}{\sigma^2} \sum_{i=1}^n (y_i - \hat{\mu}_i).$$

Η μέση τιμή και η διακύμανση δίνονται από τους τύπους (3.8) και (3.6)

$$\begin{aligned} E(Y) &= \theta = \mu, \\ \text{Var}(Y) &= \phi = \sigma^2. \end{aligned}$$

Διωνυμική κατανομή

Έστω ότι $Y \sim Bi(n, p)$ όπου $n \in \mathbb{N} = \{1, 2, \dots\}$ είναι οι δοκιμές και p η πιθανότητα επιτυχίας της κάθε δοκιμής, με συνάρτηση πιθανότητας

$$f_y(n; p) = \binom{n}{y} p^y (1-p)^{n-y}, \quad (3.27)$$

όπου $y = 0, 1, 2, \dots, n$ και $p \in [0, 1]$, και μπορεί να γραφεί στην μορφή

$$\begin{aligned}
 f_Y(y; p) &= \binom{n}{y} p^y (1-p)^{n-y} \\
 &= \exp \left\{ \log \binom{n}{y} + y \log p + (n-y) \log(1-p) \right\} \\
 &= \exp \left\{ \log \binom{n}{y} + y \log \frac{p}{1-p} + n \log(1-p) \right\} \\
 &= \exp \{c(y) + y\theta - b(\theta)\},
 \end{aligned} \tag{3.28}$$

καταλήγουμε στην επιθυμητή μορφή (3.7), όπου

$$\theta = \log \frac{p}{1-p}, \quad b(\theta) = -n \log(1-p), \quad c(y) = \log \binom{n}{y} \quad \text{και}$$

$$D = 2 \sum_{i=1}^n n_i \left\{ y_i \ln \left(\frac{y_i}{\mu_i} \right) + (n_i - y_i) \ln \left(\frac{n_i - y_i}{1 - \mu_i} \right) \right\}.$$

Η μέση τιμή και η διακύμανση δίνονται από τους τύπους (3.8) και (3.6)

$$\begin{aligned}
 E(Y) &= \frac{ne^\theta}{1+e^\theta} = np, \\
 \text{Var}(Y) &= np(1-p).
 \end{aligned}$$

Κατανομή Poisson

Έστω ότι $Y \sim \text{Poisson}(\lambda)$ με παράμετρο $\lambda > 0$ με συνάρτηση πιθανότητας

$$\begin{aligned}
 f_Y(y; \lambda) &= e^{-\lambda} \frac{\lambda^y}{y!}, \quad y = 0, 1, 2, \dots \\
 &= \exp \{-\lambda + y \log \lambda - \log y!\} \\
 &= \exp \{y\theta - e^\theta - \log y!\} \\
 &= \exp \{y\theta - b(\theta) + c(y)\},
 \end{aligned} \tag{3.29}$$

καταλήγουμε στην επιθυμητή μορφή (3.7), όπου

$$\theta = \log \lambda, \quad b(\theta) = e^\theta, \quad c(y) = -\log y! \quad \text{και}$$

$$D = 2 \sum_{i=1}^n n_i \left\{ y_i \ln \left(\frac{y_i}{\mu_i} \right) (y_i - \hat{\mu}_i) \right\}.$$

Η μέση τιμή και η διακύμανση δίνονται από τους τύπους (3.8) και (3.6)

$$\begin{aligned}
 E(Y) &= e^\theta = \mu, \\
 \text{Var}(Y) &= e^\theta = \mu.
 \end{aligned}$$

Αρνητική διωνυμική κατανομή

Έστω ότι $Y \sim \text{NB}(a, p)$ με παράμετρο $a > 0$ και $0 < p < 1$ η πιθανότητα επιτυχίας.
Η συνάρτηση πιθανότητας

$$\begin{aligned}
 f_Y(y; a, p) &= \frac{\Gamma(a+y)}{\Gamma(a)y!} (1-p)^a p^y, \quad y = 0, 1, 2, \dots \\
 &= \exp \left\{ \log \frac{\Gamma(a+y)}{\Gamma(a)y!} + a \log(1-p) + y \log p \right\} \\
 &= \exp \left\{ \log \frac{\Gamma(a+y)}{\Gamma(a)y!} + \frac{ya^{-1} \log p - (-\log(1-p))}{a^{-1}} \right\} \\
 &= \exp \left\{ c(y, \phi) + \frac{y\theta - b(\theta)}{\alpha(\phi)} \right\},
 \end{aligned} \tag{3.30}$$

καταλήγουμε στην επιθυμητή μορφή (3.7), όπου

$$\theta = a^{-1} \log p, \quad b(\theta) = -\log(1-p) = -\log(1-e^{a\theta}), \quad \phi = a^{-1}, \quad \alpha(\phi) = \phi, \quad c(y) = \log \frac{\Gamma(\phi^{-1} + y)}{\Gamma(\phi^{-1})y!} \quad \text{και}$$

$$D = 2 \sum_{i=1}^n n_i \left\{ y_i \ln \left(\frac{y_i}{\mu_i} \right) + \left(y_i + \frac{1}{\kappa} \right) \ln \left(\frac{y_i + 1/\kappa}{\mu_i + 1/\kappa} \right) \right\}.$$

Η μέση τιμή και η διακύμανση δίνονται απο τους τύπους (3.8) και (3.6)

$$\begin{aligned}
 E(Y) &= \frac{ae^{a\theta}}{1-e^{a\theta}} = \frac{ap}{1-p}, \\
 \text{Var}(Y) &= \frac{a^2 e^{a\theta}}{(1-e^{a\theta})^2} \frac{1}{a} = \frac{ap}{(1-p)^2}.
 \end{aligned}$$

Κατανομή γάμμα

Έστω ότι $Y \sim \mathbb{G}(\mu, a)$ με παραμέτρους $a, \mu > 0$. Η συνάρτηση πιθανότητας

$$\begin{aligned}
 f_Y(y; a, \mu) &= \frac{y^{a-1} e^{-\frac{ay}{\mu}}}{\Gamma(a) \left(\frac{\mu}{a}\right)^a}, \quad y > 0 \\
 &= \exp \left\{ -\log \Gamma(a) - a \log \mu + a \log a + (a-1) \log y - \frac{ay}{\mu} \right\} \\
 &= \exp \left\{ [(a-1) \log y - \log \Gamma(a + a \log a)] + \frac{y \left(-\frac{1}{\mu} - \log \mu \right)}{a^{-1}} \right\} \\
 &= \exp \left\{ c(y, \phi) + \frac{y\theta - b(\theta)}{\alpha(\phi)} \right\},
 \end{aligned} \tag{3.31}$$

καταλήγουμε στην επιθυμητή μορφή (3.7), όπου

$$\theta = -\frac{1}{\mu}, \quad b(\theta) = \log \mu = -\log(-\theta), \quad \phi = a^{-1}, \quad \alpha(\phi) = \phi,$$

$$c(y, \phi) = (\phi^{-1} - 1) \log y - \log \Gamma(\phi^{-1}) - \phi^{-1} \log \phi$$

$$\text{και } D = 2\nu \sum_{i=1}^n \left\{ -\ln\left(\frac{y_i}{\hat{\mu}_i}\right) + \frac{y_i - \hat{\mu}_i}{\hat{\mu}_i} \right\}.$$

Η μέση τιμή και η διακύμανση δίνονται απο τους τύπους (3.8) και (3.6)

$$E(Y) = -\frac{1}{\theta} = \mu,$$

$$Var(Y) = a^{-1} \frac{1}{\theta^2} = \frac{\mu^2}{a}.$$

Inverse Gaussian

Έστω ότι $Y \sim \text{IG}(\mu, \lambda)$. Η συνάρτηση πιθανότητας

$$\begin{aligned} f(y) &= \sqrt{\frac{\lambda}{2\pi y^3}} \exp\left\{-\frac{\lambda(y-\mu)^2}{2\mu^2 y}\right\}, \quad y > 0, \quad \lambda, \mu > 0, \\ &= \exp\left\{-\frac{1}{2} \ln(2\pi y^3) - \log\sqrt{\lambda} - \frac{1}{2y} \left(\frac{y-\mu}{\mu\sqrt{\lambda}}\right)^2\right\} \\ &= \exp\left\{-\frac{y}{2\mu^2\lambda} + \frac{1}{\mu\lambda} - \frac{1}{2} \ln(2\pi y^3) - \log\sqrt{\lambda}\right\} \\ &= \exp\left\{c(y, \phi) + \frac{y\theta - b(\theta)}{a(\phi)} + \text{όροι που περιέχουν μόνο } y \text{ και } \lambda\right\}, \end{aligned} \tag{3.32}$$

καταλήγουμε στην επιθυμητή μορφή (3.7), όπου

$$\theta = -\frac{1}{2\mu^2}, \quad b(\theta) = -\sqrt{-2\theta} = -\log(1 - e^{a\theta}), \quad \phi = \lambda$$

$$\text{και } D = \frac{1}{\lambda} \sum_{i=1}^n \left\{ \frac{(y_i - \hat{\mu}_i)^2}{\hat{\mu}_i^2 y_i} \right\}.$$

Η μέση τιμή και η διακύμανση δίνονται απο τους τύπους (3.8) και (3.6)

$$E(Y) = \frac{1}{\sqrt{-2\theta}} = \mu,$$

$$Var(Y) = \frac{\lambda}{(-2\theta)^{\frac{3}{2}}} = \lambda\mu^3.$$

Κεφάλαιο 4

Αποθεματοποίηση με χρήση των γενικευμένων γραμμικών μοντέλων

Σε αυτό το κεφάλαιο θα δούμε την εφαρμογή των γενικευμένων γραμμικών μοντέλων στην αποθεματοποίηση ζημιών. Με την χρήση κατανομών που ανήκουν στην εκθετική οικογένεια και σε συνδιασμό με την κατάλληλη συνάρτηση σύνδεσης, μπορούμε να παράγουμε ένα ικανοποιητικό αριθμό μοντέλων. Σκοπός αυτών των μοντέλων σε συνδιασμό με τα τρίγωνα εξέλιξης είναι να εκτιμήσουμε τις μελλοντικές απαιτήσεις ενός χαρτοφυλακίου ζημιών μιας ασφαλιστικής εταιρείας.

Μοντελοποίηση απαιτήσεων

Αν η εξαρτημένη μεταβλητή εκφράζει τις απαιτήσεις που εμφανίζονται τότε θέτουμε $X_{i,j}$, $0 \leq i, j \leq I = J$ τις προσαυξητικές αιτήσεις της ζημίας που συνέβη το

έτος i για το έτος εξέλιξης j , και $C_{i,j} = \sum_{k=0}^j X_{i,k}$ είναι οι αθροιστικές απαιτήσεις

για το έτος ατυχήματος i και έτος εξέλιξης j . Σε αυτή την περίπτωση οι κατανομές που χρησιμοποιούνται ευρέως είναι η κατανομή Poisson και η αρνητική διωνυμική. Αυτές οι κατανομές μπορούν να χρησιμοποιηθούν, όπως παρουσιάσαν στο σύγγραμμά τους οι England and Verrall [4].

4.1 Υπερσκεδαστικό μοντέλο Poisson

Το υπερσκεδαστικό μοντέλο Poisson είναι ένα στατιστικό μοντέλο το οποίο έχει αρκετά ικανοποιητική εφαρμογή στην αποθεματοποίηση. Αποτελεί μια γενίκευση όπως θα λέγαμε και πιο σαφή ερμηνεία του μοντέλου Poisson. Το υπερσκεδαστικό μοντέλο Poisson είχε προταθεί από τους Renshaw and Verrall το 1998 [12]. Το μοντέλο αυτό παρέχει μια εκτίμηση για το μέσο απόθεμα

και για την διακύμανση γύρω από το απόθεμα, αλλά σε καμία περίπτωση δεν δίνει ακριβείς πληροφορίες για την κατανομή του αποθέματος. Το μοντέλο παρασκευάζεται από τα γενικευμένα γραμμικά μοντέλα που όπως αναφέρεται και πιο πάνω είναι μια οικογένεια από μοντέλα και αποτελεί μια ευκολοπροσάρμοστη γενίκευση της γραμμικής παλινδρόμησης. Τα γενικευμένα γραμμικά μοντέλα ταιριάζουν απόλυτα όπως θα λέγαμε σε δεδομένα τριγώνων αποθεματοποίησης με την χρήση στατιστικών πακέτων, τέτοια παραδείγματα θα παρουσιαστούν στο επόμενο κεφάλαιο. Παρόλο τους περιορισμούς που έχουν τα γενικευμένα γραμμικά μοντέλα, ανάλογα με την φύση των δεδομένων, έχουν μια καλή εφαρμογή στην αποθεματοποίηση.

Μαθηματική σύνθεση του υπερσχεδαστικού μοντέλου Poisson

Έστω η συνάρτηση πιθανότητας της κατανομής Poisson

$$f(y) = \frac{\mu^y}{y!} e^{-\mu}, \quad y = 0, 1, 2, \dots, \mu > 0.$$

Προκειμένου να εργαστούμε με την μέθοδο αυτή κάνουμε τις κάτωθι υποθέσεις.

Βασικές υποθέσεις

Στο υπερσχεδαστικό μοντέλο Poisson υποθέτουμε ότι όλες οι τυχαίες μεταβλητές είναι ανεξάρτητες μεταξύ τους και με μια ελεύθερη παράμετρο κλίμακας ϕ η οποία χρησιμοποιείται στην διακύμανση.

$$\begin{aligned} E(X_{i,j}) &= \mu_{i,j} = x_i y_i, \\ \text{Var}(X_{i,j}) &= \phi \mu_{i,j} = \phi E(X_{i,j}) = \phi x_i y_i. \end{aligned} \quad (4.1)$$

Το x_i εδώ παριστάνει την αναμενόμενη απόλυτη όπως θα λέγαμε απαίτηση στο τελευταίο αναπτυσσόμενο σημείο του τριγώνου ενώ το y_i παριστάνει την προτεινόμενη απόλυτη απαίτηση η οποία αναμένεται να εμφανιστεί στην αναπτυσσόμενη περίοδο j .

Στο υπερσχεδαστικό μοντέλο Poisson όπως και στο απλό γραμμικό μοντέλο πρέπει να χρησιμοποιήσουμε λογάριθμο. Έτσι με την βοήθεια της λογαριθμικής συνάρτησης σύνδεσης, η γραμμική συνάρτηση πρόβλεψης $\eta_{i,j}$ συνδέεται με την μέση τιμή $\mu_{i,j}$ ως κάτωθι

$$\begin{aligned} \log \mu_{i,j} &= n_{i,j}, \\ n_{i,j} &= c + \alpha_i + \beta_j, \end{aligned}$$

όπου $\alpha_0 = \beta_0 = 0$.

Οι παράμετροι α_i, β_j προσδιορίζουν την επίδραση του i έτους ζημίας και j έτους εξέλιξης στην αναμενόμενη τιμή των προσαυξητικών απαιτήσεων $X_{i,j}$, $\forall i, j$.

Με την βοήθεια πινάκων η σχέση (2) γίνεται

$$\eta_{i,j} = \mathbf{\Gamma}_{i,j}^T \boldsymbol{\beta}, \quad (4.2)$$

όπου $\mathbf{\Gamma}_{i,j} = (1, \delta_{1,i}, \dots, \delta_{I,i}, \delta_{1,j}, \dots, \delta_{I,j})^T$ και

$$\delta_{i,j} = \begin{cases} 0, & i \neq j \\ 1, & i = j, \end{cases} \quad (4.3)$$

Ο πίνακας β των παραμέτρων έχει την μορφή

$$\beta = (c, \alpha_1, \dots, \alpha_I, \beta_1, \dots, \beta_I)^T.$$

Με την βοήθεια μαθηματικών προγραμματιστικών πακέτων υπολογίζεται εύκολα η εκτίμηση $\hat{\beta}$ των συντελεστών και μέσω αυτής υπολογίζουμε την εκτίμηση $\hat{\mu}_{i,j}$ της μέσης τιμής των $X_{i,j}$, $\forall i, j$. Οι εκτιμήσεις που έχουμε υπολογίσει ως τώρα θα μας βοηθήσουν να βρούμε μια εκτίμηση για την παράμετρο ϕ . Αυτό επιτυγχάνεται είτε μέσω του Pearson χ^2 test είτε αν διαιρέσουν την διακύμανση του μοντέλου με την διαφορά του αριθμού των παρατηρήσεων, έστω $n = \frac{(I+1)(I+2)}{2}$, με τον αριθμό των παραμέτρων του μοντέλου, έστω $p = 2I + 1$. Με την χρήση του Pearson χ^2 test οδηγούμαστε στον κάτωθι τύπο

$$\hat{\phi} = \frac{1}{n-p} \sum_{i=0}^I \sum_{j=0}^{I-i} \frac{(X_{i,j} - \hat{\mu}_{i,j})^2}{Var(\hat{\mu}_{i,j})}, \quad (4.4)$$

όπου $n - p = \frac{I(I-1)}{2}$.

Έχοντας ορίσει το μοντέλο μας θα πρέπει να δούμε πόσο καλά εφαρμόζεται στα δεδομένα μας και αρκεί να υπολογίσουμε την διακύμανση του μοντέλου. Στην δικιά μας περίπτωση, όπου δουλεύουμε με την υπερσκεδαστική κατανομή Poisson, θα χρησιμοποιήσουμε τον κάτωθι τύπο

$$D = \frac{2}{\hat{\phi}} \sum_{i=0}^I \sum_{j=0}^{I-i} \left[X_{i,j} \log \left(\frac{X_{i,j}}{\hat{\mu}_{i,j}} \right) - (X_{i,j} - \hat{\mu}_{i,j}) \right]. \quad (4.5)$$

Η διακύμανση ακολουθεί ασυμπτωτικά την $\chi_{\frac{I(I-1)}{2}}^2$ κατανομή και αν παίρνει τιμές απο την δεξιά ουρά τότε συμπεραίνουμε ότι το μοντέλο μας δεν εφαρμόζεται καλά στα δεδομένα [3].

Προκειμένου η όλη διαδικασία να γίνει ευκολότερα κατανοητή θα δούμε μια εφαρμογή αυτής.

4.1.1 Εφαρμογή της διαδικασίας

Στο μοντέλο της υπερσκεδαστικής Poisson (ODP), οι προσαυξητικές απαιτήσεις X_{ij} , σε κάθε κελί του τριγώνου θεωρούμαι ότι κατανέμονται ως ανεξάρτητες ODP

μεταβλητές με μέση τιμή και διακύμανση όπως ορίσαμε παραπάνω.

Ο λόγος που ορίσαμε ότι $a_0 = \beta_0 = 0$ είναι προκειμένου να αποφευχθεί η υπερπαραμετροποίηση. Τα a_i αποτελούν παράγοντες που σχετίζονται με τις γραμμές του τριγώνου και τα β_j είναι παράγοντες που αφορούν τις στήλες του τριγώνου. Τέλος, με c ορίζουμε μια σταθερά που εφαρμόζεται σε όλα τα κελία [6]. Αυτό γίνεται ευκολότερα κατανοητό αν δούμε την δομή της διαδικασίας μοντελοποίησης σε ένα απλό 3×3 τρίγωνο.

Πίνακας 4.1: Το Μοντέλο Υπερκεδαστικής Poisson

Λογάριθμοι Πληρωτέων Ζημιών ανά Έτος Εξέλιξης			
Έτος Ατυχήματος	0	1	2
0	c	$c + \beta_1$	$c + \beta_2$
1	$c + a_1$	$c + a_1 + \beta_1$	
2	$c + a_2$		

Αφού υπολογίσουμε τις τιμές των c, a_i, β_i , επειδή χρησιμοποιήσαμε την λογαριθμική συνάρτηση σύνδεσης, μπορούμε να υπολογίσουμε τις προσαυξητικές απαιτήσεις μέσω της $e^{n_i, j} = e^{c+a_i+\beta_j}$.

Παρόλο που το ODP μοντέλο μπορεί να εφαρμοστεί σε ένα τρίγωνο εξέλιξης με μικρό αριθμό αρνητικών προσαυξητικών απαιτήσεων, είναι απαραίτητο το άθροισμα των στοιχείων κάθε στήλης και κάθε γραμμής να είναι θετικό. Επίσης, οι απαιτήσεις που θα υπολογιστούν μέσω αυτού του μοντέλου θα είναι πάντα θετικές. Γίνεται έτσι αντιληπτό ότι το μοντέλο της υπερκεδαστικής Poisson δεν λειτουργεί για οποιαδήποτε βάση δεδομένων.

Η εφαρμογή του μοντέλου στα δεδομένα προκειμένου να εκτιμήσουμε το τελικό απόθεμα που απαιτείται καθώς και το σχετικό σφάλμα της εκτίμησης αυτής περιέχει τέσσερα βασικά στάδια

1. Εκτίμηση των παραμέτρων μέσω της βασικής διαδικασίας εφαρμογής των γενικευμένων γραμμικών μοντέλων και μετέπειτα τον έλεγχο καλής προσαρμογής.
2. Με την βοήθεια αυτών των παραμέτρων εκτιμούμε τις μελλοντικές απαιτήσεις και το τελικό απόθεμα.
3. Υπολογίζουμε τα σφάλματα εκτίμησης για κάθε ένα μελλοντικό κελί του τριγώνου.
4. Υπολογίζουμε το σφάλμα εκτίμησης του αποθέματος για κάθε έτος ζημίας καθώς και για το συνολικό.

Τα στάδια της διαδικασίας θα γίνουν εύκολα κατανοητά με ένα απλό παράδειγμα εφαρμογής του μοντέλου σε ένα τρίγωνο 3×3 .

Βήμα 1. Εφαρμογή του ODP μοντέλου

Θα χρησιμοποιήσουμε το κάτωθι τρίγωνο εξέλιξης επαυξητικών ζημιών 3 ετών για 3 χρόνια εξέλιξης.

Προσαυξητικές Απαιτήσεις ανά Έτος Εξέλιξης			
Έτος Ατυχήματος	0	1	2
0	$X_{0,0}$	$X_{0,1}$	$X_{0,2}$
1	$X_{1,0}$	$X_{1,1}$	
2	$X_{2,0}$		

Το μοντέλο μας τροποποιείται σε μορφή πινάκων μέσω της σχέσης

$$\log(\mathbf{E}[\mathbf{x}]) = \mathbf{\Gamma}\boldsymbol{\beta}, \quad (4.6)$$

όπου $\mathbf{\Gamma}$ ορίζουμε τον πίνακα σχεδιασμού. Έτσι έχουμε

$$\mathbf{X} = \begin{bmatrix} X_{0,0} \\ X_{1,0} \\ X_{2,0} \\ X_{0,1} \\ X_{1,1} \\ X_{0,2} \end{bmatrix}, \quad \mathbf{\Gamma} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 \\ 1 & 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 1 \end{bmatrix}, \quad \boldsymbol{\beta} = \begin{bmatrix} c \\ a_1 \\ a_2 \\ \beta_1 \\ \beta_2 \end{bmatrix}. \quad (4.7)$$

Προκειμένου να προχωρήσουμε στην διαδικασία εκτίμησης των παραμέτρων $\hat{\boldsymbol{\beta}}$ ορίζουμε τον πίνακα βαρών \mathbf{W} . Ο πίνακας αυτός στην περίπτωση που εργαζόμαστε με την κατανομή Poisson είναι ένας διαγώνιος πίνακας με στοιχεία τις εκτιμήσεις των απαιτήσεων ως κάτωθι

$$\mathbf{W} = \begin{bmatrix} e^c & 0 & 0 & 0 & 0 & 0 \\ 0 & e^{c+a_1} & 0 & 0 & 0 & 0 \\ 0 & 0 & e^{c+a_2} & 0 & 0 & 0 \\ 0 & 0 & 0 & e^{c+\beta_1} & 0 & 0 \\ 0 & 0 & 0 & 0 & e^{c+a_1+\beta_1} & 0 \\ 0 & 0 & 0 & 0 & 0 & e^{c+\beta_2} \end{bmatrix}. \quad (4.8)$$

Με την βοήθεια του (3.20), όπου στην συγκεκριμένη περίπτωση ορίσαμε σαν πίνακα σχεδιασμού τον $\mathbf{\Gamma}$ και μέσω επαναληπτικής διαδικασίας θα υπολογίσουμε

το διάνυσμα $\hat{\beta}$. Θα πρέπει να σημειωθεί ότι για τον υπολογισμό του z με την βοήθεια της (3.19) είναι

$$z = \Gamma\beta + \frac{X - e^{\Gamma\beta}}{e^{\Gamma\beta}}. \quad (4.9)$$

Αν ορίσουμε σαν $W^{(0)}$ και $\beta^{(0)}$ τους αρχικούς πίνακες της επαναληπτικής διαδικασίας για τον υπολογισμό των παραμέτρων $\hat{\beta}$ τότε για το παράδειγμά μας θα έχουν την κάτωθι μορφή

$$W^{(0)} = \begin{bmatrix} X_{0,0} & 0 & 0 & 0 & 0 & 0 \\ 0 & X_{1,0} & 0 & 0 & 0 & 0 \\ 0 & 0 & X_{2,0} & 0 & 0 & 0 \\ 0 & 0 & 0 & X_{0,1} & 0 & 0 \\ 0 & 0 & 0 & 0 & X_{1,1} & 0 \\ 0 & 0 & 0 & 0 & 0 & X_{0,2} \end{bmatrix}, \beta^{(0)} = \begin{bmatrix} \log(X_{0,0}) \\ \log(X_{1,0}) \\ \log(X_{2,0}) \\ \log(X_{0,1}) \\ \log(X_{1,1}) \\ \log(X_{0,2}) \end{bmatrix}, \quad (4.10)$$

και αντικαθιστώντας στον (4.9) έχουμε $z^{(0)} = \beta^{(0)}$. Αντικαθιστώντας τι τιμές που θα βρούμε στον τύπο (3.20) για την δεύτερη επανάληψη και θα βρούμε την τιμή του $\beta^{(1)}$. Η διαδικασία θα σταματήσει όταν ικανοποιηθεί η σχέση (3.21), όταν δηλαδή πετύχουμε την ελάχιστη απόσταση μεταξύ του $\beta^{(m)}$ από το προηγούμενο του, έστω $\beta^{(m-1)}$.

Βήμα 2. Υπολογισμός μελλοντικών απαιτήσεων και του τελικού αποθέματος

Έχοντας υπολογίσει τον πίνακα $\hat{\beta}$ και αντικαθιστώντας στον

$$\hat{X}_{i,j} = \hat{\mu}_{i,j} = e^{\hat{c} + \hat{a}_i + \hat{\beta}_j} \quad (4.11)$$

συμπληρώνουμε ολόκληρο τον πίνακα απαιτήσεων.

Οι εκτιμήσεις του συνολικού αποθέματος κάθε έτους εξέλιξης υπολογίζονται με την βοήθεια του τύπου (2.2), δηλαδή

$$\hat{R}_i = \sum_{k=I-i+1}^J \hat{X}_{i,k},$$

και το συνολικό απόθεμα $\hat{R} = \sum_{i=1}^I \hat{R}_i$.

Βήμα 3. Σφάλμα πρόβλεψης των μελλοντικών επαυξητικών απαιτήσεων

Στα προηγούμενα βήματα εκτιμήσαμε τις τιμές των μελλοντικών κελιών του τριγώνου. Το επόμενο στάδιο είναι να εκτιμήσουμε τα σφάλματα αυτών των

προβλέψεων. Για μια συγκεκριμένη μελλοντική επαυξητική απαίτηση $X_{i,j}$ ο τύπος (2.29) γίνεται

$$MSEP[\hat{X}_{i,j}] = E \left[\left(X_{i,j} - \hat{X}_{i,j} \right)^2 \right]. \quad (4.12)$$

Μπορούμε να πούμε ότι

$$\left(X_{i,j} - \hat{X}_{i,j} \right) = \left(X_{i,j} - E[X_{i,j}] \right) - \left(\hat{X}_{i,j} - E[X_{i,j}] \right),$$

επομένως ο (4.12) γίνεται

$$\begin{aligned} MSEP[\hat{X}_{i,j}] = & E \left[\left(X_{i,j} - E[X_{i,j}] \right)^2 \right] - E \left[\left(\hat{X}_{i,j} - E[X_{i,j}] \right)^2 \right] \\ & - 2E \left[\left(X_{i,j} - E[X_{i,j}] \right) \left(\hat{X}_{i,j} - E[X_{i,j}] \right) \right]. \end{aligned} \quad (4.13)$$

Επειδή έχουμε υποθέσει ότι οι μελλοντικές απαιτήσεις είναι ανεξάρτητες των προγενέστερων απαιτήσεων και επειδή το $\hat{X}_{i,j}$ βασίζεται μόνο στις προγενέστερες απαιτήσεις ο όρος $-2E \left[\left(X_{i,j} - E[X_{i,j}] \right) \left(\hat{X}_{i,j} - E[X_{i,j}] \right) \right] = 0$.

Σημειώνεται ότι $E \left[\left(X_{i,j} - E[X_{i,j}] \right)^2 \right] = Var[X_{i,j}]$.

Ο Renshaw [11] απέδειξε ότι $E \left[\left(\hat{X}_{i,j} - E[X_{i,j}] \right)^2 \right] \approx Var[\hat{\eta}_{i,j}] \mu_{i,j}^2$. Έτσι η σχέση (4.12) γίνεται

$$MSEP[\hat{X}_{i,j}] \approx \phi \mu_{i,j} + Var[\hat{\eta}_{i,j}] \mu_{i,j}^2, \quad (4.14)$$

ή

$$MSEP[\hat{X}_{i,j}] = Var[X_{i,j}] + Var[\hat{X}_{i,j}]. \quad (4.15)$$

Αν στην σχέση (4.14) αντικαταστήσουμε τις $\phi, \mu_{i,j}$ με τις εκτιμήσεις αυτών θα λάβουμε τον τύπο για την εκτίμηση του μέσου τετραγωνικού σφάλματος

$$\widehat{MSEP}[\hat{X}_{i,j}] \approx \hat{\phi} \hat{\mu}_{i,j} + Var[\hat{\eta}_{i,j}] \hat{\mu}_{i,j}^2. \quad (4.16)$$

Για τον υπολογισμό του $\hat{\phi}$ θα χρησιμοποιήσουμε τον τύπο (4.4). Έχοντας ορίσει τον τρόπο υπολογισμού του $MSEP$ για μια μεμονωμένη επαυξητική τιμή θα πρέπει να δείξουμε και πώς υπολογίζεται για το συνολικό απόθεμα που θα εκτιμήσουμε. Αυτό αποτελεί και το τελευταίο στάδιο της διαδικασίας.

Βήμα 4. Υπολογισμός του σφάλματος πρόβλεψης των αποθεμάτων

Αυτό το στάδιο της διαδικασίας αποτελεί ένα βήμα επιπλέον από τις τυπικές εφαρμογές γενικευμένων γραμμικών μοντέλων. Σκοπός είναι να υπολογίσουμε τις τιμές του $MSEP(\hat{R}_i)$ και του $MSEP(\hat{R})$. Στην βιβλιογραφία, όπως για παράδειγμα

στο σύγγραμμα των England and Verall [4] καθώς και σε αυτό του Renshaw [11], πολλές φορές θα βρούμε τις κάτωθι σχέσεις

$$\begin{aligned} MSEP[\hat{R}_i] &\approx \sum_{j \in \Delta_i} \hat{\phi} \hat{\mu}_{i,j} + \sum_{j \in \Delta_i} \hat{\mu}^2 Var[\hat{\eta}_{i,j}] \\ &+ 2 \sum_{\substack{j_1, j_2 \in \Delta_i, \\ j_2 > j_1}} \hat{\mu}_{i,j_1} \hat{\mu}_{i,j_2} Cov[\hat{\eta}_{1,j_1}, \hat{\eta}_{i,j_2}], \end{aligned} \quad (4.17)$$

όπου Δ_i αντιπροσωπεύει τα κελιά της γραμμής i που αφορούν το μέλλον (αναφερόμαστε δηλαδή στο κάτω δεξιά τρίγωνο). Κατ' επέκταση σαν Δ ορίζουμε όλα τα κελιά του κάτω δεξιά τριγώνου και έχουμε

$$\begin{aligned} MSEP[\hat{R}] &\approx \sum_{i,j \in \Delta} \hat{\phi} \hat{\mu}_{i,j} + \sum_{i,j \in \Delta} \hat{\mu}_{i,j} Var[\hat{\eta}_{i,j}] \\ &+ 2 \sum_{\substack{i_1, j_1 \in \Delta, \\ i_2, j_2 \in \Delta, i_1, j_1 \neq i_2, j_2}} \hat{\mu}_{i_1, j_1} \hat{\mu}_{i_2, j_2} Con[\hat{\eta}_{i_1, j_1}, \hat{\eta}_{i_2, j_2}]. \end{aligned} \quad (4.18)$$

Το πρώτο άθροισμα και στις δύο σχέσεις αντιπροσωπεύουν την διακύμανση της διαδικασίας και τα επόμενα δύο αθροίσματα αντιπροσωπεύουν την διακύμανση της εκτίμησης. Δεν είναι πολύ εύκολο να εφαρμοστούν στην πράξη διότι θέλει ιδιαίτερη προσοχή στις τιμές των δεικτών των αθροισμάτων. Ευτύχως είναι πιο εύκολο στην κατανόηση και την εφαρμογή στην πράξη αν οι δύο αυτές σχέσεις εκφραστούν με σχέσεις πινάκων και πιο συγκεκριμένα στον πίνακα 3×3 που μελετάμε.

Αρχικά ορίζουμε τον τύπο για την εκτίμηση του πίνακα διακύμανσης/συνδιακύμανσης (variance/covariance) των παραμέτρων του μοντέλου

$$\hat{\Sigma} = \hat{\phi} (\mathbf{\Gamma}^T \mathbf{W} \mathbf{\Gamma})^{-1}. \quad (4.19)$$

Για τον πίνακα 3×3 που μελετάμε, ο οποίος έχει πέντε παραμέτρους, ο $\hat{\Sigma}$ είναι ένας 5×5 πίνακας όπου τα στοιχεία της διαγωνίου του αντιπροσωπεύουν την εκτιμώμενη διακύμανση των παραμέτρων.

Για τον υπολογισμό του $MSEP$ είδαμε ότι μας ενδιαφέρουν τα κελιά του κάτω δεξιά τριγώνου, επομένως είναι σημαντικό να ορίσουμε και εμείς τον πίνακα σχεδιασμού για τα κελιά αυτά. Έστω

$$\mathbf{\Gamma}_F = \begin{bmatrix} 1 & 1 & 0 & 0 & 1 \\ 1 & 0 & 1 & 1 & 0 \\ 1 & 0 & 1 & 0 & 1 \end{bmatrix}. \quad (4.20)$$

Οι εξισώσεις που αναφέραμε για το $MSEP$ περιέχουν τις τιμές της διακύμανσης και συνδιακύμανσης των συναρτήσεων σύνδεσης $\hat{\eta}_{i,j}$. Για να υπολογίσουμε τις τιμές αυτές, με την βοήθεια πινάκων, ορίζουμε τον 3×3 πίνακα \mathbf{N} ως

$$\mathbf{N} = \mathbf{\Gamma}_F \hat{\Sigma} (\mathbf{\Gamma}_F)^T = \begin{bmatrix} Cov(\hat{\eta}_{1,2}, \hat{\eta}_{1,2}) & Cov(\hat{\eta}_{2,1}, \hat{\eta}_{1,2}) & Cov(\hat{\eta}_{2,2}, \hat{\eta}_{1,2}) \\ Cov(\hat{\eta}_{1,2}, \hat{\eta}_{2,1}) & Cov(\hat{\eta}_{2,1}, \hat{\eta}_{2,1}) & Cov(\hat{\eta}_{2,2}, \hat{\eta}_{2,1}) \\ Cov(\hat{\eta}_{1,2}, \hat{\eta}_{2,2}) & Cov(\hat{\eta}_{2,1}, \hat{\eta}_{2,2}) & Cov(\hat{\eta}_{2,2}, \hat{\eta}_{2,2}) \end{bmatrix}. \quad (4.21)$$

Στις σχέσεις (4.17) και (4.18) στο δεύτερο και τρίτο άθροισμα υπάρχει το γινόμενο της $\hat{\mu}_{i,j}$ με το $Var[\hat{\eta}_{i,j}]$ και το $Cov[\hat{\eta}_{i,j1}, \hat{\eta}_{i,j2}]$. Συνεπώς ορίζουμε

$$diag(\hat{\mu})Ndiag(\hat{\mu}), \quad (4.22)$$

όπου

$$diag(\hat{\mu}) = \begin{bmatrix} \hat{\mu}_{1,2} & 0 & 0 \\ 0 & \hat{\mu}_{2,1} & 0 \\ 0 & 0 & \hat{\mu}_{2,2} \end{bmatrix} = \begin{bmatrix} e^{\hat{c}+\hat{a}_1+\hat{\beta}_2} & 0 & 0 \\ 0 & e^{\hat{c}+\hat{a}_2+\hat{\beta}_1} & 0 \\ 0 & 0 & e^{\hat{c}+\hat{a}_2+\hat{\beta}_2} \end{bmatrix}. \quad (4.23)$$

Αν στην (4.22) προσθέσουμε και τον διαγώνιο πίνακα $\hat{\phi}diag(\hat{\mu})$, ο οποίος εκφράζει το πρώτο άθροισμα στις (4.17) και (4.18), τότε είμαστε σε θέση να υπολογίσουμε την εκτίμηση του μέσου τετραγωνικού σφάλματος [6].

Η παραπάνω διαδικασία θα γίνει πιο εύκολα αντιληπτή με την εφαρμογή της σε ένα απλό παράδειγμα με γνωστές τις προσαυξητικές απαιτήσεις.

4.1.2 Εφαρμογή του ODP μοντέλου.

Έστω το τρίγωνο εξέλιξης προσαυξητικών απαιτήσεων.

Προσαυξητικές Απαιτήσεις ανά Έτος Εξέλιξης			
Έτος Ατυχήματος	0	1	2
0	420	150	70
1	300	120	
2	340		

Βήμα 1.

Αρχικά υπολογίζουμε τους λογάριθμους των απαιτήσεων αυτών επειδή σαν συνάρτηση σύνδεσης έχουμε ορίσει την λογαριθμική.

Λογάριθμοι Προσαυξητικών Απαιτήσεων ανά Έτος Εξέλιξης			
Έτος Ατυχήματος	0	1	2
0	6.040255	5.010635	4.248495
1	5.703782	4.787492	
2	5.828946		

Οι πίνακες που χρησιμοποιούμε στο πρώτο στάδιο της μεθόδου με την βοήθεια των (4.7) και (4.10) έχουμε

$$\mathbf{X} = \begin{bmatrix} 420 \\ 300 \\ 340 \\ 150 \\ 120 \\ 70 \end{bmatrix}, \quad \mathbf{W}^{(0)} = \begin{bmatrix} 420 & 0 & 0 & 0 & 0 & 0 \\ 0 & 300 & 0 & 0 & 0 & 0 \\ 0 & 0 & 340 & 0 & 0 & 0 \\ 0 & 0 & 0 & 150 & 0 & 0 \\ 0 & 0 & 0 & 0 & 120 & 0 \\ 0 & 0 & 0 & 0 & 0 & 70 \end{bmatrix}, \quad \boldsymbol{\beta}^{(0)} = \begin{bmatrix} 6.040255 \\ 5.703782 \\ 5.828946 \\ 5.010635 \\ 4.787492 \\ 4.248495 \end{bmatrix}.$$

Στην συνέχεια θα πρέπει να υπολογίσουμε τους όρους της (3.20). Συνεπώς,

$$\boldsymbol{\Gamma}^T \mathbf{W}^{(0)} \boldsymbol{\Gamma} = \begin{bmatrix} 1400 & 420 & 340 & 270 & 70 \\ 420 & 420 & 0 & 120 & 0 \\ 340 & 0 & 340 & 0 & 0 \\ 270 & 120 & 0 & 270 & 0 \\ 70 & 0 & 0 & 0 & 70 \end{bmatrix}$$

και ο αντίστροφος αυτού

$$\begin{aligned} & \left(\boldsymbol{\Gamma}^T \mathbf{W}^{(0)} \boldsymbol{\Gamma} \right)^{-1} = \\ = & \begin{bmatrix} 0.002107280 & -0.0017241379 & -0.002107280 & -0.0013409962 & -0.002107280 \\ -0.001724138 & 0.0041379310 & 0.001724138 & -0.0001149425 & 0.001724138 \\ -0.002107280 & 0.0017241379 & 0.005048456 & 0.0013409962 & 0.002107280 \\ -0.001340996 & -0.0001149425 & 0.001340996 & 0.0050957854 & 0.001340996 \\ -0.002107280 & 0.0017241379 & 0.002107280 & 0.0013409962 & 0.016392994 \end{bmatrix}. \end{aligned}$$

Επειδή όπως εξηγήσαμε $\mathbf{z}^{(0)} = \boldsymbol{\beta}^{(0)}$, έχουμε

$$\mathbf{W}^{(0)} \mathbf{z}^{(0)} = \begin{bmatrix} 2536.9070 \\ 1711.1347 \\ 1981.8415 \\ 751.5953 \\ 574.4990 \\ 297.3947 \end{bmatrix}.$$

Αντικαθιστώντας τους πίνακες που έχουμε υπολογίσει στην (3.20) παίρνουμε τον

$$\hat{\boldsymbol{\beta}}^{(1)} = \begin{bmatrix} 6.0272284 \\ 1711.1347 \\ -0.3052092 \\ -0.1982828 \\ -0.9801195 \\ -1.7787332 \end{bmatrix}.$$

Με την βοήθεια των $\hat{\boldsymbol{\beta}}^{(1)}$ παραμέτρων μέσω της (4.8) θα υπολογίσουμε τον $\mathbf{W}^{(1)}$ και μέσω της (4.9) θα υπολογίσουμε τον $\mathbf{z}^{(1)}$. Η επαναληπτική διαδικασία θα

συνεχιστεί μέχρι να καταλήξουμε στον $\hat{\beta}^{(4)}$ τις τελικές εκτιμημένες παραμέτρους

$$\hat{\beta}^{(4)} = \begin{bmatrix} \hat{c} = 6.02718 \\ \hat{a}_1 = -0.30538 \\ \hat{a}_2 = -0.19824 \\ \hat{\beta}_1 = -0.98083 \\ \hat{\beta}_2 = -1.77869 \end{bmatrix} \quad (4.24)$$

Θα είχαμε το ίδιο αποτέλεσμα αν χρησιμοποιούσαμε και την συνάρτηση `glmReserve` του πακέτου `ChainLadder` της *R*.

Βήμα 2.

Ο πίνακας με τις εκτιμώμενες επαυξητικές απαιτήσεις ανά έτος εξέλιξης, χρησιμοποιώντας την (4.11) θα έχει την κάτωθι μορφή.

Εκτιμώμενες Επαυξητικές Απαιτήσεις ανά Έτος Εξέλιξης			
Έτος Ατυχήματος	0	1	2
0	$e^{\hat{c}}$	$e^{\hat{c}+\hat{\beta}_1}$	$e^{\hat{c}+\hat{\beta}_2}$
1	$e^{\hat{c}+\hat{a}_1}$	$e^{\hat{c}+\hat{a}_1+\hat{\beta}_1}$	$e^{\hat{c}+\hat{a}_1+\hat{\beta}_2}$
2	$e^{\hat{c}+\hat{a}_2}$	$e^{\hat{c}+\hat{a}_2+\hat{\beta}_1}$	$e^{\hat{c}+\hat{a}_2+\hat{\beta}_2}$

Τώρα μπορούμε να συμπληρώσουμε ολόκληρο τον πίνακα των εκτιμώμενων επαυξητικών απαιτήσεων και να υπολογίσουμε τις εκτιμήσεις του συνολικού αποθέματος κάθε έτους εξέλιξης και συνολικά.

Εκτιμώμενες Επαυξητικές Απαιτήσεις ανά Έτος Εξέλιξης				Απόθεματα
Έτος Ατυχήματος	0	1	2	
0	414.545	165.495	70	0
1	305.454	114.546	52	52
2	340	128	57	185
Συνολικό Απόθεμα				237

Στον παραπάνω πίνακα φαίνονται οι εκτιμώμενες απαιτήσεις τόσο των προηγούμενων ετών όσο και των μελλοντικών. Σημειώνεται ότι οι τιμές που εκτιμήσαμε στο πάνω δεξιά κελί και κάτω αριστερά είναι ίσες με τις πραγματικές επαυξητικές απαιτήσεις. Αυτό προκύπτει από το γεγονός ότι αρχικά ορίσαμε ότι $a_0 = \beta_0 = 0$.

Βήμα 3-4

Αρχικά θα υπολογίσουμε την παράμετρο κλίμακας $\hat{\phi}$ απο τον τύπο (4.4). Απο τον πίνακα των αρχικών τιμών των απαιτήσεων και των τιμών των εκτιμημένων απαιτήσεων, με $n = 6$ και $p = 5$ έχουμε $\hat{\phi} = 0.6203009$.

Στη συνέχεια θα υπολογίσουμε τον πίνακα διακύμανσης/συνδιακύμανσης των εκτιμώμενων παραμέτρων $\hat{\Sigma}$ μέσω της (4.19). Θα χρειαστούμε τον πίνακα βαρών ο οποίος θα υπολογιστεί μέσω της (4.8) και θα είναι ένας 5×5 διαγώνιος πίνακας με στοιχεία της διαγωνίου τις τιμές των απαιτήσεων που έχουμε εκτιμήσει. Έτσι έχουμε

$$\mathbf{W} = \begin{bmatrix} 414.545 & 0 & 0 & 0 & 0 & 0 \\ 0 & 305.454 & 0 & 0 & 0 & 0 \\ 0 & 0 & 340 & 0 & 0 & 0 \\ 0 & 0 & 0 & 165.495 & 0 & 0 \\ 0 & 0 & 0 & 0 & 114.546 & 0 \\ 0 & 0 & 0 & 0 & 0 & 70 \end{bmatrix}. \quad (4.25)$$

Χρησιμοποιώντας τον πίνακα $\mathbf{\Gamma}$

$$\mathbf{\Gamma}^T \mathbf{W} \mathbf{\Gamma} = \begin{bmatrix} 1400.0355 & 420.0155 & 340 & 270.0095 & 70 \\ 420.0155 & 420.0155 & 0.000 & 114.5496 & 0.000 \\ 340 & 0.000 & 340 & 0.000 & 0.000 \\ 270.0095 & 114.5496 & 0.000 & 270.0095 & 0.000 \\ 70 & 0.000 & 0.000 & 00.000 & 70 \end{bmatrix} \quad (4.26)$$

Ο πίνακα διακύμανσης/συνδιακύμανσης των εκτιμώμενων παραμέτρων, $\hat{\Sigma}$ όπως τον ορίσαμε στην (4.19), είναι ο κάτωθι 5×5 πίνακας

$$\begin{aligned} \hat{\Sigma} &= \hat{\phi} (\mathbf{\Gamma}^T \mathbf{W} \mathbf{\Gamma})^{-1} = \\ &= \begin{bmatrix} 0.001323163 & -0.001088209 & -0.001323163 & -0.000861498 & -0.001323163 \\ -0.001088209 & 0.002565061 & 0.001088209 & 0.000000000 & 0.001088209 \\ -0.001323163 & 0.001088209 & 0.003147577 & 0.000861498 & 0.001323163 \\ -0.000861498 & 0.000000000 & 0.000861498 & 0.003158828 & 0.000861498 \\ -0.001323163 & 0.001088209 & 0.001323163 & 0.000861498 & 0.010184604 \end{bmatrix}. \end{aligned} \quad (4.27)$$

Για τον υπολογισμό του πίνακα διακύμανσης/συνδιακύμανσης της συνάρτησης πρόβλεψης θα χρειαστούμε τον πίνακα σχεδιασμού $\mathbf{\Gamma}_F$ για τα κελιά του κάτω δεξιά τριγώνου τον οποίο ορίσαμε στην (4.20) και επομένως η (4.21) γίνεται

$$\begin{aligned} \mathbf{N} &= \mathbf{\Gamma}_F \hat{\Sigma} (\mathbf{\Gamma}_F)^T = \\ &= \begin{bmatrix} 0.01142650 & 0.000000000 & 0.009949650 \\ 0.000000000 & 0.004983242 & 0.002685912 \\ 0.00994965 & 0.002685912 & 0.012009019 \end{bmatrix}. \end{aligned} \quad (4.28)$$

Ο πίνακας $\mathbf{diag}(\hat{\mu})$ (4.23) για τα δεδομένα μας είναι

$$\mathbf{diag}(\hat{\mu}) = \begin{bmatrix} 52 & 0 & 0 \\ 0 & 128 & 0 \\ 0 & 0 & 27 \end{bmatrix} \quad (4.29)$$

και αντικαθιστώντας στην (4.22)

$$\mathbf{diag}(\hat{\mu})\mathbf{Ndiag}(\hat{\mu}) = \begin{bmatrix} 30.89726 & 0.00000 & 29.49076 \\ 0.00000 & 81.64544 & 19.59642 \\ 29.49076 & 19.59642 & 39.01730 \end{bmatrix}. \quad (4.30)$$

Προκειμένου να βρούμε το μέσο τετραγωνικό σφάλμα θα πρέπει να υπολογίσουμε και την ποσότητα $\hat{\phi}\mathbf{diag}(\hat{\mu})$

$$\hat{\phi}\mathbf{diag}(\hat{\mu}) = \begin{bmatrix} 32.25565 & 0.00000 & 0.00000 \\ 0.00000 & 79.39852 & 0.00000 \\ 0.00000 & 0.00000 & 35.35715 \end{bmatrix}. \quad (4.31)$$

Έχοντας υπολογίσει όλους τους όρους των σχέσεων (4.17) και (4.18) έχουμε

$$\hat{\phi}\mathbf{diag}(\hat{\mu}) + \mathbf{diag}(\hat{\mu})\mathbf{Ndiag}(\hat{\mu}) = \begin{bmatrix} 63.15291 & 0.00000 & 29.49076 \\ 0.00000 & 161.04395 & 19.59642 \\ 29.49076 & 19.59642 & 74.37445 \end{bmatrix} \quad (4.32)$$

Σε αυτό το σημείο της διαδικασίας έχουν πραγματοποιηθεί όλοι οι υπολογισμοί των ποσοτήτων που χρειαζόμαστε προκειμένου να υπολογίσουμε τα σφάλματα των παραμέτρων (standard errors) του μοντέλου, το μέσο τετραγωνικό σφάλμα των επαυξητικών απαιτήσεων και το μέσο τετραγωνικό σφάλμα για το απόθεμα κάθε έτους ατυχήματος αλλά και για το συνολικό.

Τα σφάλματα των παραμέτρων του ODP μοντέλου είναι η τετραγωνική ρίζα των στοιχείων της διαγωνίου του πίνακα $\hat{\Sigma}$, σχέση (4.27), ως κάτωθι.

Παράμετροι	Standard Error
c	0.03637530
a_1	0.05064644
a_2	0.05610327
β_1	0.05620345
β_2	0.10091880

Σχετικά με τον υπολογισμό του μέσου τετραγωνικού σφάλματος των επαυξητικών απαιτήσεων και του αποθέματος για κάθε έτος ατυχήματος του τριγώνου μπορούμε να χρησιμοποιήσουμε τον πίνακα (4.32). Πιο συγκεκριμένα το μέσο τετραγωνικό

σφάλμα για το πρώτο έτος ατυχήματος, έστω $MSEP(\hat{X}_{1,2})$, που στον πίνακα είναι το 63.15291 μας δίνει εκτιμώμενο σφάλμα 7.898965.

Για το μέσο τετραγωνικό σφάλμα του εκτιμώμενου αποθέματος για το δεύτερο έτος ατυχήματος του τριγώνου εξέλιξης, έστω $MSEP(\hat{R}_2)$, αν προσθέσουμε τις τιμές του πίνακα (4.32) $161.04395 + 2 \times 19.59642 + 74.37445 = 274.61124$ εκτιμούμε ότι το σφάλμα είναι 16.571614. Τέλος αν προσθέσουμε όλα τα στοιχεία του πίνακα αυτού θα πάρουμε την τιμή 396.74567 και έτσι το σφάλμα ισούται με 19.898225. Τα αποτελέσματα συνοψίζονται στον κάτωθι πίνακα όπου με SE είναι τα σφάλματα.

Έτος ατυχήματος	Απόθεμα	SE	CV
1	52	7.898965	15%
2	185	16.571614	9%
Συνολικό	237	19.898225	8%

Όπως είδαμε η όλη διαδικασία είναι ιδιαίτερος απαιτητική, αν σκεφτούμε τον χρόνο που απαιτήθηκε για να υπολογιστεί το απόθεμα για ένα αρκετά απλό τρίγωνο 3×3 . Για αυτό τον λόγο είναι απαραίτητη η χρήση λογισμικών πορογραμμάτων, όπως το πακέτο R όπου με την χρήση της εντολής `glmReserve(.)` του πακέτου `ChainLadder` βρίσκουμε τα ίδια αποτελέσματα [6].

4.1.3 Γενίκευση του ODP μοντέλου

Το μοντέλο της υπερσκεδαστικής Poisson που αναλυτικά είδαμε παραπάνω, αποτελεί μια περίπτωση ενός πιο γενικού μοντέλου των γενικευμένων γραμμικών μοντέλων όπου

$$Var[X_{i,j}] = \phi \mu_{i,j}^z. \quad (4.33)$$

Ο πιο πάνω τύπος είναι μια γενική μορφή για τα γενικευμένα γραμμικά μοντέλα όπου για τις διάφορες τιμές του z προκύπτουν τα ακόλουθα μοντέλα. Σημειώνεται ότι η διαδικασία είναι η ίδια με αυτή που παρουσιάστηκε παραπάνω [6].

Κανονική κατανομή

Για $z = 0$ έχουμε μοντελοποίηση με την χρήση της κανονικής κατανομής με συνάρτηση πυκνότητας πιθανότητας με μέση τιμή μ και διακύμανση σ^2 η οποία είναι η ακόλουθη

$$f(y) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{(y - \mu)^2}{2\sigma^2}\right\}, \quad y \in \mathbb{R},$$

με

$$\begin{aligned} E(X_{i,j}) &= \mu_{i,j}, \\ Var(X_{i,j}) &= \phi \mu^0 = \phi = \sigma^2. \end{aligned}$$

Η συνάρτηση σύνδεσης είναι ίδια για όλες τις περιπτώσεις που ακολουθούν και ίση με $n_{i,j} = c + \alpha_i + \beta_j$ όπου $\alpha_0 = \beta_0 = 0$.

Η συνάρτηση απόκλισης (Deviance) είναι η ακόλουθη

$$D = \frac{1}{\sigma^2} \sum_{i=0}^I \sum_{j=0}^{I-i} (X_{i,j} - \hat{\mu}_{i,j})^2 \stackrel{\phi=\sigma^2}{=} \frac{1}{\hat{\phi}} \sum_{i=0}^I \sum_{j=0}^{I-i} (X_{i,j} - \hat{\mu}_{i,j})^2.$$

Η ισότητα ισχύει λόγω της σχέσης (4.33) βάζοντας όπου $z = 0$. Σημειώνεται ότι μικρές τιμές του D εκφράζουν καλή προσαρμογή.

Poisson κατανομή

Για $z = 1$ έχουμε μοντελοποίηση με την χρήση της Poisson όπως αναλυτικά αναφέρθηκε παραπάνω.

Gamma κατανομή

Για $z = 2$ έχουμε μοντελοποίηση με την χρήση της κατανομής Gamma με συνάρτηση πυκνότητας πιθανότητας

$$f(y) = \frac{y^{a-1} e^{-\frac{ay}{\mu}}}{\Gamma(a) \left(\frac{\mu}{a}\right)^a}, \quad y > 0,$$

με

$$E(X_{i,j}) = \mu_{i,j},$$

$$Var(X_{i,j}) = \phi \mu_{i,j}^2 = \frac{\mu_{i,j}^2}{a},$$

όπου $\phi = \frac{1}{a}$. Η συνάρτηση Deviance είναι η ακόλουθη

$$D = \frac{2}{\hat{\phi}} \sum_{i=0}^I \sum_{j=0}^{I-1} \left\{ -\log\left(\frac{X_{i,j} - \hat{\mu}_{i,j}}{\hat{\mu}_{i,j}}\right) + \frac{X_{i,j} - \hat{\mu}_{i,j}}{\hat{\mu}_{i,j}} \right\}.$$

Inverse Gaussian κατανομή

Για $z = 3$ έχουμε μοντελοποίηση με την χρήση της κατανομής Inverse Gaussian με συνάρτηση πυκνότητας πιθανότητας

$$f(y) = \sqrt{\frac{\lambda}{2\pi y^3}} \exp\left\{-\frac{\lambda(y - \mu)^2}{2\mu^2 y}\right\}, \quad y > 0, \quad \lambda, \mu > 0,$$

με

$$\begin{aligned} E(X_{i,j}) &= \mu_{i,j}, \\ \text{Var}(X_{i,j}) &= \phi \mu_{i,j}^3. \end{aligned}$$

όπου $\phi = \frac{1}{\lambda}$. Η συνάρτηση Deviance είναι η ακόλουθη

$$D = \frac{1}{\hat{\phi}} \sum_{i=0}^I \sum_{j=0}^{I-1} \left[\frac{(X_{i,j} \hat{\mu}_{i,j})^2}{\hat{\mu}_{i,j}^2 X_{i,j}} \right].$$

4.2 Υπερσκεδαστική αρνητική διωνυμική κατανομή

Το συγκεκριμένο μοντέλο διαισθητικά πλησιάζει περισσότερο την τεχνική Chain Ladder καθώς οι παράμετροι μοιάζουν με τους συντελεστές εξέλιξης αυτής. Η εφαρμογή του μοντέλου γίνεται είτε σε δεδομένα που αφορούν προσαυξητικές απαιτήσεις είτε σε αθροιστικές απαιτήσεις.

Η αρνητική διωνυμική κατανομή, σύμφωνα με τον Verall [16], προκύπτει αν υποθέσουμε ότι οι παράμετροι μ της Poisson κατανομής ακολουθούν την Gamma και η συνάρτηση πυκνότητας πιθανότητας είναι

$$f(y) = \frac{\Gamma(y+r)}{y! \Gamma(r)} \left(\frac{r}{r+\mu} \right)^r \left(\frac{\mu}{r+\mu} \right)^y, \quad y = 0, 1, 2, \dots, \quad r, \mu > 0. \quad (4.34)$$

Έστω οι αθροιστικές απαιτήσεις $C_{i,j}$ ακολουθούν την υπερσκεδαστική αρνητική διωνυμική κατανομή με

$$\begin{aligned} E(C_{i,j}|C_{i,j-1}) &= \mu_{i,j} = \lambda_j C_{i,j-1}, \text{ και} \\ \text{Var}(C_{i,j}|C_{i,j-1}) &= \phi \left(\mu_{i,j} + \frac{\mu_{i,j}^2}{r} \right). \end{aligned} \quad (4.35)$$

Η κανονική συνάρτηση σύνδεσης είναι η $g(\mu) = \log \left[\frac{\mu}{\mu+r} \right]$. Η γραμμική συνάρτηση πρόβλεψης που σχετίζεται με την μέση τιμή $\mu_{i,j}$ είναι

$$\eta_{i,j} = \log(\mu_{i,j}) = c + a_{j-1} + \log(C_{i,j-1}). \quad (4.36)$$

Παρατηρούμε ότι, σε αντίθεση με τα προηγούμενα μοντέλα κατανομών, οι παράμετροι του μοντέλου αφορούν μόνο τα έτη εξέλιξης των ζημιών. Με την βοήθεια πινάκων η παραπάνω σχέση γράφεται

$$\eta_{i,j} = \mathbf{\Gamma}^T \boldsymbol{\beta} + \log(C_{i,j-1}), \quad j \geq 1, \quad (4.37)$$

όπου ο πίνακας $\mathbf{\Gamma}$ είναι ο πίνακας σχεδιασμού με στοιχεία

$$\mathbf{\Gamma}_j = (1, \delta_{1,j-1}, \dots, \delta_{I-1,j-1})^T \quad (4.38)$$

και

$$\boldsymbol{\beta} = (c, a_1, \dots, a_{I-1})^T. \quad (4.39)$$

Προκειμένου να υπολογίσουμε την εκτίμηση κλίμακας $\hat{\phi}$ εφαρμόζουμε την ίδια σχέση με πριν (4.4).

Για να ελέγξουμε πόσο καλά εφαρμόζεται στα δεδομένα μας το μοντέλο αρκεί να υπολογίσουμε την διακύμανση D ,

$$D = \frac{2}{\hat{\phi}} \sum_{i=0}^{I-1} \sum_{j=1}^{I-1} \left[X_{i,j} \log \left(\frac{\mu_{i,j}}{r + \mu_{i,j}} \right) + r \log \left(\frac{1}{r + \mu_{i,j}} \right) + \kappa \right]. \quad (4.40)$$

Παρατηρούμε ότι σε περίπτωση που κάποιο απο τα $\lambda_j < 1$ τότε η διακύμανση θα είναι αρνητική και κατ' επέκταση το μοντέλο μας δεν δουλεύει. Ωστόσο, όπως παρουσίασαν οι England and Verrall [4], μπορούμε να ξεπεράσουμε το πρόβλημα αυτό με την βοήθεια της κανονικής προσέγγισης.

Υποθέτουμε ότι οι αθροιστικές απαιτήσεις $C_{i,j}$ ακολουθούν την κανονική κατανομή με

$$\begin{aligned} E(C_{i,j}|C_{i,j-1}) &= \mu_{i,j} = \lambda_j, \\ Var(C_{i,j}|C_{i,j-1}) &= \phi_j C_{i,j-1}. \end{aligned} \quad (4.41)$$

Αν θεωρήσουμε $f_{i,j} = \frac{C_{i,j}}{w_{i,j}}$ τότε έχουμε

$$\begin{aligned} E(f_{i,j}|C_{i,j-1}) &= \mu_{i,j} = \lambda_j, \\ Var(f_{i,j}|C_{i,j-1}) &= \frac{\phi_j}{w_{i,j}}. \end{aligned} \quad (4.42)$$

Η συνάρτηση πρόβλεψης σε αυτή τν περίπτωση είναι

$$\eta_{i,j} = \mu_{i,j} + a_{j-1}, \quad j \geq 1 \text{ και } a_0 = 0. \quad (4.43)$$

Το συγκεκριμένο μοντέλο είναι ένα γενικευμένο γραμμικό μοντέλο με βάρη $W_{i,j} = \frac{w_{i,j}}{\phi_j}$. Σημειώνεται ότι η παράμετρος κλίμακας εξαρτάται μόνο απο το έτος εξέλιξης και έτσι οδηγούμαστε σε μια ειδική μέθοδο για την εκτίμηση των παραμέτρων του μοντέλου. Ο ενδιαφερόμενος αναγνώστης μπορεί να δει αναλυτικά την μέθοδο στο σύγγραμμα των McCullagh and Nelder [9].

Κεφάλαιο 5

Εφαρμογή σε πραγματικά δεδομένα

5.1 Δεδομένα

Στο κεφάλαιο αυτό θα δούμε την εφαρμογή των γενικευμένων γραμμικών μοντέλων που αναπτύξαμε παραπάνω σε πραγματικά δεδομένα. Τα δεδομένα που θα μελετήσουμε, Πίνακας 5.1, είναι αυτά που χρησιμοποίησαν και οι Ashe and Taylor το 1983 [13] καθώς και ο Verall το 1990 [14] καθώς και στο σύγγραμά του το 1991 [15]). Είναι τα ίδια δεδομένα που χρησιμοποίησαμε και στην εφαρμογή της μεθόδου Chain Ladder στο κεφάλαιο της αποθεματοποίησης.

Πίνακας 5.1: Αθροιστικές απαιτήσεις $C_{i,j}$

	0	1	2	3	4	5	6	7	8	9
0	357848	1124788	1735330	2217820	2745596	3319994	3466336	3606286	3833515	3901463
1	352118	1236139	2170033	3353322	3799067	4120063	4647867	4914039	5339085	
2	290507	1292306	2218525	3235179	3985995	4132918	4628910	4909315		
3	310608	1418858	2195047	3757447	4029929	4381982	4588268			
4	443160	1136350	2128333	2897821	3402672	3873311				
5	396132	1333217	2180715	2985752	3691712					
6	440832	1288463	2419861	3483130						
7	359480	1421128	2864498							
8	376686	1363294								
9	344014									

Τα δεδομένα μας αφορούν απαιτήσεις ζημιών 10 ετών και παρουσιάζουν την εξέλιξη των απαιτήσεων αυτών σε βάρδος 10 έτων.

Η εφαρμογή των μοντέλων θα γίνει με την βοήθεια της R και με την χρήση της συνάρτησης glmReserve του πακέτου Chain Ladder.

5.2 Εφαρμογή της συνάρτησης βάση των μοντέλων μας

Στο σημείο αυτό είμαστε σε θέση να δούμε πως εφαρμόζονται τα μοντέλα που παρουσιάστηκαν παραπάνω στα δεδομένα μας.

5.2.1 Μοντέλο με την χρήση της ODP

Η σύνταξη της `glmReserve` είναι η κάτωθι

$$ODP = glmReserve(GenIns, var.power = 1). \quad (5.1)$$

Θέτωντας $var.power = 1$ θεωρούμε ότι στην σχέση (4.33) έχουμε $Var[X_{i,j}] = \phi \mu_{i,j}^z$, $z = 1$. Τρέχοντας την εντολή το αποτέλεσμα που θα πάρουμε είναι ο κάτωθι Πίνακας 5.2.

Πίνακας 5.2: Εκτιμήσεις του ODP μοντέλου

```
> # Poisson distribution, z = 1
> ODP=glmReserve(data,var.power=1)
> ODP
```

	Latest	Dev.To.Date	Ultimate	IBNR	S.E	CV
2	5339085	0.98258394	5433719	94634	110099.9	1.1634283
3	4909315	0.91271125	5378826	469511	216043.4	0.4601455
4	4588268	0.86605312	5297906	709638	260872.1	0.3676129
5	3873311	0.79727286	4858200	984889	303550.0	0.3082073
6	3691712	0.72228301	5111171	1419459	375013.9	0.2641949
7	3483130	0.61531018	5660771	2177641	495378.0	0.2274838
8	2864498	0.42219349	6784799	3920301	789961.1	0.2015052
9	1363294	0.24162172	5642266	4278972	1046513.8	0.2445713
10	344014	0.06922055	4969825	4625811	1980101.4	0.4280550
total	30456627	0.61982473	49137483	18680856	2945660.9	0.1576834

Όπου

- Latest: Είναι τα στοιχεία της κυρίας διαγωνίου των δεδομένων μας. Εκφράζει δηλαδή την τελευταία πληρωμή της ασφαλιστικής εταιρείας για την απαίτηση του i έτους.
- Dev.To.Date: Εκφράζει το ποσοστό μεταξύ των απαιτήσεων που έχουν πληρωθεί και αυτών που εκτιμήθηκαν για τα έτη εξέλιξης που μελετάμε, δηλαδή ως $\frac{Latest}{Ultimate}$.
- Ultimate: Είναι το συνολικό ποσό απαίτησης που εκτιμήθηκε για το τελικό έτος εξέλιξης (για το 10 έτος). Εκφράζει δηλαδή την ποσότητα $\hat{C}_{i,10}$.
- IBNR: Εκφράζει το απόθεμα για τις απαιτήσεις που έγιναν αλλά δεν έχουν αναγγελθεί ακόμα στην εταιρεία. Επί της ουσίας είναι η διαφορά της τελικής

απαίτησης που εκτιμήθηκε για το 10ο έτος εξέλιξης και της τελευταίας πληρωμής της ασφαλιστικής για την απαίτηση του έτους i .

- **S.E.:** Είναι η τετραγωνική ρίζα του μέσου τετραγωνικού σφάλματος πρόβλεψης και υπολογίζεται όπως είδαμε στο προηγούμενο κεφάλαιο στα δύο τελευταία στάδια της εφαρμογής τόσο για κάθε έτος ατυχήματος όσο και για το συνολικό απόθεμα.
- **CV:** Είναι ο συντελεστής διασποράς (Coefficient of Variance) που ορίζεται ως $\frac{S.E.}{IBNR}$.

Σύμφωνα με το μοντέλο μας το συνολικό απόθεμα είναι 18680856 το οποίο, όπως αναμέναμε, είναι ίδιο με αυτό της μεθόδου Chain Ladder.

Στη συνέχεια με την εντολή `summary(ODP, type = "model")` θα λάβουμε τον κάτωθι Πίνακα 5.3 που σχετίζεται με τους συντελεστές παλινδρόμησης του μοντέλου μας.

```

> summary(ODP, type = "model")

Call:
glm(formula = value ~ factor(origin) + factor(dev), family = fam,
     data = ldaFit, offset = offset)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-464.88  -123.73   -21.75   116.21   494.35

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  12.506405   0.172924  72.323 < 2e-16 ***
factor(origin)2  0.331272   0.153537   2.158  0.03771 *
factor(origin)3  0.321119   0.157719   2.036  0.04916 *
factor(origin)4  0.305960   0.160736   1.903  0.06499 .
factor(origin)5  0.219316   0.167970   1.306  0.19994
factor(origin)6  0.270077   0.170756   1.582  0.12247
factor(origin)7  0.372208   0.174451   2.134  0.03976 *
factor(origin)8  0.553333   0.186525   2.967  0.00532 **
factor(origin)9  0.368934   0.239181   1.542  0.13170
factor(origin)10 0.242033   0.427562   0.566  0.57485
factor(dev)2     0.912526   0.148850   6.131 4.65e-07 ***
factor(dev)3     0.958831   0.152569   6.285 2.90e-07 ***
factor(dev)4     1.025997   0.156883   6.540 1.33e-07 ***
factor(dev)5     0.435276   0.183914   2.367  0.02344 *
factor(dev)6     0.080057   0.214770   0.373  0.71152
factor(dev)7    -0.006381   0.238290  -0.027  0.97878
factor(dev)8    -0.394452   0.310289  -1.271  0.21180
factor(dev)9     0.009378   0.320249   0.029  0.97680
factor(dev)10   -1.379907   0.896690  -1.539  0.13258
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for Tweedie family taken to be 52601.93)

Null deviance: 10699464  on 54  degrees of freedom
Residual deviance: 1903014  on 36  degrees of freedom
AIC: NA

Number of Fisher Scoring iterations: 4

```

Πίνακας 5.3: Summary of ODP model

Απο το συγκεκριμένο αποτέλεσμα που λάβαμε ο Πίνακας 5.4 παρουσιάζει συνοπτικά τις παραμέτρους του μοντέλου και τα αντίστοιχα σφάλματα

Πίνακας 5.4: Παράμετροι μοντέλου

Παράμετροι		Standard Error
c	12.506405	0.172924
a_1	0.331272	0.153537
a_2	0.321119	0.157719
a_3	0.305960	0.160736
a_4	0.219316	0.167970
a_5	0.270077	0.170756
a_6	0.372208	0.174451
a_7	0.553333	0.186525
a_8	0.368934	0.239181
a_9	0.242033	0.427562
β_1	0.912526	0.148850
β_2	0.958831	0.152569
β_3	1.025997	0.156883
β_4	0.435276	0.183914
β_5	0.080057	0.214770
β_6	-0.006381	0.238290
β_7	-0.394452	0.310289
β_8	0.009378	0.320249
β_9	-1.379907	0.896690

και η εκτίμηση της παραμέτρου κλίμακας $\hat{\phi} = 52601.93$.

Τέλος με την εντολή `ODP$FullTriangle` θα λάβουμε το κάτωθι συμπληρωμένο Πίνακα 5.5 με τις εκτιμήσεις των αθροιστικών απαιτήσεων και παρατηρούμε ότι είναι το ίδιο με αυτό που λάβαμε απο την μέθοδο Chain Ladder.

Πίνακας 5.5: Συμπληρωμένο τρίγωνο εξέλιξης αθροιστικών ζημιών

	0	1	2	3	4	5	6	7	8	9
0	357848	1124788	1735330	2218270	2745596	3319994	3466336	3606286	3833515	3901463
1	352118	1236139	2170033	3353322	3799067	4120063	4647867	4914039	5339085	5433719
2	290507	1292306	2218525	3235179	3985995	4132918	4628910	4909315	5285148	5378826
3	310608	1418858	2195047	3757447	4029929	4381982	4588268	4835458	5205637	5297906
4	443160	1136350	2128333	2897821	3402672	3873311	4207459	4434133	4773589	5297906
5	396132	1333217	2180715	2985752	3691712	4074999	4426546	4665023	5022155	4858200
6	440832	1288463	2419861	3483130	4088678	4513179	4902528	5166649	5562182	5111171
7	359480	1421128	2864498	4174756	4900545	5409337	5875997	6192562	6666635	5660771
8	376686	1363294	2382128	3471744	4075313	4498426	4886502	5149760	5544000	6784799
9	344014	1200818	2098228	3057984	3589620	3962307	4304132	4536015	4883270	4969825

5.2.2 Μοντέλο με την χρήση της κανονικής κατανομής

Η σύνταξη της `glmReserve` είναι η κάτωθι

$$N = \text{glmReserve}(\text{GenIns}, \text{var.power} = 0). \quad (5.2)$$

Θέτωντας `var.power = 0` θεωρούμαι ότι στην σχέση (4.33) έχουμε $\text{Var}[X_{i,j}] = \phi \mu_{i,j}^z$, $z = 0$. Τρέχοντας την εντολή το αποτέλεσμα που θα πάρουμε είναι τον κάτωθι Πίνακα 5.6.

Πίνακας 5.6: Εκτιμήσεις του Κανονικού μοντέλου

```
> N
      Latest Dev.To.Date Ultimate      IBNR      S.E      CV
2      5339085  0.98144404  5440030  100945  312863.1  3.0993420
3      4909315  0.90805535  5406405  497090  389780.8  0.7841252
4      4588268  0.85051875  5394670  806402  460596.2  0.5711744
5      3873311  0.79916112  4846721  973410  464891.3  0.4775904
6      3691712  0.72934376  5061690  1369978  522446.4  0.3813539
7      3483130  0.61955894  5621951  2138821  604381.5  0.2825769
8      2864498  0.41194158  6953651  4089153  793616.9  0.1940786
9      1363294  0.23639381  5767046  4403752  1003683.0  0.2279154
10     344014  0.06696178  5137468  4793454  2628436.4  0.5483387
total 30456627  0.61367826  49629633  19173006  4205117.4  0.2193249
```

Σύμφωνα με το μοντέλο μας το συνολικό εκτιμώμενο απόθεμα είναι 19173006.

Στη συνέχεια με την εντολή `summary(N, type = "model")` θα λάβουμε θα λάβουμε τον κάτωθι Πίνακα 5.7 που σχετίζεται με τους συντελεστές παλινδρόμησης του μοντέλου μας.

```

> summary(N, type = "model")

Call:
glm(formula = value ~ factor(origin) + factor(dev), family = fam,
     data = ldaFit, offset = offset)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-329668  -73675   -9887   116129  419612

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  12.40848    0.20660  60.061 < 2e-16 ***
factor(origin)2  0.39584    0.15718   2.518 0.016377 *
factor(origin)3  0.39257    0.15906   2.468 0.018472 *
factor(origin)4  0.45477    0.15668   2.903 0.006282 **
factor(origin)5  0.24200    0.17008   1.423 0.163373
factor(origin)6  0.29588    0.16811   1.760 0.086894 .
factor(origin)7  0.42291    0.16395   2.579 0.014125 *
factor(origin)8  0.65390    0.16688   3.918 0.000382 ***
factor(origin)9  0.45660    0.21866   2.088 0.043921 *
factor(origin)10 0.33996    0.54765   0.621 0.538669
factor(dev)2     0.93300    0.17193   5.427 4.06e-06 ***
factor(dev)3     0.99616    0.17345   5.743 1.53e-06 ***
factor(dev)4     1.08573    0.17570   6.179 4.00e-07 ***
factor(dev)5     0.44501    0.21303   2.089 0.043845 *
factor(dev)6     0.04021    0.27670   0.145 0.885278
factor(dev)7     0.03045    0.29713   0.102 0.918942
factor(dev)8    -0.32999    0.45897  -0.719 0.476789
factor(dev)9     0.08932    0.40422   0.221 0.826371
factor(dev)10   -1.28198    2.57611  -0.498 0.621762
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for Tweedie family taken to be 30442414652)

Null deviance: 6.5982e+12  on 54  degrees of freedom
Residual deviance: 1.0959e+12  on 36  degrees of freedom
AIC: NA

Number of Fisher Scoring iterations: 7

```

Πίνακας 5.7: Summary of Normal model

Στο παραπάνω αποτέλεσμα που λάβαμε εντοπίζουμε τι τιμές των παραμέτρων του μοντέλου μας καθώς και τα αντίστοιχα σφάλματα αυτών.

Η εκτίμηση της παραμέτρου κλίμακας είναι $\hat{\phi} = 30442414652$.

Τέλος με την εντολή `N$FullTriangle` θα λάβουμε το κάτωθι συμπληρωμένο Πίνακα 5.8 με τις εκτιμήσεις των αθροιστικών απαιτήσεων.

Πίνακας 5.8: Συμπληρωμένο τρίγωνο εξέλιξης αθροιστικών ζημιών

	0	1	2	3	4	5	6	7	8	9
0	357848	1124788	1735330	2218270	2745596	3319994	3466336	3606286	3833515	3901463
1	352118	1236139	2170033	3353322	3799067	4120063	4647867	4914039	5339085	5440030
2	290507	1292306	2218525	3235179	3985995	4132918	4628910	4909315	5305789	5406405
3	310608	1418858	2195047	3757447	4029929	4381982	4588268	4865679	5287597	5394670
4	443160	1136350	2128333	2897821	3402672	3873311	4194869	4419113	4760168	4846720
5	396132	1333217	2180715	2985752	3691712	4034397	4373755	4610412	4970347	5061690
6	440832	1288463	2419861	3483130	4066401	4455505	4840831	5109545	5518235	5621951
7	359480	1421128	2864498	4259085	4993911	5484119	5969568	6308103	6822986	6953651
8	376686	1363294	2410082	3554959	4158210	4560643	4959169	5237087	5659777	5767046
9	344014	1218542	2150085	3168918	3705755	4063883	4418534	4665855	5042009	5137468

5.2.3 Μοντέλο με την χρήση της Gamma

Η σύνταξη της `glmReserve` είναι η κάτωθι

$$G = \text{glmReserve}(\text{GenIns}, \text{var.power} = 2). \quad (5.3)$$

Θέτωντας `var.power = 0` θεωρούμε ότι στην σχέση (4.33) έχουμε $\text{Var}[X_{i,j}] = \phi \mu_{i,j}^z$, $z = 2$. Τρέχοντας την εντολή το αποτέλεσμα που θα πάρουμε είναι ο κάτωθι Πίνακας 5.9.

Πίνακας 5.9: Εκτιμήσεις του Gamma μοντέλου

```
> # Gamma distribution, z = 2
> G=glmReserve(GenIns, var.power=2)
> G
```

	Latest	Dev.To.Date	Ultimate	IBNR	S.E	CV
2	5339085	0.98282233	5432401	93316	45166.37	0.4840153
3	4909315	0.91663147	5355822	446507	160557.17	0.3595849
4	4588268	0.88245851	5199415	611147	177624.61	0.2906414
5	3873311	0.79610317	4865338	992027	254470.93	0.2565161
6	3691712	0.71756209	5144798	1453086	351334.26	0.2417849
7	3483130	0.61438536	5669292	2186162	526287.87	0.2407360
8	2864498	0.43869627	6529570	3665072	941322.25	0.2568360
9	1363294	0.24851783	5485699	4122405	1175945.87	0.2852572
10	344014	0.07078338	4860096	4516082	1667392.40	0.3692122
total	30456627	0.62742277	48542432	18085805	2702709.78	0.1494382

Σύμφωνα με το μοντέλο μας το συνολικό εκτιμώμενο απόθεμα είναι 18085805.

Στη συνέχεια με την εντολή `summary(G, type = "model")` θα λάβουμε να λάβουμε τον κάτωθι Πίνακα 5.10 σχετίζονται με τους συντελεστές παλινδρόμησης του μοντέλου μας.

```

> summary(G, type = "model")

Call:
glm(formula = value ~ factor(origin) + factor(dev), family = fam,
     data = ldaFit, offset = offset)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-0.89880  -0.16494  -0.02739   0.12880   0.65508

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  12.55954    0.15681  80.093 < 2e-16 ***
factor(origin)2  0.31725    0.15306   2.073 0.045414 *
factor(origin)3  0.28342    0.16007   1.771 0.085102 .
factor(origin)4  0.16543    0.16773   0.986 0.330584
factor(origin)5  0.23059    0.17684   1.304 0.200526
factor(origin)6  0.27302    0.18838   1.449 0.155915
factor(origin)7  0.35231    0.20405   1.727 0.092815 .
factor(origin)8  0.46192    0.22733   2.032 0.049585 *
factor(origin)9  0.30715    0.26729   1.149 0.258086
factor(origin)10 0.18890    0.36057   0.524 0.603562
factor(dev)2     0.90857    0.15306   5.936 8.46e-07 ***
factor(dev)3     0.93156    0.16007   5.820 1.21e-06 ***
factor(dev)4     0.99753    0.16773   5.947 8.17e-07 ***
factor(dev)5     0.41453    0.17684   2.344 0.024707 *
factor(dev)6     0.11082    0.18838   0.588 0.560011
factor(dev)7    -0.05421    0.20405  -0.266 0.792021
factor(dev)8    -0.44967    0.22733  -1.978 0.055613 .
factor(dev)9    -0.05944    0.26729  -0.222 0.825276
factor(dev)10   -1.43304    0.36057  -3.974 0.000325 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for Tweedie family taken to be 0.1054213)

Null deviance: 20.0996 on 54 degrees of freedom
Residual deviance: 4.0235 on 36 degrees of freedom
AIC: NA

Number of Fisher Scoring iterations: 7

```

Πίνακας 5.10: Summary of Gamma model

Στο παραπάνω αποτέλεσμα που λάβαμε εντοπίζουμε τι τιμές των παραμέτρων του μοντέλου με τα αντίστοιχα σφάλματα αυτών και την εκτίμηση της παραμέτρου κλίμακας $\hat{\phi} = 0.1054213$.

Τέλος με την εντολή *G\$FullTriangle* θα λάβουμε τον κάτωθι συμπληρωμένο Πίνακα 5.11 με τις εκτιμήσεις των αθροιστικών απαιτήσεων.

Πίνακας 5.11: Συμπληρωμένο τρίγωνο εξέλιξης αθροιστικών ζημιών

	0	1	2	3	4	5	6	7	8	9
0	357848	1124788	1735330	2218270	2745596	3319994	3466336	3606286	3833515	3901463
1	352118	1236139	2170033	3353322	3799067	4120063	4647867	4914039	5339085	5432401
2	290507	1292306	2218525	3235179	3985995	4132918	4628910	4909315	5265610	5355822
3	310608	1418858	2195047	3757447	4029929	4381982	4588268	4802603	5119244	5199416
4	443160	1136350	2128333	2897821	3402672	3873311	4213044	4441809	4779769	4865338
5	396132	1333217	2180715	2985752	3691712	4109771	4464230	4702911	5055520	5144798
6	440832	1288463	2419861	3483130	4096290	4548849	4932559	5190938	5572646	5669292
7	359480	1421128	2864498	4090157	4774347	5279332	5707492	5995802	6421728	6529570
8	376686	1363294	2346176	3396082	3982163	4414736	4781501	5028469	5393320	5485698
9	344014	1197431	2070700	3003519	3524239	3908570	4234432	4453858	4778020	4860096

5.2.4 Μοντέλο με την χρήση της Inverse Gaussian

Η σύνταξη της `glmReserve` είναι η κάτωθι

$$IG = glmReserve(GenIns, var.power = 3). \quad (5.4)$$

Θέτωντας $var.power = 0$ θεωρούμε ότι στην σχέση (4.33) έχουμε $Var[X_{i,j}] = \phi \mu_{i,j}^z$, $z = 3$. Τρέχοντας την εντολή το αποτέλεσμα που θα πάρουμε είναι ο κάτωθι Πίνακας 5.12.

Πίνακας 5.12: Εκτιμήσεις του Inverse Gaussian μοντέλου

```
> # Inverse-Gaussian, z = 3
> IG=glmReserve(GenIns, var.power=3)
> IG
```

	Latest	Dev.To.Date	Ultimate	IBNR	S.E	CV
2	5339085	0.98133436	5440638	101553	25066.3	0.2468297
3	4909315	0.91501869	5365262	455947	140964.3	0.3091681
4	4588268	0.89847652	5106720	518452	121041.0	0.2334661
5	3873311	0.80123682	4834165	960854	230418.4	0.2398059
6	3691712	0.71580291	5157442	1465730	372674.2	0.2542584
7	3483130	0.61780156	5637943	2154813	604526.8	0.2805472
8	2864498	0.46186652	6202004	3337506	1103054.3	0.3305026
9	1363294	0.25636091	5317870	3954576	1318209.8	0.3333378
10	344014	0.07208386	4772414	4428400	1549837.4	0.3499768
total	30456627	0.63670891	47834460	17377833	2760957.4	0.1588781

Σύμφωνα με το μοντέλο μας το συνολικό εκτιμώμενο απόθεμα είναι 17377833.

Στη συνέχεια με την εντολή `summary(IG, type = "model")` θα λάβουμε να λάβουμε τον κάτωθι Πίνακα 5.13 που σχετίζεται με τους συντελεστές παλινδρόμησης του μοντέλου μας.

```

> summary(IG, type = "model")

Call:
glm(formula = value ~ factor(origin) + factor(dev), family = fam,
     data = ldaFit, offset = offset)

Deviance Residuals:
      Min       1Q   Median       3Q      Max
-1.813e-03 -2.280e-04 -2.056e-05  1.083e-04  9.348e-04

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  12.56589    0.14443  87.002 < 2e-16 ***
factor(origin)2  0.40184    0.14658   2.741  0.00947 **
factor(origin)3  0.35938    0.15481   2.321  0.02604 *
factor(origin)4  0.06528    0.16010   0.408  0.68590
factor(origin)5  0.27471    0.18981   1.447  0.15648
factor(origin)6  0.31821    0.21391   1.488  0.14558
factor(origin)7  0.37372    0.24468   1.527  0.13541
factor(origin)8  0.39010    0.26648   1.464  0.15189
factor(origin)9  0.28598    0.28558   1.001  0.32332
factor(origin)10 0.18255    0.31765   0.575  0.56908
factor(dev)2     0.91904    0.18403   4.994 1.53e-05 ***
factor(dev)3     0.90928    0.19334   4.703 3.71e-05 ***
factor(dev)4     0.99677    0.21022   4.742 3.30e-05 ***
factor(dev)5     0.37406    0.17973   2.081  0.04458 *
factor(dev)6     0.15957    0.17955   0.889  0.38007
factor(dev)7    -0.15948    0.17639  -0.904  0.37192
factor(dev)8    -0.53780    0.17892  -3.006  0.00480 **
factor(dev)9    -0.13526    0.22857  -0.592  0.55770
factor(dev)10   -1.43939    0.19150  -7.517 6.98e-09 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for Tweedie family taken to be 2.326768e-07)

Null deviance: 4.7187e-05 on 54 degrees of freedom
Residual deviance: 1.0266e-05 on 36 degrees of freedom
AIC: NA

Number of Fisher Scoring iterations: 7

```

Πίνακας 5.13: Summary of Inverse Gaussian model

Στο παραπάνω αποτέλεσμα που λάβαμε εντοπίζουμε τι τιμές των παραμέτρων του μοντέλου μας με τα αντίστοιχα σφάλματα αυτών και την εκτίμηση της παραμέτρου κλίμακας που στην συγκεκριμένη περίπτωση είναι πάρα πολύ μικρή, σχεδόν μηδενική, $\hat{\phi} = 2.326768e - 07$.

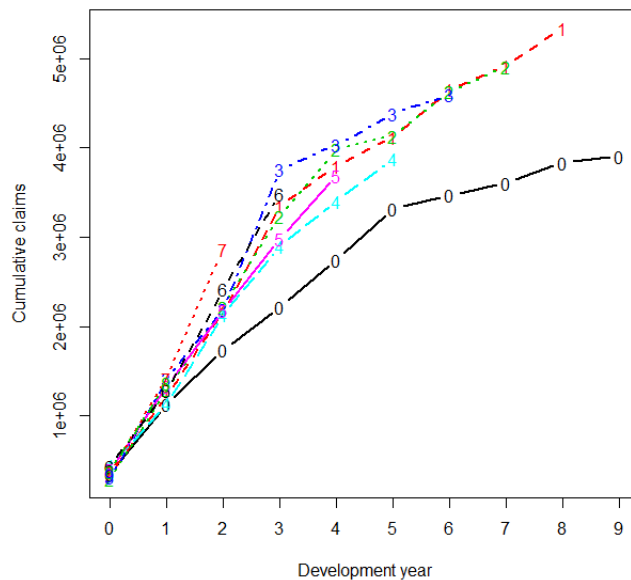
Τέλος με την εντολή `IG$FullTriangle` θα λάβουμε το κάτωθι συμπληρωμένο Πίνακα 5.14 με τις εκτιμήσεις των αθροιστικών απαιτήσεων.

Πίνακας 5.14: Συμπληρωμένο τρίγωνο εξέλιξης αθροιστικών ζημιών

	0	1	2	3	4	5	6	7	8	9
0	357848	1124788	1735330	2218270	2745596	3319994	3466336	3606286	3833515	3901463
1	352118	1236139	2170033	3353322	3799067	4120063	4647867	4914039	5339085	5440638
2	290507	1292306	2218525	3235179	3985995	4132918	4628910	4909315	5267931	5365263
3	310608	1418858	2195047	3757447	4029929	4381982	4588268	4766950	5034189	5106720
4	443160	1136350	2128333	2897821	3402672	3873311	4194926	4415236	4744736	4834165
5	396132	1333217	2180715	2985752	3691712	4153869	4489782	4719887	5064036	5157441
6	440832	1288463	2419861	3483130	4088545	4577083	4932171	5175411	5539205	5637942
7	359480	1421128	2864498	4011607	4627020	5123626	5484578	5731835	6101637	6202005
8	376686	1363294	2310379	3344061	3898621	4346122	4671383	4894191	5227426	5317869
9	344014	1206418	2060442	2992554	3492623	3896152	4189453	4390368	4690859	4772415

5.3 Σχολιασμός μοντέλων

Αρχικά θα πρέπει να σημειωθεί ότι οι αθροιστικές απαιτήσεις αυξάνονται με την πάροδο του χρόνου όπως φαίνεται στο Σχήμα 5.9.

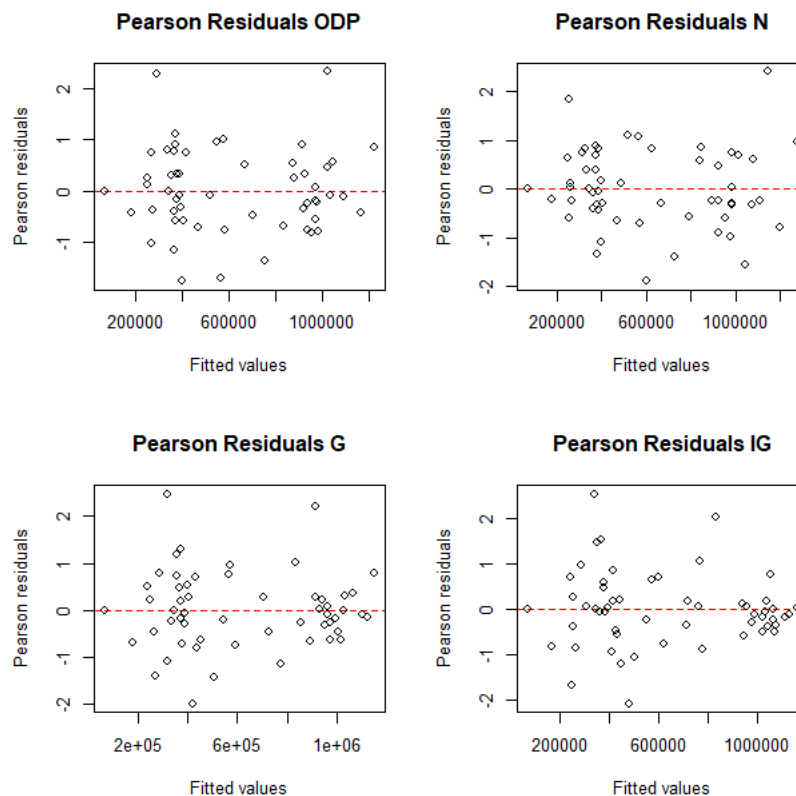


Σχήμα 5.1: Δεδομένα

Πιο κάτω δίνουμε τα σχήματα για τα υπόλοιπα Pearson και των Deviance υπολοίπων. Στο σημείο αυτό θα πρέπει να σημειωθεί ότι ο λόγος που χρησιμοποιούμε γραφήματα των σφαλμάτων έναντι των προσαρμοσμένων τιμών και όχι των παρατηρούμενων, είναι γιατί τα σφάλματα σχετίζονται με τις παρατηρούμενες τιμές και όχι με τις προσαρμοσμένες. Τα γραφήματα μας δείχνουν μια πρώτη εικόνα για την αξιολόγηση των διαφόρων μοντέλων που παρουσιάσαμε παραπάνω. Τα γραφήματα μοιάζουν όμοια λόγω της λογαριθμικής συνάρτησης σύνδεσης. Όπως αναφέραμε και παραπάνω, η λογαριθμική συνάρτηση σύνδεσης

χρησιμοποιείται ευρέως για την εκτίμηση αποθέματος κυρίως εξ' αιτίας της απλής της ερμηνείας.

Απο τα διαγράμματα υπολοίπων (Σχήμα 5.2) έχουμε μια πρώτη εικόνα σχετικά με την απόκλιση των μοντέλων μας. Τα υπόλοιπα θα μπορούσαμε να μπούμε ότι είναι κατά κάποιο τρόπο οι παρατηρούμενες τιμές των (άγνωστων) σφαλμάτων.

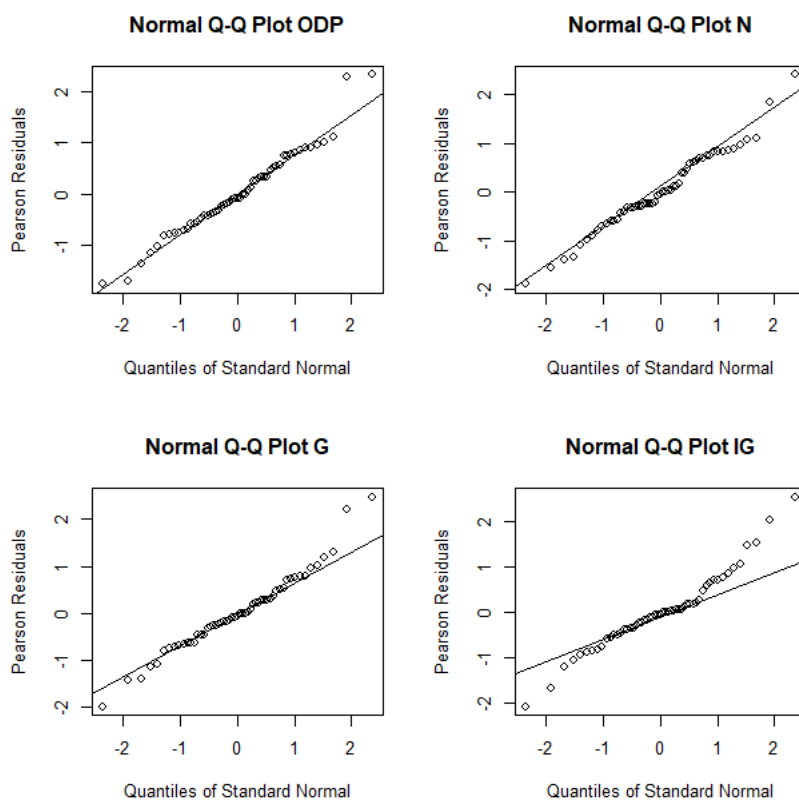


Σχήμα 5.2: Υπόλοιπα Pearson των μοντέλων

Ικανοποιητικό θεωρείται ένα μοντέλο όπου τα υπόλοιπα συγκεντρώνονται τυχαία γύρω από την ευθεία $\epsilon = 0$. Από τα διαγράμματα των υπολοίπων μπορούμε να κάνουμε και έναν πρώτο έλεγχο της σταθερότητας ή μη της διασποράς της τυχαίας μεταβλητής. Αν για παράδειγμα, το διάγραμμα υπολοίπων έχει μορφή τραπέζιου (ανοιχτής βεντάλιας) η πιο πιθανή αιτία αυτής της απόκλισης από την τυχαία συγκέντρωση των σημείων γύρω από την ευθεία $\epsilon = 0$ είναι η μη σταθερότητα της διασποράς των τυχαίων σφαλμάτων. Αυτό παρατηρείται σε πολλές οικονομικές και εμπορικές εφαρμογές όπου η μεταβολή της διασποράς με τις τυχαίες μεταβλητές δίνει διαγράμματα υπολοίπων όπου αυξανόμενων των τυχαίων μεταβλητών αυξάνει και η διασπορά ή αντίστροφα. Αυτό συμβαίνει διότι τέτοιες εφαρμογές ακολουθούν πολλαπλασιαστικά μοντέλα.

Στα μοντέλα μας παρατηρούμε ότι η διακύμανση μεταβάλλεται καθώς αυξάνονται οι αθροιστικές απαιτήσεις. Συνεπώς αντιλαμβανόμαστε ότι η διασπορά των

υπολοίπων δεν είναι σταθερή, δηλαδή δεν υπάρχει ομοσκεδαστικότητα. Στα διαγράμματά των υπολοίπων παρατηρούμε ότι υπάρχουν και ακραίες παρατηρήσεις (outliers) που οφείλονται σε λανθασμένη παρατήρηση ή πιθανόν σε απότομη στιγμιαία διαταραχή του συστήματος που παρατηρούμε. Στα διαγράμματα υπολοίπων των μοντέλων μας εντοπίζουμε την ύπαρξη τέτοιων τιμών. Σημειώνεται ότι ποτέ δεν πρέπει να απορρίπτουμε μια ακραία παρατήρηση αν δεν είμαστε βέβαιοι ότι πρόκειται για λάθος ή απότομη στιγμιαία διαταραχή καθώς έγκυρες ακραίες παρατηρήσεις μπορεί να αποδειχθούν οι πλέον ενδιαφέρουσες. Το $Q - Q$ norm αποτελεί συγκριτικό γράφημα ποσοστιαίων σημείων μιας υποτιθέμενης κατανομής σε σχέση με την παρατηρηθείσα. Αν οι παρατηρήσεις βρίσκονται πάνω στην ευθεία τότε η μεταβλητή ακολουθεί την κανονική κατανομή. Αποτελεί δηλαδή έναν έλεγχο κανονικότητας των τυποποιημένων υπολοίπων.



Σχήμα 5.3: Normal Q-Q plot των μοντέλων

Στα γραφήματα των μοντέλων μας (Σχήμα 5.3), εκτός του μοντέλου της Inverse Gaussian, φαίνεται τα υπόλοιπα να ακολουθούν την κανονική κατανομή εκτός από κάποια σημεία (δυο-τρία) που βρίσκονται πάνω δεξιά. Παρόλα αυτά, σχετικά με τα υπόλοιπα μοντέλα, μεγάλο μέρος τους βρίσκεται κοντά στο σώμα της εμπειρικής/θεωρητικής κατανομής. Από τα διαγράμματα των μοντέλων μας δεν μπορούμε να εκφράσουμε μια ξεκάθαρη

επιλογή σχετικά με το ποιο μοντέλο λειτουργεί καλύτερα.

Μια διαφορετική προσέγγιση για τον έλεγχο καλής προσαρμογής των μοντέλων μας είναι ελέγξουμε τις τιμές του Deviance και του Pearson χ^2 test και των βαθμών ελευθερίας (Df).

Μοντέλο	Df	χ^2	D
ODP	36	1893649	1903014
Normal	36	$1.095923e + 12$	$1.095923e + 12$
Gamma	36	3.795166	4.023484
Inverse Gaussian	36	$8.383561e - 06$	$1.026559e - 05$

Συνοπτικά, με την βοήθεια της R, παρουσιάζονται συνοπτικά στον Πίνακα 5.15 τα αποθέματα που εκτιμήσαμε για κάθε μοντέλο.

Πίνακας 5.15: Εκτιμήσεις των αποθεμάτων κάθε έτους \hat{R}_i

	i	2	3	4	5	6	7	8	9	10	Total \hat{R}
ODP	\hat{R}_i	94634	469511	709638	984889	1419459	2177641	3920301	4278972	4625811	18680856
N	\hat{R}_i	100945	497090	806402	973410	1369978	2138821	4089153	4403752	4793454	19173006
G	\hat{R}_i	93316	446507	611147	992027	1453086	2186162	3665072	4122405	4516082	18085805
IG	\hat{R}_i	101553	455947	518452	960854	1465730	2154813	3337506	3954576	4428400	17377833

Παράρτημα

```

1 rm(list=ls(all=TRUE))
2                                     CHAIN LADDER EXAMPLE
3
4 #PACKAGES
5 library(actuar)
6 library(ChainLadder)
7
8 #TRIANGLE
9 data_cl=t(matrix(c(357848,766940,610542,482940,527326,574398,146342,139950,
10 227229,67948,352118,884021,933894,1183289,445745,320996,527804,266172,425046,NA,
11 290507,1001799,926219,1016654,750816,146923,495992,280405,NA,NA,310608,1108250,
12 776189,1562400,272482,352053,206286,NA,NA,NA,443160,693190,991983,769488,504851,
13 470639,NA,NA,NA,NA,396132,937085,847498,805037,705960,NA,NA,NA,NA,NA,440832,
14 847631,1131398,1063269,NA,NA,NA,NA,NA,NA,359480,1061648,1443370,NA,NA,NA,NA,
15 NA,NA,NA,376686,986608,NA,NA,NA,NA,NA,NA,NA,NA,344014,NA,NA,NA,NA,NA,NA,NA,
16 NA,NA),ncol=10,dimnames=list(dev=0:9,origin=0:9))
17 data_cl
18 m=dim(data_cl)[1]
19 m
20 n=dim(data_cl)[2]
21 n
22 #CUMMULATIVE TRIANGLE
23 cum.data_cl=t(apply(data_cl,1,cumsum))
24 cum.data_cl
25
26 #LATEST PAID
27 latest.paid=cum.data_cl[row(cum.data_cl)==n-col(cum.data_cl)+1]
28 latest.paid
29
30 #DEVELOPMENT FACTOR
31 dev.fac<- sapply(1:(n-1),
32 function(i){
33 sum(cum.data_cl[c(1:(n-i)),i+1])/sum(cum.data_cl[c(1:(n-i)),i])
34 }
35 )
36 dev.fac
37
38 #PREDICTED CLAIM VALUES
39 full.data_cl=cum.data_cl
40 for(k in 1:(n-1)){
41 full.data_cl[(n-k+1):n,k+1]=full.data_cl[(n-k+1):n,k]*dev.fac[k]
42 }
43 full.data_cl
44
45 #ULTIMATE RESERVE
46 ultimate.paid=full.data_cl[,n]
47 ultimate.paid
48 CL.reserve1=sum(ultimate.paid-latest.paid)
49 CL.reserve1
50
51                                     ODP EXAMPLE
52
53 rm(list=ls(all=TRUE))
54 #PACKAGES
55 library(ChainLadder)
56
57 #DATA
58 claims=matrix(c(420,150,70,300,120,NA,340,NA,NA),nrow=3,ncol=3,byrow=TRUE)
59 claims
60 log.claims=log(claims)
61 log.claims
62 X=matrix(c(420,300,340,150,120,70),ncol=1)
63 X
64 W=diag(c(420,300,340,150,120,70))
65 W
66 z=log(X)
67 z
68 G=matrix(c(1,0,0,0,0,1,1,0,0,0,1,0,0,1,0,0,1,0,0,1,0,0,1,0,0,1,0,0,0,0,1),ncol=5,
69 nrow=6,byrow=TRUE)
70 G
71 a=t(G)%*%W%*%G
72 a
73 solve(a)
74 Wz=W%*%z
75 Wz
76 B1=solve(a)%*%t(G)%*%Wz

```

```

77 B1
78 claims2=triangle (c(420,570,640,300,420,340))
79 claims2
80 m1=glmReserve(claims2, var.power=1)
81 m1
82 summary(m1, type="model")
83 m1$FullTriangle
84 claims1=matrix(c(414.5601174,155.4599279,70,305.4658535,114.5496095,NA,340,NA
85 ,NA),nrow=3,ncol=3,)
86 claims1
87 W1=diag(c(414.5601174,305.4658535,340,155.4599279,114.5496095,70))
88 W1
89 A1=t(G)%*%W1%*%G
90 A1
91 f=0.6203009
92 S=f*solve(A1)
93 S
94 s=round(S, digits = 9)
95 s
96 Gf=matrix(c(1,1,0,0,1,1,0,1,1,0,1,0,1,0,1),ncol=5,
97 nrow=3,byrow=TRUE)
98 Gf
99 N=Gf%*%S%*%t(Gf)
100 N
101 n=round(N, digits = 9)
102 n
103 diagm=diag(c(52,128,57))
104 diagm
105 diagm%*%N%*%diagm
106 round(diagm%*%N%*%diagm, digits = 9)
107 f*diagm
108 MSEP=(f*diagm)+diagm%*%N%*%diagm
109 round(MSEP,digits = 9)
110 SE=(diag(S))^(1/2)
111 SE
112 summary(m1, type="model")
113
114
115                                     GLM in RESERVINGS
116 rm(list=ls(all=TRUE))
117 #PACKAGES
118 library(ChainLadder)
119 library(data.table)
120
121 #DATA
122 data_c1=t(matrix(c(357848,766940,610542,482940,527326,574398,146342,139950,
123 227229,67948,
124 352118,884021,933894,1183289,445745,320996,527804,266172,425046,NA,
125 290507,1001799,926219,1016654,750816,146923,495992,280405,NA,NA,
126 310608,1108250,776189,1562400,272482,352053,206286,NA,NA,NA,
127 443160,693190,991983,769488,504851,470639,NA,NA,NA,NA,
128 396132,937085,847498,805037,705960,NA,NA,NA,NA,NA,
129 440832,847631,1131398,1063269,NA,NA,NA,NA,NA,NA,
130 359480,1061648,1443370,NA,NA,NA,NA,NA,NA,NA,
131 376686,986608,NA,NA,NA,NA,NA,NA,NA,NA,
132 344014,NA,NA,NA,NA,NA,NA,NA,NA,NA),ncol=10,dimnames=list(dev=0:9,origin=0:9))
133 data_c1
134 data=t(apply(data_c1,1,cumsum))
135 data
136 data=as.triangle(data)
137 data_incr=as.triangle(data_c1)
138 data=triangle(c(357848,1124788,1735330,2218270,2745596,3319994,3466336,3606286,
139 3833515,3901463,352118,1236139,2170033,3353322,3799067,4120063,4647867,4914039,
140 5339085,290507,1292306,2218525,3235179,3985995,4132918,4628910,4909315,310608,
141 1418858,2195047,3757447,4029929,4381982,4588268,443160,1136350,2128333,2897821,
142 3402672,3873311,396132,1333217,2180715,2985752,3691712,440832,1288463,2419861,
143 3483130,359480,1421128,2864498,376686,1363294,344014))
144 data
145
146 #PLOT CLAIMS
147 x <- seq(0,10,1)
148 # Cumulative data
149 plot(data, xlab="Development year",
150 ylab="Cumulative claims",lwd="2", pch=c(48:57), xaxt="n")
151 axis(1, at = x, labels = x-1)
152
153 #GLM IN RESERVINGS

```

```

154
155 #Poisson distribution, z = 1
156 ODP=glmReserve(data,var.power=1)
157 ODP
158 ODP_glm<-ODP$model
159 ODP_glm
160 summary(ODP, type = "model")
161 ODP$FullTriangle
162 ODP_residPearson <-residuals(ODP,type=c("pearson"))
163 ODP_residPearson
164 plot(fitted.values(ODP_glm),ODP_residPearson,ylab="Pearson residuals",
165 xlab="Fitted values",abline(h=0, "lty"=2, col="red"))
166 ODP_residDeviance <-residuals(ODP,type=c("deviance"))
167 ODP_residDeviance
168 plot(fitted.values(ODP_glm),ODP_residDeviance,ylab="Deviance residuals",
169 xlab="Fitted values",abline(h=0, "lty"=2, col="red"))
170 plot(ODP, which = 5, xlab = "Quantiles of Standard Normal",
171 ylab = "Pearson Residuals")
172 #DEVIANCE
173 ODP_D <- sum(residuals(ODP_glm, "deviance")^2)
174 ODP_D
175 # PEARSON CHISQ STATISTIC
176 ODP_P <- sum(residuals(ODP_glm, "pearson")^2)
177 ODP_P
178
179 #Normal distribution, z = 0
180 N=glmReserve(data,var.power=0)
181 N
182 N_glm<-N$model
183 summary(N, type = "model")
184 N$FullTriangle
185 N_residPearson <-residuals(N,type=c("pearson"))
186 N_residPearson
187 plot(fitted.values(N_glm),N_residPearson,ylab="Pearson residuals",
188 xlab="Fitted values",abline(h=0, "lty"=2, col="red"))
189 N_residDeviance <-residuals(N,type=c("deviance"))
190 N_residDeviance
191 plot(fitted.values(N_glm),N_residDeviance,ylab="Deviance residuals",
192 xlab="Fitted values",abline(h=0, "lty"=2, col="red"))
193 plot(N, which = 5, xlab = "Quantiles of Standard Normal",
194 ylab = "Pearson Residuals")
195 #DEVIANCE
196 N_D <- sum(residuals(N_glm, "deviance")^2)
197 N_D
198 # PEARSON CHISQ STATISTIC
199 N_P <- sum(residuals(N_glm, "pearson")^2)
200 N_P
201
202 #Gamma distribution, z = 2
203 G=glmReserve(data,var.power=2)
204 G
205 G_glm<-G$model
206 summary(G, type = "model")
207 G$FullTriangle
208 G_residPearson <-residuals(G,type=c("pearson"))
209 G_residPearson
210 plot(fitted.values(G_glm),G_residPearson,ylab="Pearson residuals",
211 xlab="Fitted values",abline(h=0, "lty"=2, col="red"))
212 G_residDeviance <-residuals(G,type=c("deviance"))
213 G_residDeviance
214 plot(fitted.values(G_glm),G_residDeviance,ylab="Deviance residuals",
215 xlab="Fitted values",abline(h=0, "lty"=2, col="red"))
216 plot(G, which = 5, xlab = "Quantiles of Standard Normal",
217 ylab = "Pearson Residuals")
218 #DEVIANCE
219 G_D <- sum(residuals(G_glm, "deviance")^2)
220 G_D
221 # PEARSON CHISQ STATISTIC
222 G_P <- sum(residuals(G_glm, "pearson")^2)
223 G_P
224 round(2.326768e-07,digits=10)
225
226 #Inverse-Gaussian, z = 3
227 IG=glmReserve(data,var.power=3)
228 IG
229 summary(IG, type = "model")
230 IG_glm<-IG$model

```

```

231 summary(IG, type = "model")
232 IG$FullTriangle
233 IG_residPearson <-residuals(IG,type=c("pearson"))
234 IG_residPearson
235 plot(fitted.values(IG_glm),IG_residPearson,ylab="Pearson residuals",
236 xlab="Fitted values",abline(h=0, "lty"=2, col="red"))
237 IG_residDeviance <-residuals(IG,type=c("deviance"))
238 IG_residDeviance
239 plot(fitted.values(IG_glm),IG_residDeviance,ylab="Deviance residuals",
240 xlab="Fitted values",abline(h=0, "lty"=2, col="red"))
241 plot(IG, which = 5, xlab = "Quantiles of Standard Normal",
242 ylab = "Pearson Residuals")
243 #DEVIANCE
244 IG_D <- sum(residuals(IG_glm, "deviance")^2)
245 IG_D
246 # PEARSON CHISQ STATISTIC
247 IG_P <- sum(residuals(IG_glm, "pearson")^2)
248 IG_P
249
250 #Plots
251 #Plot Pearson Residuals
252 par(mfrow=c(2,2))
253 plot(fitted.values(ODP_glm),main="Pearson Residuals ODP",ODP_residPearson,
254 ylab="Pearson residuals", xlab="Fitted values",abline(h=0, "lty"=2, col="red"))
255 plot(fitted.values(N_glm),main="Pearson Residuals N",N_residPearson,
256 ylab="Pearson residuals", xlab="Fitted values",abline(h=0, "lty"=2, col="red"))
257 plot(fitted.values(G_glm),main="Pearson Residuals G",G_residPearson,
258 ylab="Pearson residuals", xlab="Fitted values",abline(h=0, "lty"=2, col="red"))
259 plot(fitted.values(IG_glm),main="Pearson Residuals IG",IG_residPearson,
260 ylab="Pearson residuals", xlab="Fitted values",abline(h=0, "lty"=2, col="red"))
261
262 #Plot Deviance Residuals
263 par(mfrow=c(2,2))
264 plot(fitted.values(ODP_glm),main="Deviance Residuals ODP",ODP_residDeviance,
265 ylab="Pearson residuals", xlab="Fitted values",abline(h=0, "lty"=2, col="red"))
266 plot(fitted.values(N_glm),main="Deviance N Gaussian",N_residDeviance,
267 ylab="Pearson residuals", xlab="Fitted values",abline(h=0, "lty"=2, col="red"))
268 plot(fitted.values(G_glm),main="Deviance Residuals G",G_residDeviance,
269 ylab="Pearson residuals", xlab="Fitted values",abline(h=0, "lty"=2, col="red"))
270 plot(fitted.values(IG_glm),main="Deviance Residuals IG",IG_residDeviance,
271 ylab="Pearson residuals", xlab="Fitted values",abline(h=0, "lty"=2, col="red"))
272
273 #Q-Q Plot
274 par(mfrow=c(2,2))
275 plot(ODP, which = 5,main="Normal Q-Q Plot ODP",
276 xlab = "Quantiles of Standard Normal", ylab = "Pearson Residuals")
277 plot(N, which = 5,main="Normal Q-Q Plot N",
278 xlab = "Quantiles of Standard Normal",ylab = "Pearson Residuals")
279 plot(G, which = 5,main="Normal Q-Q Plot G",
280 xlab = "Quantiles of Standard Normal",ylab = "Pearson Residuals")
281 plot(IG, which = 5,main="Normal Q-Q Plot IG",
282 xlab = "Quantiles of Standard Normal",ylab = "Pearson Residuals")

```

Βιβλιογραφία

Ξενόγλωσση

- [1] Ronald L Bornhuetter and Ronald E Ferguson. “The actuary and IBNR”. In: *Proceedings of the casualty actuarial society*. Vol. 59. 112. 1972, pp. 181–195.
- [2] Alessandro Carrato and G McGuire. “A practitioner’s introduction to stochastic reserving”. In: *en. In* (2016), p. 75.
- [3] Piet De Jong and Gillian Z Heller. *Generalized Linear Models for Insurance Data 2008*.
- [4] Peter D England and Richard J Verrall. “Stochastic claims reserving in general insurance”. In: *British Actuarial Journal* 8.3 (2002), pp. 443–518.
- [5] James W Hardin et al. *Generalized linear models and extensions*. Stata press, 2007.
- [6] David Hindley. *Claims Reserving in General Insurance*. Cambridge University Press, 2017.
- [7] Annette J.Dobson. *An Introduction To Generalized Linear Models Second Edition*. 2002. URL: <http://www.math.mcgill.ca/~dstephens/OldCourses/556-2006/Math556-ModesOfConvergence.pdf>.
- [8] Thomas Mack. “Distribution-free calculation of the standard error of chain ladder reserve estimates”. In: *ASTIN Bulletin: The Journal of the IAA* 23.2 (1993), pp. 213–225.
- [9] PM McCullough and JA Nelder. *Generalized linear models 2nd edition*. 1989.
- [10] John Ashworth Nelder and Robert WM Wedderburn. “Generalized linear models”. In: *Journal of the Royal Statistical Society: Series A (General)* 135.3 (1972), pp. 370–384.
- [11] Arthur E Renshaw. “Modelling the claims process in the presence of covariates”. In: *ASTIN Bulletin: The Journal of the IAA* 24.2 (1994), pp. 265–285.

- [12] Arthur E Renshaw and Richard J Verrall. “A stochastic model underlying the chain-ladder technique”. In: *British Actuarial Journal* 4.4 (1998), pp. 903–923.
- [13] Greg C Taylor and FR Ashe. “Second moments of estimates of outstanding claims”. In: *Journal of Econometrics* 23.1 (1983), pp. 37–61.
- [14] Richard J Verrall. “Bayes and empirical Bayes estimation for the chain ladder model”. In: *ASTIN Bulletin: The Journal of the IAA* 20.2 (1990), pp. 217–243.
- [15] Richard J Verrall. “On the estimation of reserves from loglinear models”. In: *Insurance: mathematics and economics* 10.1 (1991), pp. 75–80.
- [16] RJ Verrall and PD England. “Comments on: “A comparison of stochastic models that reproduce chain ladder reserve estimates”, by Mack and Venter”. In: *Insurance: mathematics and economics* 26.1 (2000), pp. 109–111.
- [17] Mario V Wüthrich and Michael Merz. *Stochastic claims reserving methods in insurance*. Vol. 435. John Wiley & Sons, 2008.

Ελληνική

- [18] Γεώργιος Ηλιόπουλος. *Σημειώσεις Γενικευμένα Γραμμικά Μοντέλα*. 2017.
- [19] Γεώργιος Πιτσέλης. *Σημειώσεις Θεωρία Αξιοπιστίας Χαρτοφυλακίου και Αποθεματοποίηση, Π.Μ.Σ. «Αναλογιστική Επιστήμη & Διοικητική Κινδύνου»*. 2016.