

ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΕΙΡΑΙΩΣ
Σχολή Χρηματοοικονομικής και Στατιστικής



Τμήμα Στατιστικής και Ασφαλιστικής Επιστήμης

**ΜΕΤΑΠΤΥΧΙΑΚΟ ΠΡΟΓΡΑΜΜΑ ΣΠΟΥΔΩΝ
ΣΤΗΝ ΑΝΑΛΟΓΙΣΤΙΚΗ ΕΠΙΣΤΗΜΗ ΚΑΙ
ΔΙΑΧΕΙΡΙΣΗ ΚΙΝΔΥΝΩΝ**

**ΜΕΛΕΤΗ ΜΟΝΤΕΛΩΝ ΘΝΗΣΙΜΟΤΗΤΑΣ
ΜΕ ΒΑΣΗ ΤΟ ΠΛΑΙΣΙΟ ΤΩΝ
ΓΕΝΙΚΕΥΜΕΝΩΝ ΓΡΑΜΜΙΚΩΝ
ΜΟΝΤΕΛΩΝ**

Δημήτριος Α. Παπακώστας

Διπλωματική Εργασία

που υποβλήθηκε στο Τμήμα Στατιστικής και Ασφαλιστικής
Επιστήμης του Πανεπιστημίου Πειραιώς ως μέρος των
απαιτήσεων για την απόκτηση του Μεταπτυχιακού
Διπλώματος Ειδίκευσης στην *Αναλογιστική Επιστήμη και
Διαχείριση Κινδύνων*.

Πειραιάς
Σεπτέμβριος 2023



ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΕΙΡΑΙΩΣ

ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΕΙΡΑΙΩΣ
Σχολή Χρηματοοικονομικής και Στατιστικής



Τμήμα Στατιστικής και Ασφαλιστικής Επιστήμης

**ΜΕΤΑΠΤΥΧΙΑΚΟ ΠΡΟΓΡΑΜΜΑ ΣΠΟΥΔΩΝ
ΣΤΗΝ ΑΝΑΛΟΓΙΣΤΙΚΗ ΕΠΙΣΤΗΜΗ ΚΑΙ
ΔΙΑΧΕΙΡΙΣΗ ΚΙΝΔΥΝΩΝ**

**ΜΕΛΕΤΗ ΜΟΝΤΕΛΩΝ ΘΝΗΣΙΜΟΤΗΤΑΣ
ΜΕ ΒΑΣΗ ΤΟ ΠΛΑΙΣΙΟ ΤΩΝ
ΓΕΝΙΚΕΥΜΕΝΩΝ ΓΡΑΜΜΙΚΩΝ
ΜΟΝΤΕΛΩΝ**

Δημήτριος Α. Παπακώστας

Διπλωματική Εργασία

που υποβλήθηκε στο Τμήμα Στατιστικής και Ασφαλιστικής
Επιστήμης του Πανεπιστημίου Πειραιώς ως μέρος των
απαιτήσεων για την απόκτηση του Μεταπτυχιακού
Διπλώματος Ειδίκευσης στην Αναλογιστική Επιστήμη και
Διαχείριση Κινδύνων.

Πειραιάς
Σεπτέμβριος 2023

ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΕΙΡΑΙΩΣ

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

Τα μέλη της Επιτροπής ήταν:

- Πολίτης Κωνσταντίνος, Αναπληρωτής Καθηγητής (Επιβλέπων)
- Βερροπούλου Γεωργία, Καθηγήτρια
- Μπερσίμης Σωτήριος, Καθηγητής

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

UNIVERSITY OF PIRAEUS
School of Finance and Statistics



Department of Statistics and Insurance Science

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

**STUDY OF MORTALITY MODELS BASED
ON THE FRAMEWORK OF GENERALIZED
LINEAR MODELS**

By

Dimitrios A. Papakostas

MSc Dissertation

submitted to the Department of Statistics and Insurance
Science of the University of Piraeus in partial fulfilment of
the requirements for the degree of Master of Science in
Actuarial Science & Risk Management.

Piraeus, Greece
September 2023

ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΕΙΡΑΙΩΣ

*Στους γονείς μου
Αποστόλη και Χαρούλα*

*Στην αδερφή μου
Σοφία*

Στην Έφη

ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΕΙΡΑΙΩΣ

Ευχαριστίες

Αρχικά, θα ήθελα να ευχαριστήσω θερμά τον Αναπληρωτή Καθηγητή του τμήματος Στατιστικής και Ασφαλιστικής Επιστήμης κύριο Πολίτη Κωνσταντίνο για τη δυνατότητα που μου έδωσε να ασχοληθώ με το αντικείμενο της διπλωματικής μου εργασίας και την πολύτιμη καθοδήγησή του κατά τη διάρκεια της εκπόνησής της. Επιπλέον, θα ήθελα να ευχαριστήσω θερμά τον Επίκουρο Καθηγητή του τμήματος Στατιστικής και Ασφαλιστικής Επιστήμης κύριο Μποζικά Απόστολο για την σημαντική καθοδήγηση και την πολύτιμη βοήθεια που μου παρείχε καθ' όλη τη διάρκεια της συγγραφής της παρούσας διπλωματικής εργασίας. Επιπλέον ευχαριστώ την Καθηγήτρια κυρία Βερροπούλου Γεωργία και τον Αναπληρωτή Καθηγητή κύριο Μπερσίμη Σωτήριο που μου έκαναν την τιμή να είναι μέλη της συμβουλευτικής μου επιτροπής. Ευχαριστώ τους γονείς μου για την στήριξη και τη βοήθειά τους κατά τη διάρκεια εκπόνησης της εργασίας μου. Τέλος, ευχαριστώ την Έφη καθώς συνέβαλε με την παρουσία της, την συμπαράσταση και την εμπύχωσή της.

Περίληψη

Στην παρούσα διπλωματική εργασία θα γίνει παρουσίαση της δημοφιλούς μεθόδου μοντελοποίησης και πρόβλεψης θνησιμότητας Lee – Carter. Θα περιγραφούν οι βασικές υποθέσεις στις οποίες στηρίζεται το μοντέλο Lee – Carter καθώς και ορισμένες βασικές μέθοδοι εκτίμησης των παραμέτρων του μοντέλου. Στη συνέχεια, θα γίνει εισαγωγή στη θεωρία των γενικευμένων γραμμικών μοντέλων και θα περιγραφούν τα μοντέλα των κατανομών Poisson, Αρνητικής Διωνυμικής και Διωνυμικής. Εν συνεχεία, θα μελετήσουμε το μοντέλο Lee – Carter υπό το πρίσμα της θεωρίας των γενικευμένων γραμμικών μοντέλων. Σε επόμενο κεφάλαιο θα γίνει εφαρμογή της παραπάνω θεωρίας μέσω του στατιστικού πακέτου της R στα δεδομένα θνησιμότητας της Ελλάδας για τα έτη 1981 – 2019 και τις ηλικίες 0 – 85. Για την άντληση συμπερασμάτων θα προχωρήσουμε σε σύγκριση των αποτελεσμάτων ορισμένων κριτηρίων καλής προσαρμογής όπως είναι το κριτήριο της απόκλισης, τα κριτήρια πληροφορίας των Akaike και Schwarz καθώς και στη σύγκριση των διαγραμμάτων κανονικοποιημένων καταλοίπων έναντι των προσαρμοσμένων τιμών των θανάτων. Σε τελικό βήμα, θα προχωρήσουμε στη σύγκριση των αποτελεσμάτων μας με αποτελέσματα από τη διεθνή βιβλιογραφία.

ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΕΙΡΑΙΩΣ

Abstract

In the present thesis, the popular Lee – Carter method for modelling and forecasting mortality will be presented. The basic assumptions on which the Lee – Carter model is based will be described as well as some basic methods for estimating model's parameters. Then, the theory of generalized linear models will be introduced and Poisson, Negative Binomial, and Binomial models will be described. Next, we will study the Lee – Carter model in the light of the theory of generalized linear models. In the next chapter, the above theory will be applied through the statistical package of R to the mortality data of Greece for the years 1981 – 2019 and the ages 0 – 85. To draw conclusions, we will compare the results of some goodness of fit criteria such as the deviance criterion, Akaike and Schwarz information criteria as well as the comparison of plots of normalized residuals against adjusted death values. In the final step, we will compare the results of current thesis with those of the international literature.

ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΕΙΡΑΙΩΣ

Περιεχόμενα

Περίληψη.....	- 1 -
Abstract	- 3 -
Περιεχόμενα	- 5 -
Κατάλογος Πινάκων.....	- 9 -
Κατάλογος Διαγραμμάτων	- 11 -
ΚΕΦΑΛΑΙΟ 1.....	- 15 -
Εισαγωγή.....	- 15 -
1.1 Περιγραφή Δεδομένων	- 19 -
1.2 Δεδομένα θανάτων ελληνικού πληθυσμού.....	- 20 -
1.3 Έκθεση στον κίνδυνο.....	- 24 -
1.4 Ειδικός κατά ηλικία δείκτης θνησιμότητας.....	- 25 -
Κ Ε Φ Α Λ Α Ι Ο 2	- 30 -
Μοντέλο θνησιμότητας Lee - Carter.....	- 30 -
2.1 Προβολή θνησιμότητας.....	- 30 -
2.2 Περιγραφή μοντέλου θνησιμότητας Lee-Carter.....	- 31 -
2.3 Μέθοδοι εκτίμησης παραμέτρων μοντέλου Lee - Carter.....	- 32 -
2.3.1 Συνήθης Μέθοδος Ελαχίστων Τετραγώνων - Ordinary Least Squares (OLS).....	- 33 -
2.3.2 Μέθοδος Σταθμισμένων Ελαχίστων Τετραγώνων - Weighted Least Squares (WLS).....	- 36 -
2.3.3 Μέθοδος Μέγιστης Πιθανοφάνειας Poisson.....	- 37 -
2.4 Μειονεκτήματα του μοντέλου Lee - Carter	- 39 -
2.5 Εκτίμηση χρονολογικών σειρών και πρόβλεψη μέσω του μοντέλου Lee – Carter	- 39 -
2.5.1 Μεθοδολογία Box - Jenkins.....	- 40 -
2.5.2 Τυχαίος περίπατος με παράμετρο μετατόπισης	- 40 -
Κ Ε Φ Α Λ Α Ι Ο 3	- 42 -
Πλαίσιο Γενικευμένων Γραμμικών μοντέλων	- 42 -
3.1 Εκθετική Οικογένεια κατανομών	- 42 -
3.1.1 Poisson κατανομή.....	- 44 -
3.1.2 Αρνητική Διωνυμική κατανομή.....	- 46 -
3.1.3 Διωνυμική κατανομή.....	- 50 -
3.2 Βασική δομή Γενικευμένων Γραμμικών μοντέλων	- 51 -

3.3	Σύνδεση μεταξύ Γενικευμένων Γραμμικών Μοντέλων και Κλασικού Μοντέλου Γραμμικής Παλινδρόμησης.....	53 -
3.4	Εκτίμηση παραμέτρων Γενικευμένων Γραμμικών Μοντέλων	55 -
3.4.1	Μέθοδος Newton – Raphson	55 -
3.4.2	Αλγόριθμος Επανασταθμιζόμενων Ελαχίστων Τετραγώνων (IRLS).....	56 -
3.5	Διαγνωστικοί έλεγχοι καλής προσαρμογής	60 -
3.5.1	Απόκλιση - Deviance	60 -
3.5.2	Ανάλυση καταλοίπων.....	61 -
3.5.3	Κατάλοιπα Pearson	61 -
3.5.4	Κατάλοιπα απόκλισης (Deviance)	62 -
3.5.5	Έλεγχος χ^2 του Pearson	63 -
3.5.6	Κριτήρια Πληροφορίας	63 -
3.5.7	Συντελεστής Προσδιορισμού R^2	64 -
3.6	Poisson γενικευμένο γραμμικό μοντέλο.....	65 -
3.7	Υπερδιασπορά.....	69 -
3.8	Αρνητικό Διωνυμικό γενικευμένο γραμμικό μοντέλο	70 -
3.9	Διωνυμικό γενικευμένο γραμμικό μοντέλο.....	72 -
3.10	Γενικευμένα μη-γραμμικά μοντέλα	74 -
	Εφαρμογή στα ελληνικά δεδομένα	78 -
4.1	Προσαρμογή του κλασικού μοντέλου Lee-Carter μέσω της R	78 -
4.2	Σύγκριση παρατηρούμενου και εκτιμώμενου αριθμού θανάτων μέσω του κλασικού μοντέλου Lee – Carter.....	80 -
4.3	Εκτίμηση παραμέτρου k_t του ελληνικού πληθυσμού για τα έτη 2020 - 2050	82 -
4.4	Προσαρμογή μοντέλων Lee - Carter στα ελληνικά δεδομένα.....	84 -
4.5	Απόκλιση, κριτήρια πληροφορίας και καλής προσαρμογής	89 -
	ΚΕΦΑΛΑΙΟ 4.....	93 -
	Σύγκριση με αποτελέσματα από τη διεθνή βιβλιογραφία.....	93 -
	ΚΕΦΑΛΑΙΟ 5.....	97 -
	Συμπεράσματα.....	97 -
	ΒΙΒΛΙΟΓΡΑΦΙΑ.....	98 -
	Ελληνική	98 -
	Αγγλόφωνη.....	98 -
	Διαδικτυακές πηγές.....	100 -
	ΠΑΡΑΡΤΗΜΑΤΑ.....	101 -

Παράρτημα Α: Κώδικας R και πίνακες αποτελεσμάτων για το Κεφάλαιο 1. - 101 -
Παράρτημα Β: Κώδικας R και πίνακες αποτελεσμάτων για το Κεφάλαιο 4..... - 109 -

ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΕΙΡΑΙΩΣ

ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΕΙΡΑΙΩΣ

Κατάλογος Πινάκων

Πίνακας 1.1: Αριθμός θανάτων για ηλικία 4 ετών (έτος 2006) και 7 ετών (έτος 2019).....	- 20 -
Πίνακας 4.1: Προβλέψεις παραμέτρου κ μέσω των μοντέλων ARIMA για τα έτη 2020 – 2050.	- 83 -
Πίνακας 4.2: Παράμετρος διασποράς φ	- 89 -
Πίνακας 4.3: Απόκλιση και κριτήρια πληροφορίας.....	- 90 -
Πίνακας 5.1: Αποτελέσματα απόκλισης τριών μοντέλων Lee - Carter για 11 χώρες (1981 - 2019).....	- 94 -
Πίνακας 5.2: Κριτήρια πληροφορίας AIC και BIC για τα δεδομένα του Βελγίου.....	- 95 -
Πίνακας 5.3: Παράμετρος διασποράς φ για τα δεδομένα του Βελγίου.....	- 96 -

ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΕΙΡΑΙΩΣ

Κατάλογος Διαγραμμάτων

Διάγραμμα 1.1: (α) Προσδόκιμο Ανδρών και Γυναικών, (β) Αριθμός γεννήσεων στην Ελλάδα για τα έτη 1981 - 2019.	- 16 -
Διάγραμμα 1.2: Αριθμός θανάτων: (α) ανδρικού πληθυσμού ηλικίας 0 - 85 για τα έτη 1981 - 2019, (β) γυναικείου πληθυσμού ηλικίας 0 - 85 για τα έτη 1981 - 2019.	- 21 -
Διάγραμμα 1.3: Αριθμός θανάτων: (α) ανδρικού πληθυσμού ηλικίας 67 ετών, (β) ανδρικού πληθυσμού ανά ηλικία για τα έτη 1981 και 2019.	- 22 -
Διάγραμμα 1.4: Αριθμός θανάτων: (α) γυναικείου πληθυσμού ηλικίας 67 ετών, (β) γυναικείου πληθυσμού ανά ηλικία για τα έτη 1981 και 2019.	- 23 -
Διάγραμμα 1.5: Αριθμός θανάτων γυναικείου και ανδρικού πληθυσμού για το έτος 2019.-	24 -
Διάγραμμα 1.6: Έκθεση στον κίνδυνο: (α) ανδρικού πληθυσμού για τα έτη 1981 και 2019, (β) γυναικείου πληθυσμού για τα έτη 1981 και 2019.	- 25 -
Διάγραμμα 1.7: (α) Ειδικός κατά ηλικία δείκτης θνησιμότητας συνολικού πληθυσμού για τα έτη 1981 και 2019, (β) Λογάριθμος ειδικού κατά ηλικία δείκτη θνησιμότητας συνολικού πληθυσμού για τα έτη 1981 και 2019.	- 27 -
Διάγραμμα 1.8: Λογάριθμος δείκτη θνησιμότητας ανδρικού πληθυσμού: (α) ηλικίας 67 ετών, (β) ανά ηλικία για το έτος 2019.	- 28 -
Διάγραμμα 1.9: Λογάριθμος δείκτη θνησιμότητας γυναικείου πληθυσμού: (α) ηλικίας 67 ετών, (β) ανά ηλικία συγκριτικά για τα έτη 1981 και 2019.	- 29 -
Διάγραμμα 1.10: Λογάριθμος δείκτη θνησιμότητας ανδρικού έναντι γυναικείου πληθυσμού για το έτος 2019.	- 29 -
Διάγραμμα 4.1: Διάγραμμα δείκτη θνησιμότητας ανά ηλικία: (α) του ανδρικού πληθυσμού, (β) του γυναικείου πληθυσμού, (γ) του συνολικού πληθυσμού.	- 79 -
Διάγραμμα 4.2: Εκτίμηση παραμέτρου: (α) α_x , (β) β_x , (γ) k_t για άνδρες, γυναίκες και συνολικό πληθυσμό ως προς την ηλικία.	- 80 -
Διάγραμμα 4.3: Θνησιμότητα ετών 1981 και 2019 έναντι μέσης θνησιμότητας για όλες τις ηλικίες για (α) άνδρες και (β) γυναίκες.	- 81 -
Διάγραμμα 4.4: Εκτιμήσεις θνησιμότητας (α) ανδρών, (β) γυναικών ετών 1981 και 2019 έναντι αντίστοιχων παρατηρούμενων τιμών.	- 82 -
Διάγραμμα 4.5: Πρόβλεψη τιμών παραμέτρου k_t (α) του συνολικού πληθυσμού, (β) των ανδρών, (γ) των γυναικών για τα έτη 2020 - 2050.	- 82 -

Διάγραμμα 4.6: Πρόβλεψη της παραμέτρου kt για άνδρες, γυναίκες και συνολικό πληθυσμό της Ελλάδας για τα έτη 2020 - 2050.	- 84 -
Διάγραμμα 4.7: Κανονικοποιημένα κατάλοιπα έναντι εκτιμήσεων ανδρικού πληθυσμού για τα μοντέλα: (α) Poisson Lee - Carter, (β) Αρνητικό Διωνυμικό Lee - Carter και (γ) Διωνυμικό Lee - Carter.....	- 85 -
Διάγραμμα 4.8: Γραφική παράσταση των δειγματικών ποσοστημορίων ως προς τα θεωρητικά ποσοστημόρια της Κανονικής Κατανομής για τα μοντέλα ανδρικού πληθυσμού: (α) Poisson Lee - Carter, (β) Αρνητικό Διωνυμικό Lee - Carter και (γ) Διωνυμικό Lee - Carter.	- 85 -
Διάγραμμα 4.9: Κανονικοποιημένα κατάλοιπα έναντι εκτιμήσεων γυναικείου πληθυσμού για τα μοντέλα: (α) Poisson Lee - Carter, (β) Αρνητικό Διωνυμικό Lee - Carter και (γ) Διωνυμικό Lee - Carter.....	- 86 -
Διάγραμμα 4.10: Γραφική παράσταση των δειγματικών ποσοστημορίων ως προς τα θεωρητικά ποσοστημόρια της Κανονικής Κατανομής για τα μοντέλα γυναικείου πληθυσμού: (α) Poisson Lee - Carter, (β) Αρνητικό Διωνυμικό Lee - Carter και (γ) Διωνυμικό Lee - Carter.	- 87 -
Διάγραμμα 4.11: Κανονικοποιημένα κατάλοιπα έναντι εκτιμήσεων συνολικού πληθυσμού για τα μοντέλα: (α) Poisson Lee - Carter, (β) Αρνητικό Διωνυμικό Lee - Carter και (γ) Διωνυμικό Lee - Carter.....	- 88 -
Διάγραμμα 4.12: Γραφική παράσταση των δειγματικών ποσοστημορίων ως προς τα θεωρητικά ποσοστημόρια της Κανονικής Κατανομής για τα μοντέλα συνολικού πληθυσμού: (α) Poisson Lee - Carter, (β) Αρνητικό Διωνυμικό Lee - Carter και (γ) Διωνυμικό Lee - Carter.	- 88 -
Διάγραμμα 4.13: Σύγκριση εκτίμησης αριθμού θανάτων ατόμων ηλικίας 67 ετών του κλασικού μοντέλου Lee – Carter έναντι του Poisson, Αρνητικού Διωνυμικού και Διωνυμικού για: (α) Άνδρες και (β) Γυναίκες.....	- 91 -
Διάγραμμα 4.14: Σύγκριση εκτίμησης αριθμού θανάτων έτους 2019 του κλασικού μοντέλου Lee – Carter έναντι του Poisson, Αρνητικού Διωνυμικού και Διωνυμικού για: (α) Άνδρες και (β) Γυναίκες.....	- 92 -
Διάγραμμα 4.15: Σύγκριση εκτίμησης παραμέτρου ax για το έτος 2019 του κλασικού μοντέλου Lee – Carter έναντι του Poisson, Αρνητικού Διωνυμικού και Διωνυμικού για: (α) Άνδρες και (β) Γυναίκες.....	- 92 -
Διάγραμμα 5.1: Λογάριθμος θνησιμότητας συνολικού πληθυσμού Ελλάδας και Βελγίου για τα έτη 1981 και 2019.	- 95 -

Διάγραμμα 5.2: Σύγκριση μεταξύ μοντέλων εκτίμησης λογαρίθμου αριθμού θανάτων ατόμων ηλικίας 67 ετών για τα δεδομένα του Βελγίου για: (α) Άνδρες και (β) Γυναίκες..... - 96 -

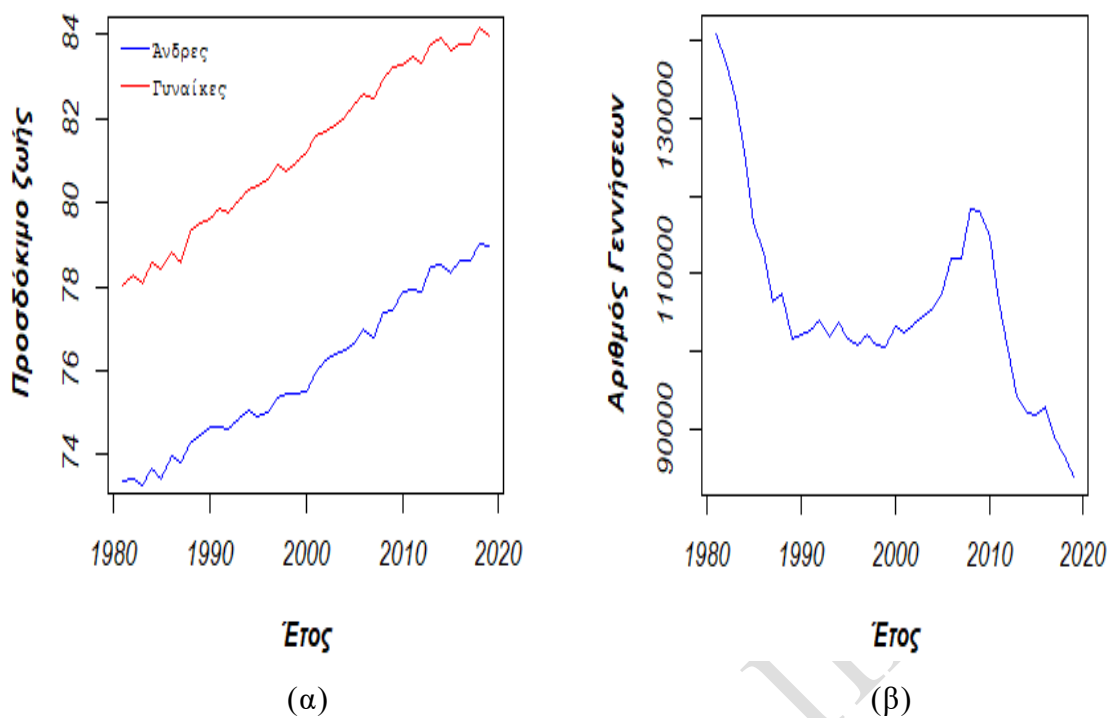
ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΕΙΡΑΙΩΣ

ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΕΙΡΑΙΩΣ

ΚΕΦΑΛΑΙΟ 1

Εισαγωγή

Τα μοντέλα πρόβλεψης θνησιμότητας έχουν απασχολήσει για αρκετές δεκαετίες τους αναλογιστές. Ειδικότερα από το 2019 και μετά με την εμφάνιση της πανδημίας του κορονοϊού (Covid-19) η οποία είχε ως αποτέλεσμα την αύξηση του αριθμού των θανάτων παγκοσμίως. Σε πολλές ανεπτυγμένες χώρες έχει παρατηρηθεί η διαρκώς αυξανόμενη τάση του μέσου προσδόκιμου ζωής τις τελευταίες δεκαετίες. Βάσει στοιχείων της Eurostat το μέσο προσδόκιμο ζωής ενός νεογέννητου παιδιού στην Ευρωπαϊκή Ένωση έχει αυξηθεί κατά 1,5 έτος από το έτος 2010 μέχρι το 2019. Ενώ, το 2020 αυτή η αύξηση του προσδόκιμου ζωής των χωρών της Ευρωπαϊκής Ένωσης διακόπτεται από τις επιπτώσεις της εμφάνισης της πανδημίας του κορονοϊού (Covid-19) τις οποίες παρατηρούμε μέχρι σήμερα και η οποία οδήγησε στη ραγδαία αύξηση του αριθμού των θανάτων παγκοσμίως. Παρά τις επιπτώσεις της πανδημίας, βάσει ερευνών οι χώρες της Ευρωπαϊκής Ένωσης είναι εκείνες που θα αντιμετωπίσουν έντονα τις συνέπειες του φαινομένου της γήρανσης του πληθυσμού. Αναφορικά με την Ελλάδα, παρατηρούμε αύξηση του μέσου προσδόκιμου ζωής για τα έτη 2010 - 2019 κατά ένα έτος, η οποία συνοδεύεται από μείωση του αριθμού των θανάτων.



Διάγραμμα 1.1: (α) Προσδόκιμο Ανδρών και Γυναικών, (β) Αριθμός γεννήσεων στην Ελλάδα για τα έτη 1981 - 2019.

Η μακροζωία είναι το αποτέλεσμα μιας πολύπλοκης αλληλεπίδρασης διαφορετικών παραγόντων όπως είναι ο υγιεινός τρόπος ζωής και η βελτίωση της ποιότητάς της, η καλύτερη εκπαίδευση, η βελτίωση του οικονομικού επιπέδου και η ραγδαία τεχνολογική εξέλιξη που εξυπηρετεί την πρόοδο της διάγνωσης και της ιατρικής θεραπείας αρκετών ασθενειών. Το φαινόμενο της αύξησης του προσδόκιμου ζωής ή αλλιώς της μακροζωίας αποτελεί μια θετική εξέλιξη για την κοινωνία αλλά παράλληλα αποτελεί και κίνδυνο ο οποίος δημιουργεί προκλήσεις όχι μόνο για τον άνθρωπο που χρειάζεται εισόδημα για διάστημα μεγαλύτερο από το αναμενόμενο μετά την συνταξιοδότηση, αλλά και για τα δημόσια ασφαλιστικά ταμεία, τα ιδιωτικά ταμεία επαγγελματικών συνταξιοδοτικών παροχών καθώς και για τις ασφαλιστικές εταιρίες. Σε ατομικό επίπεδο, οι πολίτες θα κληθούν να καλύψουν την υγειονομική και ιατροφαρμακευτική τους περίθαλψη για μεγαλύτερο χρονικό διάστημα. Καθώς η ιατρική έρευνα προχωράει με την βοήθεια της τεχνολογίας, όλο και περισσότερες θεραπείες ασθενειών ανακαλύπτονται, ενώ πολλές από τις ήδη υπάρχουσες θεραπείες βελτιώνονται με αποτέλεσμα να γίνονται όλο και πιο κοστοβόρες. Παράλληλα, οι

πολίτες θα πρέπει να μεριμνήσουν για την εξασφάλιση εισοδημάτων καθώς θα απαιτηθεί να επιβιώσουν για μεγαλύτερο χρονικό διάστημα με εισοδήματα που προέρχονται αποκλειστικά από τη σύνταξή τους και τα οποία θα είναι συρρικνωμένα συγκριτικά με τις απολαβές τους ως εργαζόμενοι. Αναφορικά με το σύστημα κοινωνικής ασφάλισης, η αύξηση του μέσου προσδόκιμου ζωής σε συνδυασμό με τους αυξανόμενους δείκτες γήρανσης του πληθυσμού της Ελλάδας, δηλαδή της αναλογίας των ηλικιωμένων ως προς τον συνολικό πληθυσμό και της συνεχόμενης μείωσης της γεννητικότητας θα επιφέρει σημαντικές επιπτώσεις στα δημόσια ασφαλιστικά ταμεία λόγω των αυξανόμενων υγειονομικών δαπανών και ασφαλιστικών παροχών που θα πρέπει να καλυφθούν για μεγαλύτερο χρονικό διάστημα.

Ο κίνδυνος μακροζωίας για μια ασφαλιστική εταιρία είναι ο κίνδυνος το ασφαλισμένο άτομο να ξεπεράσει το αναμενόμενο προσδόκιμο ζωής και αποτελείται από δύο συνιστώσες. Η πρώτη συνιστώσα είναι ο ατομικός κίνδυνος μακροζωίας κατά τον οποίο το άτομο θα ξεπεράσει το προσδόκιμο ζωής και ο οποίος σε ένα σύνολο όμοιων κινδύνων μπορεί να αντισταθμιστεί, ενώ η δεύτερη συνιστώσα και η σημαντικότερη είναι ο συλλογικός κίνδυνος μακροζωίας που είναι ο κίνδυνος να υποεκτιμηθεί συνολικά το αναμενόμενο προσδόκιμο ζωής. Οι αναλογιστές χρησιμοποιούν αναλογιστικούς πίνακες επιβίωσης προκειμένου να καθορίσουν τα ασφάλιστρα ασφαλίσεων ζωής και συνταξιοδοτικών παροχών. Στην περίπτωση υποεκτίμησης του αναμενόμενου προσδόκιμου ζωής στους εν λόγω πίνακες, οι ασφαλιστικές εταιρίες θα κληθούν να αντλήσουν κεφάλαια για να καλύψουν τα μη επαρκή αποθεματικά τους για τη κάλυψη ασφαλιστικών παροχών επιβίωσης και συνταξιοδότησης. Κατ' επέκταση η ορθή εκτίμηση της μακροζωίας και αντίστοιχα της θνησιμότητας αποτελούν παράγοντες που συνδέονται άμεσα με την ακριβή τιμολόγηση και διατήρηση επαρκών αποθεμάτων για την κάλυψη ασφαλιστικών παροχών επιβίωσης, θανάτου και συνταξιοδότησης.

Συνδυαστικά με τα ανωτέρω, η εφαρμογή από το 2018 της νέας Ευρωπαϊκής νομοθεσίας Φερεγγυότητα II (Solvency II), η οποία δίνει έμφαση στο τρόπο αντιμετώπισης των κινδύνων από τις ασφαλιστικές και αντασφαλιστικές εταιρίες και πώς αυτές διαχειρίζονται τις δυσμενείς οικονομικές και δημογραφικές καταστάσεις με σκοπό την προστασία των ασφαλισμένων, δυσχεραίνει το έργο των ασφαλιστικών εταιριών οι οποίες υποχρεούνται να εναρμονιστούν με τη νέα νομοθεσία και να

διατηρούν υψηλά αποθεματικά. Οι παραπάνω λόγοι καθιστούν πιο επιτακτική από ποτέ τη δυνατότητα πρόβλεψης της θνησιμότητας, κατ' επέκταση και της μακροζωίας καθώς πρόκειται για παράγοντες που συνδέονται άμεσα με τη σωστή τιμολόγηση και διατήρηση αποθεματικών ασφαλιστικών προϊόντων με καλύψεις επιβίωσης, θανάτου καθώς και συνταξιοδοτικών παροχών.

Το μοντέλο που αναπτύχθηκε από τους Lee και Carter (1992) αποτέλεσε μια δημοφιλή μέθοδο μοντελοποίησης και πρόβλεψης διαχρονικά της τάσης του δείκτη θνησιμότητας, κυρίως λόγω της ευκολίας στην εφαρμογή του. Κατά καιρούς έχουν γίνει αρκετές δημοσιεύσεις για στοχαστικά μοντέλα πρόβλεψης θνησιμότητας, κάποια από τα οποία αποτελούν προεκτάσεις του μοντέλου Lee Carter και επτά από τα πιο δημοφιλή συγκρίνονται στη δημοσίευσή του Cairns et al. (2009) ως προς την απόδοσή τους συγκριτικά με αυτό των Lee και Carter.

Σκοπός της παρούσας διπλωματικής εργασίας είναι η παρουσίαση του δημοφιλούς μοντέλου θνησιμότητας των Lee - Carter και η μελέτη του μέσα από το πλαίσιο της θεωρίας των γενικευμένων γραμμικών μοντέλων. Στο πρώτο Κεφάλαιο θα προηγηθεί παρουσίαση των ελληνικών δεδομένων που θα χρησιμοποιηθούν στη διπλωματική εργασία καθώς και ο προσδιορισμός του παρατηρούμενου κατά ηλικία δείκτη θνησιμότητας. Στη συνέχεια, στο Κεφάλαιο 2 θα μελετήσουμε τη μέθοδο πρόβλεψης θνησιμότητας των Lee και Carter καθώς και τις υποθέσεις στις οποίες βασίζεται. Κατόπιν ορισμού των παραμέτρων του μοντέλου, θα προχωρήσουμε στην παρουσίαση μερικών από τις πιο δημοφιλείς μεθόδους εκτίμησής τους. Ολοκληρώνοντας με το Κεφάλαιο 2, θα γίνει αναφορά στη θεωρία μοντέλων πρόβλεψης χρονοσειρών την οποία εισήγαγαν οι Box και Jenkins.

Στο Κεφάλαιο 3 θα αναφερθούμε στη βασική δομή των Γενικευμένων Γραμμικών Μοντέλων τα οποία αναπτύχθηκαν από τους Nelder και Wedderburn το 1972. Στο κεφάλαιο αυτό θα μελετήσουμε τρεις γνωστές κατανομές οι οποίες ανήκουν στην κλάση της Εκθετικής Οικογένειας Κατανομών και πώς μοντελοποιούνται υπό το πλαίσιο των γενικευμένων γραμμικών μοντέλων. Επιπλέον, θα γίνει αναφορά στις πιο γνωστές μεθόδους εκτίμησης των παραμέτρων των μοντέλων τα οποία ανήκουν στην οικογένεια των γενικευμένων γραμμικών μοντέλων καθώς και σε ορισμένα κριτήρια καλής προσαρμογής όπως είναι η απόκλιση (deviance) και τα κριτήρια πληροφορίας των Akaike και Schwarz. Εν συνέχεια, θα μελετήσουμε το μοντέλο των Lee - Carter

υπό το πρίσμα των γενικευμένων γραμμικών μοντέλων και πιο συγκεκριμένα ως γενικευμένο μη-γραμμικό μοντέλο λόγω του πολλαπλασιαστικού παράγοντα που περιλαμβάνει. Τέλος, θα προσαρμόσουμε το γενικευμένο μη- γραμμικό μοντέλο Lee - Carter μέσω των μοντέλων της Poisson, της Αρνητικής Διωνυμικής και της Διωνυμικής κατανομής.

Στο Κεφάλαιο 4 θα εφαρμόσουμε στα ελληνικά δεδομένα με τη βοήθεια του στατιστικού πακέτου ανοιχτού κώδικα της R το απλό μοντέλο Lee - Carter μέσω της εντολής lca() του πακέτου demography – version 2.0 (Hyndman et al, 2023). Επιπλέον, θα εκτιμηθούν μέσω της μεθόδου ARIMA οι τιμές της παραμέτρου k_t που αναφέρεται στο γενικό επίπεδο θνησιμότητας του μοντέλου για τα έτη 2020 - 2050. Στη συνέχεια, θα εφαρμόσουμε στα δεδομένα του ελληνικού πληθυσμού τα μοντέλα Poisson Lee - Carter, Αρνητικό Διωνυμικό Lee – Carter, Διωνυμικό Lee – Carter και θα συγκρίνουμε τα αποτελέσματά τους. Για την άντληση περαιτέρω συμπερασμάτων, θα προχωρήσουμε σε σύγκριση των αποτελεσμάτων ορισμένων κριτηρίων καλής προσαρμογής όπως είναι η απόκλιση deviance, τα κριτήρια πληροφορίας των Akaike και Schwarz καθώς και στη σύγκριση των διαγραμμάτων κανονικοποιημένων καταλοίπων έναντι των προσαρμοσμένων τιμών των θανάτων.

Στο Κεφάλαιο 5 γίνεται αναφορά σε αποτελέσματα από τη διεθνή βιβλιογραφία και σύγκριση αποτελεσμάτων με την παρούσα μελέτη. Τέλος, καταλήγουμε στο συμπέρασμα ότι κατόπιν προσαρμογής των τριών μοντέλων, το Αρνητικό Διωνυμικό μοντέλο Lee - Carter υπερτερεί έναντι του Διωνυμικού μοντέλου Lee - Carter και του Poisson μοντέλου Lee - Carter καθώς παρουσιάζει καλύτερη προσαρμογή στα δεδομένα της Ελλάδας.

1.1 Περιγραφή Δεδομένων

Η εξαγωγή των δεδομένων για την παρούσα διπλωματική εργασία πραγματοποιήθηκε από τη Βάση Δεδομένων (Human Mortality Database, www.mortality.gr), η οποία δημιουργήθηκε και διατηρείται από το τμήμα Δημογραφίας του Πανεπιστημίου της Καλιφόρνια των Η.Π.Α. (Department of Demography at the University of California, Berkeley, USA) και το Max Planck Ερευνητικό Ινστιτούτο Δημογραφίας του Ροστόκ της Γερμανίας (Max Planck Institute for Demographic Research in Rostock, Germany) με σκοπό να παρέχει πρόσβαση σε λεπτομερή δημογραφικά δεδομένα και δεδομένα

θνησιμότητας για 41 χώρες μέχρι αυτή τη στιγμή. Συγκεκριμένα, τα δημογραφικά δεδομένα που χρησιμοποιήθηκαν στην παρούσα διπλωματική εργασία είναι ο αριθμός θανάτων και ο πληθυσμός της Ελλάδας (αναφέρεται και ως έκθεση στον κίνδυνο). Το σύνολο των δεδομένων αφορά την χρονική περίοδο 1981 - 2019 και είναι ταξινομημένο ανά έτος και φύλο για τις ηλικίες από 0 έως 85 ετών.

1.2 Δεδομένα θανάτων ελληνικού πληθυσμού

Στην παράγραφο αυτή θα συγκρίνουμε και θα αναλύσουμε κάποια στοιχεία που αφορούν τον συνολικό αριθμό θανάτων του ελληνικού πληθυσμού ανά φύλο, ενώ παράλληλα θα παραθέσουμε ορισμένα διαγράμματα τα οποία και θα σχολιάσουμε. Αξίζει να αναφερθεί ότι στα δεδομένα του γυναικείου πληθυσμού για τις ηλικίες 4 και 7 για τα έτη 2006 και 2019 αντίστοιχα, καταγράφεται μηδενικός αριθμός θανάτων σε σύνολο πληθυσμού 52382 και 51331, αντίστοιχα.

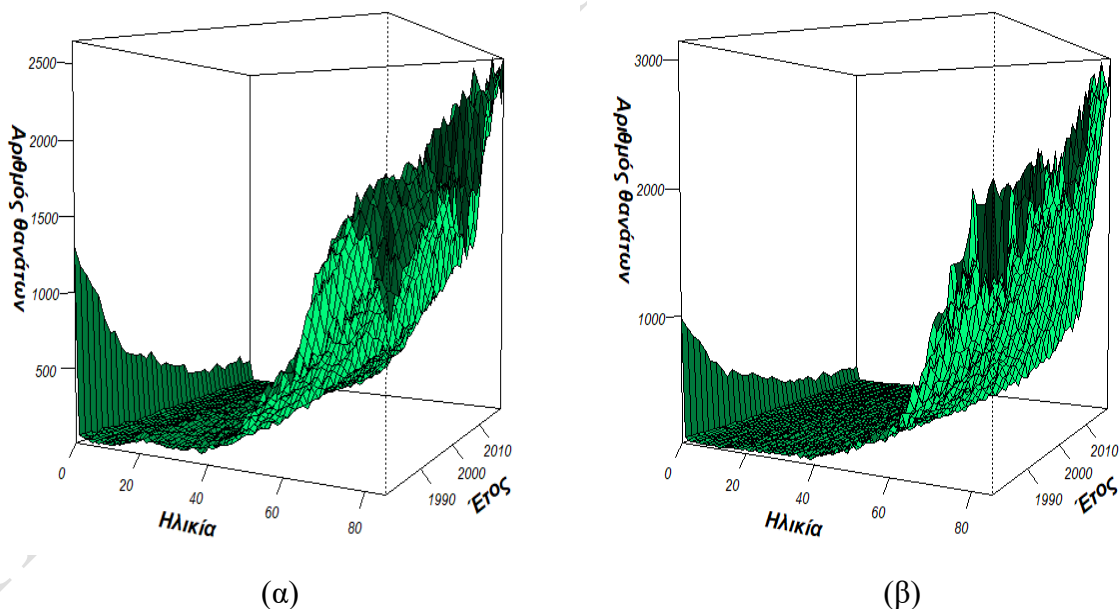
Ο ειδικός κατά ηλικία δείκτης θνησιμότητας $m_{x,t}$ που θα μελετήσουμε ακολούθως ορίζεται ως το πηλίκο του παρατηρούμενου αριθμού θανάτων για κάθε ηλικία x ως προς τον συνολικό πληθυσμό στο μέσον του t έτους (συχνά αναφέρεται ως έκθεση στον κίνδυνο). Καθώς η μελέτη των μοντέλων που θα δούμε στη συνέχεια βασίζεται στον λογάριθμο του κατά ηλικία δείκτη θνησιμότητας υποθέτουμε ότι: α) ο παρατηρούμενος αριθμός θανάτων στην ηλικία 4, το έτος 2006 για τον γυναικείο πληθυσμό ισούται με τον μέσο όρο των παρατηρούμενων αριθμών θανάτων των ετών 2005 και 2007 στην ηλικία των 4 χρόνων, β) ο αριθμός των θανάτων στην ηλικία 7 το έτος 2019 για τον γυναικείο πληθυσμό ισούται με τον παρατηρούμενο αριθμό θανάτων του έτους 2018 της ίδιας ηλικιακής ομάδας όπως παρουσιάζεται στον ακόλουθο πίνακα.

Πίνακας 1.1: Αριθμός θανάτων για ηλικία 4 ετών (έτος 2006) και 7 ετών (έτος 2019).

Ηλικία	Έτος	Αριθμός θανάτων	Έκθεση στον Κίνδυνο	Πληθυσμός
4	2005	10	52503	52373
4	2006	8	52019	52382
4	2007	5	51905	52148
7	2018	4	52277	53431
7	2019	4	50731	51331
7	2020	<i>Δεν υπάρχουν διαθέσιμα δεδομένα</i>		

Στο ακόλουθο Διάγραμμα 1.2(α) για τον συνολικό αριθμό των θανάτων του ανδρικού πληθυσμού για τις ηλικίες 0 έως 85 και για τα έτη 1981 έως 2019 παρατηρείται υψηλή βρεφική θνησιμότητα το έτος 1981 η οποία σταδιακά φθίνει με την πάροδο των ετών, με τον αριθμό να μειώνεται κάτω από 500 θανάτους ανά έτος από το έτος 2000 και μετά. Πιο συγκεκριμένα, βάσει των επίσημων στοιχείων της ελληνικής Στατιστικής Αρχής (ΕΛ.ΣΤΑΤ.) ο συνολικός αριθμός θανάτων αρρένων ηλικίας έως ενός έτους τη χρονιά 2000 είναι 323. Επίσης, παρατηρείται ότι με την πάροδο των ετών οι υψηλότεροι αριθμοί θανάτων σημειώνονται σε μεγαλύτερες ηλικίες λόγω του αυξανόμενου προσδόκιμου ζωής.

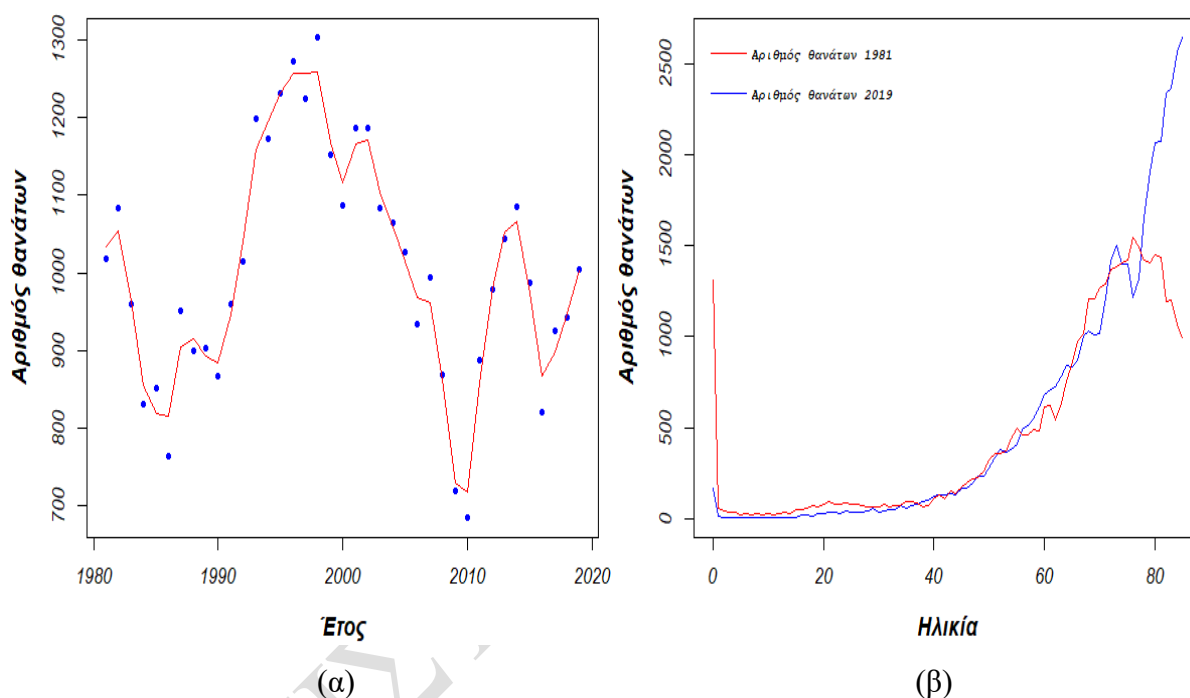
Αντίστοιχα, στο Διάγραμμα 2(β) για τον συνολικό αριθμό των θανάτων του γυναικείου πληθυσμού παρατηρείται επίσης υψηλή βρεφική θνησιμότητα το έτος 1981, η οποία σταδιακά φθίνει με την πάροδο των ετών, με τον αριθμό των θανάτων να πέφτει κάτω από τους 500 από το 1995 και μετά. Βάσει των στοιχείων της ΕΛ.ΣΤΑΤ ο αριθμός των θανάτων των θηλέων ηλικίας έως ενός έτους την χρονιά 2000 ήταν 238. Αναφορικά με την βρεφική θνησιμότητα και συγκριτικά με τα δύο διαγράμματα διαπιστώνεται ότι αυτή του ανδρικού πληθυσμού ξεπερνά αυτή του γυναικείου το έτος 1981 με την διαφορά τους να εξαλείφεται και τους αριθμούς να κυμαίνονται στα ίδια επίπεδα από το έτος 1990 και έπειτα.



Διάγραμμα 1.2: Αριθμός θανάτων: (α) ανδρικού πληθυσμού ηλικίας 0 - 85 για τα έτη 1981 - 2019, (β) γυναικείου πληθυσμού ηλικίας 0 - 85 για τα έτη 1981 - 2019.

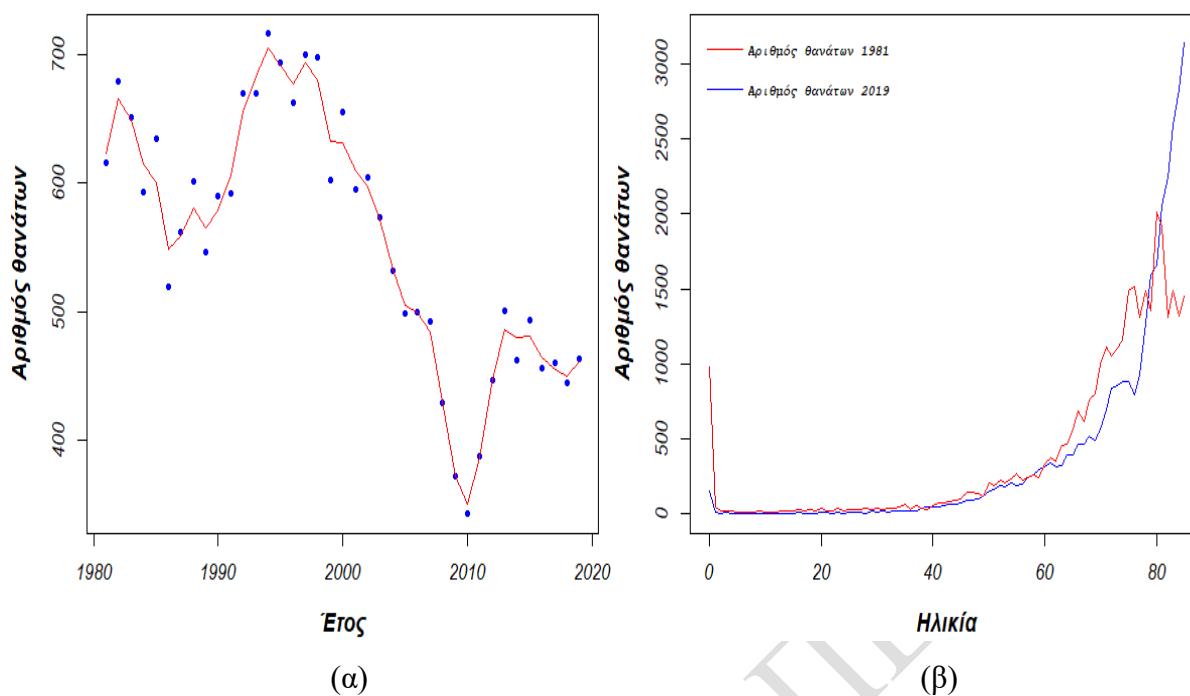
Στο Διάγραμμα 1.3(α) παρατηρούμε σημαντική μείωση στον αριθμό των θανάτων ανδρικού πληθυσμού ηλικίας 67 ετών στο χρονικό διάστημα από το έτος 2000 έως 2010 όπου και

καταγράφεται η ελάχιστη τιμή του 684 (μέγιστη τιμή 1303 το έτος 1998), ενώ ο αριθμός αυτός αυξάνεται εκ νέου από το 2010 και για την επόμενη πενταετία, περίοδος κατά την οποία από τα στοιχεία της ΕΛ.ΣΤΑΤ προκύπτει αύξηση των περιστατικών νεοπλασμάτων καθώς και άλλων μη καθορισμένων βίαιων πράξεων. Παράλληλα, στο Διάγραμμα 1.3(β) διαπιστώνουμε όπως είναι αναμενόμενο ότι οι θάνατοι για το έτος 2019 αυξάνονται γραμμικά μέχρι την ηλικία των 60 όπου ο αριθμός παρουσιάζει απότομη αύξηση, με μια στασιμότητα στις ηλικίες 77-79 όπου και παρατηρείται εντονότερη αύξηση με τον μέγιστο αριθμό να καταγράφεται κοντά στην ηλικία των 85 ετών.



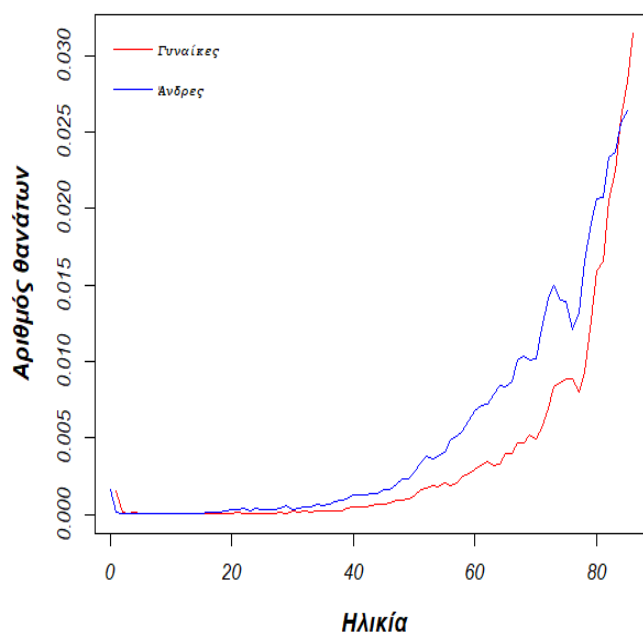
Διάγραμμα 1.3: Αριθμός θανάτων: (α) ανδρικού πληθυσμού ηλικίας 67 ετών, (β) ανδρικού πληθυσμού ανά ηλικία για τα έτη 1981 και 2019.

Στο διάγραμμα 1.4(α) παρατηρούμε ότι ο αριθμός θανάτων του γυναικείου πληθυσμού ηλικίας 67 ετών μειώνεται συνεχόμενα από το 1995 μέχρι το έτος 2010 όπου και καταγράφεται η ελάχιστη τιμή 343 (μέγιστη τιμή 716 το έτος 1994) για το χρονικό διάστημα που μελετάμε. Μεταξύ των ετών 2010 – 2012 παρατηρείται αύξηση των θανάτων στον εν λόγω πληθυσμό με την τάση να είναι και πάλι πτωτική μετά το 2012. Από το διάγραμμα 1.4(β) παρατηρούμε έντονες διαφορές στην βρεφική θνησιμότητα η οποία έχει μειωθεί σημαντικά το 2019. Παράλληλα, εντοπίζεται μείωση του αριθμού των θανάτων από την ηλικία των 62 ετών μέχρι την ηλικία των 78 ετών λόγω της αύξησης του προσδόκιμου ζωής, ενώ ο αριθμός των θανάτων αυξάνεται απότομα από την ηλικία των 78 ετών και μετέπειτα για το έτος 2019.



Διάγραμμα 1.4: Αριθμός θανάτων: (α) γυναικείου πληθυσμού ηλικίας 67 ετών, (β) γυναικείου πληθυσμού ανά ηλικία για τα έτη 1981 και 2019.

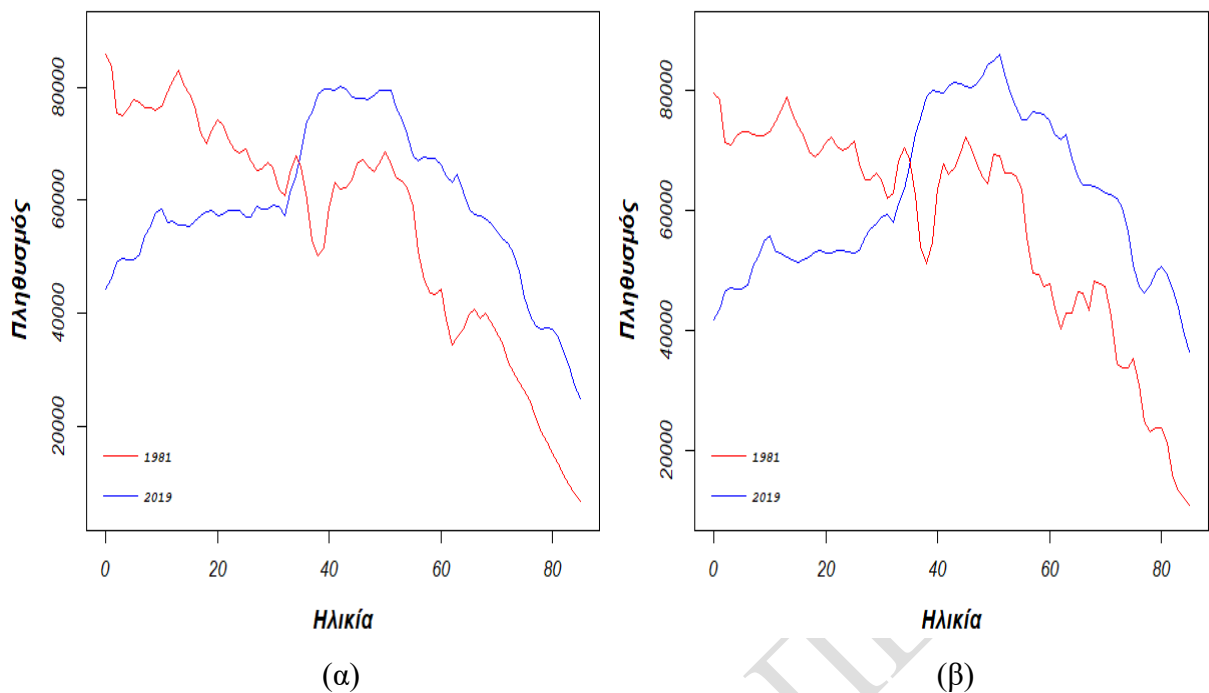
Στο Διάγραμμα 1.5 παρουσιάζεται ο αριθμός θανάτων γυναικείου και ανδρικού πληθυσμού ανά 100.000 για το έτος 2019. Παρατηρείται σημαντική αύξηση μετά την ηλικία των 40 ετών στους θανάτους του ανδρικού πληθυσμού έναντι αυτού των γυναικών, με τους περισσότερους να ταξινομούνται σε παθολογικά αίτια (ΕΛ.ΣΤΑΤ.). Ο μέγιστος αριθμός θανάτων για το έτος 2019 καταγράφεται στην ηλικία των 85 ετών με τον αριθμό για τις γυναίκες να κυμαίνεται στους 3144 και για τους άνδρες στους 2642, αντίστοιχα. Αξίζει να σημειωθεί ότι οι ελάχιστες τιμές θανάτων για το έτος 2019 και για τα δύο φύλα παρατηρούνται μεταξύ των ηλικιών 8 έως 10.



Διάγραμμα 1.5: Αριθμός θανάτων γυναικείου και ανδρικού πληθυσμού για το έτος 2019.

1.3 Έκθεση στον κίνδυνο

Στην παράγραφο αυτή θα αναλύσουμε και θα συγκρίνουμε κάποια διαγράμματα που αφορούν τον συνολικό αριθμό έκθεσης του ελληνικού πληθυσμού ηλικίας 0-85 στον κίνδυνο θανάτου ανά φύλο για τα έτη 1981 – 2019. Ως έκθεση στον κίνδυνο αναφέρεται ο συνολικός πληθυσμός στο μέσο του έτους. Από το Διάγραμμα 1.6(α) των ανδρών εντοπίζεται ότι ο πληθυσμός ηλικίας 40-85 στο μέσο του 2019 είναι αισθητά αυξημένος λόγω της βελτίωσης του προσδόκιμου ζωής και της γήρανσης του πληθυσμού ενώ αντίθετα ο πληθυσμός βρεφικής ηλικίας, δηλαδή μέχρι ενός έτους για το 2019 είναι ο μισός από τον αντίστοιχο πληθυσμό του 1981 λόγω του μειωμένου αριθμού γεννήσεων. Αντίστοιχα, στο Διάγραμμα 1.6(β) καταλήγουμε στα ίδια συμπεράσματα για τον γυναικείο πληθυσμό.



Διάγραμμα 1.6: Έκθεση στον κίνδυνο: (α) ανδρικού πληθυσμού για τα έτη 1981 και 2019, (β) γυναικείου πληθυσμού για τα έτη 1981 και 2019.

1.4 Ειδικός κατά ηλικία δείκτης θνησιμότητας

Πριν προχωρήσουμε στον ορισμό του ειδικού κατά ηλικία δείκτη θνησιμότητας $m_{x,t}$ για την καλύτερη κατανόηση του αναγνώστη θα χρειαστεί να εισάγουμε τις ακόλουθες ποσότητες. Ας συμβολίσουμε με $T(x)$ τον υπολοιπό χρόνο ζωής ενός ατόμου ηλικίας x το έτος t . Επομένως, ένα άτομο ηλικίας x θα επιβιώσει μέχρι την ηλικία $x + T(x)$ κατά το έτος $t + T(x)$. Στη συνέχεια, θα ορίσουμε την ποσότητα $q_x(t)$ η οποία είναι η πιθανότητα ένα άτομο ηλικίας x να αποβιώσει πριν φτάσει στην ηλικία $x+t$, δηλαδή

$$q_x(t) = P[T(x) \leq t]$$

Αντίστοιχα, ορίζεται η πιθανότητα $p_x(t)$ ένα άτομο ηλικίας x το έτος t να επιβιώσει μέχρι την ηλικία $x+1$, δηλαδή

$$p_x(t) = P[T(x) > t]$$

Οι πιθανότητες $p_x(t)$ και $q_x(t)$ συνδέονται με την ακόλουθη σχέση:

$$p_x(t) = 1 - q_x(t)$$

Ιδιαίτερο ενδιαφέρον για τους αναλογιστές παρουσιάζει η έννοια της έντασης θνησιμότητας (force of mortality), η οποία για την ηλικία $x+t$ δίνεται από την ακόλουθη σχέση:

$$\mu_{x,t} = \lim_{dt \rightarrow 0^+} \frac{P[T(x) \leq t + dt | T(x) > t]}{dt}$$

Η ένταση θνησιμότητας είναι μια έννοια η οποία χρησιμοποιείται εκτεταμένα στην Αναλογιστική επιστήμη και τη Δημογραφία για να περιγράψει τον στιγμιαίο ρυθμό θανάτου ενός ατόμου ή ενός πληθυσμού ηλικίας x την χρονική στιγμή t και συμβολίζεται με $\mu_{x,t}$. Η ένταση θνησιμότητας χρησιμοποιείται για την κατασκευή πινάκων θνησιμότητας, τον υπολογισμό του προσδόκιμου ζωής και της πιθανότητας επιβίωσης ενός ατόμου ή ενός πληθυσμού. Για την απλοποίηση των υπολογισμών της παρούσας διπλωματικής εργασίας και σύμφωνα με τη δημοσίευση του Cairns et al. (2009), υπό την υπόθεση ότι η ένταση θνησιμότητας παραμένει σταθερή για χρονικά διαστήματα μικρότερα του έτους για ακέραιες ηλικίες, οι τιμές του ειδικού κατά ηλικία δείκτη θνησιμότητας $m_{x,t}$ και της έντασης θνησιμότητας $\mu_{x,t}$ συμπίπτουν. Παράλληλα, βάσει της προηγούμενης υπόθεσης προκύπτουν η παρακάτω σχέση με την πιθανότητα θανάτου $q_x(t)$:

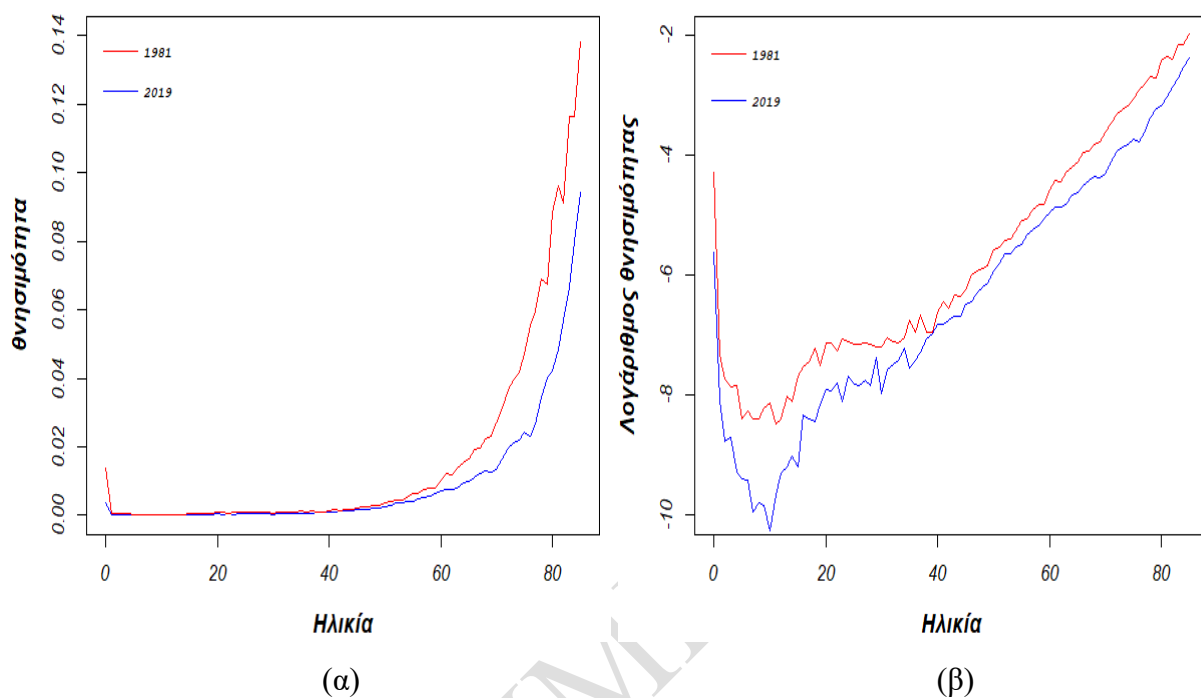
$$q_x(t) = 1 - \exp(-\mu_{x,t}) = 1 - \exp(-m_{x,t}) \quad (1.1)$$

Η σχέση (1.1) χρησιμοποιείται εκτεταμένα στην ανάλυση δεδομένων θνησιμότητας, Brouhns et al. (2002). Ο ειδικός κατά ηλικία δείκτης θνησιμότητας $m_{x,t}$ εκφράζεται ως το πηλίκο του αριθμού των θανόντων $D_{x,t}$ ηλικίας x σε ένα ημερολογιακό έτος t ως προς τον συνολικό πληθυσμό ηλικίας x στο μέσο του ημερολογιακού έτους t ή όπως ορίσαμε παραπάνω την συνολική έκθεση στον κίνδυνο $e_{x,t}$ ανά φύλο. Επομένως, ο ειδικός κατά ηλικία δείκτης θνησιμότητας δίνεται από τον ακόλουθο τύπο:

$$m_{x,t} = \frac{D_{x,t}}{e_{x,t}} \quad (1.2)$$

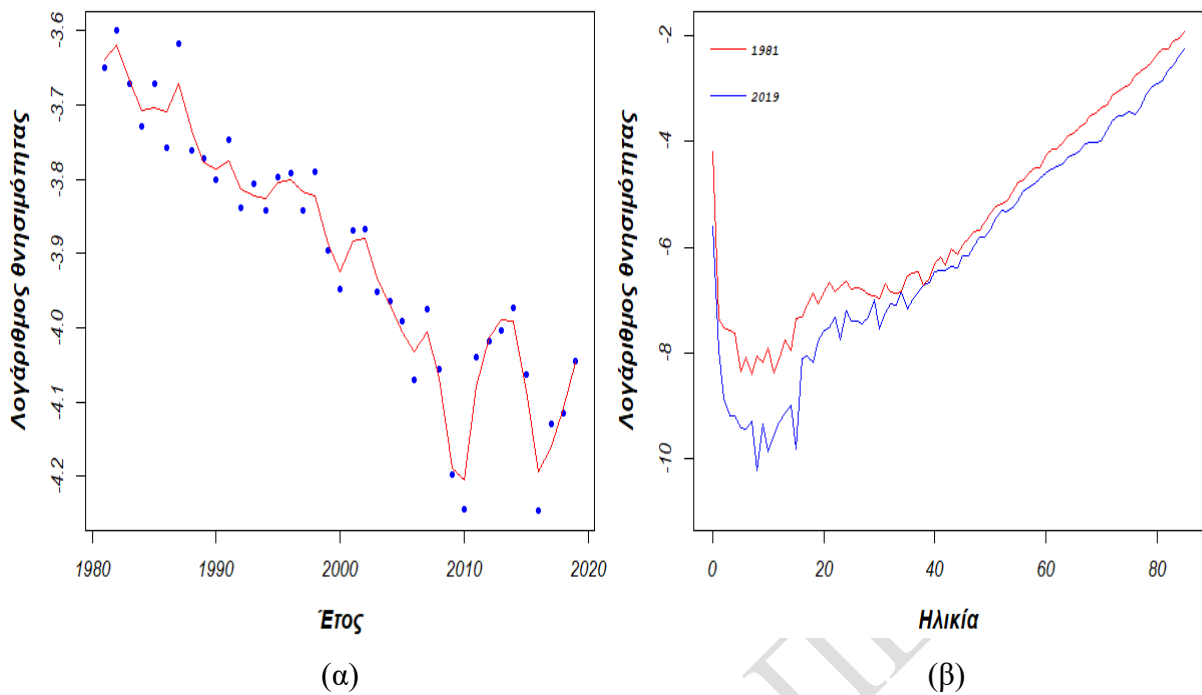
Ο εν λόγω δείκτης εκφράζει το ποσοστό του πληθυσμού ηλικίας x το οποίο θα αποβιώσει εντός του ημερολογιακού έτους t . Χρησιμοποιείται ευρέως στη δημογραφία και είναι ένα χρήσιμο

μέτρο για την κατανόηση των προτύπων θνησιμότητας σε διαφορετικές ηλικιακές ομάδες. Για την καλύτερη μελέτη του ειδικού κατά ηλικία δείκτη θνησιμότητας $m_{x,t}$ χρησιμοποιείται ο λογάριθμός του, ο οποίος όπως παρατηρείται και στα Διαγράμματα 1.7(α) και (β) παρουσιάζει μεγαλύτερη ευαισθησία στις ηλικίες 1 - 55.



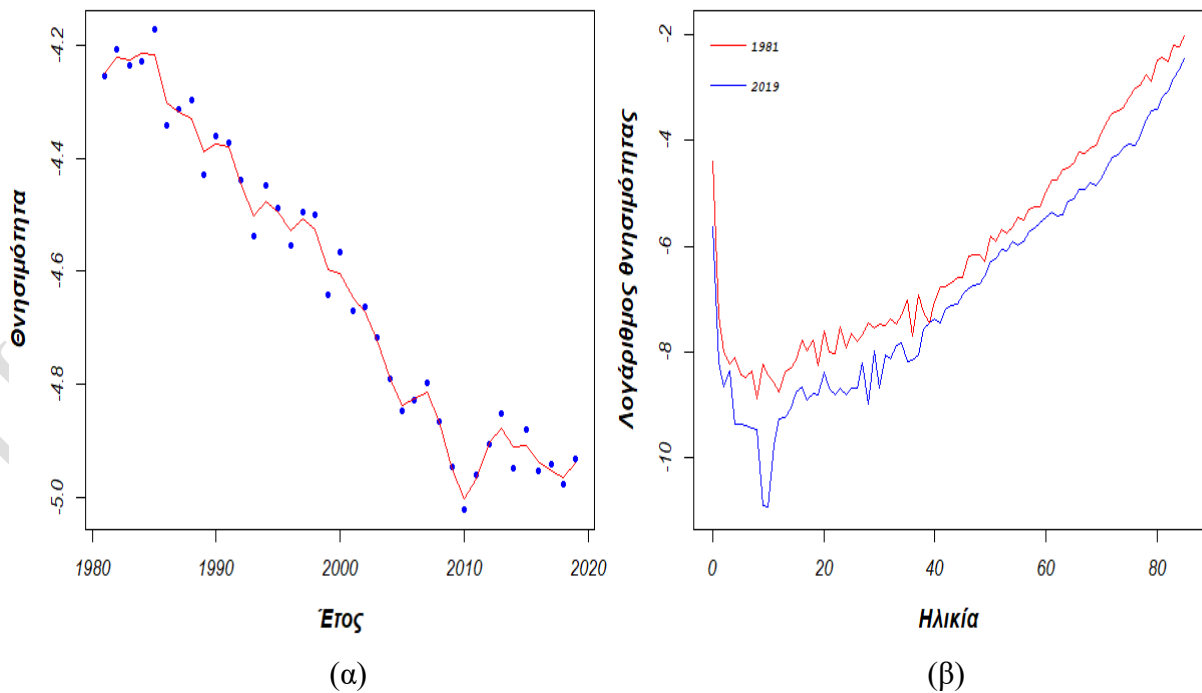
Διάγραμμα 1.7: (α) Ειδικός κατά ηλικία δείκτης θνησιμότητας συνολικού πληθυσμού για τα έτη 1981 και 2019, (β) Λογάριθμος ειδικού κατά ηλικία δείκτη θνησιμότητας συνολικού πληθυσμού για τα έτη 1981 και 2019.

Ακολούθως, στο Διάγραμμα 1.8(α) προκύπτει ότι η θνησιμότητα των ανδρών ηλικίας 67 ετών, ηλικία που με τα σημερινά δεδομένα είναι ηλικία έναρξης συνταξιοδότησης, μειώνεται διαχρονικά με την μεγαλύτερη μείωση να συμβαίνει το έτος 2010, έτος κατά το οποίο σημειώνεται και ο μικρότερος αριθμός θανάτων ανδρών. Μετά το έτος 2010 παρατηρείται μια σημαντική αύξηση μέχρι το 2015, συνεπεία στις αύξησης του αριθμού των θανάτων λόγω αύξησης περιστατικών νεοπλασμάτων καθώς και άλλων μη καθορισμένων βίαιων πράξεων, σύμφωνα με τα δημοσιευμένα στοιχεία της ΕΛ.ΣΤΑΤ. Παράλληλα, στο Διάγραμμα 1.8(β) γίνεται εμφανής η βελτίωση στις θνησιμότητας των ανδρών το έτος 2019 συγκριτικά με την αντίστοιχη θνησιμότητα του 1981, με την σημαντικότερη βελτίωση να παρατηρείται στις ηλικίες 1 έως 18.



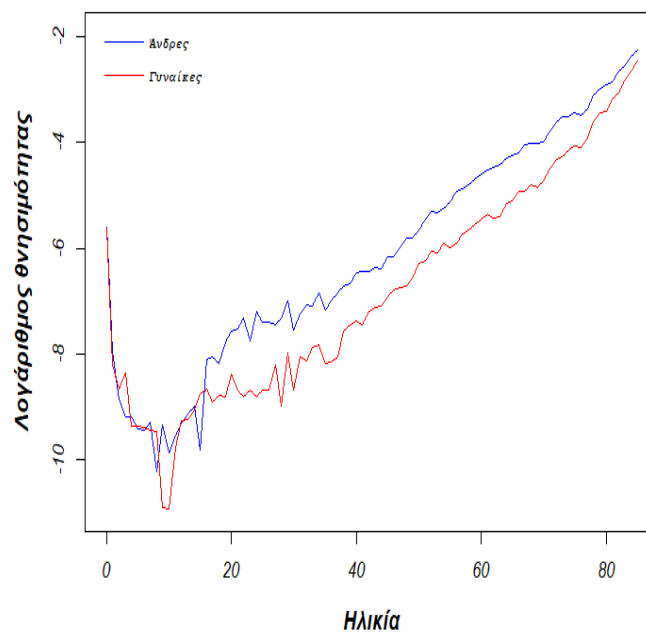
Διάγραμμα 1.8: Λογάριθμος δείκτη θνησιμότητας ανδρικού πληθυσμού: (α) ηλικίας 67 ετών, (β) ανά ηλικία για το έτος 2019.

Αντίστοιχα, για τον γυναικείο πληθυσμό, το διάγραμμα 1.9(α) απεικονίζει την πτωτική τάση της θνησιμότητας των γυναικών ηλικίας 67 ετών. Από το διάγραμμα 1.9(β) προκύπτει ότι η θνησιμότητα των γυναικών για το έτος 2019 είναι βελτιωμένη συγκριτικά με το 1981 με την σημαντικότερη βελτίωση να καταγράφεται στην ηλικία των 15 ετών.



Διάγραμμα 1.9: Λογάριθμος δείκτη θνησιμότητας γυναικείου πληθυσμού: (α) ηλικίας 67 ετών, (β) ανά ηλικία συγκριτικά για τα έτη 1981 και 2019.

Στο Διάγραμμα 1.10 προχωρήσαμε στη σύγκριση του λογαρίθμου των δεικτών θνησιμότητας του ανδρικού και του γυναικείου πληθυσμού για το έτος 2019. Από το εν λόγω διάγραμμα παρατηρείται ότι η ανδρική θνησιμότητα αυξάνεται έντονα από την ηλικία των 18 ετών, όπου σύμφωνα με στοιχεία της ΕΛ. ΣΤΑΤ. η κυριότερη αιτία θανάτου είναι τα ατυχήματα, με τα ατυχήματα μεταφοράς να αποτελούν την πλειοψηφία τους. Τέλος, η θνησιμότητα των ανδρών παραμένει σε υψηλότερα επίπεδα μέχρι και την ηλικία των 85 ετών.



Διάγραμμα 1.10: Λογάριθμος δείκτη θνησιμότητας ανδρικού έναντι γυναικείου πληθυσμού για το έτος 2019.

ΚΕΦΑΛΑΙΟ 2

Μοντέλο θνησιμότητας Lee - Carter

2.1 Προβολή θνησιμότητας

Η μέθοδος προβολής μελλοντικών δεικτών θνησιμότητας αναφέρεται στη διαδικασία της παρέκτασης (extrapolation) του επιπέδου θνησιμότητας βάσει της τάσης ιστορικών δεδομένων. Η διαδικασία της παρέκτασης χρησιμοποιείται εκτεταμένα στην Αναλογιστική επιστήμη καθώς και στην Δημογραφία με σκοπό την εκτίμηση μελλοντικών δεικτών θνησιμότητας καθώς και του προσδόκιμου ζωής.

Τα μοντέλα θνησιμότητας τα οποία χρησιμοποιούν τη μέθοδο της παρέκτασης για την πρόβλεψη μελλοντικών τιμών βασίζονται σε μια ποικιλία μεθόδων εκτίμησης, όπως η ανάλυση παλινδρόμησης, η ανάλυση χρονοσειρών και τα στοχαστικά μοντέλα. Το προσδόκιμο ζωής καθώς και ο ειδικός κατά ηλικία δείκτης θνησιμότητας αποτελούν σημαντικά δημογραφικά μεγέθη για διάφορες εφαρμογές όπως, η κατασκευή αναλογιστικών πινάκων θνησιμότητας, σχεδιασμός συνταξιοδοτικών προγραμμάτων καθώς και προγραμμάτων ασφαλιστικών καλύψεων ζωής και υγείας.

Εντούτοις, θα πρέπει να αναφερθεί ότι η μέθοδος της παρέκτασης υπόκειται σε ορισμένους περιορισμούς. Όπως έχουμε ήδη αναφέρει η μέθοδος, βασιζόμενη στην τάση του επιπέδου θνησιμότητας παρελθοντικών δεδομένων, δύναται να εκτιμήσει μελλοντικές τιμές δεικτών θνησιμότητας, οι εκτιμήσεις των οποίων μπορεί να μην είναι αξιόπιστες καθώς δεν λαμβάνει υπόψη οποιοδήποτε άλλο παράγοντα που μπορεί να επηρεάσει τη θνησιμότητα. Παράγοντες οι οποίοι μπορεί να επηρεάσουν τους δείκτες θνησιμότητας είναι η βελτίωση του κοινωνικοοικονομικού επιπέδου ενός πληθυσμού, οι εξελίξεις στην ιατρική επιστήμη και η ανακάλυψη νέων θεραπειών αλλά και η εμφάνιση ασθενειών. Μερικά παραδείγματα που αναφέρουν στη δημοσίευσή τους οι Booth και Tickle (2008) είναι η αύξηση του αριθμού των θανάτων ενήλικων ανδρών λόγω του ιού του AIDS και η περίοδος στασιμότητας της θνησιμότητας του πληθυσμού της Αυστραλίας τη δεκαετία του 1960 (Taylor et al., 1998).

Παράλληλα, τα μοντέλα παρέκτασης επηρεάζονται έντονα από ακραίες τιμές, οι οποίες μπορεί να οδηγήσουν σε εκτιμήσεις που δεν είναι αξιόπιστες. Για την αντιμετώπιση αυτών των περιορισμών, προτάθηκαν μοντέλα θνησιμότητας τα οποία ενσωματώνουν επιπλέον παράγοντες όπως είναι οι αλλαγές στην υγειονομική περίθαλψη, οι βελτιώσεις του βιοτικού επιπέδου και οι αλλαγές του τρόπου ζωής που μπορεί να επηρεάσουν τα επίπεδα θνησιμότητας. Κάποια από αυτά τα μοντέλα αποτελούν επεκτάσεις του κλασικού μοντέλου θνησιμότητας των Lee και Carter.

2.2 Περιγραφή μοντέλου θνησιμότητας Lee-Carter

Το 1992 οι Lee και Carter δημοσίευσαν μια νέα μέθοδο πρόβλεψης του επιπέδου θνησιμότητας, βασιζόμενοι στον συνδυασμό μιας στοχαστικής προσέγγισης της ηλικιακής κατανομής της θνησιμότητας και της θεωρίας χρονολογικών σειρών. Η μέθοδος εφαρμόστηκε για πρώτη φορά με σκοπό την περιγραφή της μεταβολής της θνησιμότητας των Ηνωμένων Πολιτειών για τα έτη 1933 – 1987 (Lee & Carter, 1992). Το μοντέλο Lee – Carter έγινε ιδιαίτερα δημοφιλές καθώς υπερτερεί άλλων μοντέλων πρόβλεψης θνησιμότητας εξαιτίας της απλότητάς του να εφαρμοστεί και της εύκολης ερμηνείας των παραμέτρων του. Η εν λόγω μέθοδος βασίζεται στην υπόθεση ότι ο λογάριθμος του ειδικού κατά ηλικία δείκτη θνησιμότητας δύναται να εκφραστεί ως άθροισμα μιας ηλικιακής παραμέτρου η οποία περιγράφει τη μέση θνησιμότητα ανά ηλικία και είναι ανεξάρτητη από τον χρόνο και ενός διγραμμικού παράγοντα ο οποίος περιγράφει πόσο έντονα μεταβάλλεται η θνησιμότητα σε κάθε ηλικία καθώς το γενικό επίπεδό της αλλάζει. Το στοχαστικό μοντέλο των Lee-Carter κάνει χρήση ιστορικών δεδομένων και βασίζεται στη μέθοδο παρέκτασης για την εκτίμηση της μελλοντικής τάσης του επιπέδου θνησιμότητας. Λαμβάνοντας τον λογάριθμο του ειδικού κατά ηλικία δείκτη θνησιμότητας εξασφαλίζεται ότι οι παρατηρούμενες τιμές και οι προβλέψεις του μοντέλου θα είναι πάντα θετικές. Παράλληλα, οι εκτιμήσεις του μοντέλου προκύπτουν από την συνήθη μέθοδο Ελαχίστων Τετραγώνων. Κατόπιν εκτίμησης του παράγοντα του γενικού επιπέδου θνησιμότητας, οι Lee και Carter θεώρησαν ότι ο εν λόγω παράγοντας αποτελεί μια χρονοσειρά. Για την μοντελοποίηση και την πρόβλεψη μελλοντικών τιμών της παρακάτω χρονοσειράς χρησιμοποιήθηκε η μέθοδος εκτίμησης χρονοσειρών ARIMA. Το μοντέλο αποτελείται από τρεις παραμέτρους α_x , β_x , k_t οι οποίες συνδέονται με την ακόλουθη σχέση:

$$m_{x,t} = \exp(\alpha_x + \beta_x k_t + \varepsilon_{x,t}) \quad (2.1)$$

ή

$$\ln m_{x,t} = \alpha_x + \beta_x k_t + \varepsilon_{x,t} \quad (2.2)$$

όπου $m_{x,t}$ είναι ο ειδικός κατά ηλικία δείκτης θνησιμότητας στην ηλικία x , τη χρονική περίοδο t , α_x είναι η παράμετρος που εκφράζει τη μέση θνησιμότητα ανά ηλικία, k_t είναι παράμετρος που περιγράφει το γενικό επίπεδο θνησιμότητας τη χρονική στιγμή t , ενώ β_x είναι η παράμετρος που εκφράζει τη μεταβολή της μέσης θνησιμότητας στην ηλικία x , καθώς μεταβάλλεται το γενικό επίπεδο θνησιμότητας k_t . Τέλος, τα σφάλματα $\varepsilon_{x,t}$ ακολουθούν κανονική κατανομή με μέση τιμή 0 και σταθερή διακύμανση σ_ε^2 και εκφράζουν την αδυναμία του μοντέλου να ερμηνεύσει επιδράσεις που σχετίζονται με διάφορες ηλικίες και περιόδους προσαρμογής.

Προκειμένου να εξάγουμε από το παραπάνω μοντέλο μία μοναδική λύση προτάθηκαν οι εξής παρακάτω περιορισμοί για τις παραμέτρους β_x , k_t :

$$\sum_t k_t = 0 \quad (2.3)$$

και

$$\sum_x \beta_x^2 = 1 \quad (2.4)$$

Στη σημείο αυτό αξίζει να σημειωθεί ότι η μέθοδος των Lee – Carter βασίζεται στην παρέκταση της τάσης ιστορικών δεδομένων επιπέδου θνησιμότητας και δεν ενσωματώνει υποθέσεις για την εξέλιξη της ιατρικής, ή οικονομικές αλλαγές που ενδέχεται να προκύψουν μελλοντικά. Κατ' επέκταση το μοντέλο των Lee – Carter αποτυγχάνει να προβλέψει μεταβολές στα επίπεδα θνησιμότητας λόγω ανακάλυψης νέων ιατρικών θεραπειών ή ακόμη και την εμφάνιση κάποιας νέας πανδημίας όπως για παράδειγμα αυτή του Covid-19, η οποία δύναται να επηρεάσει τους δείκτες θνησιμότητας του πληθυσμού.

2.3 Μέθοδοι εκτίμησης παραμέτρων μοντέλου Lee - Carter

Η ιδιαιτερότητα του μοντέλου Lee - Carter έγκειται στο γεγονός ότι οι παράμετροι του μοντέλου δεν μπορούν να εκτιμηθούν με τις συνήθεις μεθόδους παλινδρόμησης (Currie, 2013) καθώς στο δεξιό μέλος της εξίσωσης (2.2) υπάρχουν μόνο άγνωστες παράμετροι προς εκτίμηση. Ακολούθως, παρουσιάζονται και αναλύονται οι τρεις δημοφιλέστερες στη διεθνή

βιβλιογραφία μέθοδοι εκτίμησης των παραμέτρων του εν λόγω μοντέλου, οι οποίες είναι η συνήθης μέθοδος Ελαχίστων Τετραγώνων – Ordinary Least Squares κάνοντας χρήση της μεθόδου Διάσπασης Ιδιαζουσών Τιμών - Singular Value Decomposition (Lee and Carter, 1992), η μέθοδος Ελαχίστων Τετραγώνων με βάρη - Weighted Least Squares (Wilmoth, 1993) και η μέθοδος Μέγιστης Πιθανοφάνειας Maximum Likelihood Estimation (Brouhns et al., 2002). Για τους σκοπούς της παρούσας διπλωματικής εργασίας και για τα δεδομένα της Ελλάδας θα γίνει χρήση της μεθόδου Διάσπασης Ιδιαζουσών Τιμών (Singular Value Decomposition) και Μέγιστης Πιθανοφάνειας. Ενδεικτικά ακολούθως αναφέρονται και οι τρεις μέθοδοι.

2.3.1 Συνήθης Μέθοδος Ελαχίστων Τετραγώνων - Ordinary Least Squares (OLS)

Οι εκτιμητές $\widehat{\alpha}_x$, $\widehat{\beta}_x$ και \widehat{k}_t του μοντέλου (2.2) δύναται να προκύψουν με την εφαρμογή της συνήθους μεθόδου Ελαχίστων Τετραγώνων (Ordinary Least Squares method) και είναι αυτοί που ελαχιστοποιούν την ακόλουθη συνάρτηση:

$$f_{x,t} = \sum_{x,t} (\ln m_{x,t} - \alpha_x - \beta_x k_t)^2 \quad (2.5)$$

Οι εκτιμήσεις της παραμέτρου α_x προκύπτουν λαμβάνοντας την μέση τιμή των λογαρίθμων των παρατηρούμενων τιμών του ειδικού κατά ηλικία δείκτη θνησιμότητας ($\ln m_{x,t}$) ή ελαχιστοποιώντας την εξίσωση (2.4) λαμβάνοντας μερικές παραγώγους ως προς α_x οι οποίες εξισώνονται με το μηδέν, όπως παρουσιάζεται ακολούθως:

$$0 = \frac{\partial f_{x,t}}{\partial \alpha_x} = -2 \sum_t (\ln m_{x,t} - \alpha_x - \beta_x k_t)$$

ή ισοδύναμα

$$\sum_t \ln m_{x,t} - \sum_t \alpha_x - \beta_x \sum_t k_t = 0.$$

Εφαρμόζοντας τον περιορισμό (2.3) λαμβάνουμε τη σχέση

$$\sum_t \alpha_x = \sum_t \ln m_{x,t}$$

Από τη σχέση αυτή προκύπτει ο εκτιμητής:

$$\widehat{\alpha}_x = \frac{1}{t_n - t_1 + 1} \sum_t \ln m_{x,t} \quad (2.6)$$

Προκειμένου να ξεπεραστεί το πρόβλημα που δημιουργείται στην εκτίμηση των παραμέτρων του μοντέλου θνησιμότητας, λόγω της παρουσίας του διγραμμικού παράγοντα $\beta_x k_t$, οι Lee και Carter έκαναν χρήση της μεθόδου Διάσπασης Ιδιαζουσών Τιμών (SVD).

Η μέθοδος διάσπασης ιδιαζουσών τιμών αποτελεί την ανάλυση ενός πίνακα σε τρεις επιμέρους πίνακες. Πιο συγκεκριμένα, η διάσπαση ιδιαζουσών τιμών ενός πίνακα M με διαστάσεις $m \times n$ είναι της μορφής $M = UdV^T$, όπου U είναι ένας ορθογώνιος πίνακας με διαστάσεις $m \times m$, V είναι ένας ορθογώνιος πίνακας με διαστάσεις $n \times n$ και V^T ο ανάστροφός του, ενώ d είναι ένας διαγώνιος πίνακας με διαστάσεις $m \times n$ και οι τιμές στην κύρια διαγώνιό του καθορίζονται μοναδικά από τον πίνακα M καθώς είναι οι ιδιάζουσες τιμές του σε φθίνουσα σειρά $\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_m \geq 0$. Διευκρινίζεται ότι ιδιάζουσα τιμή ενός πίνακα M με διαστάσεις $m \times n$ ορίζεται η τετραγωνική ρίζα της ιδιοτιμής λ_i του πίνακα $M^T M$,

$$\sigma_i = \sqrt{\lambda_i}, \quad i = 1, 2, \dots, n$$

Στην περίπτωση του μοντέλου Lee – Carter η μέθοδος ιδιαζουσών τιμών εφαρμόζεται στον πίνακα $Z_{x,t}$ με διαστάσεις $m \times n$, όπου

$$Z_{x,t} = \ln \widehat{m}_{x,t} - \widehat{\alpha}_x = \widehat{\beta}_x \widehat{k}_t \quad (2.7)$$

$$Z_{x,t} = \begin{bmatrix} Z_{x_1,t_1} & \dots & Z_{x_1,t_n} \\ \vdots & \ddots & \vdots \\ Z_{x_m,t_1} & \dots & Z_{x_m,t_n} \end{bmatrix} = \begin{bmatrix} \ln \widehat{m}_{x_1,t_1} - \widehat{\alpha}_{x_1} & \dots & \ln \widehat{m}_{x_1,t_m} - \widehat{\alpha}_{x_1} \\ \vdots & \ddots & \vdots \\ \ln \widehat{m}_{x_m,t_1} - \widehat{\alpha}_{x_m} & \dots & \ln \widehat{m}_{x_m,t_m} - \widehat{\alpha}_{x_m} \end{bmatrix}$$

Με την εφαρμογή της μεθόδου Singular Value Decomposition (SVD) θα διασπάσουμε τον πίνακα $Z_{x,t}$ σε γινόμενο τριών πινάκων.

$$Z_{x,t} = UdV^T \quad (2.8)$$

Όπου: U είναι ένας ορθογώνιος πίνακας με διαστάσεις $m \times m$, δηλαδή είναι ο

$$\begin{bmatrix} u_{1,1} & \dots & u_{1,m} \\ \vdots & \ddots & \vdots \\ u_{m,1} & \dots & u_{m,m} \end{bmatrix}$$

V είναι ένας ορθογώνιος πίνακας με διαστάσεις $n \times n$ και ο ανάστροφός του, V^T είναι ο

$$\begin{bmatrix} v_{1,1} & \dots & v_{n,1} \\ \vdots & \dots & \vdots \\ v_{1,n} & \dots & v_{n,n} \end{bmatrix}$$

d είναι ένας διαγώνιος πίνακας με διαστάσεις $m \times n$, δηλαδή είναι ο

$$\begin{bmatrix} \sigma_1 & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & \sigma_m \end{bmatrix}$$

Από τις σχέσεις (2.7) και (2.8) προκύπτει ότι:

$$\widehat{\beta}_x = U \quad \text{και} \quad \widehat{k}_t = dV^T$$

Σε επόμενο βήμα, προκειμένου να αποφευχθούν μεγάλες διαφορές μεταξύ των προβλεπόμενων και των πραγματικών θανάτων οι Lee και Carter εφάρμοσαν ένα δεύτερο βήμα εκ νέου εκτίμησης της παραμέτρου k_t . Λαμβάνοντας τις εκτιμήσεις των παραμέτρων α_x, β_x από το αρχικό βήμα με την συνήθη μέθοδο Ελαχίστων Τετραγώνων εξίσωσαν τις παρατηρούμενες τιμές των θανάτων με τις εκτιμημένες τιμές οι οποίες προκύπτουν από το μοντέλο Lee-Carter, ως ακολούθως:

$$\sum_x D_{x,t} = \sum_x E_{x,t} \exp(\widehat{\alpha}_x + \widehat{\beta}_x \widehat{k}_t')$$

Όπου, \widehat{k}_t' είναι η δεύτερη εκτίμηση της παραμέτρου k_t ,

$\widehat{\alpha}_x$ είναι η εκτίμηση της παραμέτρου α_x η οποία προκύπτει από τη σχέση (2.5)

$\widehat{\beta}_x$ είναι η εκτίμηση της παραμέτρου β_x η οποία προκύπτει από τη μέθοδο SVD

Σύμφωνα με αρκετούς ερευνητές η εκτίμηση των παραμέτρων του μοντέλου Lee – Carter μέσω της μεθόδου SVD παρουσιάζει ορισμένες αδυναμίες. Η σημαντικότερη αδυναμία της εν λόγω μεθόδου είναι η υπόθεση ότι τα σφάλματα είναι ομοσκεδαστικά και κανονικά κατανοημένα, η οποία δεν είναι ρεαλιστική για την ανθρώπινη θνησιμότητα καθώς, όσο αυξάνεται η ηλικία, τόσο μεγαλύτερες μεταβολές παρατηρούνται στους δείκτες θνησιμότητας. Επιπρόσθετα, μέσω

της μεθόδου SVD δεν χρησιμοποιείται κάποια στάθμιση για τις ηλικίες στις οποίες ο αριθμός των θανάτων είναι ιδιαίτερα αυξημένος.

2.3.2 Μέθοδος Σταθμισμένων Ελαχίστων Τετραγώνων - Weighted Least Squares

(WLS)

Στην επιστημονική του δημοσίευση ο Wilmoth (1993) προτείνει ως εναλλακτική μέθοδο εκτίμησης των παραμέτρων της εξίσωσης (2.2) τη μέθοδο Σταθμισμένων Ελαχίστων Τετραγώνων. Καθώς σε κάποιες ηλικίες ο αριθμός θανάτων δύναται να είναι μηδενικός κατ' επέκταση ο ειδικός κατά ηλικία δείκτης θνησιμότητας $m_{x,t}$ (1.2) λαμβάνει μηδενικές τιμές. Βάσει της μεθόδου Σταθμισμένων Ελαχίστων Τετραγώνων προκειμένου να αποφευχθεί ο υπολογισμός λογαρίθμων μηδενικών τιμών προτάθηκε η ελαχιστοποίηση της ποσότητας:

$$\sum_{x,t} w_{x,t} [\ln m_{x,t} - \alpha_x - \beta_x k_t]^2 \quad (2.9)$$

όπου τα βάρη $w_{x,t}$ θα είναι ίσα με τον αριθμό των θανάτων $D_{x,t}$.

Εφαρμόζοντας την μέθοδο Σταθμισμένων Ελαχίστων Τετραγώνων, θέτουμε τις μερικές παραγώγους της ποσότητας (2.9) ίσες με μηδέν και λύνουμε ως προς την παράμετρο που μας ενδιαφέρει, έτσι ελαχιστοποιώντας την παραπάνω ποσότητα παίρνουμε αριθμητικά τις εκτιμήσεις των παραμέτρων $\widehat{\alpha}_x, \widehat{\beta}_x, \widehat{k}_t$

$$\widehat{\alpha}_x = \frac{\sum_t D_{x,t} (\ln m_{x,t} - \widehat{\beta}_x \widehat{k}_t)}{\sum_t D_{x,t}} \quad (2.10)$$

$$\widehat{\beta}_x = \frac{\sum_t D_{x,t} \widehat{k}_t (\ln m_{x,t} - \widehat{\alpha}_x)}{\sum_t D_{x,t} \widehat{k}_t^2} \quad (2.11)$$

$$\widehat{k}_t = \frac{\sum_t D_{x,t} \widehat{\beta}_x (\ln m_{x,t} - \widehat{\alpha}_x)}{\sum_t D_{x,t} \widehat{\beta}_x^2} \quad (2.12)$$

Επιπροσθέτως, η ανωτέρω μέθοδος εκτίμησης του Wilmoth δεν διασφαλίζει μόνο τον υπολογισμό λογαρίθμων μη μηδενικών δεικτών θνησιμότητας αλλά παράλληλα θέτει

συγκεκριμένη στάθμιση για κάθε ηλικιακή ομάδα. Όπως αναφέρεται στη δημοσίευση (Basellini, 2022), η μέθοδος εκτίμησης μέσω της SVD αμφισβητείται από αρκετούς ερευνητές λόγω των μη ρεαλιστικών υποθέσεων της μεθόδου και την έλλειψη στάθμισης των δεικτών θνησιμότητας. Η ύπαρξη στάθμισης είναι αναγκαία ειδικά στις μεγαλύτερες ηλικίες όπου παρατηρείται ότι οι δείκτες θνησιμότητας είναι περισσότερο ευμετάβλητοι, καθώς στο κλάσμα του δείκτη θνησιμότητας ο αριθμός των θανάτων που βρίσκεται στον αριθμητή τείνει να αυξάνεται και η έκθεση στον κίνδυνο που βρίσκεται στον παρονομαστή τείνει να μειώνεται λόγω της θνησιμότητας. Κατ' επέκταση στη μέθοδο Σταθμισμένων Ελαχίστων Τετραγώνων χρησιμοποιήθηκε ως στάθμιση ο αριθμός θανάτων $D_{x,t}$ για την καλύτερη εκτίμηση των παραμέτρων του μοντέλου Lee -Carter (Basellini, 2022).

2.3.3 Μέθοδος Μέγιστης Πιθανοφάνειας Poisson

Μια εναλλακτική μέθοδος εκτίμησης των παραμέτρων α_x , β_x , k_t προτάθηκε από τον Brouhns et al. (2002) ο οποίος αξιοποιώντας την υπόθεση ότι ο αριθμός των θανάτων $D_{x,t}$ στην ηλικία x και το έτος t μπορεί να προσεγγιστεί ικανοποιητικά από την κατανομή Poisson (Brillinger, 1986) ως ακολούθως:

$$D_{x,t} \sim \text{Poisson}(e_{x,t}m_{x,t}) \quad \text{όπου} \quad m_{x,t} = \exp(\alpha_x + \beta_x k_t)$$

και

$$E(D_{x,t}) = e_{x,t}m_{x,t} \tag{2.13}$$

Διευκρινίζουμε ότι η ποσότητα $e_{x,t}$ είναι η συνολική έκθεση στον κίνδυνο όπως την έχουμε ορίσει στο Κεφάλαιο 2.

Στην ακόλουθη εξίσωση (2.14) περιγράφεται η συνάρτηση πυκνότητας πιθανότητας του αριθμού των θανάτων $D_{x,t}$ υπό την υπόθεση ότι ακολουθούν την κατανομή Poisson:

$$f(D_{x,t}) = \frac{(e_{x,t}m_{x,t})^{D_{x,t}} \exp(-e_{x,t}m_{x,t})}{D_{x,t}!} \tag{2.14}$$

Έτσι, χρησιμοποιώντας την μέθοδο Μέγιστης Πιθανοφάνειας Poisson δύναται να εκτιμηθούν οι παράμετροι του μοντέλου Lee-Carter. Η συνάρτηση πιθανοφάνειας με βάση την (2.14) έχει την ακόλουθη μορφή:

$$L(\alpha_x, \beta_x, k_t) = \prod f(D_{x,t}) = \prod \left[\frac{(e_{x,t}m_{x,t})^{D_{x,t}} \exp(-e_{x,t}m_{x,t})}{D_{x,t}!} \right] \tag{2.15}$$

Ο λογάριθμος της συνάρτησης πιθανοφάνειας (2.15) έχει την παρακάτω μορφή:

$$\begin{aligned}
 l(\alpha_x, \beta_x, k_t) &= \ln L(\alpha_x, \beta_x, k_t) = \sum_{x,t} \ln \left[\frac{(e_{x,t} m_{x,t})^{D_{x,t}} \exp(-e_{x,t} m_{x,t})}{D_{x,t}!} \right] \\
 &= \sum_{x,t} [D_{x,t} \ln(e_{x,t} m_{x,t}) - (e_{x,t} m_{x,t}) - \ln(D_{x,t}!)] \\
 &= \sum_{x,t} [D_{x,t} (\ln e_{x,t} - \ln m_{x,t}) - (e_{x,t} m_{x,t}) - \ln(D_{x,t}!)] \\
 &= \sum_{x,t} [D_{x,t} \ln m_{x,t} - (e_{x,t} m_{x,t}) - \ln(D_{x,t}!) + D_{x,t} \ln e_{x,t}] \quad (2.16)
 \end{aligned}$$

Αντικαθιστώντας στην ανωτέρω σχέση (2.16) σύμφωνα με τη (1.2) καταλήγουμε στο ακόλουθο αποτέλεσμα:

$$\begin{aligned}
 l(\alpha_x, \beta_x, k_t) &= \sum_{x,t} [D_{x,t} \ln(\exp(\alpha_x + \beta_x k_t)) - e_{x,t} \exp(\alpha_x + \beta_x k_t)] \\
 \Leftrightarrow l(\alpha_x, \beta_x, k_t) &= \sum_{x,t} [D_{x,t} (\alpha_x + \beta_x k_t) - e_{x,t} \exp(\alpha_x + \beta_x k_t)] \quad (2.17)
 \end{aligned}$$

Μεγιστοποιώντας τη σχέση (2.17) προκύπτουν οι εκτιμήσεις των παραμέτρων α_x, β_x, k_t .

Σε αυτή την προσέγγιση, το μοντέλο Lee – Carter αποτελεί ένα μοντέλο Poisson παλινδρόμησης με εξαρτημένη μεταβλητή των αριθμό των θανάτων $D_{x,t}$ και την έκθεση στον κίνδυνο $e_{x,t}$ ως βοηθητική μεταβλητή (offset). Αναφορικά με τη συγκεκριμένη μέθοδο εκτίμησης αξίζει να αναφερθεί ότι σε αντίθεση με τις προαναφερόμενες μεθόδους, τα σφάλματα του μοντέλου υπολογίζονται απευθείας για την εκτίμηση του αριθμού των θανάτων και ο συνολικός αριθμός των παρατηρούμενων θανάτων ισούται με τον συνολικό αριθμό των προβλεπόμενων. Κατ' επέκταση δεν απαιτείται το δεύτερο στάδιο επανεκτίμησης της παραμέτρου k_t .

2.4 Μειονεκτήματα του μοντέλου Lee - Carter

Η εισαγωγή του μοντέλου Lee - Carter το 1992 αποτέλεσε καινοτομία μεταξύ των μέχρι τότε μοντέλων πρόβλεψης θνησιμότητας, παρέχοντας μια απλή και ταυτόχρονα ισχυρή στοχαστική προσέγγιση. Ωστόσο, παρόλα τα πλεονεκτήματα του μοντέλου η μέθοδος Lee - Carter παρουσιάζει και ορισμένα μειονεκτήματα. Το βασικό μειονέκτημα της μεθόδου Lee - Carter προκύπτει από την εκτίμηση των παραμέτρων του μοντέλου μέσω της μεθόδου SVD. Κατά τη διαδικασία εκτίμησης των παραμέτρων του μοντέλου μέσω της SVD στην οποία δεν χρησιμοποιούνται βάρη, χρησιμοποιείται η υπόθεση της ομοσκεδαστικότητας των σφαλμάτων $\varepsilon_{x,t}$. Αυτό σημαίνει ότι τα σφάλματα $\varepsilon_{x,t}$ ακολουθούν την κανονική κατανομή με μέση τιμή μηδέν και σταθερή διακύμανση σ_ε^2 . Η υπόθεση της ομοσκεδαστικότητας των σφαλμάτων δεν είναι ρεαλιστική για δεδομένα ανθρώπινης θνησιμότητας, Alho (2000), καθώς όπως παρατηρούμε κι από το Διάγραμμα 1.7(β) ο λογάριθμος του κατά ηλικία δείκτη θνησιμότητας στις μεγαλύτερες ηλικίες παρουσιάζει μεγαλύτερη μεταβλητότητα συγκριτικά με τις μικρότερες ηλικίες.

2.5 Εκτίμηση χρονολογικών σειρών και πρόβλεψη μέσω του μοντέλου Lee - Carter

Όπως έχουμε ήδη αναφέρει στην Παράγραφο 2.1, το μοντέλο θνησιμότητας των Lee και Carter, το οποίο περιγράφεται από τη σχέση (2.2), για την πρόβλεψη του λογαρίθμου του ειδικού κατά ηλικία δείκτη θνησιμότητας βασίζεται στη πρόβλεψη μέσω της μεθόδου παρέκτασης παρελθοντικών δεδομένων. Έτσι κατόπιν εκτίμησης των παραμέτρων του μοντέλου α_x, β_x, k_t , οι Lee και Carter για την πρόβλεψη των δεικτών θνησιμότητας για μελλοντικά έτη θεώρησαν ότι η παράμετρος β_x παραμένει σταθερή ως προς τον χρόνο και η παράμετρος k_t είναι μια χρονολογική σειρά. Για την εκτίμηση των μελλοντικών τιμών της χρονολογικής σειράς k_t χρησιμοποιήθηκε η μέθοδος αυτοπαλινδρομικών μοντέλων κινητού μέσου όρου Autoregressive Integrated Moving Average - ARIMA(p, d, q). Οι Lee και Carter αφού εξέτασαν αρκετά μοντέλα ARIMA κατέληξαν στο συμπέρασμα ότι το καταλληλότερο μοντέλο είναι το ARIMA (0,1,0) το οποίο είναι ένας τυχαίος περίπατος με παράμετρο μετατόπισης (Random Walk with Drift).

2.5.1 Μεθοδολογία Box - Jenkins

Η μέθοδος Box – Jenkins αναπτύχθηκε το 1976 και είναι μία τεχνική πρόβλεψης χρονοσειρών. Η εν λόγω μέθοδος βασίζεται στη θεωρία των αυτοπαλίνδρομων μοντέλων κινητού μέσου όρου (Autoregressive Integrated Moving Average ARIMA). Ένα μοντέλο ARIMA (p,d,q) αποτελείται από ένα συνδυασμό δύο διαφορετικών μοντέλων χρονολογικών σειρών, ένα μοντέλο αυτοπαλίνδρομης χρονοσειράς παραμέτρου p (Autoregressive AR(p) model) και ένα μοντέλο χρονοσειράς κινητού μέσου με παράμετρο q (Moving Average MA(q) model). Η μέθοδος Box-Jenkins περιλαμβάνει τέσσερα στάδια ανάλυσης χρονοσειρών και εφαρμόζεται σε χρονοσειρές οι οποίες δεν είναι στάσιμες. Με τον όρο στάσιμη χαρακτηρίζεται μία χρονοσειρά η οποία δεν εμφανίζει καμία τάση ή εποχικότητα και οι στατιστικές ιδιότητές της όπως ο μέσος όρος και η διακύμανση παραμένουν σταθερές με την πάροδο του χρόνου. Στο πρώτο στάδιο της ταυτοποίησης (identification) γίνεται έλεγχος της στασιμότητας της χρονοσειράς και η ανάλυση των δεδομένων της για τυχόν τάσεις ή εποχικότητα μέσω της συνάρτησης αυτοσυσχέτισης (ACF) και την συνάρτηση μερικής αυτοσυσχέτισης (PACF). Στο δεύτερο στάδιο της εκτίμησης (estimation) αφού εξασφαλιστεί η στασιμότητας της χρονοσειράς, εκτιμώνται οι παράμετροι p, d και q οι οποίες καθορίζουν τα προτεινόμενα μοντέλα ARIMA (p,d,q). Στη συνέχεια τα μοντέλα αξιολογούνται βάσει των κριτηρίων πληροφορίας του Akaike (AIC) και του Swartz (BIC) και επιλέγεται το μοντέλο με την καλύτερη προσαρμογή και τις λιγότερες δυνατές παραμέτρους. Στο επόμενο στάδιο της διάγνωσης (diagnostic checking) γίνεται έλεγχος των καταλοίπων για τυχόν τάσεις ή ενδείξεις συσχέτισης.

2.5.2 Τυχαίος περίπατος με παράμετρο μετατόπισης

Ένας τυχαίος περίπατος (Random Walk) είναι μια διαδικασία κατά την οποία η μελλοντική τιμή μιας μεταβλητής προσδιορίζεται από την ακριβώς παλαιότερη τιμή για την οποία υπάρχουν διαθέσιμα δεδομένα προσθέτοντας έναν όρο σφάλματος που ακολουθεί κανονική κατανομή με μέση τιμή μηδέν και διακύμανση σ^2 . Ο τυχαίος περίπατος με κανονικά σφάλματα περιγράφεται από την ακόλουθη σχέση:

$$y_t = y_{t-1} + \xi_t \quad \text{όπου } \xi_t \sim N(0, \sigma^2)$$

όπου y_t είναι η τιμή της μεταβλητής y στον χρόνο t που θέλουμε να προβλέψουμε, y_{t-1} είναι η τιμή μέχρι την οποία έχουμε δεδομένα και ξ_t είναι ο όρος του σφάλματος. Με την προσθήκη της παραμέτρου θ (drift) στο παραπάνω μοντέλο, το μοντέλο του τυχαίου περιπάτου παρουσιάζει μια τάση ή οποία μπορεί να είναι ανοδική αν $\theta > 0$ ή καθοδική αν $\theta < 0$. Να σημειώσουμε ότι ο τυχαίος περίπατος με παράμετρο μετατόπισης προκύπτει από τη θεωρία των μοντέλων ARIMA(p,d,q) ορίζοντας τις παραμέτρους p,d,q του μοντέλου ίσες με 0,1,0 αντίστοιχα, δηλαδή πρόκειται για ένα μοντέλο ARIMA(0,1,0). Σύμφωνα με όσα έχουν ήδη αναφερθεί στην θεωρία των μοντέλων ARIMA κατά την εφαρμογή του τυχαίου περιπάτου απαιτείται η στασιμότητα της χρονοσειράς. Ο τυχαίος περίπατος με παράμετρο μετατόπισης περιγράφεται από την ακόλουθη σχέση:

$$y_t = y_{t-1} + \theta + \xi_t \quad \text{όπου } \xi_t \sim N(0, \sigma^2)$$

Χρησιμοποιώντας τον παραπάνω τύπο του τυχαίου περιπάτου με μετατόπιση, η πρόβλεψη των μελλοντικών τιμών της χρονολογικής σειράς \hat{k}_t δίνεται από την ακόλουθη σχέση:

$$\hat{k}_t = \hat{k}_{t-1} + \theta \quad (2.18)$$

Ο εκτιμητής μέγιστης πιθανοφάνειας ο οποίος εξαρτάται από την πρώτη και την τελευταία εκτίμηση της παραμέτρου k_t δίνεται από τον ακόλουθο τύπο:

$$\hat{\theta} = \frac{\hat{k}_T - \hat{k}_1}{T - 1} \quad (2.19)$$

και

$$\hat{\sigma}^2 = \frac{1}{(T - 1) \sum_{t=2}^T (k_t - k_{t-1} - \hat{\theta})^2} \quad (2.20)$$

ΚΕΦΑΛΑΙΟ 3

Πλαίσιο Γενικευμένων Γραμμικών μοντέλων

3.1 Εκθετική Οικογένεια κατανομών

Η Εκθετική Οικογένεια Κατανομών είναι μια ευρεία κλάση κατανομών πιθανότητας στην οποία περιλαμβάνεται πληθώρα γνωστών κατανομών, ενδεικτικά αναφέρονται η Κανονική, Εκθετική, Γάμμα, Bernouli, Διωνυμική, Αρνητική Διωνυμική, Poisson, Βήτα και Weibull. Παρατηρείται ότι στην Εκθετική Οικογένεια Κατανομών ανήκουν τόσο συνεχείς όσο και διακριτές κατανομές.

Έστω μια τυχαία μεταβλητή Y της οποίας η κατανομή είναι μονοπαραμετρική και ανήκει στην προαναφερόμενη κλάση κατανομών τότε η συνάρτηση πιθανότητας ή συνάρτηση πυκνότητας πιθανότητας δύναται να γραφεί στην παρακάτω μορφή:

$$f(y; \theta) = \exp \left[\frac{y\theta - b(\theta)}{a(\varphi)} + c(y, \varphi) \right] \quad (3.1)$$

Η παράμετρος θ καλείται φυσική παράμετρος ή παράμετρος κεντρικής τάσης (location parameter) ενώ η παράμετρος φ καλείται παράμετρος κλίμακας (dispersion parameter) και συνδέονται με τη διασπορά της τ.μ. Y με τον ακόλουθο τύπο:

$$V(Y) = a(\varphi)b''(\theta) \quad (3.2)$$

Στις κατανομές που χρησιμοποιούνται για την μοντελοποίηση μετρήσιμων δεδομένων όπως είναι το μοντέλο της Poisson η παράμετρος φ ισούται με τη μονάδα. Η συνάρτηση $b(\cdot)$ είναι σημαντική για τα Γενικευμένα Γραμμικά Μοντέλα καθώς περιγράφει τη σχέση μεταξύ της μέσης τιμής και της διακύμανσης της κατανομής.

Η συνάρτηση πιθανοφάνειας οποιασδήποτε κατανομής η οποία ανήκει στην Εκθετική Οικογένεια Κατανομών έχει την ακόλουθη μορφή:

$$L(\theta; y, \varphi) = \prod_{i=1}^N \exp \left\{ \frac{y_i \theta - b(\theta)}{a(\varphi)} + c(y_i, \varphi) \right\} \quad (3.3)$$

Παίρνοντας λογαρίθμους και στα δύο μέρη της εξίσωσης (3.3) καταλήγουμε ακολούθως στη λογαριθμική συνάρτηση πιθανοφάνειας:

$$l(\theta; y, \varphi) = \log L(\theta; y, \varphi) = \sum_{i=1}^N \left\{ \frac{y_i \theta - b(\theta)}{a(\varphi)} + c(y_i, \varphi) \right\} \quad (3.4)$$

Από τη μαθηματική στατιστική ισχύουν οι ακόλουθες σχέσεις (3.5) και (3.6):

$$E \left(\frac{\partial l}{\partial \theta} \right) = 0 \quad (3.5)$$

και

$$E \left(\frac{\partial^2 l}{\partial \theta^2} \right) + E \left[\left(\frac{\partial l}{\partial \theta} \right)^2 \right] = 0 \quad (3.6)$$

Παραγωγίζοντας μερικώς ως προς θ τη σχέση (3.4) καταλήγουμε στα ακόλουθα αποτελέσματα:

$$\frac{\partial l}{\partial \theta} = \frac{[y - b'(\theta)]}{a(\varphi)} \quad (3.7)$$

και

$$\frac{\partial^2 l}{\partial \theta^2} = \frac{-b''(\theta)}{a(\varphi)} \quad (3.8)$$

Από τις σχέσεις (3.5) και (3.7) προκύπτει ότι:

$$E\left(\frac{\partial l}{\partial \theta}\right) = E\left\{\frac{[y - b'(\theta)]}{a(\varphi)}\right\} = 0$$

$$\Leftrightarrow E(Y) = \mu = b'(\theta) \quad (3.9)$$

Κατ' επέκταση από την ανωτέρω σχέση (3.9) προκύπτει ότι η μέση τιμή της τ.μ. Y δίνεται από την πρώτη παράγωγο ως προς θ της συνάρτησης b .

Παράλληλα, αντικαθιστώντας από τη σχέση (3.8) στη σχέση (3.6) καταλήγουμε ότι:

$$E\left(\frac{\partial^2 l}{\partial \theta^2}\right) + E\left[\left(\frac{\partial l}{\partial \theta}\right)^2\right] = \frac{-b''(\theta)}{a(\varphi)} + \frac{\text{Var}(y)}{a^2(\varphi)} = 0 \Leftrightarrow V(Y) = a(\varphi)b''(\theta) = \sigma^2 \quad (3.10)$$

Από την σχέση (3.10) συμπεραίνουμε ότι η διακύμανση είναι συνάρτηση της μέσης τιμής μ , Ulf Olsson (2002).

3.1.1 Poisson κατανομή

Η κατανομή Poisson είναι μονοπαραμετρική κατανομή και χρησιμοποιείται όταν η τυχαία μεταβλητή Y είναι διακριτή και αντιπροσωπεύει αριθμό γεγονότων όπως είναι για παράδειγμα ο αριθμός συμβάντων, ατυχημάτων και θανάτων που συμβαίνουν σε συγκεκριμένο χώρο ή χρόνο. Με τον χαρακτηρισμό μονοπαραμετρική εννοείται ότι υπάρχει μόνο μία παράμετρος μ προς εκτίμηση, η οποία συνήθως αναφέρεται ως μέσος ρυθμός εμφάνισης συμβάντων και είναι σταθερή, μη - αρνητική. Μια μοναδική ιδιότητα της κατανομής Poisson ανάμεσα στις γνωστές κατανομές είναι ότι η μέση τιμή της είναι ίση με τη διασπορά της και η οποία αναφέρεται στη διεθνή βιβλιογραφία με τον όρο ισομεταβλητότητα (equidispersion).

Η συνάρτηση πιθανότητας της κατανομής Poisson δίνεται από τον ακόλουθο τύπο:

$$f(y) = \frac{\mu^y e^{-\mu}}{y!} \quad y = 0, 1, 2, \dots \quad (3.11)$$

Βάσει της ιδιότητας της ισομεταβλητότητας, η μέση τιμή και η διασπορά της κατανομής Poisson έχουν την παρακάτω μορφή:

$$E(y) = \text{Var}(y) = \mu$$

Η κατανομή Poisson ανήκει στην εκθετική οικογένεια κατανομών επομένως βάσει της σχέσης (3.1) η συνάρτηση πιθανοφάνειάς της έχει την ακόλουθη μορφή:

$$L(\mu; y) = \prod_{i=1}^N \exp\{y_i \log \mu - \mu - \log(y_i!)\} \quad (3.12)$$

Παίρνοντας λογάριθμους και στις δύο πλευρές της εξίσωσης (3.12) είναι η εξής:

$$l(\mu; y) = \sum_{i=1}^N \{y_i \log \mu - \mu - \log(y_i!)\} \quad (3.13)$$

Επομένως, από τη σχέση (3.13) και την θεωρία της εκθετικής οικογένειας κατανομών προκύπτει ότι η φυσική παράμετρος θ για την κατανομή Poisson δίνεται από την ακόλουθη εξίσωση:

$$\theta = \log \mu \quad (3.14)$$

και

$$b(\theta) = \mu \quad (3.15)$$

Λύνοντας την εξίσωση (3.14) ως προς μ και από τη σχέση (3.15) προκύπτει η μέση τιμή της τ.μ. Y :

$$E(Y) = \mu = e^\theta \quad (3.16)$$

Παράλληλα, ο τύπος της διακύμανσης της τ.μ. Y δίνεται από τον ακόλουθο τύπο:

$$V(Y) = (e^\theta)' = e^\theta \quad (3.17)$$

Επιπλέον, από τον τύπο (3.13) και την θεωρία της εκθετικής οικογένειας κατανομών προκύπτει ότι η παράμετρος κλίμακας ϕ για την κατανομή Poisson ισούται με τη μονάδα. Τέλος, από τις σχέσεις (3.16) και (3.17) επιβεβαιώνεται η ιδιότητα της ισομεταβλητότητας της κατανομής.

3.1.2 Αρνητική Διωνυμική κατανομή

Η Αρνητική Διωνυμική κατανομή είναι διπαραμετρική κατανομή και εφαρμόζεται όταν η τ.μ. Y είναι διακριτή και αντιπροσωπεύει τον αριθμό των αποτυχημένων δοκιμών Bernoulli μέχρι την a επιτυχία. Η Αρνητική Διωνυμική κατανομή προκύπτει από τον συνδυασμό αρκετών κατανομών ένας από τους οποίους είναι η μίξη των κατανομών της Poisson με τη Γάμμα, με την παράμετρο διασποράς a της Αρνητικής Διωνυμικής, η οποία περιγράφει την επιπλέον διασπορά στο μοντέλο, να αποτελεί τη παράμετρο σχήματος της κατανομής Γάμμα (Hilbe, 2012). Για την καλύτερη κατανόηση να αναφέρουμε συνοπτικά ότι η κατανομή Γάμμα ανήκει στην κλάση της οικογένειας των εκθετικών κατανομών και η συνάρτηση πυκνότητας μιας τ.μ. X η οποία ακολουθεί την κατανομή Γάμμα με παραμέτρους a, λ δίνεται από τον ακόλουθο τύπο:

$$f(x) = \frac{\lambda^a}{\Gamma(a)} x^{a-1} e^{-\lambda x}, \quad \text{όπου } x \geq 0, \quad a > 0 \quad \text{και} \quad \lambda > 0 \quad (3.18)$$

Ο τύπος της μέσης τιμής και της διακύμανσης της κατανομής Γάμμα είναι οι ακόλουθοι:

$$E(X) = a\lambda$$

και

$$Var(X) = a\lambda^2$$

Η διακύμανση της Αρνητικής Διωνυμικής κατανομής δίνεται από τον τύπο $\mu + a\mu^2$ ή $\mu(1 + a\mu)$, όπου a είναι η παράμετρος διασποράς της κατανομής, η οποία παρατηρείται ότι όσο μικραίνει και τείνει στο μηδέν η κατανομή προσεγγίζει την κατανομή Poisson.

Η συνάρτηση πιθανότητας μιας τ.μ. Y η οποία ακολουθεί την Αρνητική Διωνυμική κατανομή είναι η εξής:

$$f(y; p, a) = \binom{y+a-1}{a-1} p^a (1-p)^y \quad (3.19)$$

Μία εναλλακτική μορφή της εξίσωσης (3.19) είναι η ακόλουθη:

$$f(y; p, a) = \frac{\Gamma(y+a)}{y! \Gamma(a)} p^a (1-p)^y \quad (3.20)$$

όπου $y = 0, 1, 2, \dots$, $r = 1, 2, \dots$, και $p \in [0, 1]$

Η μέση τιμή και η διασπορά της τ.μ. Y υπολογίζονται από τους παρακάτω τύπους:

$$E(Y) = \mu = \frac{a(1-p)}{p} \quad (3.21)$$

και

$$V(Y) = \frac{a(1-p)}{p^2} \quad (3.22)$$

Στη θεωρία των γενικευμένων γραμμικών μοντέλων που θα αναλύσουμε στο κεφάλαιο 3 χρησιμοποιείται για λόγους ευκολίας η παρακάτω παραμετροποιημένη συνάρτηση πιθανότητας η οποία προκύπτει ως εξής:

Λύνοντας την εξίσωση (3.21) ως προς p και θέτοντας όπου $\varphi = \frac{1}{a}$ καταλήγουμε στην παρακάτω εξίσωση:

$$p = \frac{1}{1 + \varphi\mu} \quad (3.23)$$

Από τη σχέση (3.23) αντικαθιστώντας στη σχέση (3.20) η συνάρτηση πυκνότητας διαμορφώνεται ως εξής:

$$f(y; \mu, \varphi) = \frac{\Gamma\left(y + \frac{1}{\varphi}\right)}{y! \Gamma\left(\frac{1}{\varphi}\right)} \left(\frac{1}{1 + \varphi\mu}\right)^{\frac{1}{\varphi}} \left(\frac{\varphi\mu}{1 + \varphi\mu}\right)^y \quad (3.24)$$

με μέση τιμή $E(Y) = \mu$ και διακύμανση $V(Y) = \mu \left(1 + \frac{\mu}{\varphi}\right)$

Όπως έχει ήδη αναφερθεί, η διωνυμική κατανομή προκύπτει από τη μίξη των κατανομών Poisson και Γάμμα, όπως παρουσιάζεται ακολούθως (Hilbe, 2012):

Υποθέτουμε ότι η τ.μ. Y ακολουθεί την κατανομή Poisson με την ακόλουθη συνάρτηση πιθανότητας:

$$f(y; \mu, \theta) = \frac{e^{-\mu\theta} (\mu\theta)^y}{y!}$$

Παράλληλα υποθέτουμε ότι η τ.μ. θ ακολουθεί την κατανομή Γάμμα με συνάρτηση πυκνότητας την (3.18).

Από το θεώρημα ολικής πιθανότητας προκύπτει ότι:

$$f(y; \theta) = \int_0^{\infty} f(y|\theta) f(\theta) d\theta \quad (3.25)$$

Επομένως από τις σχέσεις (3.18) και (3.25) προκύπτει ότι η πυκνότητα της Y μπορεί να γραφεί στη μορφή:

$$f(y; \theta, \mu, \alpha, \lambda) = \int_0^{\infty} \frac{e^{-\mu\theta} (\mu\theta)^y}{y!} \frac{\lambda^\alpha}{\Gamma(\alpha)} \theta^{\alpha-1} e^{-\lambda\theta} d\theta \quad (3.26)$$

Αν υποθέσουμε ότι $\alpha=\lambda$, τότε η κατανομή Γάμμα είναι μονοπαραμετρική και η εξίσωση (3.26) διαμορφώνεται ως εξής:

$$f(y; \theta, \mu, \alpha) = \int_0^{\infty} \frac{e^{-\mu\theta} (\mu\theta)^y}{y!} \frac{\alpha^\alpha}{\Gamma(\alpha)} \theta^{\alpha-1} e^{-\alpha\theta} d\theta$$

$$f(y; \theta, \mu, \alpha) = \frac{\mu^y \alpha^\alpha}{y! \Gamma(\alpha)} \int_0^{\infty} \theta^{y+\alpha-1} e^{-(\alpha+\mu)\theta} d\theta$$

$$f(y; \mu, \alpha) = \frac{\mu^y \alpha^\alpha \Gamma(y + \alpha)}{y! \Gamma(\alpha) (\alpha + \mu)^{y+\alpha}}$$

$$f(y; \mu, \alpha) = \frac{\mu^y}{y!} \frac{\alpha^\alpha}{\Gamma(\alpha)} \Gamma(y + \alpha) \left(\frac{\mu}{\alpha + \mu} \right)^y \frac{1}{\mu^y} \left(\frac{\alpha}{\alpha + \mu} \right)^\alpha \frac{1}{\alpha^\alpha}$$

$$f(y; \mu, \alpha) = \frac{\Gamma(y + \alpha)}{y! \Gamma(\alpha)} \left(\frac{1}{1 + \frac{\mu}{\alpha}} \right)^\alpha \left(1 - \frac{1}{1 + \frac{\mu}{\alpha}} \right)^y \quad (3.27)$$

Θέτοντας $\alpha = \frac{1}{\varphi}$ στη σχέση (3.27) προκύπτει η σχέση (3.24):

$$f(y; \mu, \varphi) = \frac{\Gamma\left(y + \frac{1}{\varphi}\right)}{y! \Gamma\left(\frac{1}{\varphi}\right)} \left(\frac{1}{1 + \varphi\mu}\right)^{\frac{1}{\varphi}} \left(\frac{\varphi\mu}{1 + \varphi\mu}\right)^y$$

Η Αρνητική Διωνυμική κατανομή ανήκει στην Εκθετική Οικογένεια Κατανομών, αν θεωρήσουμε γνωστή την παράμετρο φ , όπως προκύπτει από την παρακάτω μορφή της συνάρτησης πιθανότητας:

$$l(y; \mu, \varphi) = \sum \left\{ y \log\left(\frac{\varphi\mu}{1 + \varphi\mu}\right) - \frac{1}{\varphi} \log(1 + \varphi\mu) + \log\left(y + \frac{1}{\varphi}\right) - \log y! - \log\left(\frac{1}{\varphi}\right) \right\} \quad (3.28)$$

Από τη σχέση (3.24) και τη θεωρία της εκθετικής οικογένειας κατανομών προκύπτει ότι η φυσική παράμετρος θ έχει την ακόλουθη μορφή:

$$\theta = \log\left(\frac{\varphi\mu}{1 + \varphi\mu}\right) \quad (3.29)$$

Επιλύοντας τη σχέση (3.29) ως προς μ καταλήγουμε στο ακόλουθο αποτέλεσμα:

$$\mu = \frac{e^\theta}{\varphi(1 - e^\theta)} \quad (3.30)$$

Επιπλέον, από τις σχέσεις (3.24) και (3.30) προκύπτει ότι:

$$b(\theta) = -\frac{1}{\varphi} \log(1 + \varphi\mu) \quad (3.31)$$

Ο τύπος της μέσης τιμής της τ.μ. Y δίνεται από τον παρακάτω τύπο:

$$b'(\theta) = \frac{1}{1 + \varphi\mu} \mu(1 + \varphi\mu) = \mu$$

Παράλληλα, ο τύπος της διακύμανσης της τ.μ. Y προκύπτει ως εξής:

$$\text{Var}(Y) = b''(\theta) = \frac{\partial^2 b}{\partial \mu^2} \left(\frac{\partial \mu}{\partial \theta} \right)^2 + \frac{\partial b}{\partial \mu} \frac{\partial^2 \mu}{\partial \theta^2} = \mu + \varphi \mu^2$$

3.1.3 Διωνυμική κατανομή

Η Διωνυμική κατανομή είναι μονοπαραμετρική κατανομή, αν θεωρήσουμε γνωστή την παράμετρο p και χρησιμοποιείται όταν η τυχαία μεταβλητή Y είναι διακριτή και αντιπροσωπεύει τον αριθμό των επιτυχημένων δοκιμών Bernoulli σε n ανεξάρτητες δοκιμές με πιθανότητα επιτυχίας p .

Η συνάρτηση πιθανότητας της κατανομής είναι η εξής:

$$f(y) = \binom{n}{y} p^y (1-p)^{n-y} \quad (3.32)$$

$$\text{όπου } y = 0, 1, 2, \dots, n \text{ και } p \in [0, 1]$$

$$\text{με μέση τιμή } E(Y) = np \text{ και διακύμανση } V(Y) = np(1-p)$$

Μια διωνυμική τυχαία μεταβλητή μετατρέπεται σε αναλογία αν τη διαιρέσουμε με το συνολικό αριθμό δοκιμών n , κατ' επέκταση η νέα μεταβλητή $\frac{y}{n}$ που προκύπτει ονομάζεται διωνυμική αναλογία.

Η Διωνυμική κατανομή ανήκει στην Εκθετική Οικογένεια Κατανομών όπως προκύπτει από την παρακάτω μορφή της συνάρτησης πιθανοφάνειας:

$$\begin{aligned} l(y; p) &= \sum_{i=1}^N \left\{ y_i \ln p_i + n \ln(1-p_i) - y_i \ln(1-p_i) + \ln \binom{n}{y_i} \right\} \\ &= \sum_{i=1}^N \left\{ y_i \ln \left(\frac{p_i}{1-p_i} \right) + n \ln(1-p_i) + \ln \binom{n}{y_i} \right\} \end{aligned} \quad (3.33)$$

Επομένως, από τη σχέση (3.33) και την θεωρία της εκθετικής οικογένειας κατανομών προκύπτει ότι η φυσική παράμετρος θ για την Διωνυμική κατανομή δίνεται από την ακόλουθη εξίσωση:

$$\theta = \ln \left(\frac{p}{1-p} \right) \quad (3.34)$$

Επιλύοντας τη σχέση (3.34) ως προς p λαμβάνουμε το ακόλουθο αποτέλεσμα:

$$p = \frac{e^{\theta}}{1 + e^{\theta}} \quad (3.35)$$

Σύμφωνα με τη θεωρία της εκθετικής οικογένειας κατανομών και τη σχέση (3.33) προκύπτει το ακόλουθο αποτέλεσμα λαμβάνοντας υπόψη τη σχέση (3.35):

$$b(\theta) = n \ln(1-p) = -n \ln(1 + e^{\theta}) \quad (3.36)$$

Παραγωγίζοντας ως προς θ τη σχέση (3.36) και λαμβάνοντας υπόψη τη σχέση (3.33) προκύπτει ο τύπος της μέσης τιμής της τ.μ. Y :

$$E(Y) = \frac{ne^{\theta}}{1 + e^{\theta}} = np \quad (3.37)$$

Ακολούθως, παραγωγίζοντας δύο φορές ως προς θ τη σχέση (3.36) προκύπτει ο τύπος της διακύμανσης της διακύμανσης της τ.μ. Y :

$$V(Y) = \frac{ne^{\theta}}{(1 + e^{\theta})^2} = np(1 - p) \quad (3.38)$$

3.2 Βασική δομή Γενικευμένων Γραμμικών μοντέλων

Το 1972 οι Nelder και Wedderburn ανέπτυξαν τη θεωρία των Γενικευμένων Γραμμικών Μοντέλων (GLM) η οποία ενοποιεί την ήδη υπάρχουσα μέχρι τότε θεωρία γραμμικών μοντέλων, η οποία διευρύνεται σε μια ευρεία κλάση κατανομών η οποία ονομάζεται Εκθετική Οικογένεια Κατανομών. Στην πραγματικότητα η θεωρία αυτή συμπεριλαμβάνει για πρώτη φορά τις περιπτώσεις των logit μοντέλων τα οποία θα δούμε αναλυτικά στην συνέχεια και τα οποία είναι χρήσιμα για την γενικότερη διαχείριση διωνυμικών και μετρήσιμων δεδομένων, αντίστοιχα. Μια σημαντική συνεισφορά της θεωρίας Γενικευμένων Γραμμικών Μοντέλων είναι η μέθοδος Επανασταθμιζόμενου Αλγορίθμου Ελαχίστων Τετραγώνων (Iterated Least Squares algorithm - IRLS) για την εύρεση του εκτιμητή μέγιστης πιθανοφάνειας των παραμέτρων στα μοντέλα. Στην Αναλογιστική επιστήμη η προαναφερόμενη θεωρία έχει χρησιμοποιηθεί στον κλάδο ασφάλισης περιουσίας και αστικής ευθύνης για την εκτίμηση συχνότητας και έντασης αποζημιώσεων, την εκτίμηση του δείκτη κέρδους / ζημίας (loss ratio) καθώς και την τιμολόγηση ασφαλιστικών παροχών Goldburd, et al. (2016), τη μοντελοποίηση της πιθανότητας θανάτου Naufal N., et al(2019) και την διαδικασία αποθεματοποίησης.

Στο κλασσικό γραμμικό μοντέλο παλινδρόμησης, η εξαρτημένη μεταβλητή Y εκφράζεται μέσω της ακόλουθης σχέσης :

$$E(Y_i) = \mu_i = x_i^T \beta \quad \text{όπου} \quad Y_i \sim N(\mu_i, \sigma^2)$$

Όπου β είναι ένα διάνυσμα άγνωστων παραμέτρων, x είναι ένα διάνυσμα γνωστών εξαρτημένων μεταβλητών και ε είναι η παράμετρος των σφαλμάτων. Στο κλασσικό γραμμικό μοντέλο υποθέτουμε ότι η μέση τιμή των σφαλμάτων είναι μηδέν και η διακύμανσή τους είναι σταθερή, έτσι καταλήγουμε ότι $\mu = E(Y) = x^T \beta$.

Ένα γενικευμένο γραμμικό μοντέλο αποτελείται από τρεις συνιστώσες:

1. Μια τυχαία μεταβλητή, που προσδιορίζει τη δεσμευμένη κατανομή της μεταβλητής απόκρισης Y για την οποία ισχύει ότι $E(Y) = \mu$ και $\text{Var}(Y) = \sigma^2$. Η κατανομή της μεταβλητής Y η οποία μπορεί να είναι είτε συνεχής είτε διακριτή τυχαία μεταβλητή θα πρέπει να ανήκει στην Εκθετική Οικογένεια Κατανομών χωρίς να είναι απαραίτητα η κανονική κατανομή.

2. Μια γραμμική πρόβλεψη, που είναι γραμμική συνάρτηση των ερμηνευτικών μεταβλητών x_i

$$\eta_i = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_k x_k = \sum_{i=1}^k x_i \beta_i \quad (3.39)$$

3. Μια αντιστρέψιμη γραμμική συνάρτηση σύνδεσης $g(\cdot)$, η οποία μετασχηματίζει την αναμενόμενη τιμή της μεταβλητής απόκρισης $\mu_i = E(Y_i)$ στη γραμμική πρόβλεψη:

$$g(\mu_i) = \eta_i = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_k x_k \quad (3.40)$$

και

$$g[E(Y_i)] = \eta_i = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_k x_k \quad (3.41)$$

Επειδή υποθέτουμε ότι η συνάρτηση σύνδεσης είναι αντιστρέψιμη η τελευταία σχέση μπορεί να γραφεί και ως εξής:

$$\mu_i = g^{-1}(\eta_i) = g^{-1}(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_k x_k) \quad (3.42)$$

Η συνάρτηση σύνδεσης περιγράφει πως ο μέσος όρος της απόκρισης και ένας γραμμικός συνδυασμός των προβλέψεων, ενώ σύμφωνα με τον Lindsey (2000) θα πρέπει να είναι μονότονη και παραγωγίσιμη συνάρτηση.

Στον ακόλουθο Πίνακα 3.1 περιλαμβάνονται κάποιες βασικές κατανομές της κλάσης της εκθετικής οικογένειας κατανομών συνοδευτικά με τις αντίστοιχες πληροφορίες:

Πίνακας 3.1: Υπό εξέταση κατανομές και οι συναρτήσεις σύνδεσής τους.

Κατανομή	Συνάρτηση Σύνδεσης $g(\mu)$	Όνομα Συνάρτησης Σύνδεσης $g(\mu)$	$V(Y)$	Παράμετρος Κλίμακας ϕ
Poisson	$\log \mu$	Log	μ	1
Διωνυμική	$\log\left(\frac{p}{1-p}\right)$	Logit	$np(1-p)$	1
Αρνητική Διωνυμική	$\log\left(\frac{\mu}{\mu+a}\right)$	Log	$\mu + a\mu^2$	$\frac{1}{a}$

3.3 Σύνδεση μεταξύ Γενικευμένων Γραμμικών Μοντέλων και Κλασικού Μοντέλου Γραμμικής Παλινδρόμησης

Στην παράγραφο αυτή θα αναλύσουμε την σχέση μεταξύ των γενικευμένων γραμμικών μοντέλων και του κλασικού γραμμικού μοντέλου παλινδρόμησης και υπό ποιες προϋποθέσεις τα γενικευμένα γραμμικά μοντέλα αποτελούν επέκταση της θεωρίας της γραμμικής παλινδρόμησης. Όπως έχουμε ήδη αναφέρει το κλασικό γραμμικό μοντέλο παλινδρόμησης βασίζεται στην υπόθεση της κανονικής κατανομής των σφαλμάτων. Η συνάρτηση πυκνότητας της κανονικής κατανομής ή κατανομής Gauss δίνεται από την παρακάτω σχέση:

$$f(y; \mu, \sigma^2) = \frac{1}{\sqrt{(2\pi\sigma^2)}} \exp\left[-\frac{(y - \mu)^2}{2\sigma^2}\right] \quad (3.43)$$

Σύμφωνα με τη θεωρία των γενικευμένων γραμμικών μοντέλων, η συνάρτηση πυκνότητας πιθανότητας της εξαρτημένης μεταβλητής θα πρέπει να ανήκει στην Εκθετική Οικογένεια

Κατανομών. Η συνάρτηση της κανονικής κατανομής δύναται να γραφεί στην μορφή (3.1) όπως παρουσιάζεται ακολούθως:

$$\begin{aligned}
 f(y; \mu, \sigma^2) &= \exp \left[-\frac{(y - \mu)^2}{2\sigma^2} - \frac{1}{2} \log(2\pi\sigma^2) \right] \\
 &= \exp \left[-\frac{1}{2} \left(\frac{y^2}{\sigma^2} + \frac{\mu^2}{\sigma^2} - \frac{2y\mu}{\sigma^2} + \log(2\pi\sigma^2) \right) \right] \\
 &= \exp \left[\frac{y\mu - \frac{\mu^2}{2}}{\sigma^2} - \frac{y^2}{2\sigma^2} - \frac{1}{2} \log(2\pi\sigma^2) \right] \quad (3.44)
 \end{aligned}$$

Από τη σχέση (3.44) προκύπτει ότι η κανονική κατανομή ανήκει στην κλάση της εκθετικής οικογένειας κατανομών ενώ επιπλέον προκύπτει ότι η φυσική παράμετρος θ για την δίνεται από την ακόλουθη σχέση:

$$\theta = \mu \quad (3.45)$$

Από τη θεωρία της εκθετικής οικογένειας κατανομών και τις σχέσεις (3.44) και (3.45) καταλήγουμε ότι:

$$b(\theta) = \frac{\mu^2}{2} = \frac{\theta^2}{2} \quad (3.46)$$

Η μέση τιμή της τ.μ. Y δίνεται από την ακόλουθη σχέση:

$$E(Y) = b'(\theta) = \theta = \mu \quad (3.47)$$

Επιπλέον, από τη σχέση (3.44) προκύπτει ότι η παράμετρος κλίμακας φ είναι ίση με:

$$\alpha(\varphi) = \sigma^2 \quad (3.48)$$

Από τις σχέσεις (3.10) και (3.48) καταλήγουμε στον τύπο της διακύμανσης της τ.μ. Y :

$$V(Y) = \alpha(\varphi)b''(\theta) = \sigma^2 \quad (3.49)$$

Η πιο γνωστή ειδική περίπτωση Γενικευμένων Γραμμικών Μοντέλων είναι της μορφής:

$$g(\mu_i) = g[E(Y_i)] = \mu_i = \eta_i = \sum_i x_i \beta_i \quad (3.50)$$

Η ανωτέρω σχέση (3.46) προκύπτει όταν η εξαρτημένη μεταβλητή Y ακολουθεί την κανονική κατανομή με μέση τιμή μ και διακύμανση σ^2 .

$$E(Y_i) = \mu_i = x_i^T \beta \quad \text{όπου} \quad Y_i \sim N(\mu_i, \sigma^2)$$

Επιπλέον, παρατηρούμε ότι η συνάρτηση σύνδεσης είναι η $g(\mu_i) = \mu_i$, η οποία ονομάζεται ταυτοτική και είναι η συνήθης συνάρτηση σύνδεσης η οποία χρησιμοποιείται στην περίπτωση της Κανονικής κατανομής. Με τη χρήση της ταυτοτικής συνάρτησης σύνδεσης η γραμμική πρόβλεψη είναι ίση με τη μέση τιμή της τ.μ. Y . Παράλληλα, στο μοντέλο (3.46) ισχύει η υπόθεση της κανονικής κατανομής των σφαλμάτων ε με μέση τιμή μηδέν και σταθερή διακύμανση σ^2 . Επομένως, με βάση τα ανωτέρω από το Γενικευμένο Γραμμικό Μοντέλο με τις παραπάνω υποθέσεις προκύπτει το κλασικό μοντέλο γραμμικής παλινδρόμησης.

3.4 Εκτίμηση παραμέτρων Γενικευμένων Γραμμικών Μοντέλων

3.4.1 Μέθοδος Newton – Raphson

Η μέθοδος Newton – Raphson αποτελεί έναν αλγόριθμο αριθμητικής βελτιστοποίησης για την εύρεση των ριζών μιας διαφοροποιήσιμης συνάρτησης. Όσον αφορά τα γενικευμένα γραμμικά μοντέλα, η μέθοδος Newton – Raphson χρησιμοποιείται για την εύρεση των εκτιμητών μέγιστης πιθανοφάνειας των παραμέτρων ενός μοντέλου. Ο αλγόριθμος ξεκινά με μια αρχική εκτίμηση των παραμέτρων και επαναλαμβάνεται μέχρι να επιτευχθεί σύγκλιση. Η εκτίμηση του αλγορίθμου βασίζεται σε μια τροποποίηση της σειράς Taylor. Η σειρά Taylor έχει την εξής μορφή:

$$f(x_0) + (x_1 - x_0)f'(x_0) + \frac{(x_1 - x_0)^2}{2!} f''(x_0) + \frac{(x_1 - x_0)^3}{3!} f'''(x_0) + \dots + \frac{(x_1 - x_0)^k}{k!} f^{(k)}(x_0) \quad (3.51)$$

Λαμβάνοντας τους δύο πρώτους όρους της σχέσης (3.51) έχουμε :

$$0 = f(x_0) + (x_1 - x_0)f'(x_0) \Leftrightarrow x_1 = x_0 - \frac{f(x_0)}{f'(x_0)} \quad (3.52)$$

Ακολούθως, θεωρώντας τη λογαριθμική συνάρτηση πιθανοφάνειας $l(\beta; y)$ και σύμφωνα με τη μέθοδο εκτίμησης των Newton – Raphson η ανωτέρω σχέση (3.52) χρησιμοποιείται για την εκτίμηση παραμέτρων των γενικευμένων γραμμικών μοντέλων ως εξής:

$$\beta_j = \beta_{j-1} - \frac{\frac{\partial l}{\partial \beta}(\beta_{j-1})}{\frac{\partial^2 l}{\partial^2 \beta}(\beta_{j-1})} \quad (3.53)$$

Όπου β_{j-1} και β_j είναι εκτιμήσεις των παραμέτρων στις επαναλήψεις $j - 1$ και j , αντίστοιχα. Στη σχέση (3.49) η ποσότητα

$$\frac{\partial^2 l}{\partial^2 \beta}(\beta_{j-1}) \quad (3.54)$$

καλείται Εσσιανός πίνακας (Hessian matrix) ενώ ο αντίστροφός του ο οποίος απαιτείται στη σχέση (3.53) καλείται πίνακας πληροφορίας του Fisher:

$$\left(\frac{\partial^2 l}{\partial^2 \beta}(\beta_{j-1}) \right)^{-1} \quad (3.55)$$

3.4.2 Αλγόριθμος Επανασταθμιζόμενων Ελαχίστων Τετραγώνων (IRLS)

Ο αλγόριθμος επανασταθμιζόμενων ελαχίστων τετραγώνων είναι μια αριθμητική μέθοδος η οποία χρησιμοποιείται για την εκτίμηση παραμέτρων στη θεωρία των Γενικευμένων Γραμμικών Μοντέλων. Ο εν λόγω αλγόριθμος αποτελεί μια παραλλαγή της κλασικής μεθόδου εκτίμησης παραμέτρων των ελαχίστων τετραγώνων, η οποία λαμβάνει υπόψη τη μη γραμμική σχέση μεταξύ της συνάρτησης σύνδεσης και των ερμηνευτικών μεταβλητών του μοντέλου. Αποτελεί ένα χρήσιμο εργαλείο για τη θεωρία των γενικευμένων γραμμικών μοντέλων καθώς δύναται να προχωρήσει σε εκτιμήσεις και για μη γραμμικές σχέσεις μεταξύ της μεταβλητής απόκρισης και των μεταβλητών πρόβλεψης. Όπως έχει αναφερθεί στην παράγραφο 2.2 στη θεωρία γενικευμένων γραμμικών μοντέλων γίνεται χρήση του Αλγορίθμου Επανασταθμιζόμενων Ελαχίστων Τετραγώνων, ο οποίος είναι μια απλοποίηση της μεθόδου

εκτίμησης μέγιστης πιθανοφάνειας και χρησιμοποιείται αποκλειστικά για κατανομές οι οποίες ανήκουν στην εκθετική οικογένεια κατανομών.

Στην εκθετική οικογένεια κατανομών, η λογαριθμική συνάρτηση της πιθανοφάνειας έχει την παρακάτω μορφή, η οποία περιγράφεται από τη σχέση (3.4):

$$\frac{\partial l}{\partial \beta} = \sum \left[\frac{y\theta - b(\theta)}{a(\varphi)} + c(y, \varphi) \right]$$

Για την παραγωγή της l ως προς τις παραμέτρους β γίνεται χρήση από τον απειροστικό λογισμό του κανόνα της αλυσίδας Ambrosio και Dal Maso (1990) λαμβάνοντας υπόψη τις κατάλληλες υποθέσεις παραγωγής

$$\frac{\partial l}{\partial \beta} = \frac{\partial l}{\partial \theta} \frac{\partial \theta}{\partial \mu} \frac{\partial \mu}{\partial \eta} \frac{\partial \eta}{\partial \beta} \quad (3.56)$$

Παραγωγίζοντας μερικώς ως προς θ τη σχέση (3.4) και λαμβάνοντας υπόψη τη σχέση (3.9) προκύπτει ο ακόλουθος τύπος:

$$\frac{\partial l}{\partial \theta} = \sum \left[\frac{y\theta - b'(\theta)}{a(\varphi)} \right] = \sum \left[\frac{y - \mu}{a(\varphi)} \right] \quad (3.57)$$

Παράλληλα, παραγωγίζοντας ως προς θ τη σχέση (3.9) καταλήγουμε ότι:

$$\frac{\partial \mu}{\partial \theta} = \frac{\partial b'(\theta)}{\partial \theta} = b''(\theta) = V(\mu) \quad (3.58)$$

Επομένως, αντιστρέφοντας τη σχέση (3.57) καταλήγουμε στο ακόλουθο αποτέλεσμα:

$$\frac{\partial \theta}{\partial \mu} = \frac{1}{V(\mu)} \quad (3.59)$$

Από τη σχέση (3.39) της γραμμικής συνάρτησης σύνδεσης η οποία αναφέρθηκε στην παράγραφο 4.2 γνωρίζουμε ότι:

$$\eta_i = \sum_i x_i \beta_i$$

Παραγωγίζοντας μερικώς ως προς θ τη σχέση (3.39) προκύπτει ότι:

$$\frac{\partial \eta}{\partial \beta} = \frac{\partial (x\beta)}{\partial \beta} = x \quad (3.60)$$

Επιπλέον, παραγωγίζοντας τη σχέση (3.41) ως προς μ ισχύουν τα κάτωθι:

$$\frac{\partial \eta}{\partial \mu} = [g^{-1}(\eta)]' = \frac{1}{\frac{\partial \eta}{\partial \mu}} = \frac{1}{g^{-1}(\mu)} \quad (3.61)$$

Κατ' επέκταση η σχέση (3.56) παίρνει την ακόλουθη μορφή:

$$\frac{\partial l}{\partial \beta} = \sum \left[\frac{(y - \mu)}{a(\varphi)} \frac{1}{V(\mu)} x \frac{1}{g^{-1}(\mu)} \right] = \sum \left[\frac{(y - \mu)x}{a(\varphi)V(\mu)g^{-1}(\mu)} \right] = 0 \quad (3.62)$$

Ο αλγόριθμος για τα γενικευμένα γραμμικά μοντέλα αντικαθιστά την πληροφορία Fisher I , η οποία χρησιμοποιείται για την εκτίμηση μέγιστης πιθανοφάνειας, με τον Εσσιανό πίνακα H ο οποίος είναι ένας τετράγωνος πίνακας μερικών παραγώγων δεύτερης τάξης της λογαριθμικής συνάρτησης l .

$$I = -E \left[\frac{\partial^2 l}{\partial \beta \partial \beta'} \right] = E \left[\frac{\partial l}{\partial \beta} \frac{\partial l}{\partial \beta'} \right] \quad (3.63)$$

Αντικαθιστώντας από τη σχέση (3.62) στη σχέση (3.63) προκύπτει το ακόλουθο αποτέλεσμα:

$$I = \frac{\vartheta}{\partial \beta_j} \left[\frac{(y - \mu)x_j}{a(\varphi)V(\mu)} \frac{\partial \mu}{\partial \eta} \right] * \frac{\partial l}{\partial \beta_k} \left[\frac{(y - \mu)x_k}{a(\varphi)V(\mu)} \frac{\vartheta \mu}{\partial \eta} \right]$$

$$I = \frac{(y - \mu)^2 x_j x_k}{a(\varphi)V(\mu)^2} \left[\frac{\partial \mu}{\partial \eta} \right]^2 \quad (3.64)$$

Από τη θεωρία της εκθετικής οικογένειας κατανομών:

$$V(y) = a(\varphi)V(\mu) = (y - \mu)^2 \quad (3.65)$$

Αντικαθιστώντας στη σχέση (3.64) από τις σχέσεις (3.61) και (3.65) προκύπτει ότι:

$$I = \frac{x_j x_k}{V(y)} \left[\frac{\partial \mu}{\partial \eta} \right]^2 = \frac{x_j x_k}{V(y) g'^2} \quad (3.66)$$

Αντικαθιστώντας στη σχέση (3.53) ο τύπος του αλγορίθμου του Newton – Raphson διαμορφώνεται ως εξής:

$$\beta_r = \beta_{r-1} - \left[\frac{x_j x_k}{V(y)} \left[\frac{\partial \mu}{\partial \eta} \right]^2 \right]^{-1} \left[\frac{x_k (y - \mu)}{V(y)} \left[\frac{\partial \mu}{\partial \eta} \right] \right] \quad (3.67)$$

και

$$\left[\frac{x_j x_k}{V(y)} \left[\frac{\partial \mu}{\partial \eta} \right]^2 \right] \beta_r = \left[\frac{x_j x_k}{V(y)} \left[\frac{\partial \mu}{\partial \eta} \right]^2 \right] \beta_{r-1} + \left[\frac{x_k (y - \mu)}{V(y)} \left[\frac{\partial \mu}{\partial \eta} \right] \right] \quad (3.68)$$

Θέτοντας τις ακόλουθες ποσότητες:

$$W^{-1} = \left(\frac{\partial \eta}{\partial \mu} \right)^2 V \quad (3.69)$$

και

$$z = \eta + (y - \mu) \left(\frac{\partial \eta}{\partial \mu} \right) \quad (3.70)$$

Καταλήγουμε ότι η σχέση (3.68) λόγω των σχέσεων (3.69) και (3.70) διαμορφώνεται ως εξής:

$$[X'WX] \beta_r = X'Wz \quad (3.71)$$

και

$$\beta_r = [X'WX]^{-1} X'Wz \quad (3.72)$$

όπου β_r είναι ο πίνακας της παλινδρόμησης με βάρη ο οποίος χρησιμοποιείται στον αλγόριθμο για την εκτίμηση των παραμέτρων.

3.5 Διαγνωστικοί έλεγχοι καλής προσαρμογής

Οι έλεγχοι καλής προσαρμογής χρησιμοποιούνται για την αξιολόγηση της προσαρμογής των Γενικευμένων Γραμμικών Μοντέλων στα δεδομένα. Ακολούθως, θα περιγράψουμε τους δημοφιλέστερες διαγνωστικούς ελέγχους και μέτρα καλής προσαρμογής όπως τα κατάλοιπα, η απόκλιση, ο έλεγχος χ^2 του Pearson, τα κριτήρια πληροφορίας του Akaike (AIC) και του Schwarz (BIC) καθώς και τον συντελεστή προσδιορισμού R^2 .

3.5.1 Απόκλιση - Deviance

Η απόκλιση (deviance) είναι ένα στατιστικό μέτρο αξιολόγησης της προσαρμογής ενός μοντέλου στα παρατηρούμενα δεδομένα και χρησιμοποιείται ευρέως στη θεωρία των γενικευμένων γραμμικών μοντέλων. Η απόκλιση αποτελεί τη μέτρηση της απόστασης ενός μοντέλου από το κορεσμένο, το οποίο είναι ένα μοντέλο που ταιριάζει απόλυτα στα δεδομένα καθώς χρησιμοποιεί τόσες παραμέτρους όσες και οι παρατηρήσεις. Η απόκλιση ακολουθεί ασυμπτωτικά την κατανομή χ^2 με $n-p$ βαθμούς ελευθερίας όπου n είναι ο αριθμός των παρατηρήσεων και p είναι ο αριθμός των παραμέτρων του υπό εξέταση μοντέλου. Όσο μικρότερη είναι η τιμή της απόκλισης από το κορεσμένο μοντέλο τόσο καλύτερη είναι η προσαρμογή του μοντέλου στα δεδομένα. Επομένως, μια τιμή του στατιστικού μέτρου ίση με μηδέν υποδηλώνει τέλεια προσαρμογή στα δεδομένα ενώ μεγάλες τιμές απόκλισης υποδηλώνουν κακή προσαρμογή του μοντέλου. Η απόκλιση προσδιορίζεται από τον ακόλουθο τύπο:

$$D = -2 \ln \left(\frac{\hat{L}}{\hat{L}_{max}} \right) = -2(\hat{l} - \hat{l}_{max}) \quad (3.73)$$

όπου:

\hat{L} : η συνάρτηση πιθανοφάνειας του υπό εξέταση μοντέλου

\hat{L}_{max} : η συνάρτηση πιθανοφάνειας του κορεσμένου μοντέλου

\hat{l} : η λογαριθμική συνάρτηση πιθανοφάνειας του υπό εξέταση μοντέλου

\hat{l}_{max} : η λογαριθμική συνάρτηση πιθανοφάνειας του κορεσμένου μοντέλου, η οποία είναι ίση με μηδέν

Συνολικά, η απόκλιση είναι ένα χρήσιμο στατιστικό μέτρο για την θεωρία των γενικευμένων γραμμικών μοντέλων καθώς μας δίνει αξιόπιστα αποτελέσματα για την αξιολόγηση της προσαρμογής ενός μοντέλου και τη σύγκριση της προσαρμογής διαφορετικών μοντέλων μεταξύ τους, λαμβάνοντας υπόψη την πολυπλοκότητά τους.

3.5.2 Ανάλυση καταλοίπων

Η ανάλυση καταλοίπων είναι μια συνήθης στατιστική τεχνική για την αξιολόγηση της προσαρμογής ενός υπό εξέταση μοντέλου στα δεδομένα. Στη θεωρία της απλής γραμμικής παλινδρόμησης τα κατάλοιπα ορίζονται ως η διαφορά ή απόκλιση της i πρόβλεψης του μοντέλου από την αντίστοιχη i παρατηρούμενη τιμή. Ακολούθως δίνεται ο μαθηματικός τύπος υπολογισμού:

$$r_i = y_i - \hat{\mu}_i \quad (3.74)$$

όπου:

y_i : η i παρατηρούμενη τιμή

$\hat{\mu}_i$: η i πρόβλεψη του μοντέλου

Στην περίπτωση των γενικευμένων γραμμικών μοντέλων δεν εφαρμόζονται τα απλά κατάλοιπα για την αξιολόγηση της καταλληλότητας ενός μοντέλου καθώς δεν λαμβάνουν υπόψη στον υπολογισμό τους την διακύμανση η οποία συνήθως δεν είναι σταθερή. Εξαιρεση αποτελεί η ειδική περίπτωση της κανονικής κατανομής την οποία αναλύσαμε στην παράγραφο 3.3. Ακολούθως παρουσιάζονται τα πιο σημαντικά μέτρα αξιολόγησης καταλληλότητας των γενικευμένων γραμμικών μοντέλων μέσω των καταλοίπων.

3.5.3 Κατάλοιπα Pearson

Τα κατάλοιπα Pearson χρησιμοποιούνται ευρέως για την αξιολόγηση της καταλληλότητας ενός μοντέλου το οποίο ανήκει στην κλάση των γενικευμένων γραμμικών μοντέλων. Υπολογίζονται ως η διαφορά της i πρόβλεψης του μοντέλου από την αντίστοιχη i παρατηρούμενη τιμή, όπως στη σχέση (3.73) για τα απλά κατάλοιπα, διαιρεμένη με την τετραγωνική ρίζα της διακύμανσης της i πρόβλεψης του μοντέλου. Ο μαθηματικός τύπος είναι ο εξής:

$$r_i^p = \frac{y_i - \hat{\mu}_i}{\sqrt{V(\hat{\mu}_i)}} \quad (3.75)$$

όπου:

y_i : η i παρατηρούμενη τιμή

$\hat{\mu}_i$: η i πρόβλεψη του μοντέλου

$V(\mu_i)$: η διακύμανση της i πρόβλεψης του μοντέλου

Κατά τον υπολογισμό του τύπου (3.75), εάν η τιμή που θα προκύψει είναι ίση ή προσεγγίζει το μηδέν τότε υπάρχουν ισχυρές ενδείξεις ότι το υπό εξέταση μοντέλο προσαρμόζεται ικανοποιητικά στα δεδομένα. Αντίστοιχα, εάν προκύψουν μεγάλες τιμές μέσω του τύπου (3.75) τότε υπάρχουν ενδείξεις ότι είτε δεν είναι ικανοποιητική η προσαρμογή του μοντέλου στα δεδομένα είτε παρατηρούνται ακραίες τιμές στα δεδομένα.

3.5.4 Κατάλοιπα απόκλισης (Deviance)

Τα κατάλοιπα απόκλισης (deviance) είναι ένα στατιστικό μέτρο το οποίο χρησιμοποιείται στο πλαίσιο των γενικευμένων γραμμικών μοντέλων για την αξιολόγηση καταλληλότητας ενός μοντέλου. Τα εν λόγω κατάλοιπα συνδέονται άμεσα με την απόκλιση για την οποία έγινε αναφορά στην παράγραφο 3.5.1. Ο μαθηματικός τύπος υπολογισμού τους είναι ο ακόλουθος:

$$r_i^D = \text{sign}(y_i - \hat{\mu}_i) \sqrt{d_i} \quad (3.76)$$

όπου:

y_i : η i παρατηρούμενη τιμή

$\hat{\mu}_i$: η i πρόβλεψη του μοντέλου

d_i : η τιμή της απόκλισης της i πρόβλεψης του μοντέλου από την αντίστοιχη παρατηρούμενη τιμή

Η συνάρτηση sign λαμβάνει την τιμή +1 αν $y_i > \hat{\mu}_i$, -1 αν $y_i < \hat{\mu}_i$ και 0 αν $y_i = \hat{\mu}_i$.

Παράλληλα ισχύουν οι παρακάτω σχέσεις:

$$\sum_i (r_i^D)^2 = D$$

και

$$\sum_i d_i = D$$

όπου D είναι η τιμή απόκλισης του μοντέλου.

Όσο η τιμή των καταλοίπων deviance προσεγγίζει το μηδέν τόσο καλύτερη είναι η προσαρμογή του μοντέλου στα δεδομένα. Τα κατάλοιπα deviance αποτελούν ένα χρήσιμο στατιστικό εργαλείο για τα γενικευμένα γραμμικά μοντέλα καθώς παρέχουν τη δυνατότητα εντοπισμού, μέσω της εκτιμώμενης μεταβλητότητας, παρατηρήσεων που έχουν μεγαλύτερη απόκλιση από τις αναμενόμενες τιμές, ειδικά για τις περιπτώσεις κατανομών όπου τα σφάλματα κατανέμονται κανονικά.

3.5.5 Έλεγχος χ^2 του Pearson

Ένας στατιστικός έλεγχος καλής προσαρμογής του μοντέλου στα δεδομένα είναι ο έλεγχος χ^2 του Pearson ο οποίος υπολογίζει την διαφορά μεταξύ των παρατηρούμενων και των εκτιμημένων τιμών του μοντέλου. Ο εν λόγω έλεγχος υπολογίζεται από το άθροισμα των τετραγώνων των κανονικοποιημένων καταλοίπων. Όσο μικρότερη είναι η τιμή του ελέγχου τόσο καλύτερη είναι η προσαρμογή του μοντέλου στα δεδομένα, ενώ ακολούθως δίνεται ο τύπος υπολογισμού:

$$\chi^2 = \sum_i \left(\frac{O_i - E_i}{\sqrt{E_i}} \right)^2 \quad (3.77)$$

Οι όροι $r_i^p = \frac{O_i - E_i}{\sqrt{E_i}}$ ονομάζονται κατάλοιπα Pearson και προσδιορίζονται ως το πηλίκο των απλών καταλοίπων $O_i - E_i$ ως προς την τετραγωνική ρίζα των εκτιμώμενων τιμών της μεταβλητής απόκρισης Y , ενώ το άθροισμα των τετραγώνων των καταλοίπων Pearson από τη σχέση (3.75)

$$\sum_i (r_i^p)^2$$

αποτελούν τον έλεγχο καλής προσαρμογής χ^2 του Pearson.

3.5.6 Κριτήρια Πληροφορίας

Το κριτήριο πληροφορίας του Akaike (AIC) αναπτύχθηκε από τον στατιστικό Akaike (1973). Το κριτήριο του Akaike αποτελεί μια μαθηματική μέθοδο η οποία χρησιμοποιείται για τη σύγκριση διαφορετικών στατιστικών μοντέλων και την αξιολόγηση ως προς την εφαρμογή

τους στα δεδομένα. Για τον υπολογισμό του εν λόγω κριτηρίου χρησιμοποιείται ο αριθμός των παραμέτρων που χρησιμοποιούνται για την κατασκευή του μοντέλου και η τιμή του εκτιμητή μέγιστης πιθανοφάνειας του μοντέλου, όπως περιγράφεται στον ακόλουθο τύπο:

$$AIC = -2l + 2k \quad (3.78)$$

όπου l είναι ο λογάριθμος της συνάρτησης πιθανοφάνειας και k ο αριθμός παραμέτρων του μοντέλου.

Καθώς αυξάνονται οι παράμετροι του μοντέλου, η ποσότητα $-2l$ μικραίνει γι' αυτό το λόγο στο κριτήριο προστέθηκε ο όρος $2k$, ο οποίος αποτελεί έναν όρο ποινής και αυξάνει την τιμή του κριτηρίου με την αύξηση του αριθμού των παραμέτρων. Το μοντέλο με τη καλύτερη προσαρμογή στα δεδομένα σύμφωνα με το κριτήριο AIC είναι αυτό που εξηγεί τη μεγαλύτερη ποσότητα διακύμανσης με το μικρότερο δυνατό αριθμό παραμέτρων. Έτσι, το μοντέλο με τη μικρότερη τιμή του κριτηρίου είναι και αυτό που προσαρμόζεται καλύτερα στα δεδομένα.

Το 1978, δημιουργήθηκε από τον Gideon Schwarz το μπεϋζιανό κριτήριο πληροφορίας (BIC), το οποίο στη διεθνή βιβλιογραφία εντοπίζεται και με την ονομασία κριτήριο πληροφορίας του Schwarz (SIC – Schwarz Information Criterion). Το κριτήριο BIC δίνει μεγαλύτερη βαρύτητα στον αριθμό των παραμέτρων του μοντέλου σε σχέση με αυτό του Akaike, όπως περιγράφεται στον ακόλουθο τύπο:

$$BIC = -2l + k \log n \quad (3.79)$$

όπου l είναι ο λογάριθμος της συνάρτησης πιθανοφάνειας, k ο αριθμός παραμέτρων του μοντέλου και n το μέγεθος του δείγματος των παρατηρήσεων. Όπως και στο κριτήριο του Akaike, περιλαμβάνεται ένας όρος ποινής, $k \log n$, ο οποίος αυξάνει την τιμή του κριτηρίου με την αύξηση του αριθμού των παραμέτρων. Βάσει του κριτηρίου BIC, το βέλτιστο μοντέλο είναι αυτό με την μικρότερη τιμή του κριτηρίου.

3.5.7 Συντελεστής Προσδιορισμού R^2

Ο συντελεστής προσδιορισμού R^2 είναι ένα χρήσιμο εργαλείο από το κλασικό μοντέλο γραμμικής παλινδρόμησης. Ο συγκεκριμένος συντελεστής ορίζεται ως εξής:

$$R^2 = \frac{SSR}{SST}$$

Όπου, SSR : το άθροισμα τετραγώνων του μοντέλου γραμμικής παλινδρόμησης που εξετάζουμε

SST : το συνολικό άθροισμα τετραγώνων της γραμμικής παλινδρόμησης

Ο συντελεστής R^2 προσδιορίζει το ποσοστό της διακύμανσης της εξαρτημένης μεταβλητής που εξηγείται από το υπό μελέτη μοντέλο και λαμβάνει τιμές μεταξύ του 0 και του 1. Όσο η τιμή του συντελεστή προσεγγίζει τη μονάδα τόσο καλύτερη η προσαρμογή στα δεδομένα του μοντέλου που εξετάζουμε. Ωστόσο, ο συντελεστής προσδιορισμού R^2 δεν είναι κατάλληλος για την περίπτωση μη-γραμμικών μοντέλων. Κατ' επέκταση στην περίπτωση των Γενικευμένων Γραμμικών Μοντέλων χρησιμοποιείται ένας ανάλογος συντελεστής προσδιορισμού R_p^2 (pseudo - R^2) McFadden, (1974) ο οποίος έχει την ακόλουθη μορφή:

$$R_p^2 = 1 - \frac{\hat{l}_{max}}{\hat{l}_0} \quad (3.80)$$

\hat{l}_0 : η λογαριθμική συνάρτηση πιθανοφάνειας του μοντέλου το οποίο περιλαμβάνει μόνο τον σταθερό όρο β_0

\hat{l}_{max} : η λογαριθμική συνάρτηση πιθανοφάνειας του κορεσμένου μοντέλου

3.6 Poisson γενικευμένο γραμμικό μοντέλο

Ένα Poisson γενικευμένο γραμμικό μοντέλο είναι ένα στατιστικό μοντέλο για την μοντελοποίηση διακριτών δεδομένων το οποίο βασίζεται στην υπόθεση ότι η μεταβλητή απόκρισης ακολουθεί την κατανομή Poisson. Η κατανομή Poisson όπως έχει παρουσιαστεί και στο Κεφάλαιο 2 είναι μονοπαραμετρική και χρησιμοποιείται για τη μοντελοποίηση διακριτών δεδομένων όπως είναι για παράδειγμα ο αριθμός συμβάντων, ατυχημάτων και θανάτων. Με τον χαρακτηρισμό μονοπαραμετρική εννοείται ότι υπάρχει μόνο μία παράμετρος μ προς εκτίμηση η οποία συνήθως αναφέρεται ως μέσος ρυθμός ή παράμετρος θέσης. Μια μοναδική ιδιότητα της κατανομής Poisson είναι ότι η μέση τιμή της είναι ίση με τη διασπορά της και η οποία αναφέρεται στην διεθνή βιβλιογραφία με τον όρο ισομεταβλητότητα (equidispersion). Επομένως, βάσει της προαναφερόμενης συνθήκης όσο μεγαλύτερη είναι η μέση τιμή της κατανομής τόσο μεγαλύτερη είναι η παρατηρηθείσα διασπορά στα δεδομένα. Το πρόβλημα το οποίο παρατηρείται συνήθως και θα αναλύσουμε σε επόμενη παράγραφο είναι ότι κατά την

μοντελοποίηση πραγματικών διακριτών δεδομένων η συνθήκη της ισομεταβλητότητας σπάνια ικανοποιείται, ενώ συνηθέστερη είναι η παρουσία υπερδιασποράς στα δεδομένα. Κατ' επέκταση κατά την προσαρμογή του Poisson μοντέλου στα δεδομένα απαιτείται να ληφθεί υπόψη ο έλεγχος της ικανοποίησης της εν λόγω συνθήκης. Εκτός από την συνθήκη της ισομεταβλητότητας, για την προσαρμογή της Poisson παλινδρόμησης πρέπει να ικανοποιούνται οι ακόλουθες υποθέσεις, η κατανομή να είναι διακριτή, μονοπαραμετρική και η παράμετρος της να είναι η αναμενόμενη μέση τιμή, ενώ απαιτείται οι τιμές των δεδομένων να είναι μη αρνητικές και ανεξάρτητες μεταξύ τους. Για το Poisson γενικευμένο γραμμικό μοντέλο ο έλεγχος χ^2 του Pearson για τη διασπορά έχει τιμή προσεγγιστικά ίση με τη μονάδα, Hilbe (2012). Η τιμή του συγκεκριμένου ελέγχου προσεγγίζει την μονάδα όταν η παρατηρούμενη και η εκτιμώμενη διακύμανση είναι ίσες ή περίπου ίσες, μια υπόθεση η οποία όπως έχουμε αναφέρει θα πρέπει να ισχύει στην Poisson παλινδρόμηση. Το μοντέλο υποθέτει ότι ο λογάριθμος της αναμενόμενης μέσης τιμής της μεταβλητής απόκρισης ισούται με τον γραμμικό συνδυασμό των ερμηνευτικών μεταβλητών x_i και ότι η κατάλληλη συνάρτηση σύνδεσης είναι η λογαριθμική, επομένως το γενικευμένο γραμμικό μοντέλο Poisson έχει την μορφή:

$$\log \mu_i = \eta_i = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_k x_k \quad (3.81)$$

και

$$\mu_i = \exp(\eta_i) = \exp(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_k x_k) \quad (3.82)$$

Παρατηρούμε ότι δεν υπάρχει γραμμική σχέση μεταξύ της μέσης πρόβλεψης μ_i και των παραμέτρων όπως σε ένα γραμμικό μοντέλο αλλά η γραμμική σχέση υπάρχει μεταξύ του λογαρίθμου της μέσης πρόβλεψης μ_i και των παραμέτρων. Ακόμη, αξίζει να σημειωθεί ότι στο παραπάνω μοντέλο ο λογάριθμος εξασφαλίζει ότι οι προβλέψεις θα είναι πάντα θετικές.

Από τη σχέση (3.81) προκύπτει μια νέα σχέση για τον λογάριθμο της συνάρτησης πιθανοφάνειας :

$$l(\beta; y_i) = \sum_i \{y_i(x_i\beta) - \exp(x_i\beta) - \log(y_i!)\}$$

Στο μοντέλο της Poisson παλινδρόμησης αυτό που μας ενδιαφέρει είναι η δημιουργία μιας σχέσης η οποία θα εκτιμά τον ρυθμό εμφάνισης των συμβάντων που μελετά το μοντέλο μας είτε αυτό είναι έλευση αποζημιώσεων ατυχήματος, είτε θανάτων. Στο μοντέλο Poisson ο ρυθμός αυτός μ ικανοποιεί την σχέση $E(Y) = \text{Var}(Y) = \mu$. Όταν η διακύμανση του μοντέλου

είναι μεγαλύτερη από την μέση τιμή του τότε παρατηρείται υπερδιασπορά των δεδομένων η οποία μπορεί να ελεγχθεί με τον έλεγχο χ^2 του Pearson διαιρώντας με τους βαθμούς ελευθερίας του μοντέλου.

Για την εκτίμηση των παραμέτρων του Poisson γενικευμένου γραμμικού μοντέλου η συνήθης μέθοδος εκτίμησης είναι η μέθοδος Μέγιστης Πιθανοφάνειας.

Η συνάρτηση πιθανοφάνειας L δίνεται από τη σχέση:

$$L(y; \beta) = \prod_{i=1}^n \frac{e^{-\mu_i} \mu_i^{y_i}}{y_i!}$$

Ο λογάριθμος l της συνάρτησης πιθανοφάνειας L είναι:

$$l = \log L(y; \beta) = \log \prod_{i=1}^n \frac{e^{-\mu_i} \mu_i^{y_i}}{y_i!} = \sum_{i=1}^n \{-\mu_i + y_i \log \mu_i - \log(y_i!)\}$$

Αντικαθιστώντας $\mu_i = e^{x_i^T \beta}$ καταλήγουμε στην ακόλουθη σχέση:

$$l = \log L(y; \beta) = \sum_{i=1}^n \{-e^{x_i^T \beta} + y_i x_i^T \beta - \log(y_i!)\}$$

Παραγωγίζοντας μερικώς την παραπάνω σχέση ως προς β και θέτοντας την μερική παράγωγο ίση με μηδέν καταλήγουμε στο ακόλουθο αποτέλεσμα.

$$\frac{\partial l}{\partial \beta} = \sum_{i=1}^n \{-x_i^T e^{x_i^T \beta} + y_i x_i^T\} = \sum_{i=0}^n \{x_i^T (y_i - e^{x_i^T \beta})\} = 0$$

Από την παραπάνω σχέση καταλήγουμε ότι η εκτιμήτρια μέγιστης πιθανοφάνειας είναι η $\hat{\mu}_i = e^{x_i^T \beta}$.

Η απόκλιση όπως έχει αναφερθεί στην Παράγραφο 3.5.1 αποτελεί ένα μέτρο καλής προσαρμογής του μοντέλου στα δεδομένα μέσω της εκτίμησης της διαφοράς μεταξύ των

παρατηρούμενων και των εκτιμημένων τιμών υπολογίζοντας το διπλάσιο της διαφοράς της λογαριθμικής συνάρτησης πιθανοφάνειας του κορεσμένου μοντέλου και της αντίστοιχης λογαριθμικής συνάρτησης του προσαρμοσμένου μοντέλου. Επομένως, για το μοντέλο της Poisson υπολογίζεται από τον ακόλουθο τύπο:

$$D_{Poisson} = 2 \sum_{i=1}^n \{y_i \log(y_i) - y_i - \log(y_i!) - y_i \log \hat{\mu}_i + \hat{\mu}_i + \log(y_i!)\}$$

$$D_{Poisson} = 2 \sum_{i=1}^n \{y_i \log y_i - y_i - y_i \log \hat{\mu}_i + \hat{\mu}_i\}$$

$$D_{Poisson} = 2 \sum_{i=1}^n \{y_i \log \frac{y_i}{\hat{\mu}_i} - (y_i - \hat{\mu}_i)\}$$

Για την περίπτωση του μοντέλου της κατανομής Poisson η τιμή του ελέγχου AIC έχει την ακόλουθη μορφή:

$$AIC_{Poisson} = 2k - 2 \sum_{i=1}^n \{-\hat{\mu}_i + y_i \log \hat{\mu}_i - \log(y_i!)\}$$

όπου p είναι ο αριθμός του παραμέτρων του μοντέλου για τους οποίους ισχύει $k \leq n$.

Για την περίπτωση του μοντέλου της κατανομής Poisson η τιμή του ελέγχου BIC έχει την ακόλουθη μορφή:

$$BIC_{Poisson} = -2 \sum_{i=1}^n \{-\hat{\mu}_i + y_i \log \hat{\mu}_i - \log(y_i!)\} + k \log n$$

όπου p είναι ο αριθμός του παραμέτρων του μοντέλου για τους οποίους ισχύει $k \leq n$.

Τα κατάλοιπα Pearson για την περίπτωση του Poisson γενικευμένου γραμμικού μοντέλου σύμφωνα με τις σχέσεις (3.16), (3.17) και (3.75) ορίζονται από τον ακόλουθο τύπο:

$$r_i^p = \frac{y_i - \hat{\mu}_i}{\sqrt{\hat{\mu}_i}}$$

ενώ το άθροισμα των τετραγώνων αυτών των καταλοίπων μας δίνουν την τιμή του ελέγχου καλής προσαρμογής του Pearson για το μοντέλο της κατανομής Poisson. Στον παρονομαστή

του κλάσματος έχει υπολογιστεί η τετραγωνική ρίζα της μέσης τιμής λόγω της ιδιότητας της Poisson περί ισότητας της μέσης τιμής και της διασποράς.

3.7 Υπερδιασπορά

Όπως αναφέραμε και στην παράγραφο 2.3 μια βασική υπόθεση του μοντέλου Poisson είναι ότι η μέση τιμή της κατανομής ισούται με την διασπορά της. Αρκετές φορές όμως στα δεδομένα είναι απίθανο να συμβεί και παρατηρείται η διασπορά να είναι είτε μικρότερη της μέσης τιμής (υποδιασπορά) είτε πιο σύννητες είναι να μεγαλύτερη (υπερδιασπορά).

Η υπερδιασπορά τείνει να εμφανίζεται όταν μεταξύ των εξαρτημένων μεταβλητών υπάρχει συσχέτιση και προκαλεί σφάλματα στην εκτίμηση του μοντέλου, Hilbe (2012). Μερικές συνήθεις αιτίες για την ύπαρξη υπερδιασποράς στα δεδομένα μπορεί να είναι ότι το μοντέλο παραλείπει μερικούς από τους παράγοντες, τα δεδομένα περιλαμβάνουν ακραίες τιμές, αποτυχία ύπαρξης όρων αλληλεπίδρασης στο μοντέλο, ελλιπή δεδομένα ή επιλογή ακατάλληλης συνάρτησης σύνδεσης. Η θνησιμότητα επηρεάζεται από το εισόδημα, τον τρόπο ζωής και το μορφωτικό επίπεδο. Αυτό σημαίνει ότι κάποια άτομα τείνουν να έχουν καλύτερη ποιότητα ζωής και κατ' επέκταση να εμφανίζουν μεγαλύτερο προσδόκιμο ζωής επομένως δεν μπορούμε να υποθέσουμε ότι η ένταση θνησιμότητας είναι ίδια για όλα τα άτομα ηλικίας x σε όλες τις χρονιές t . Αυτή η ανομοιογένεια τείνει να αυξάνει την διακύμανση σε σύγκριση με τον μέσο όρο.

Η παραβίαση της ιδιότητας της κατανομής Poisson περί ισότητας της μέσης τιμής και της διασποράς αποδεικνύει συσχέτιση μεταξύ των δεδομένων. Η παράμετρος διασποράς $\hat{\phi}$ μπορεί να εκτιμηθεί από τον έλεγχο Pearson χ^2 διαιρεμένο με τους βαθμούς ελευθερίας του μοντέλου της Poisson ως ακολούθως:

$$\hat{\phi} = \frac{\text{Pearson's } \chi^2 \text{ statistic}}{n - p} = \frac{\chi^2}{n - p} = \frac{\sum_i \frac{(y_i - \hat{\mu}_i)^2}{\hat{\mu}_i}}{n - p} \quad (3.83)$$

Όπως έχουμε ορίσει στην σχέση (3.2) η διασπορά της εξαρτημένης μεταβλητής y με την παράμετρο διασποράς ϕ συνδέονται με την σχέση (3.10), επομένως όσο πιο κοντά στην μονάδα είναι η τιμή της ϕ τόσο καλύτερη η προσαρμογή του μοντέλου στα δεδομένα, ενώ στην περίπτωση που η τιμή της ϕ είναι μεγαλύτερη της μονάδας καταλήγουμε στο συμπέρασμα ότι παρατηρείται υπερδιασπορά των δεδομένων.

3.8 Αρνητικό Διωνυμικό γενικευμένο γραμμικό μοντέλο

Το αρνητικό διωνυμικό γενικευμένο γραμμικό μοντέλο χρησιμοποιείται ευρέως λόγω της ευέλικτης προσαρμογής της επιπλέον μεταβλητότητας που εντοπίζεται στο μοντέλο της μονοπαραμετρικής Poisson. Η Αρνητική Διωνυμική κατανομή αποτελείται από μία επιπλέον παράμετρο η οποία αποκαλείται παράμετρος διασποράς και αποτελεί ένα μέτρο της επιπλέον μεταβλητότητας στα δεδομένα. Στη διεθνή βιβλιογραφία για την παράμετρο διασποράς χρησιμοποιείται και ο όρος παράμετρος ανομοιογένειας ή βοηθητική παράμετρος. Όπως έχει αναφερθεί σε προηγούμενη παράγραφο η διακύμανση της Αρνητικής Διωνυμικής κατανομής δίνεται από τον τύπο $\mu + \alpha\mu^2$ ή $\mu(1 + \alpha\mu)$, όπου α είναι η παράμετρος διασποράς της κατανομής, η οποία όσο μικραίνει και τείνει στο μηδέν η κατανομή προσεγγίζει την Poisson. Για την εφαρμογή του μοντέλου της Αρνητικής Διωνυμικής θα πρέπει να πληρούνται οι εξής συνθήκες:

- η εξαρτημένη μεταβλητή Y να λαμβάνει ακέραιες και μη αρνητικές τιμές,
- η παράμετρος της κατανομής να είναι ίση με την προσδοκώμενη μέση τιμή,
- Ο έλεγχος χ^2 του Pearson για τη διασπορά να έχει τιμή προσεγγιστικά ίση με τη μονάδα.

Το μοντέλο της Αρνητικής Διωνυμικής παρέχει τη δυνατότητα μοντελοποίησης δεδομένων μεγαλύτερης διακύμανσης από το μοντέλο της Poisson. Για το λόγο αυτό, για την μοντελοποίηση Poisson δεδομένων που παρουσιάζουν αυξημένη διασπορά γίνεται χρήση του μοντέλου της Αρνητικής Διωνυμικής με τη λογαριθμική συνάρτηση σύνδεσης $\log(\mu)$.

Στο στατιστικό πακέτο της R, η κατανομή της Αρνητικής Διωνυμικής δύναται να χρησιμοποιηθεί μέσω του πακέτου MASS στο οποίο περιλαμβάνεται και η εντολή `glm.nb()` μέσω της οποίας γίνεται η εφαρμογή του Αρνητικού Διωνυμικού γενικευμένου γραμμικού μοντέλου. Στην εντολή `glm.nb()`, για την εκτίμηση των παραμέτρων του μοντέλου εφαρμόζεται η μέθοδος εκτίμησης ελαχίστων τετραγώνων, ενώ για την εκτίμηση της παραμέτρου διασποράς α χρησιμοποιείται η μέθοδος των ροπών και κατόπιν εκτίμησής της εισάγεται στον αλγόριθμο IRLS σαν σταθερά. Ακολούθως, παρουσιάζεται ο κώδικας της εντολής `glm.nb()` στην R συνοδευόμενος από τα ορίσματά του:

```
glm.nb(formula, data, init.theta, link = log) όπου init.theta=1/α
```

Διευκρινίζεται ότι, στο όρισμα `init.theta` δύναται να δοθεί προκαθορισμένη τιμή της παραμέτρου θ , ειδάλως μπορεί να παραληφθεί και να εκτιμηθεί από το μοντέλο. Το όρισμα `link` αναφέρεται στην επιθυμητή συνάρτηση σύνδεσης στο μοντέλο, όπου για το Αρνητικό Διωνυμικό μοντέλο δεν μπορεί να είναι άλλες από την λογαριθμική (\log), την τετραγωνική

(sqrt) και την ταυτοτική (identity). Όπως στο μοντέλο της κατανομής Poisson, έτσι και στο Αρνητικό Διωνυμικό μοντέλο υποθέτουμε ότι ο λογάριθμος της αναμενόμενης μέσης τιμής της μεταβλητής απόκρισης ισούται με τον γραμμικό συνδυασμό των ερμηνευτικών μεταβλητών x_i και ότι η κατάλληλη συνάρτηση σύνδεσης είναι η λογαριθμική, επομένως το Αρνητικό Διωνυμικό γενικευμένο γραμμικό μοντέλο έχει την ακόλουθη μορφή:

$$\log \mu_i = \eta_i = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_k x_k$$

και

$$\mu_i = \exp(\eta_i) = \exp(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_k x_k)$$

Ο λογάριθμος της συνάρτησης πιθανοφάνειας του εν λόγω μοντέλου είναι ο εξής:

$$l = \log L(y; \alpha, \beta)$$

$$= \sum_{i=1}^n \left\{ y_i \log \alpha + y_i (x_i^T \beta) - \left(y_i + \frac{1}{\alpha} \right) \log \left(1 + \alpha e^{x_i^T \beta} \right) + \log \Gamma \left(y_i + \frac{1}{\alpha} \right) - \log \Gamma(y_i + 1) - \log \Gamma \left(\frac{1}{\alpha} \right) \right\}$$

Η απόκλιση του μοντέλου της Αρνητικής Διωνυμικής δίνεται από τον τύπο:

$$D_{NB} = 2 \sum_{i=1}^n \left\{ y_i \ln \left(\frac{y_i}{\hat{\mu}_i} \right) - \left(\frac{1}{\varphi} + y_i \right) \ln \left(\frac{1 + \varphi y_i}{1 + \varphi \hat{\mu}_i} \right) \right\}$$

Ο τύπος του κριτηρίου AIC για την περίπτωση του Αρνητικού Διωνυμικού μοντέλου είναι ο ακόλουθος:

$$AIC_{NB} = 2k - 2 \sum_{i=1}^n \left\{ y_i \log \alpha + y_i \hat{\mu}_i - \left(y_i + \frac{1}{\alpha} \right) \log(1 + \alpha \hat{\mu}_i) + \log \Gamma \left(y_i + \frac{1}{\alpha} \right) - \log \Gamma(y_i + 1) - \log \Gamma \left(\frac{1}{\alpha} \right) \right\}$$

Αντίστοιχα, το κριτήριο του Schwartz προκύπτει από τον ακόλουθο τύπο:

$$BIC_{NB} = k \log n$$

$$- 2 \sum_{i=1}^n \left\{ y_i \log \alpha + y_i \hat{\mu}_i - \left(y_i + \frac{1}{\alpha} \right) \log(1 + \alpha \hat{\mu}_i) + \log \Gamma \left(y_i + \frac{1}{\alpha} \right) \right. \\ \left. - \log \Gamma(y_i + 1) - \log \Gamma \left(\frac{1}{\alpha} \right) \right\}$$

3.9 Διωνυμικό γενικευμένο γραμμικό μοντέλο

Το διωνυμικό γενικευμένο γραμμικό μοντέλο είναι ένα στατιστικό μοντέλο το οποίο χρησιμοποιείται για την ανάλυση διωνυμικών δεδομένων. Η μεταβλητή απόκρισης λαμβάνει δίτιμες τιμές όπως 0 ή 1, αποτυχία ή επιτυχία και ναι ή όχι. Η τ.μ. Y ακολουθεί τη Διωνυμική κατανομή με συνάρτηση πιθανότητας (3.29). Στη λογιστική παλινδρόμηση με διωνυμικά δεδομένα χρησιμοποιούμε τη συνάρτηση σύνδεσης logit για την πιθανότητα p_i και ο οποίος μας διασφαλίζει ότι οι εκτιμήσεις που θα λάβουμε θα είναι εντός του κλειστού διαστήματος $[0,1]$.

Κάνοντας χρήση της σχέσης (3.37) και σύμφωνα με τις προϋποθέσεις του μοντέλου καταλήγουμε στην κάτωθι σχέση:

$$\eta_i = g[E(Y)] = \text{logit}(n_i p_i) \\ = \log \left(\frac{n_i p_i}{n_i - n_i p_i} \right) = \log \left(\frac{p_i}{1 - p_i} \right) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_k x_k \quad (3.84)$$

και λύνοντας ως προς p_i καταλήγουμε στο μοντέλο πρόβλεψης της πιθανότητας p_i :

$$p_i = \frac{e^{\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_k x_k}}{1 + e^{\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_k x_k}} \quad (3.85)$$

Η αναλογία $\frac{p}{1-p}$ είναι γνωστή στη διεθνή βιβλιογραφία με την ονομασία odds (σχετική πιθανότητα) και μας δείχνει πόσο πιθανή είναι η επίτευξη ενός ενδεχομένου συγκριτικά με την μη επίτευξή του.

Ο λογάριθμος της πιθανοφάνειας του διωνυμικού μοντέλου ισούται με:

$$l = \log L(p_i; y_i) = \sum_i \left[\log \binom{n_i}{y_i} + y_i \log p_i + (n_i - y_i) \log(1 - p_i) \right]$$

$$= \sum_i \left[\log \binom{n_i}{y_i} + y_i \log \left(\frac{p_i}{1-p_i} \right) + n_i \log(1-p_i) \right] \quad (3.86)$$

Αντικαθιστώντας στη σχέση (3.86) σύμφωνα με τη σχέση (3.85) καταλήγουμε στο ακόλουθο αποτέλεσμα:

$$= \sum_i \left[\log \binom{n_i}{y_i} + y_i x_i^T \beta - n_i \log(1 + e^{x_i^T \beta}) \right] \quad (3.87)$$

Παίρνοντας μερικές παραγώγους έχουμε :

$$\begin{aligned} \frac{\partial l}{\partial \beta} &= \sum_i \left[y_i x_i^T - n_i x_i^T e^{x_i^T \beta} (1 + e^{x_i^T \beta})^{-1} \right] \\ \frac{\partial l}{\partial \beta} &= \sum_i \left[\{y_i - n_i e^{x_i^T \beta} (1 + e^{x_i^T \beta})^{-1}\} x_i^T \right] \\ \frac{\partial l}{\partial \beta} &= \sum_i (y_i - n_i p_i) x_i^T = 0 \end{aligned} \quad (3.88)$$

Από τη σχέση (3.88) καταλήγουμε ότι ο εκτιμητής μέγιστης πιθανογάνειας της μέσης τιμής είναι η παρακάτω ποσότητα:

$$\hat{\mu}_i = n_i \hat{p}_i$$

Το μέτρο καλής προσαρμογής της απόκλισης για την περίπτωση του Διωνυμικού γενικευμένου γραμμικού μοντέλου υπολογίζεται από τον ακόλουθο τύπο:

$$\begin{aligned} D &= 2 \sum_{i=1}^n \left\{ \log \binom{n_i}{y_i} \right. \\ &\quad \left. + y_i \log p_i + (n_i - y_i) \log(1 - p_i) - \log \binom{n_i}{y_i} - y_i \log \hat{p}_i - (n_i - y_i) \log(1 - \hat{p}_i) \right\} \\ D &= 2 \sum_{i=1}^n \left\{ y_i \log \left(\frac{p_i}{\hat{p}_i} \right) + (n_i - y_i) \log \left(\frac{1 - p_i}{1 - \hat{p}_i} \right) \right\} \\ D &= 2 \sum_{i=1}^n \left\{ y_i \log \frac{y_i}{\hat{\mu}_i} + (n_i - y_i) \log \left(\frac{n_i - y_i}{n_i - \hat{\mu}_i} \right) \right\} \end{aligned}$$

Για την περίπτωση του μοντέλου της Διωνυμικής κατανομής η τιμή του ελέγχου AIC έχει την ακόλουθη μορφή:

$$AIC_{Binomial} = -2l + 2k = -2 \left[\sum_{i=1}^n \left\{ \log \binom{n_i}{y_i} + y_i \log \hat{p}_i + (n_i - y_i) \log(1 - \hat{p}_i) \right\} \right] + 2k$$

όπου k είναι ο αριθμός των παραμέτρων του μοντέλου.

Για την περίπτωση του μοντέλου της Διωνυμικής κατανομής η τιμή του ελέγχου BIC έχει την ακόλουθη μορφή:

$$\begin{aligned} BIC_{Binomial} &= -2l + k \log n \\ &= -2 \left[\sum_{i=1}^n \left\{ \log \binom{n_i}{y_i} + y_i \log \hat{p}_i + (n_i - y_i) \log(1 - \hat{p}_i) \right\} \right] + k \log n \end{aligned}$$

Τα κατάλοιπα Pearson για την περίπτωση του μοντέλου της Διωνυμικής κατανομής ορίζονται από την σχέση:

$$r_i^p = \frac{y_i - n_i \hat{p}_i}{\sqrt{n_i \hat{p}_i (1 - \hat{p}_i)}} \quad \text{όπου} \quad \hat{p}_i = \frac{e^{x_i^T \beta}}{1 + e^{x_i^T \beta}}$$

Το άθροισμα τετραγώνων αυτών των καταλοίπων αποτελούν τον έλεγχο καλής προσαρμογής χ^2 του Pearson για το διωνυμικό μοντέλο όπως προκύπτει από τον ακόλουθο τύπο:

$$\chi^2 = \sum_{i=1}^n \frac{(y_i - n_i \hat{p}_i)^2}{n_i \hat{p}_i (1 - \hat{p}_i)}$$

3.10 Γενικευμένα μη-γραμμικά μοντέλα

Με τον όρο γενικευμένο μη γραμμικό μοντέλο (gnm - generalized non linear model) εννοούμε ένα στατιστικό μοντέλο το οποίο περιλαμβάνει κάποιες μη γραμμικές παραμέτρους και έχει την ίδια μορφή και κάποιες κοινές ιδιότητες με αυτές των γενικευμένων γραμμικών μοντέλων. Τα γενικευμένα μη γραμμικά μοντέλα χρησιμοποιούνται για να περιγράψουν την σχέση μεταξύ μιας εξαρτημένης μεταβλητής της οποίας η κατανομή ανήκει στην εκθετική οικογένεια κατανομών και των ανεξάρτητων μεταβλητών στις οποίες περιλαμβάνονται γραμμικοί και μη γραμμικοί παράγοντες. Όπως έχουμε αναφέρει σε προηγούμενο κεφάλαιο στα γενικευμένα

γραμμικά μοντέλα υπάρχει γραμμική σχέση μεταξύ της πρόβλεψης η_i και της συνάρτησης σύνδεσης $g(\mu)$. Στα γενικευμένα μη - γραμμικά μοντέλα ισχύουν όλες οι άλλες υποθέσεις εκτός από αυτή της γραμμικής σχέσης μεταξύ της πρόβλεψης η_i και της συνάρτησης σύνδεσης $g(\mu)$. Ο λόγος που συμβαίνει αυτό είναι ότι στα μη-γραμμικά μοντέλα περιλαμβάνονται παράγοντες που παραβιάζουν την γραμμικότητα όπως ο πολλαπλασιαστικός παράγοντας που περιλαμβάνεται στο μοντέλο Lee-Carter.

Στο στατιστικό πακέτο της R παρέχεται η δυνατότητα στατιστικής ανάλυσης των μοντέλων αυτών με το πακέτο `gnm` (Turner and Firth 2012) το οποίο αποτελεί μια ειδική επέκταση της θεωρίας των γενικευμένων γραμμικών μοντέλων (`glm`) και μέσω του οποίου δίνεται η δυνατότητα ανάλυσης μοντέλων που περιλαμβάνουν μη γραμμικούς παράγοντες, όπως π.χ. πολλαπλασιαστικούς μέσω της συνάρτησης `Mult()`. Από στατιστικής άποψης, τα γενικευμένα μη γραμμικά μοντέλα είναι για τα μη γραμμικά ότι είναι τα γενικευμένα γραμμικά μοντέλα για τα γραμμικά, έτσι ένα μη γραμμικό μοντέλο δύναται να προκύψει από ένα γενικευμένο γραμμικό μοντέλο θεωρώντας ότι την εκθετική κατανομή.

Στη δημοσίευση του το 1996, ο Lane χρησιμοποιεί τον όρο του γενικευμένου γραμμικού μοντέλου με σκοπό να περιγράψει ένα στατιστικό μοντέλο, το οποίο περιλαμβάνει και κάποιες μη γραμμικές παραμέτρους, αλλά διατηρεί τη μορφή γενικευμένου γραμμικού μοντέλου. Επιπλέον, ο Currie (2016) αναφέρει ότι αρκετά στατιστικά μοντέλα, τα οποία χρησιμοποιούνται για την μοντελοποίηση και την πρόβλεψη της θνησιμότητας, όπως αυτό που εξετάζουμε, ανήκουν στην κλάση των γενικευμένων μη γραμμικών μοντέλων.

3.10.1 Το μοντέλο Lee - Carter ως γενικευμένο μη - γραμμικό μοντέλο

Αν υποθέσουμε ότι η τυχαία μεταβλητή Y περιγράφει τον αριθμό των θανάτων, δηλαδή $Y = D_{x,t}$, η γραμμική πρόβλεψη έχει την μορφή $\eta = \alpha_x + \beta_x k_t$ και η συνάρτηση σύνδεσης είναι η λογαριθμική $\log(\cdot)$ τότε το μοντέλο Lee - Carter μπορεί να μελετηθεί βάσει της θεωρίας των γενικευμένων γραμμικών μοντέλων σαν ένα μοντέλο παλινδρόμησης με ένα γραμμικό όρο α_x και έναν διγραμμικό όρο ($\beta_x k_t$) όπως φαίνεται ακολούθως:

$$\eta = g(E(Y)) = g(E(D_{x,t})) = \alpha_x + \beta_x k_t \quad (3.89)$$

Έστω $l(\alpha, \beta, k)$ ο λογάριθμος της συνάρτησης πιθανοφάνειας της $D_{x,t}$, τότε μεγιστοποιώντας την $l(\alpha, \beta, k)$ δύναται να εκτιμηθούν οι παράμετροι α_x, β_x, k_t του μοντέλου Lee - Carter υπό από το πλαίσιο των γενικευμένων γραμμικών μοντέλων.

3.10.2 Poisson Lee - Carter μοντέλο

Αν υποθέσουμε ότι ο αριθμός των θανάτων $D_{x,t}$ ακολουθεί την κατανομή Poisson με μέση τιμή την έκθεση στον κίνδυνο $e_{x,t}$ πολλαπλασιασμένη με τον ειδικό κατά ηλικία δείκτη θνησιμότητας $m_{x,t}$ όπως φαίνεται ακολούθως:

$$D_{x,t} \sim \text{Poisson}(e_{x,t}m_{x,t}) \quad \text{όπου} \quad m_{x,t} = e^{(\alpha_x + \beta_x k_t)}$$

και αντικαθιστώντας από τη σχέση (1.2) προκύπτει ότι:

$$\log \frac{D_{x,t}}{e_{x,t}} = \alpha_x + \beta_x k_t$$

και

$$\log D_{x,t} = \alpha_x + \beta_x k_t + \log e_{x,t} \quad (3.90)$$

Επομένως, ο λογάριθμος της συνάρτησης πιθανοφάνειας του Poisson Lee – Carter μοντέλου είναι ο ακόλουθος:

$$l(\alpha_x, \beta_x, k_t) = \sum_{x,t} [D_{x,t}(\alpha_x + \beta_x k_t) - e_{x,t}e^{(\alpha_x + \beta_x k_t)}] \quad (3.91)$$

3.10.3 Αρνητικό Διωνυμικό Lee – Carter μοντέλο

Το μοντέλο της Poisson όπως έχουμε αναφέρει στην Παράγραφο 3.10.2 υποθέτει ότι:

$$E(D(x, t)) = \text{Var}(D(x, t)) = e(x, t) \exp(\alpha_x + \beta_x k_t)$$

όπου αυτή η υπόθεση της ισομεταβλητότητας δεν μπορεί να ισχύει καθώς ο αριθμός των θανάτων μεταβάλλεται ανάλογα με την ηλικία και το έτος λόγω διάφορων παραγόντων. Λόγω των προαναφερόμενων, ενσωματώνεται στο διγραμμικό μοντέλο Poisson ένας ακόμα

παράγοντας $\tau_{x,t}$ για την εκτίμηση του αριθμού των θανάτων (Azman et al., 2022). Μετά την προσθήκη του παράγοντα $\tau_{x,t}$ υποθέτουμε ότι ο αριθμός των θανάτων $D_{x,t}$ ακολουθεί την κατανομή Poisson με μέση τιμή

$$e_{x,t} \exp(\alpha_x + \beta_x k_t + \tau_{x,t})$$

και διακύμανση

$$Var(D(x, t)) = \delta_{x,t} + \kappa \delta_{x,t}^2 \quad \text{όπου } \kappa = Var(\exp(\tau_{x,t})).$$

Αν η ποσότητα $\exp(\tau_{x,t})$ ακολουθεί την κατανομή Γάμμα, τότε ο αριθμός θανάτων $D_{x,t}$ ακολουθεί την Αρνητική Διωνυμική κατανομή (Rawak R et al., 2022) και (Delwarde et al., 2007) με λογαριθμική συνάρτηση πιθανοφάνειας:

$$l(\alpha_x, \beta_x, k_t, \kappa) = \sum_{x,t} \left(\sum_i \left(\log\left(\frac{1}{\kappa} + D_{x,t} - i\right) - \log(D_{x,t}!) - \left(D_{x,t} - \frac{1}{\kappa}\right) \log(1 + \kappa \delta_{x,t}) + D_{x,t} \log(\kappa \delta_{x,t}) \right) \right) \quad (3.92)$$

Όπου κ παριστάνει την παράμετρο διασποράς της Αρνητικής Διωνυμικής κατανομής την οποία αναφέραμε στην Παράγραφο 3.8, ενώ για μηδενική τιμή της παραμέτρου κ προκύπτει η κατανομή Poisson.

3.10.4 Διωνυμικό Lee - Carter μοντέλο

Αν υποθέσουμε ότι $r_{x,t}$ είναι ο αριθμός του πληθυσμού ηλικίας x στην αρχή του έτους t και $q_{x,t}$ είναι η πιθανότητα θανάτου του πληθυσμού ηλικίας x στην αρχή του έτους t , επομένως για το Διωνυμικό Lee – Carter διγραμμικό μοντέλο μπορούμε να υποθέσουμε ότι ο αριθμός των θανάτων $D_{x,t}$ ακολουθεί την διωνυμική κατανομή με παραμέτρους $n = r_{x,t}$ και $p = q_{x,t}$. Η πιθανότητα $q_{x,t}$ μπορεί να γραφεί ως εξής:

$$q_{x,t} = 1 - p_{x,t} = 1 - \exp(-m_{x,t}) = 1 - \exp(-\exp(\alpha_x + \beta_x k_t)) \quad (3.93)$$

ο λογάριθμος της συνάρτησης πιθανοφάνειας είναι:

$$l(\alpha_x, \beta_x, k_t) = \sum_{x,t} \left((r_{x,t} - D_{x,t}) \log(1 - q_{x,t}) + D_{x,t} \log(q_{x,t}) \right) \quad (3.94)$$

Εφαρμογή στα ελληνικά δεδομένα

Στο κεφάλαιο που ακολουθεί θα προχωρήσουμε στην εφαρμογή της θεωρίας που περιγράψαμε στα προηγούμενα κεφάλαια κάνοντας χρήση του στατιστικού πακέτου της R, όπου θα αντλήσουμε χρήσιμα συμπεράσματα για την ικανοποιητική προσαρμογή των τριών μοντέλων στα ελληνικά δεδομένα καθώς και την εκτίμηση του αριθμού των θανάτων. Ο έλεγχος καλής προσαρμογής των μοντέλων θα προκύψει από την σύγκριση των διακυμάνσεών τους, όπου το μοντέλο με την μικρότερη διακύμανση παρουσιάζει και την καλύτερη προσαρμογή, ενώ χρήσιμα είναι τα συμπεράσματα που θα εξάγουμε από τις γραφικές παραστάσεις των κανονικοποιημένων καταλοίπων¹ έναντι του αριθμού των θανάτων καθώς από τις γραφικές παραστάσεις των δειγματικών ποσοστημορίων ως προς τα θεωρητικά ποσοστημόρια της κανονικής κατανομής.

4.1 Προσαρμογή του κλασικού μοντέλου Lee-Carter μέσω της R

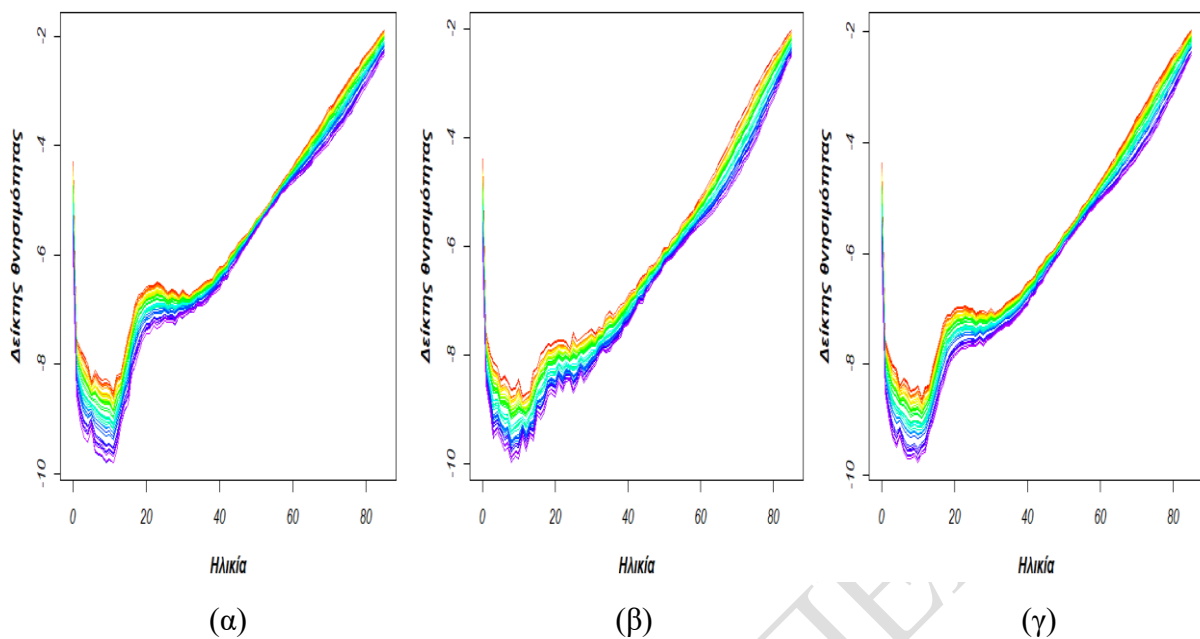
Στην παράγραφο αυτή θα προσαρμόσουμε τα ελληνικά δεδομένα στο κλασικό μοντέλο Lee – Carter, για το οποίο έγινε εκτενής αναφορά στο Κεφάλαιο 3. Μέσω του στατιστικού πακέτου της R, χρησιμοποιώντας την εντολή `lca()` του πακέτου `demography` προχωρήσαμε στην εφαρμογή του κλασικού μοντέλου Lee - Carter, το οποίο κάνει χρήση της μεθόδου SVD για την εκτίμηση των παραμέτρων του μοντέλου. Η εντολή `lca()` έχει την ακόλουθη δομή:

```
lca(data, max.age= 85, interpolate = T, adjust="none")
```

όπου `data` πρόκειται για τα δεδομένα θνησιμότητας που μελετάμε και `max.age` είναι η μέγιστη ηλικία η οποία προκύπτει από τα δεδομένα. Με το όρισμα `interpolate = T` το μοντέλο εντοπίζει τυχόν μηδενικές τιμές στα δεδομένα θνησιμότητας και τις εξισώνει με τον αριθμό θανάτων των πλησιέστερων ηλικιακών ομάδων. Τέλος, μέσω του ορίσματος `adjust` παρέχεται η δυνατότητα στο χρήστη να επιλέξει τη μέθοδο μέσω της οποίας θα εφαρμοστεί το δεύτερο βήμα εκ νέου εκτίμησης της παραμέτρου k_t όπως περιγράφεται στη Παράγραφο 2.3.1.

Στο Διάγραμμα 4.1 παρουσιάζονται οι εκτιμήσεις των δεικτών θνησιμότητας μέσω του κλασικού μοντέλου Lee - Carter για τον ανδρικό, γυναικείο και συνολικό πληθυσμό ξεχωριστά για κάθε ηλικία και έτος προσαρμογής. Η γενική εντύπωση που προκύπτει από το Διάγραμμα 4.1 είναι ότι ο δείκτης θνησιμότητας παρουσιάζει μείωση με το πέρασμα των ετών.

¹ Τα κανονικοποιημένα κατάλοιπα είναι απαραίτητα εργαλεία για τον έλεγχο των υποθέσεων κατασκευής του μοντέλου καθώς είναι ανεξάρτητα και ακολουθούν την κανονική κατανομή με μέση τιμή μηδέν και σταθερή διακύμανση, ενώ συμβάλλουν στην αναγνώριση ακραίων τιμών.



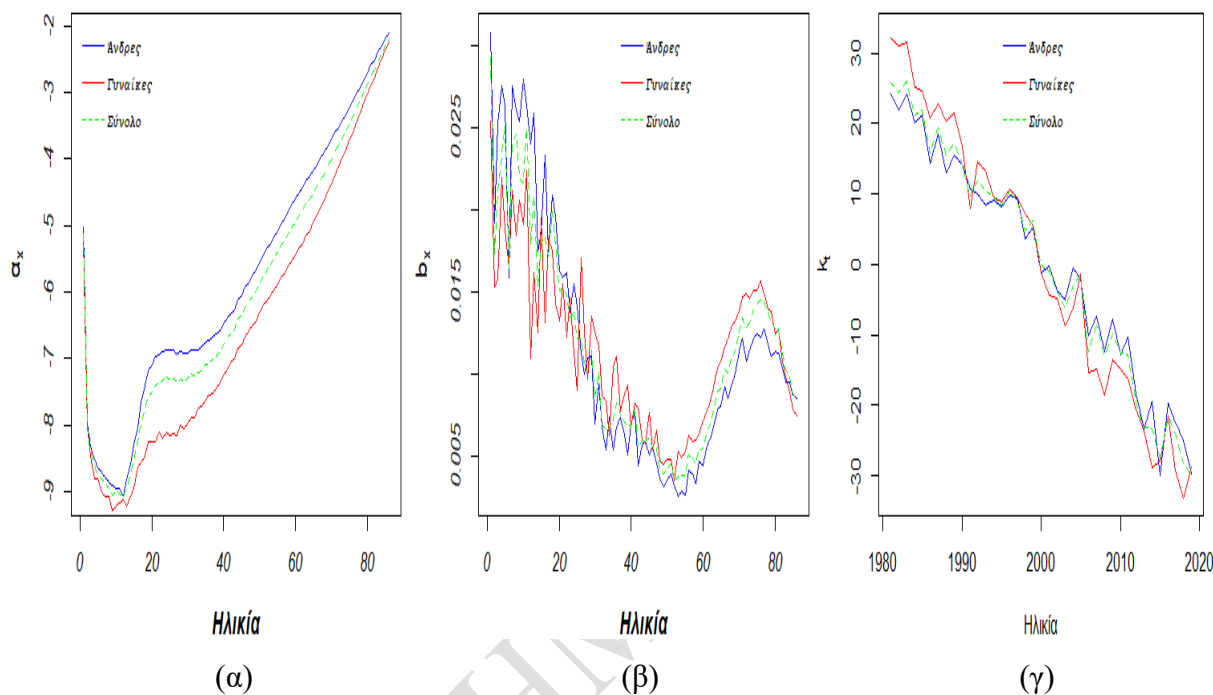
Διάγραμμα 4.1: Διάγραμμα δείκτη θνησιμότητας ανά ηλικία: (α) του ανδρικού πληθυσμού, (β) του γυναικείου πληθυσμού, (γ) του συνολικού πληθυσμού.

Στον Πίνακα Β.1 του Παραρτήματος Β παρατίθενται οι τιμές των εκτιμήσεων των παραμέτρων α_x και β_x για τις ηλικίες 0-85 και τα έτη 1981-2019, οι οποίες εκτιμήθηκαν από το κλασικό μοντέλο Lee - Carter μέσω της εντολής `lca()` του πακέτου `demography` στην R. Εναλλακτικά οι τιμές της μέσης θνησιμότητας α_x δύναται να υπολογιστούν παίρνοντας τις μέσες τιμές των γραμμών του πίνακα των λογαρίθμων των ειδικών κατά ηλικία δεικτών θνησιμότητας, εφαρμόζοντας τη σχέση 2.6.

Από το Διάγραμμα 4.2 (α) όπου έχουμε μια απεικόνιση της παραμέτρου ηλικίας α_x , η οποία αντιπροσωπεύει τη μέση θνησιμότητα ανά ηλικία, διαπιστώνουμε ότι η θνησιμότητα τόσο των γυναικών (κόκκινη γραμμή) όσο και των ανδρών (μπλε γραμμή) παρουσιάζουν γενική ανοδική τάση με την αύξηση της ηλικίας. Τέλος, παρατηρείται απότομη αύξηση της θνησιμότητας των ανδρών σε σύγκριση με αυτή των γυναικών από την ηλικία των 16 μέχρι την ηλικία των 20 ετών.

Στο Διάγραμμα 4.2 (β), η παράμετρος β_x περιγράφει τις διακυμάνσεις της θνησιμότητας στην ηλικία x καθώς μεταβάλλονται τα έτη. Από το εν λόγω διάγραμμα παρατηρούμε πιο έντονες μεταβολές της παραμέτρου για τους άνδρες και τις γυναίκες ηλικίας από 0 έως 20 ενώ μετά την ηλικία των 60 παρατηρείται έντονη αύξηση επομένως θα αναμένουμε αύξηση του αριθμού των θανάτων.

Στο διάγραμμα 4.2 (γ), η παράμετρος k_t περιγράφει τις διακυμάνσεις της θνησιμότητας ανά έτος t . Από το εν λόγω διάγραμμα παρατηρούμε ότι ενώ η θνησιμότητα ανά έτος των γυναικών το 1981 είναι πιο αυξημένη από αυτή των ανδρών, παρουσιάζει πτωτική τάση με την πάροδο των ετών φτάνοντας να είναι σημαντικά μειωμένη το έτος 2018 όπως βεβαίως αναμένουμε λόγω της μείωσης των δεικτών θνησιμότητας.

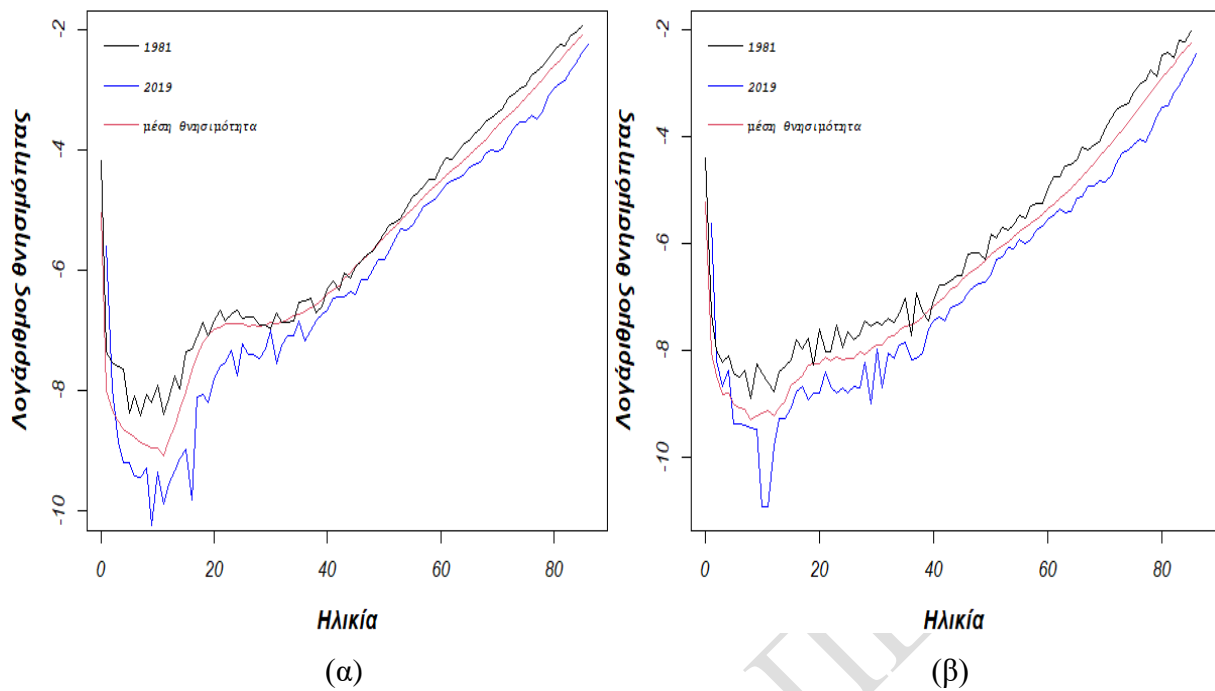


Διάγραμμα 4.2: Εκτίμηση παραμέτρου: (α) α_x , (β) β_x , (γ) k_t για άνδρες, γυναίκες και συνολικό πληθυσμό ως προς την ηλικία.

4.2 Σύγκριση παρατηρούμενου και εκτιμώμενου αριθμού θανάτων μέσω του κλασσικού μοντέλου Lee – Carter

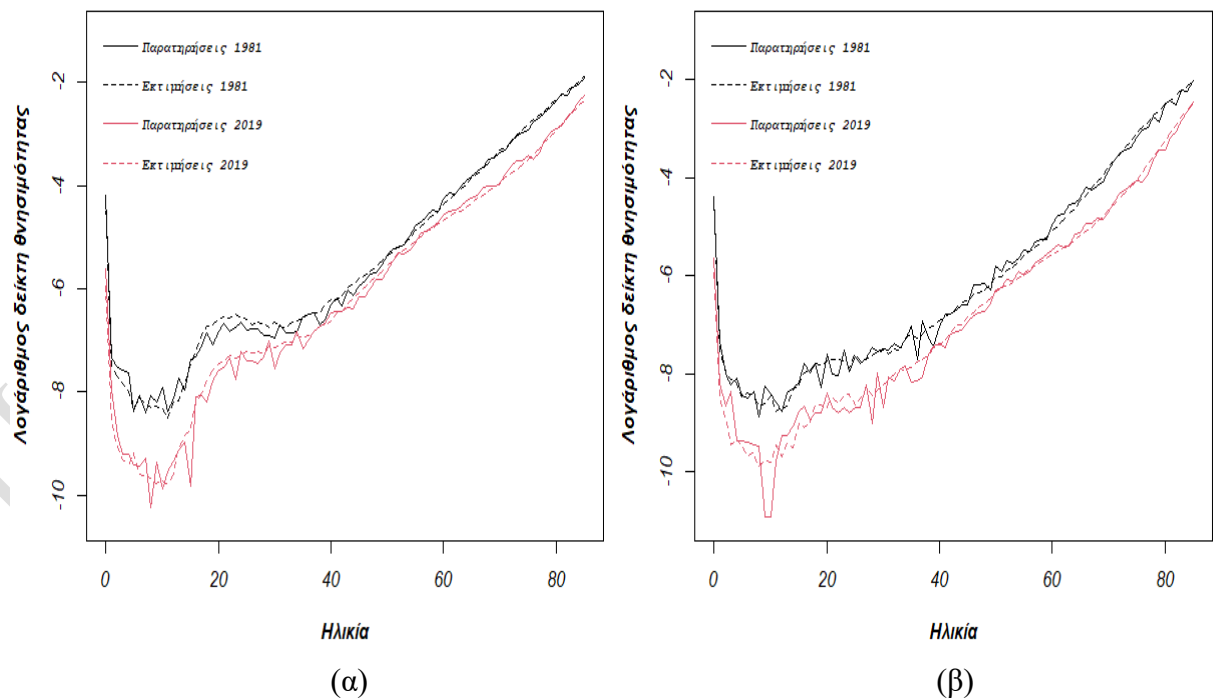
Κατόπιν εφαρμογής του κλασσικού μοντέλου Lee - Carter και εξαγωγής των εκτιμήσεων της παραμέτρου α_x θα συγκρίνουμε τα αποτελέσματά του με τις αντίστοιχες παρατηρούμενες τιμές.

Στο διάγραμμα 4.3, παρουσιάζονται οι θνησιμότητες των ετών 1981 και 2019 συγκριτικά με την μέση θνησιμότητα την οποία υπολογίσαμε μέσω του κλασσικού μοντέλου Lee - Carter και παρατηρούμε ότι η θνησιμότητα του έτους 2019 παρουσιάζει εμφανή βελτίωση σχετικά με την αντίστοιχη του έτους 1981 και για την πλειοψηφία των ηλικιών είναι μικρότερη ακόμη και από την μέση θνησιμότητα, όπως αναμενόταν.



Διάγραμμα 4.3: Θνησιμότητα ετών 1981 και 2019 έναντι μέσης θνησιμότητας για όλες τις ηλικίες για (α) άνδρες και (β) γυναίκες.

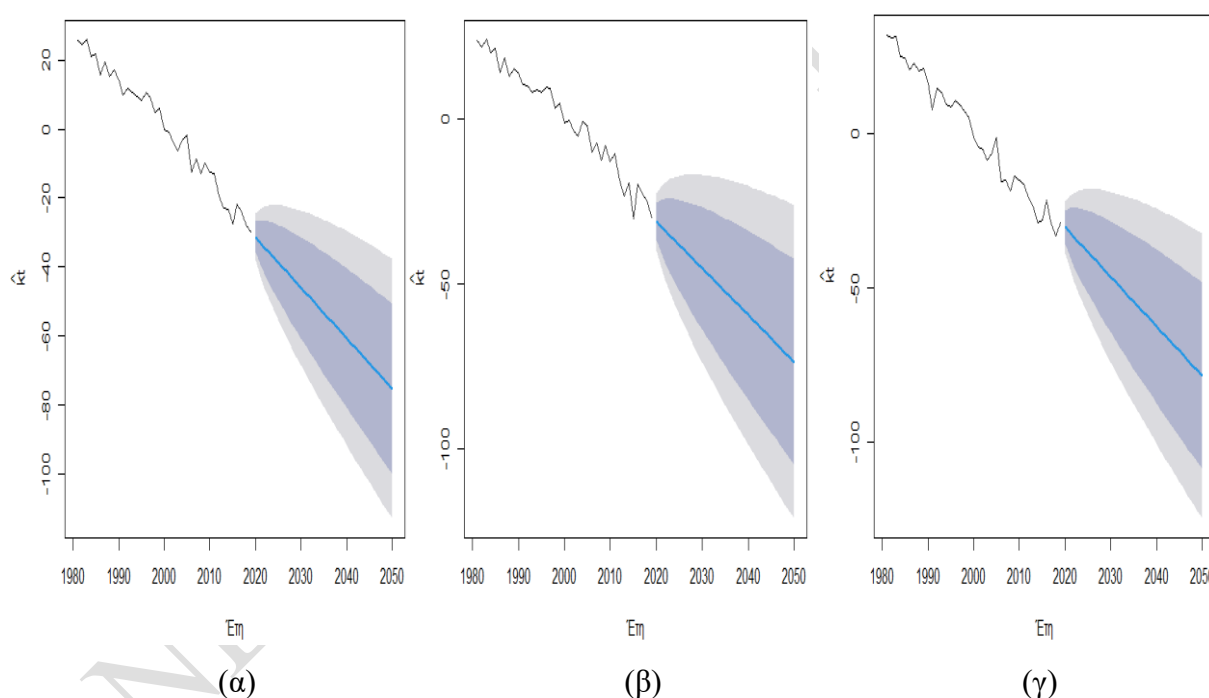
Ακολουθώς, παρουσιάζονται διαγραμματικά οι εκτιμήσεις της θνησιμότητας ανδρών και γυναικών μέσω του μοντέλου Lee – Carter (διακεκομμένη γραμμή) έναντι του αντίστοιχου ειδικού κατά ηλικία δείκτη θνησιμότητας.



Διάγραμμα 4.4: Εκτιμήσεις θνησιμότητας (α) ανδρών, (β) γυναικών ετών 1981 και 2019 έναντι αντίστοιχων παρατηρούμενων τιμών.

4.3 Εκτίμηση παραμέτρου k_t του ελληνικού πληθυσμού για τα έτη 2020 - 2050

Κάνοντας χρήση της R και μέσω του πακέτου tseries, στο οποίο περιλαμβάνεται η εντολή για την προσαρμογή μοντέλων ARIMA, και το πακέτο forecast μπορούμε να προβλέψουμε την παράμετρο k_t του μοντέλου Lee - Carter για τα μελλοντικά έτη. Στο Διάγραμμα 4.5 παρουσιάζονται οι προβλέψεις των τιμών της παραμέτρου k_t . Παρατηρούμε σημαντική πτωτική τάση στις τιμές της παραμέτρου με την πάροδο των ετών, με τις μπλε και γκρι αποχρώσεις να υποδηλώνουν τα αντίστοιχα 80% και 95% διαστήματα πρόβλεψης.



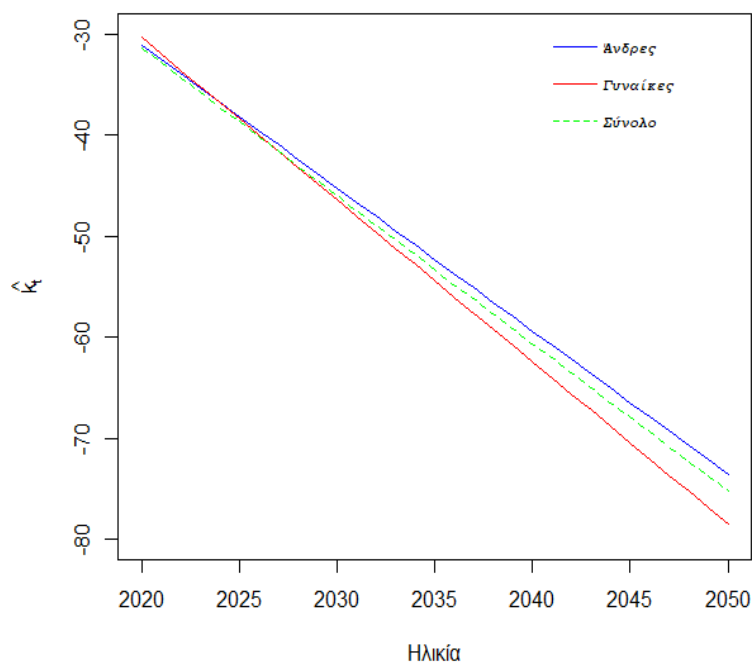
Διάγραμμα 4.5: Πρόβλεψη τιμών παραμέτρου k_t (α) του συνολικού πληθυσμού, (β) των ανδρών, (γ) των γυναικών για τα έτη 2020 - 2050.

Από τον Πίνακα 4.1 των προβλέψεων της παραμέτρου \hat{k}_t καταλήγουμε στο συμπέρασμα ότι η θνησιμότητα συνεχίζει να παρουσιάζει καθοδική τάση με την πάροδο των ετών, που σημαίνει ότι θα αναμένουμε περαιτέρω μείωση των δεικτών θνησιμότητας στο μέλλον.

Πίνακας 4.1: Προβλέψεις παραμέτρου k_t μέσω των μοντέλων ARIMA για τα έτη 2020 – 2050.

Έτος	Προβλέψεις παραμέτρου k_t για τα έτη 2020 - 2050		
	Άνδρες	Γυναίκες	Σύνολο
2020	-31.08	-30.35	-31.35
2021	-32.49	-31.95	-32.81
2022	-33.91	-33.56	-34.28
2023	-35.32	-35.16	-35.74
2024	-36.74	-36.77	-37.21
2025	-38.15	-38.37	-38.67
2026	-39.57	-39.98	-40.13
2027	-40.98	-41.58	-41.6
2028	-42.4	-43.19	-43.06
2029	-43.81	-44.79	-44.52
2030	-45.23	-46.39	-45.99
2031	-46.65	-48	-47.45
2032	-48.06	-49.6	-48.91
2033	-49.48	-51.21	-50.38
2034	-50.89	-52.81	-51.84
2035	-52.31	-54.42	-53.3
2036	-53.72	-56.02	-54.77
2037	-55.14	-57.63	-56.23
2038	-56.55	-59.23	-57.69
2039	-57.97	-60.84	-59.16
2040	-59.38	-62.44	-60.62
2041	-60.8	-64.05	-62.08
2042	-62.21	-65.65	-63.55
2043	-63.63	-67.26	-65.01
2044	-65.05	-68.86	-66.47
2045	-66.46	-70.47	-67.94
2046	-67.88	-72.07	-69.4
2047	-69.29	-73.68	-70.86
2048	-70.71	-75.28	-72.33
2049	-72.12	-76.89	-73.79
2050	-73.54	-78.49	-75.25

Παράλληλα, στο Διάγραμμα 4.6 προκύπτει ότι η θνησιμότητα μεταξύ των ετών 2020 - 2050 του γυναικείου πληθυσμού, χωρίς κάποια μεταβολή στις υποθέσεις του μοντέλου, αναμένεται να μειωθεί εντονότερα από αυτή του ανδρικού πληθυσμού.

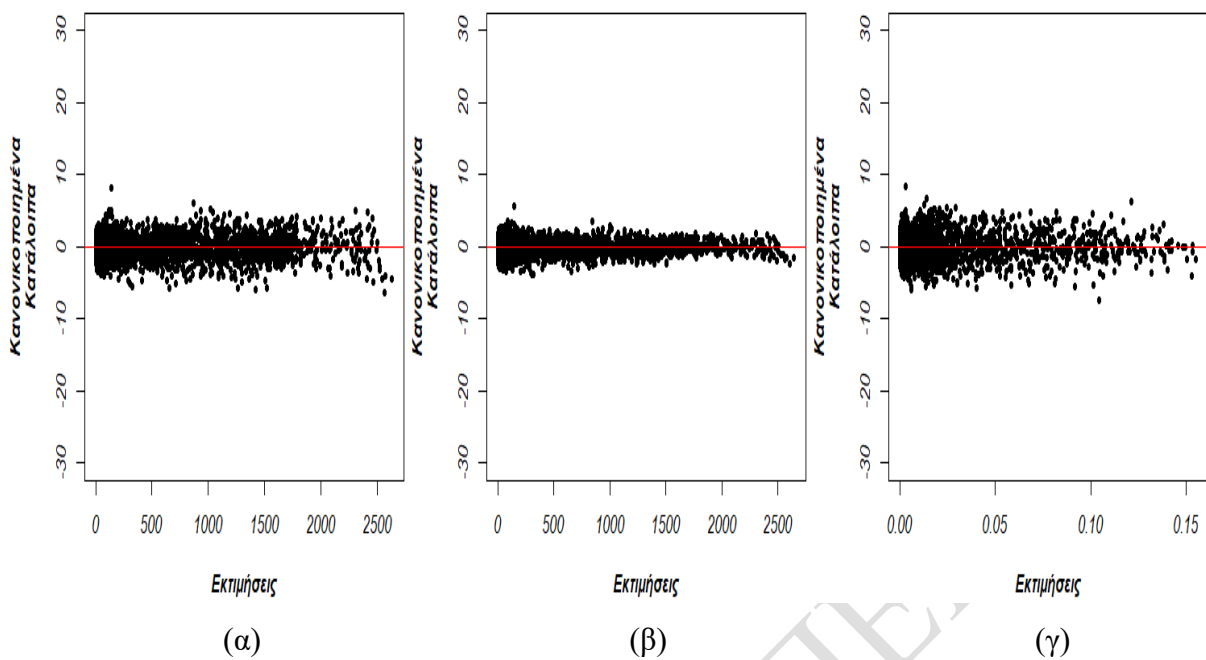


Διάγραμμα 4.6: Πρόβλεψη της παραμέτρου k_t για άνδρες, γυναίκες και συνολικό πληθυσμό της Ελλάδας για τα έτη 2020 - 2050.

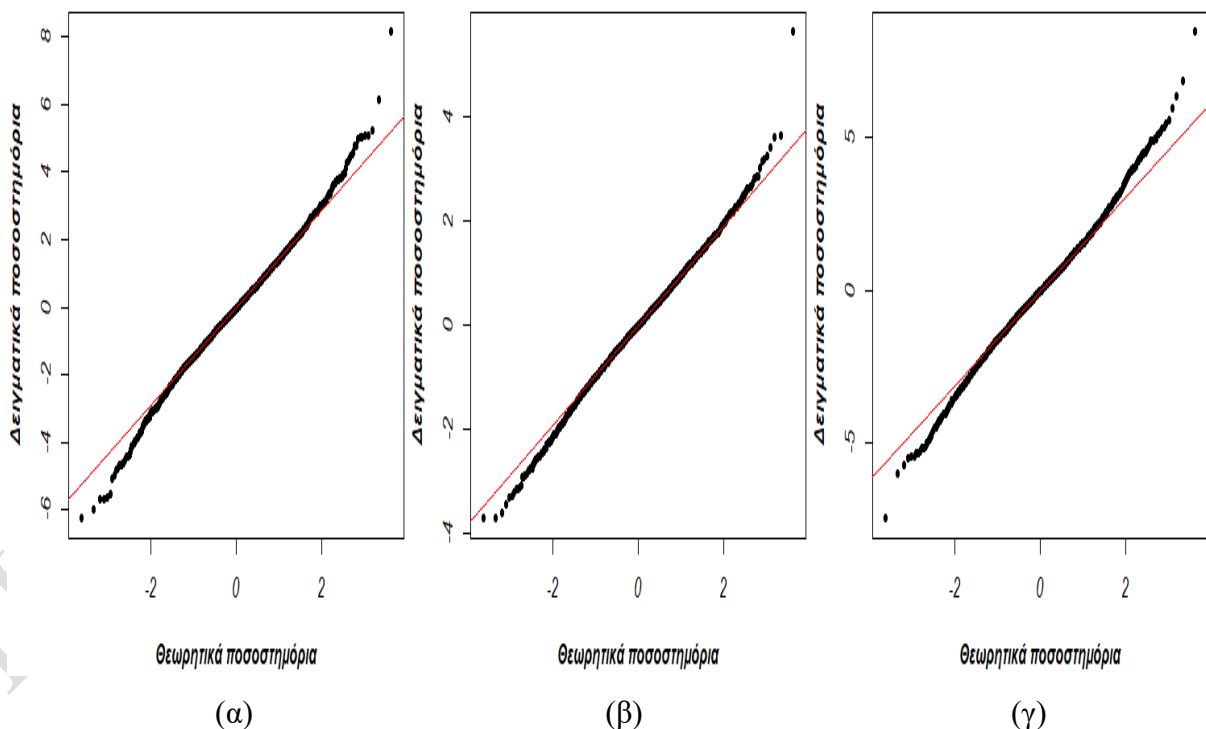
4.4 Προσαρμογή μοντέλων Lee - Carter στα ελληνικά δεδομένα

Όπως ήδη έχει αναφερθεί στην Παράγραφο 3.10, το μοντέλο Lee – Carter δύναται να μελετηθεί ως γενικευμένο μη-γραμμικό μοντέλο. Στην παράγραφο αυτή θα εφαρμόσουμε τη θεωρία, η οποία έχει αναφερθεί στην Παράγραφο 3.10 και θα εξετάσουμε την προσαρμογή των μοντέλων Poisson Lee – Carter, Αρνητικό Διωνυμικό Lee -Carter και Διωνυμικό Lee Carter στα ελληνικά δεδομένα θνησιμότητας. Για την εφαρμογή των γενικευμένων μη- γραμμικών μοντέλων μέσω της R, χρησιμοποιήθηκε η εντολή `gnm()` μέσω του ομότιτλου πακέτου, ενώ για το Αρνητικό Διωνυμικό Lee – Carter μοντέλο έγινε χρήση συμπληρωματικά του πακέτου MASS στο οποίο περιλαμβάνεται η οικογένεια κατανομών της Αρνητικής Διωνυμικής.

Από το Διάγραμμα 4.7, κατόπιν προσαρμογής των τριών μοντέλων στον πληθυσμό των ανδρών, παρατηρείται ότι οι τιμές των κανονικοποιημένων καταλοίπων για το Αρνητικό Διωνυμικό Lee - Carter μοντέλο αποκλίνουν λιγότερο από το μηδέν έναντι των δύο άλλων υπό εξέταση μοντέλων. Αυτό αποτελεί μια ένδειξη ότι το εν λόγω μοντέλο προσαρμόζεται καλύτερα στα δεδομένα του ανδρικού πληθυσμού. Παράλληλα, από το Διάγραμμα 4.8 προκύπτει ότι τα δειγματικά ποσοστημόρια του Αρνητικού Διωνυμικού Lee – Carter μοντέλου προσεγγίζουν καλύτερα την ευθεία των θεωρητικών ποσοστημορίων που αποτελεί επίσης ένδειξη καλής προσαρμογής στα δεδομένα έναντι των άλλων δύο υπό εξέταση μοντέλων.

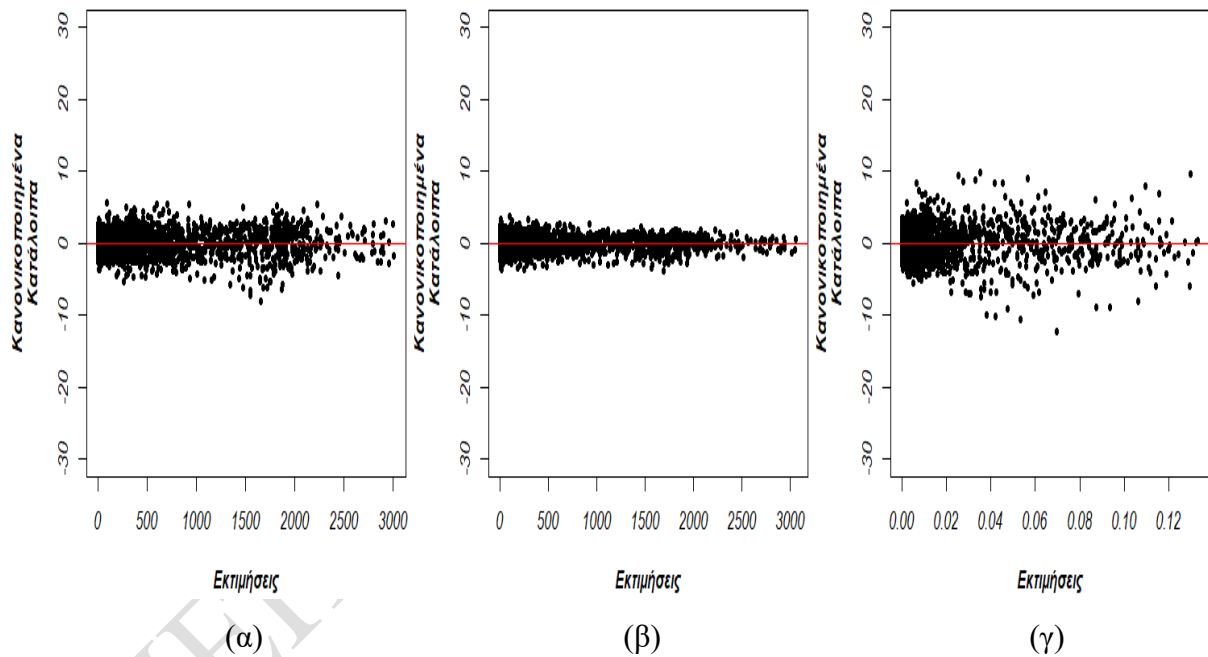


Διάγραμμα 4.7: Κανονικοποιημένα κατάλοιπα έναντι εκτιμήσεων ανδρικού πληθυσμού για τα μοντέλα: (α) Poisson Lee - Carter, (β) Αρνητικό Διωνυμικό Lee - Carter και (γ) Διωνυμικό Lee - Carter.

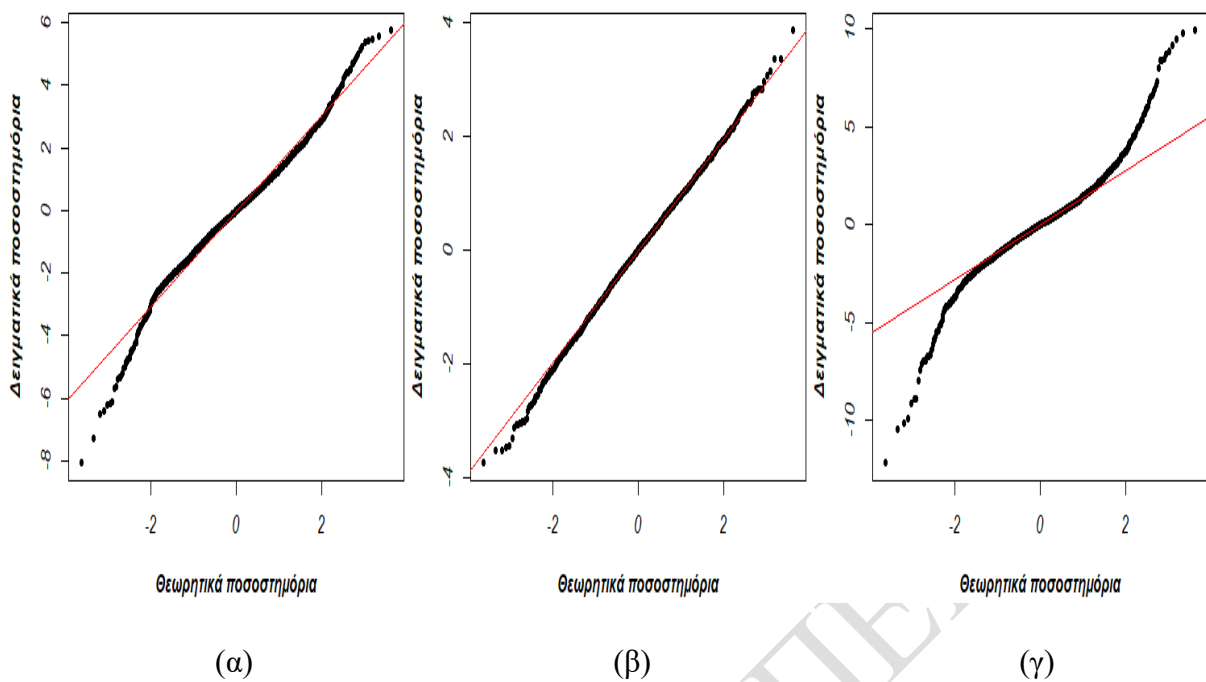


Διάγραμμα 4.8: Γραφική παράσταση των δειγματικών ποσοστημορίων ως προς τα θεωρητικά ποσοστημόρια της Κανονικής Κατανομής για τα μοντέλα ανδρικού πληθυσμού: (α) Poisson Lee - Carter, (β) Αρνητικό Διωνυμικό Lee - Carter και (γ) Διωνυμικό Lee - Carter.

Αναφορικά με την προσαρμογή των τριών μοντέλων στα δεδομένα του γυναικείου πληθυσμού καταλήγουμε βάσει των Διαγραμμάτων 4.9 (β) και 4.10 (β), ότι το Αρνητικό Διωνυμικό Lee – Carter παρουσιάζει καλύτερη προσαρμογή έναντι του Poisson Lee – Carter και του Διωνυμικού Lee – Carter μοντέλου. Αξίζει να σημειωθεί ότι ειδικά στο διωνυμικό μοντέλο του Διαγράμματος 4.9 (γ) παρατηρείται αυξημένη συσσώρευση των καταλοίπων στην αρχή του εύρους των προσαρμοσμένων τιμών που μας οδηγεί στο συμπέρασμα εμφάνισης ετεροσκεδαστικότητας στα δεδομένα. Στα διαγράμματα 4.10 (α) και (γ) παρατηρείται ότι το Poisson και το Διωνυμικό μοντέλο αποκλείουν σημαντικά από την ευθεία των θεωρητικών ποσοστημοριών, ένδειξη ότι δεν παρουσιάζουν καλή προσαρμογή σε αντίθεση με το Αρνητικό Διωνυμικό μοντέλο στο Διάγραμμα 4.10 (β), το οποίο ενώ παρουσιάζει κάποιες ακραίες τιμές φαίνεται να έχει την καλύτερη προσαρμογή.

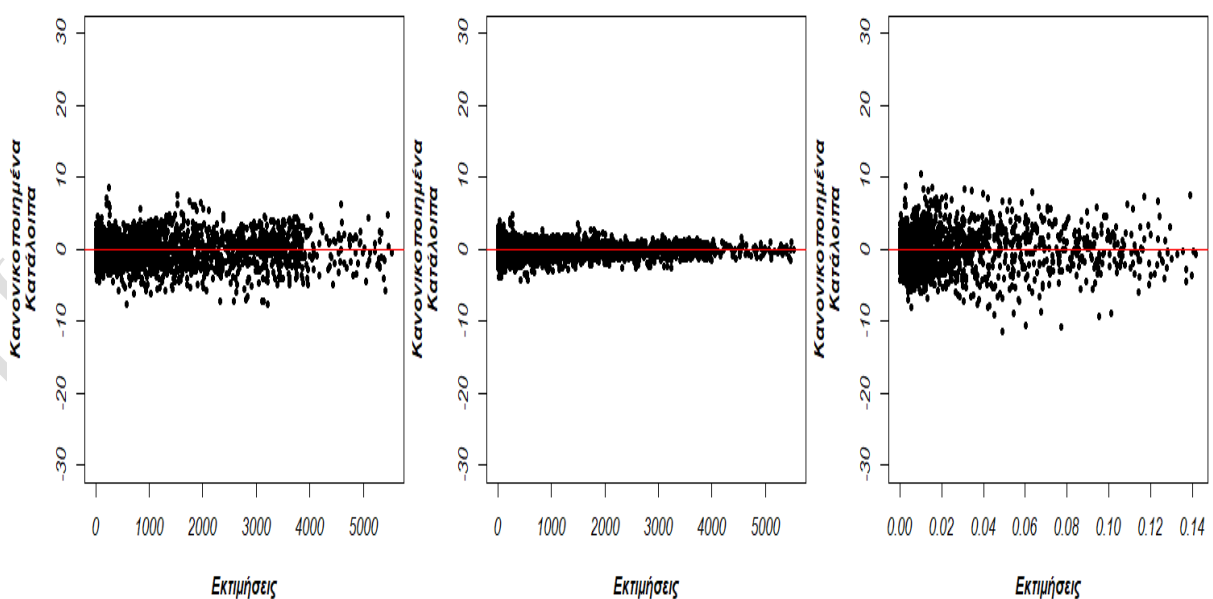


Διάγραμμα 4.9: Κανονικοποιημένα κατάλοιπα έναντι εκτιμήσεων γυναικείου πληθυσμού για τα μοντέλα: (α) Poisson Lee - Carter, (β) Αρνητικό Διωνυμικό Lee - Carter και (γ) Διωνυμικό Lee - Carter.



Διάγραμμα 4.10: Γραφική παράσταση των δειγματικών ποσοστημορίων ως προς τα θεωρητικά ποσοστημόρια της Κανονικής Κατανομής για τα μοντέλα γυναικείου πληθυσμού: (α) Poisson Lee - Carter, (β) Αρνητικό Διωνυμικό Lee - Carter και (γ) Διωνυμικό Lee - Carter.

Λαμβάνοντας υπόψη το Διάγραμμα 4.11, για την προσαρμογή των τριών μοντέλων στον συνολικό πληθυσμό, παρατηρείται ότι οι τιμές των κανονικοποιημένων καταλοίπων για το Αρνητικό Διωνυμικό Lee - Carter μοντέλο αποκλίνουν λιγότερο από το μηδέν έναντι των δύο άλλων υπό εξέταση μοντέλων. Αυτό αποτελεί μια ένδειξη ότι το εν λόγω μοντέλο προσαρμόζεται καλύτερα στα δεδομένα του συνολικού πληθυσμού.



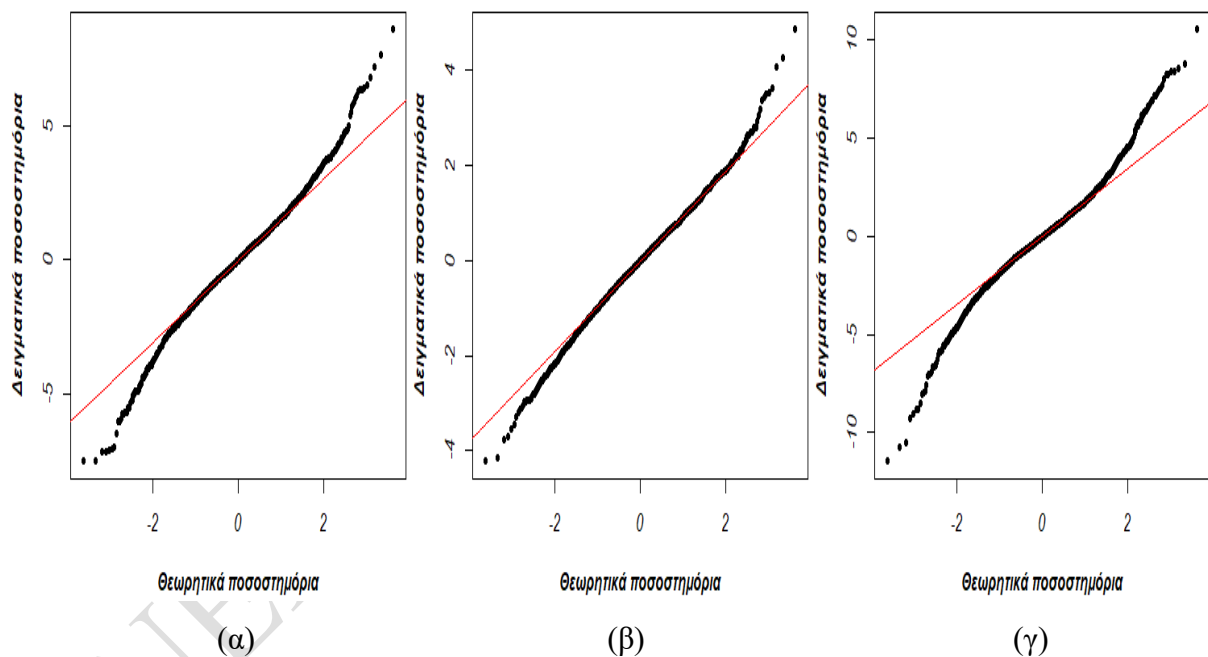
(α)

(β)

(γ)

Διάγραμμα 4.11: Κανονικοποιημένα κατάλοιπα έναντι εκτιμήσεων συνολικού πληθυσμού για τα μοντέλα: (α) Poisson Lee - Carter, (β) Αρνητικό Διωνυμικό Lee - Carter και (γ) Διωνυμικό Lee - Carter.

Από την προσαρμογή των τριών μοντέλων στα δεδομένα του συνολικού πληθυσμού προκύπτει ότι, όπως παρουσιάζεται και στο Διάγραμμα 4.12 (α) και (γ), το Poisson Lee – Carter και το Διωνυμικό Lee - Carter μοντέλο αποκλείουν σημαντικά από την ευθεία των θεωρητικών ποσοστημορίων. Επομένως δεν παρουσιάζουν ικανοποιητική προσαρμογή σε αντίθεση με το Αρνητικό Διωνυμικό Lee - Carter μοντέλο που ενώ παρουσιάζει ακραίες τιμές, υπάρχουν ενδείξεις ότι έχει ικανοποιητική προσαρμογή στα δεδομένα του συνολικού πληθυσμού.



Διάγραμμα 4.12: Γραφική παράσταση των δειγματικών ποσοστημορίων ως προς τα θεωρητικά ποσοστημόρια της Κανονικής Κατανομής για τα μοντέλα συνολικού πληθυσμού: (α) Poisson Lee - Carter, (β) Αρνητικό Διωνυμικό Lee - Carter και (γ) Διωνυμικό Lee - Carter.

Έχοντας κάνει αναφορά στην έννοια της υπερδιασποράς στην Παράγραφο 3.7 θα μελετήσουμε την ύπαρξή της στα δεδομένα της Ελλάδας. Μετά την προσαρμογή των τριών μοντέλων στα δεδομένα και κάνοντας χρήση των καταλοίπων Pearson, προχωρήσαμε στον υπολογισμό της

παραμέτρου διασποράς $\hat{\varphi}$ η οποία όπως έχει αναφερθεί εκτιμάται από τον έλεγχο Pearson χ^2 διαιρώντας με τους βαθμούς ελευθερίας του εκάστοτε μοντέλου. Σύμφωνα με τα ανωτέρω προκύπτει ο ακόλουθος Πίνακας:

Πίνακας 4.2: Παράμετρος διασποράς $\hat{\varphi}$.

	Poisson Lee - Carter	Αρνητικό Διωνυμικό Lee - Carter	Διωνυμικό Lee - Carter
Άνδρες	2.29	1.26	2.82
Γυναίκες	2.09	1.05	3.21
Σύνολο	3.03	1.22	4.53

Όπως προκύπτει από τον υπολογισμό της παραμέτρου διασποράς για το εκάστοτε μοντέλο, παρατηρείται υπερδιασπορά στα δεδομένα, με τη μικρότερη τιμή να εντοπίζεται στο μοντέλο της Αρνητικής Διωνυμικής. Η παρουσία υπερδιασποράς στα δεδομένα αποτελεί μια πρώτη ένδειξη και αναμένουμε ότι το Poisson μοντέλο δεν θα παρουσιάσει ικανοποιητική προσαρμογή.

4.5 Απόκλιση, κριτήρια πληροφορίας και καλής προσαρμογής

Στην παρούσα παράγραφο θα προχωρήσουμε αρχικά στη σύγκριση των τιμών απόκλισης των μοντέλων Poisson Lee – Carter, Αρνητικό Διωνυμικό Lee – Carter και Διωνυμικό Lee- Carter για άνδρες, γυναίκες και συνολικό πληθυσμό. Εν συνεχεία, θα σχολιάσουμε τα αποτελέσματα των κριτηρίων πληροφορίας AIC και BIC.

Στον ακόλουθο Πίνακα 4.3 παρουσιάζονται οι τιμές απόκλισης (deviance) των τριών μοντέλων και προκύπτει ότι η ελάχιστη απόκλιση για τους άνδρες, τις γυναίκες και τον συνολικό πληθυσμό εντοπίζεται στο Αρνητικό Διωνυμικό μοντέλο Lee - Carter. Σύμφωνα με όσα έχουν αναφερθεί στο Κεφάλαιο 3, βάσει του κριτηρίου της απόκλισης το Αρνητικό Διωνυμικό μοντέλο Lee – Carter προσαρμόζεται καλύτερα στα δεδομένα συγκριτικά με τα άλλα δύο μοντέλα. Παράλληλα, το μέτρο της απόκλισης συνηγορεί στα συμπεράσματα που καταλήξαμε στην Παράγραφο 4.4, δηλαδή ότι το Διωνυμικό Lee – Carter μοντέλο παρουσιάζει την λιγότερο ικανοποιητική προσαρμογή στα δεδομένα.

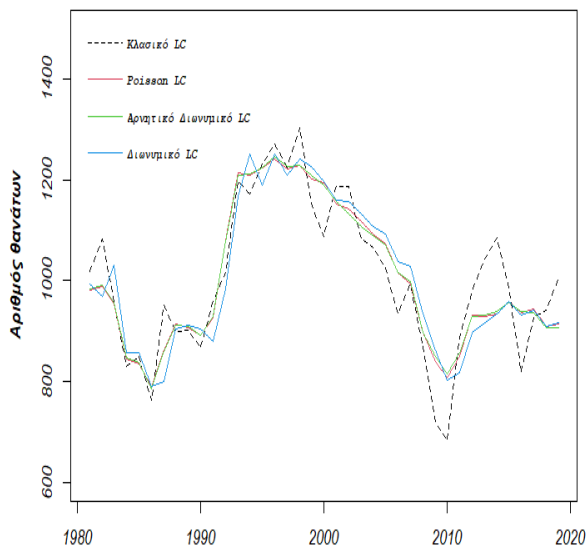
Αναφορικά με τα κριτήρια πληροφορίας του Akaike και του Schwartz, συγκρίνοντας τις τιμές των κριτηρίων του Poisson Lee-Carter και του Αρνητικού Διωνυμικού Lee - Carter μοντέλου παρατηρείται ότι το δεύτερο παρουσιάζει μικρότερες τιμές. Κατ' επέκταση το Αρνητικό

Διωνυμικό Lee - Carter μοντέλο παρουσιάζει καλύτερη προσαρμογή στα δεδομένα έναντι του Poisson Lee-Carter.

Πίνακας 4.3: Απόκλιση και κριτήρια πληροφορίας.

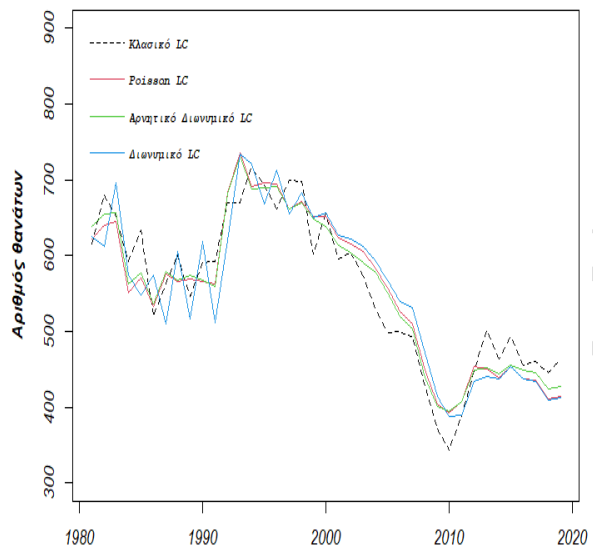
Άνδρες					
Μοντέλο	Παράμετροι Μοντέλου	Λογάριθμος Συνάρτησης Πιθανοφάνειας	Deviance	AIC	BIC
Poisson LC	209	-15344	7239	31106	32384
Αρνητικό Διωνυμικό LC	209	-14699	4002	29815	31094
Διωνυμικό LC	209	-16136	8879	32691	33969
Γυναίκες					
Μοντέλο	Παράμετροι Μοντέλου	Λογάριθμος Συνάρτησης Πιθανοφάνειας	Deviance	AIC	BIC
Poisson LC	209	-13985	6625	28387	29666
Αρνητικό Διωνυμικό LC	209	-13199	3337	26817	28095
Διωνυμικό LC	209	-15688	10068	31794	33072
Σύνολο					
Μοντέλο	Παράμετροι Μοντέλου	Λογάριθμος Συνάρτησης Πιθανοφάνειας	Deviance	AIC	BIC
Poisson LC	209	-17259	9561	34936	36215
Αρνητικό Διωνυμικό LC	209	-15787	3887	31992	33270
Διωνυμικό LC	209	-19579	14245	39575	40854

Στο Διάγραμμα 4.13 συγκρίνονται οι εκτιμήσεις του αριθμού των θανάτων στην ηλικία των 67 ετών για το κλασικό, το Poisson, το Αρνητικό Διωνυμικό και το Διωνυμικό μοντέλο (α) για τον ανδρικό και (β) για τον γυναικείο πληθυσμό. Παρατηρούμε ότι οι εκτιμήσεις του Poisson και του Αρνητικού Διωνυμικού μοντέλου παρουσιάζουν μικρές διαφορές μεταξύ τους για τους άνδρες.



Έτη

(α)

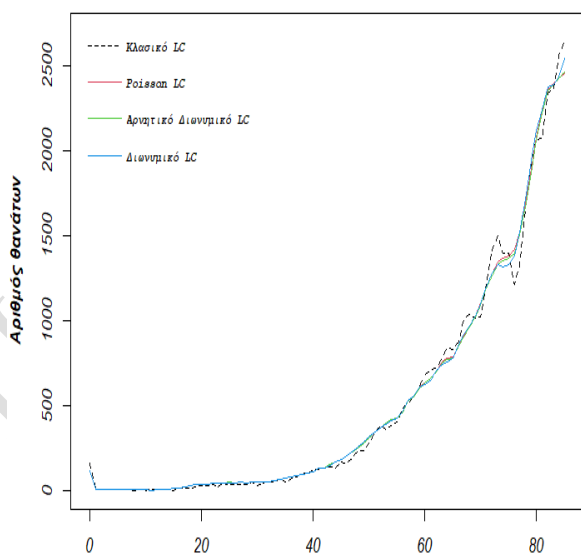


Έτη

(β)

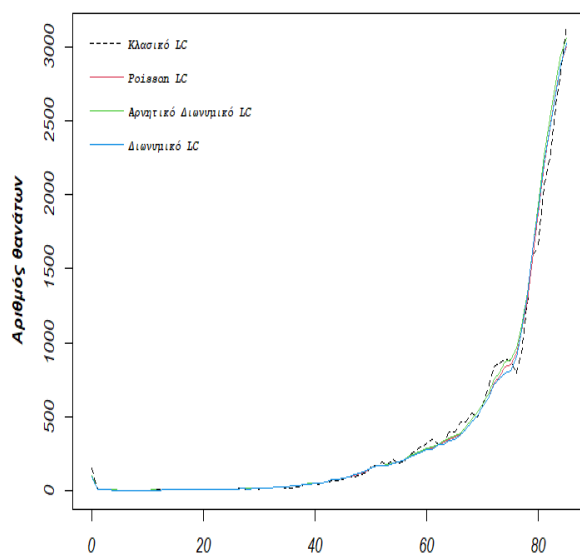
Διάγραμμα 4.13: Σύγκριση εκτίμησης αριθμού θανάτων ατόμων ηλικίας 67 ετών του κλασικού μοντέλου Lee – Carter έναντι του Poisson, Αρνητικού Διωνυμικού και Διωνυμικού για: (α) Άνδρες και (β) Γυναίκες.

Στο Διάγραμμα 4.14 συγκρίνονται οι εκτιμήσεις του αριθμού των θανάτων του έτους 2019 για το κλασσικό, το Poisson, το Αρνητικό Διωνυμικό και του Διωνυμικό μοντέλο (α) για τον ανδρικό και (β) για τον γυναικείο πληθυσμό. Παρατηρούμε ότι οι εκτιμήσεις των τεσσάρων μοντέλων είναι αρκετά κοντά, ειδικά για τον γυναικείο πληθυσμό όπου οι διαφορές είναι μικρές.



Ηλικία

(α)

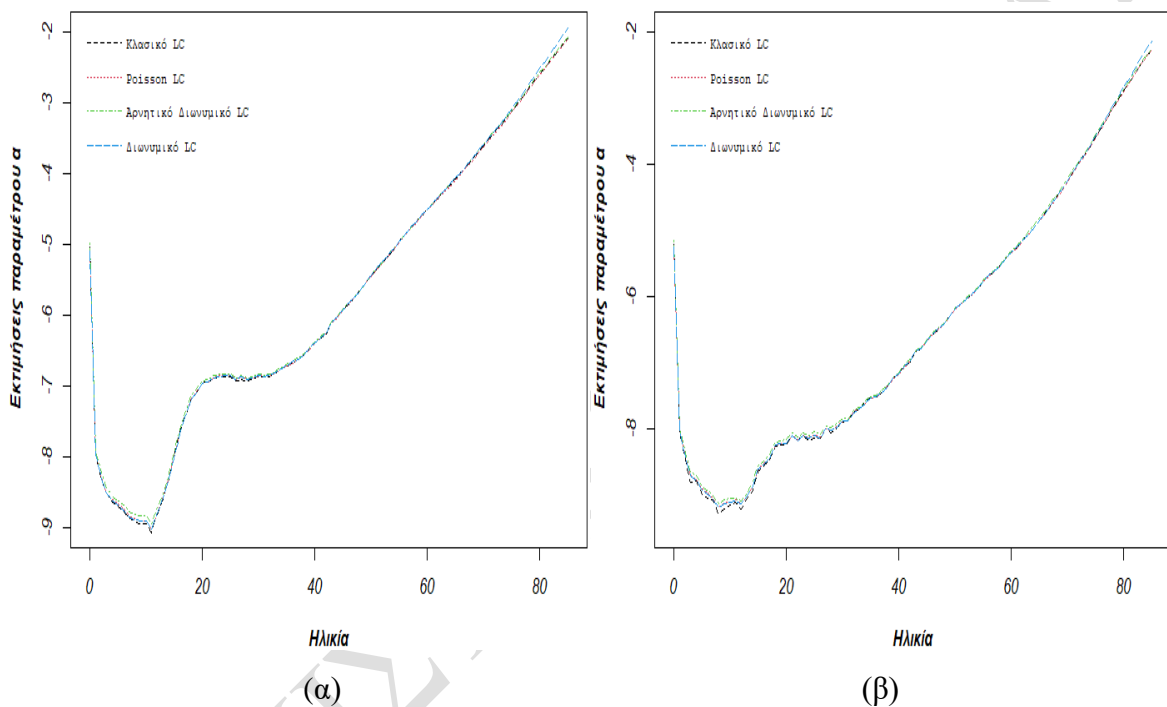


Ηλικία

(β)

Διάγραμμα 4.14: Σύγκριση εκτίμησης αριθμού θανάτων έτους 2019 του κλασικού μοντέλου Lee – Carter έναντι του Poisson, Αρνητικού Διωνυμικού και Διωνυμικού για: (α) Άνδρες και (β) Γυναίκες.

Στο Διάγραμμα 4.15 συγκρίνονται οι εκτιμήσεις της παραμέτρου α_x για το έτος 2019 για το κλασικό, το Poisson, το Αρνητικό Διωνυμικό και του Διωνυμικό μοντέλο (α) για τον ανδρικό και (β) για τον γυναικείο πληθυσμό. Παρατηρούμε ότι οι εκτιμήσεις των τεσσάρων μοντέλων είναι αρκετά κοντά, ενώ ελάχιστες διαφορές μεταξύ τους παρατηρούνται στις ηλικίες μεταξύ 0-10 και άνω των 85 ετών.



Διάγραμμα 4.15: Σύγκριση εκτίμησης παραμέτρου α_x για το έτος 2019 του κλασικού μοντέλου Lee – Carter έναντι του Poisson, Αρνητικού Διωνυμικού και Διωνυμικού για: (α) Άνδρες και (β) Γυναίκες.

ΚΕΦΑΛΑΙΟ 4

Σύγκριση με αποτελέσματα από τη διεθνή βιβλιογραφία

Σε αυτό το Κεφάλαιο θα προχωρήσουμε στη σύγκριση αποτελεσμάτων με μελέτες από τη διεθνή βιβλιογραφία. Οι Azman και Pathmanathan (2020) στη μελέτη τους προσαρμόσαν το Διωνυμικό, το Poisson και το Αρνητικό Διωνυμικό μοντέλο Lee – Carter στα δεδομένα θανάτων 10 χωρών για τις χρονολογίες 1966 έως 2016 και για τις ηλικίες από 0 μέχρι 85 ετών. Τα δεδομένα αφορούν τον συνολικό πληθυσμό καθώς και των πληθυσμό ανδρών και γυναικών ξεχωριστά των χωρών, Ιαπωνία, Η.Π.Α., Μεγάλη Βρετανία, Αυστραλία, Σουηδία, Βέλγιο, Ισπανία, Καναδάς, Ολλανδία και Βουλγαρία και αντλήθηκαν από τη Βάση Δεδομένων (Human Mortality Database, www.mortality.org). Εν συνεχεία, προχώρησαν στην εξαγωγή συμπερασμάτων για την προσαρμογή των μοντέλων βάσει του κριτηρίου της απόκλισης (deviance) και των κανονικοποιημένων καταλοίπων έναντι των προσαρμοσμένων τιμών. Τέλος, κατέληξαν στο συμπέρασμα ότι το Αρνητικό Διωνυμικό μοντέλο Lee - Carter παρουσιάζει ικανοποιητική προσαρμογή στα δεδομένα θανάτων των 10 χωρών έναντι των υπολοίπων υπό εξέταση μοντέλων.

Στην παρούσα διπλωματική εργασία για την καλύτερη σύγκριση και την εξαγωγή ασφαλέστερων συμπερασμάτων προχωρήσαμε στην άντληση δεδομένων αριθμού θανάτων, έκθεσης στον κίνδυνο και πληθυσμού των προαναφερόμενων χωρών μέσω της Βάσης Δεδομένων (Human Mortality Database, www.mortality.org). Τα δεδομένα τα οποία αντλήθηκαν αφορούν τον ανδρικό, γυναικείο και συνολικό πληθυσμό ηλικίας από 0 μέχρι 85 ετών και για τα έτη 1981 – 2019. Στη συνέχεια, προσαρμόσαμε τα τρία μοντέλα τα οποία μελετάμε στα δεδομένα των 10 χωρών και υπολογίσαμε την απόκλιση για κάθε χώρα.

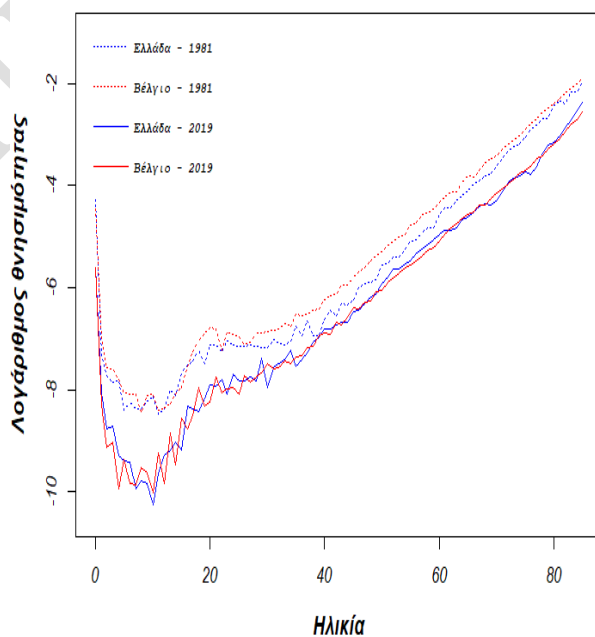
Στον ακόλουθο Πίνακα 5.1 παρουσιάζονται τα αποτελέσματα των 10 προαναφερόμενων χωρών μαζί με τις Ελλάδας προκειμένου να προχωρήσουμε στη σύγκρισή τους ως προς τις αποκλίσεις.

Πίνακας 5.1: Αποτελέσματα απόκλισης τριών μοντέλων Lee - Carter για 11 χώρες (1981 - 2019).

Χώρα	Άνδρες			Γυναίκες			Σύνολο		
	Poisson	Αρνητικό Διωνυμικό	Διωνυμικό	Poisson	Αρνητικό Διωνυμικό	Διωνυμικό	Poisson	Αρνητικό Διωνυμικό	Διωνυμικό
Ιαπωνία	19592	3522	45740	7915	4454	23086	31586	3260	72006
Η.Π.Α.	78831	3085	87923	31749	3314	37299	97052	3018	111091
Μ. Βρετανία	15115	3782	22092	10590	3171	18073	21658	3525	358
Αυσταλία	7354	3860	7660	4323	3563	4765	7732	3888	8342
Σουηδία	4035	3522	5103	3745	3278	4719	4160	3654	6231
Βέλγιο	5947	3823	9795	4582	3589	7699	6904	3855	13784
Ισπανία	21861	3156	32550	8983	4303	14971	28091	2767	44477
Καναδάς	9049	3802	10030	4813	3750	5593	9319	3982	10754
Ολλανδία	5121	3347	6773	3941	3423	5037	5022	3551	7653
Βουλγαρία	9340	3492	16133	5971	3309	12228	10882	3462	23744
Ελλάδα	7239	4002	8879	6625	3337	10068	9561	3887	14245

Παρατηρείται ότι από τα αποτελέσματα του Πίνακα 5.1, μία από τις Ευρωπαϊκές χώρες η οποία παρουσιάζει παρόμοια αποτελέσματα με αυτά της Ελλάδας ως προς την απόκλιση των μοντέλων, είναι το Βέλγιο.

Όπως προκύπτει και από στο ακόλουθο διάγραμμα, ο λογάριθμος της θνησιμότητας των 2 χωρών είναι αρκετά κοντά το 2019, ενώ μικρές διαφορές παρατηρούνται το 1981.



Διάγραμμα 5.1: Λογάριθμος θνησιμότητας συνολικού πληθυσμού Ελλάδας και Βελγίου για τα έτη 1981 και 2019.

Σύμφωνα με τους Azman και Pathmanathan (2020), το επικρατέστερο μοντέλο για τα δεδομένα του Βελγίου βάσει της απόκλισης είναι το Αρνητικό Διωνυμικό Lee – Carter. Στο ίδιο συμπέρασμα με την προαναφερόμενη έρευνα καταλήγουμε σύμφωνα με τον ανωτέρω Πίνακα 5.1. Προκειμένου να ενισχύσουμε το ανωτέρω συμπέρασμα προχωρήσαμε στον υπολογισμό των κριτηρίων πληροφορίας AIC και BIC, με τα αποτελέσματα να παρουσιάζονται στον παρακάτω πίνακα.

Πίνακας 5.2: Κριτήρια πληροφορίας AIC και BIC για τα δεδομένα του Βελγίου.

	AIC			BIC		
	Poisson Lee - Carter	Αρνητικό Διωνυμικό Lee - Carter	Διωνυμικό Lee - Carter	Poisson Lee - Carter	Αρνητικό Διωνυμικό Lee - Carter	Διωνυμικό Lee - Carter
Άνδρες	29957	29328	33742	31235	30607	35021
Γυναίκες	26809	26629	29890	28087	27908	31168
Σύνολο	32521	31444	39354	33799	32722	40632

Από τον Πίνακα 5.2 προκύπτει ότι και τα δύο κριτήρια πληροφορίας δίνουν μικρότερο αποτέλεσμα για το Αρνητικό Διωνυμικό μοντέλο έναντι του Poisson, τα οποία έχουν την ίδια συνάρτηση σύνδεσης. Αυτό σημαίνει ότι το Αρνητικό Διωνυμικό μοντέλο παρουσιάζει καλύτερη προσαρμογή.

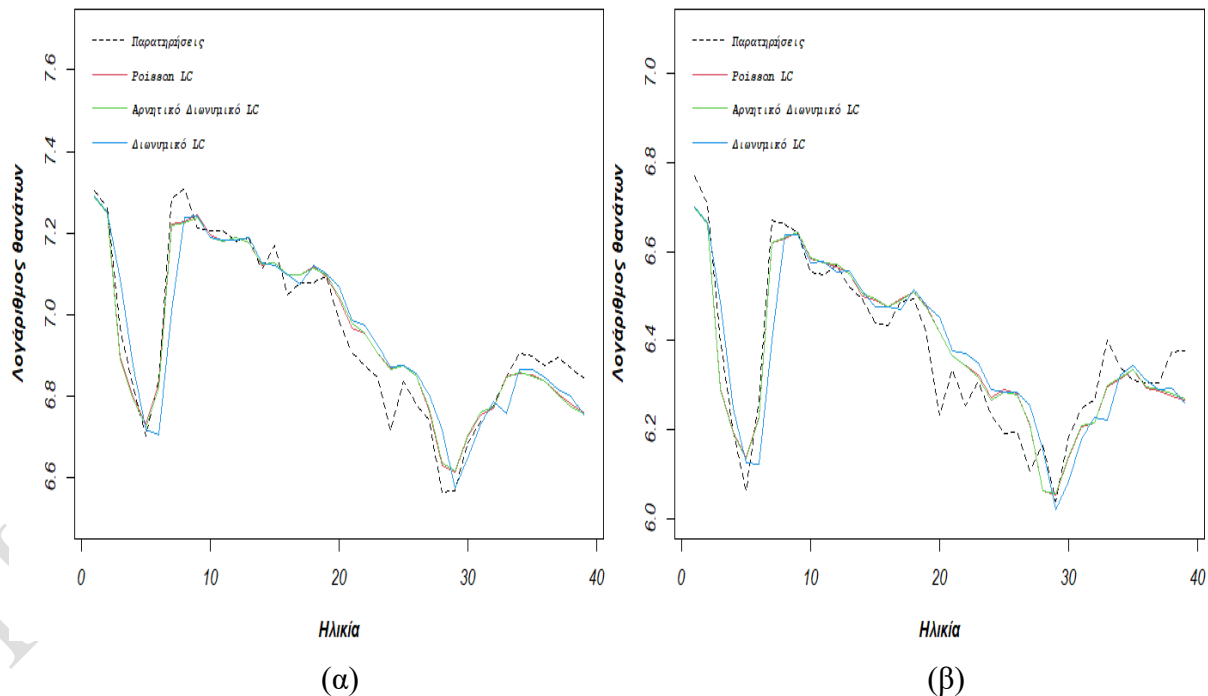
Εν συνεχεία, στην έρευνά τους οι Azman και Pathmanathan αναφέρουν ότι ως συνέπεια του μεγάλου όγκου δεδομένων θνησιμότητας είναι αναμενόμενο η μεταβλητή απόκρισης, η οποία στην προκειμένη περίπτωση είναι ο αριθμός θανάτων, να παρουσιάζει υπερδιασπορά. Λόγω της παρουσίας υπερδιασποράς στα δεδομένα καταλήγουν στο συμπέρασμα ότι το Αρνητικό Διωνυμικό μοντέλο είναι επικρατέστερο καθώς παρουσιάζει καλύτερη προσαρμογή λόγω της επιπλέον παραμέτρου υπερδιασποράς που περιλαμβάνεται στο μοντέλο. Ο εν λόγω ισχυρισμός επιβεβαιώνεται και από τον υπολογισμό της παραμέτρου υπερδιασποράς, τα αποτελέσματα του οποίου παρουσιάζονται στον Πίνακα 5.3.

Πίνακας 5.3: Παράμετρος διασποράς $\hat{\phi}$ για τα δεδομένα του Βελγίου.

	Poisson Lee - Carter	Αρνητικό Διωνυμικό Lee - Carter	Διωνυμικό Lee - Carter
Άνδρες	1.89	1.21	3.16
Γυναίκες	1.46	1.21	2.50
Σύνολο	2.20	1.23	4.48

Από τον Πίνακα 5.3 προκύπτει ότι το Αρνητικό Διωνυμικό μοντέλο παρουσιάζει μικρότερη τιμή υπερδιασποράς σε αντίθεση με τα υπόλοιπα μοντέλα.

Μελετώντας το Διάγραμμα 5.2, στο οποίο παρουσιάζονται οι εκτιμήσεις του λογαρίθμου του αριθμού των θανάτων για το Βέλγιο για κάθε μοντέλο ξεχωριστά, είναι εμφανές ότι οι εκτιμήσεις του Αρνητικού Διωνυμικού μοντέλου είναι πιο κοντά στις παρατηρούμενες τιμές. Αρκετά καλές εκτιμήσεις φαίνεται να δίνει και το Poisson μοντέλο, ενώ το Διωνυμικό μοντέλο δεν παρουσιάζει την ικανοποιητική προσαρμογή συγκριτικά με τα άλλα δύο μοντέλα.



Διάγραμμα 5.2: Σύγκριση μεταξύ μοντέλων εκτίμησης λογαρίθμου αριθμού θανάτων ατόμων ηλικίας 67 ετών για τα δεδομένα του Βελγίου για: (α) Άνδρες και (β) Γυναίκες.

ΚΕΦΑΛΑΙΟ 5

Συμπεράσματα

Συνοψίζοντας, μελετήσαμε το μοντέλου θνησιμότητας των Lee - Carter υπό το πρίσμα των γενικευμένων γραμμικών μοντέλων για τα δεδομένα της Ελλάδας, υποθέτοντας ότι η μεταβλητή απόκρισης του μοντέλου ακολουθεί μία εκ των κατανομών Poisson, Αρνητική Διωνυμική και Διωνυμική. Κατόπιν σύγκριση των τριών μοντέλων βάσει του μέτρου της απόκλισης για άνδρες, γυναίκες και συνολικό πληθυσμό, καταλήξαμε στο συμπέρασμα ότι το Αρνητικό Διωνυμικό μοντέλο Lee – Carter παρουσιάζει καλύτερη προσαρμογή στα δεδομένα. Στο εν λόγω συμπέρασμα συνηγορεί η μελέτη των κανονικοποιημένων καταλοίπων καθώς και τα κριτήρια πληροφορίας AIC και BIC. Πιθανότατα η επιπλέον παράμετρος διασποράς που περιλαμβάνει το Αρνητικό Διωνυμικό μοντέλο, να του δίνει ένα πλεονέκτημα έναντι των υπολοίπων λόγω της υπερδιασποράς που παρουσιάζουν τα δεδομένα. Συγκρίνοντας τα αποτελέσματα της Ελλάδας με αντίστοιχα αποτελέσματα από τη διεθνή βιβλιογραφία, καταλήγουμε και πάλι στο συμπέρασμα ότι το Αρνητικό Διωνυμικό μοντέλο υπερτερεί συγκριτικά με τα άλλα δύο μοντέλα.

ΒΙΒΛΙΟΓΡΑΦΙΑ

Ελληνική

- Μαυρομάτη, Ε. (2012). Μοντελοποίηση και πρόβλεψη θνησιμότητας: το μοντέλο του Lee-Carter και προεκτάσεις του (Διπλωματική εργασία). Τμήμα Στατιστικής και Ασφαλιστικής Επιστήμης, Πανεπιστήμιο Πειραιώς.
- Μποζίκας, Α. (2013). Προβολή θνησιμότητας για προγράμματα κοινωνικής ασφάλισης στην Ελλάδα. Τμήμα Στατιστικής και Ασφαλιστικής Επιστήμης, Πανεπιστήμιο Πειραιώς.
- Πολίτης, Κ. (2021). Γενικευμένα Γραμμικά Μοντέλα, Πανεπιστημιακές Σημειώσεις. Τμήμα Στατιστικής και Ασφαλιστικής Επιστήμης, Πανεπιστήμιο Πειραιώς.

Αγγλόφωνη

- Agresti, A. (2015). Foundations of linear and generalized linear models. John Wiley & Sons.
- Basellini, U., Camarda, C. G., & Booth, H. (2022). Thirty years on: A review of the Lee-Carter method for forecasting mortality. *International Journal of Forecasting* (2022).
- Booth, H., & Tickle, L. (2008). Mortality modelling and forecasting: A review of methods. *Annals of actuarial science*, 3(1-2), 3-43.
- Bozikas, A. E. (2019). Actuarial models in demography (Doctoral dissertation, University of Piraeus, Greece).
- Brouhns, N., Denuit, M., & Vermunt, J. K. (2002). A Poisson log-bilinear regression approach to the construction of projected lifetables. *Insurance: Mathematics and economics*, 31(3), 373-393.
- Cairns, A. J., Blake, D., Dowd, K., Coughlan, G. D., Epstein, D., Ong, A., & Balevich, I. (2009). A quantitative comparison of stochastic mortality models using data from England and Wales and the United States. *North American Actuarial Journal*, 13(1), 1-35.
- Coffie, E. (2015). A comparison of Poisson or Negative Binomial Regression and Lee-Carter Models of forecasting Norwegian male mortality (Master's thesis).
- Currie, I. D. (2013). Smoothing constrained generalized linear models with an application to the Lee-Carter model. *Statistical Modelling*, 13(1), 69-93.

- Delwarde, A., Denuit, M., & Partrat, C. (2007). Negative binomial version of the Lee–Carter model for mortality forecasting. *Applied Stochastic Models in Business and Industry*, 23(5), 385-401.
- Goldburd, M., Khare, A., Tevet, D., & Guller, D. (2016). Generalized linear models for insurance rating. *Casualty Actuarial Society, CAS Monographs Series*, 5.
- Hilbe, J. M. (2012). *Negative binomial regression*. 2nd edition. Cambridge, UK: Cambridge University Press.
- HMD (2019). Human Mortality Database. University of California, Berkeley (USA), and Max Planck Institute for Demographic Research (Germany). Available at www.mortality.org. University of California.
- Lane, P. W. (1996). Generalized nonlinear models. *COMPSTAT: Proceedings in Computational Statistics 12th Symposium held in Barcelona, Spain*. Physica-Verlag HD.
- Lindsey, J. K. (2000). *Applying generalized linear models*. Springer Science & Business Media.
- McCullagh, P. (2019). *Generalized linear models*. Routledge.
- Naufal, N., Devila, S., & Lestari, D. (2019). Generalized linear model (GLM) to determine life insurance premiums. In *AIP Conference Proceedings* (Vol. 2168, No. 1, p. 020036).
- Rawak, R. H., Yaacob, N. A., Pathmanathan, D., & Mohamed, I. (2022). Extending the GLM Framework of the Lee-Carter Model with Random Forest Recursive Feature Elimination Based Determinants of Mortality. *Sains Malaysiana*, 51(7), 2237-2247.
- R Development Core Team (2015). *Language and Environment for Statistical Computing*. R Foundation for Statistical Computing: Vienna, Austria R Development Core Team.
- Renshaw, A. E., & Haberman, S. (2006). A cohort-based extension to the Lee–Carter model for mortality reduction factors. *Insurance: Mathematics and economics*, 38(3), 556-570.
- Sarpong, S. A. (2013). Modeling and forecasting maternal mortality; an application of ARIMA models. *International Journal of Applied*, 3(1), 19-28.
- Taylor, R., Lewis, M., & Powles, J. (1998). The Australian mortality decline: all-cause mortality 1788–1990. *Australian and New Zealand journal of public health*, 22(1), 27-36.
- Venables, W. N. & Ripley, B. D. (2002). *Modern Applied Statistics with S*. Fourth edition. Springer.

Διαδικτυακές πηγές

<https://cran.r-project.org/web/packages/MASS/index.html>

https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Mortality_and_life_expectancy_statistics

<https://ec.europa.eu/eurostat/web/products-eurostat-news/-/ddn-20221118-1>

<https://www.statistics.gr/el/statistics/-/publication/SPO13/>

ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΕΙΡΑΙΩΣ

ΠΑΡΑΡΤΗΜΑΤΑ

Παράρτημα Α: Κώδικας R και πίνακες αποτελεσμάτων για το Κεφάλαιο 1.

```
life_exp_GRC<-read.table("HMD_GRC_Life_Exp.txt",header=TRUE)
par(mar=c(4.2,4,0.8,0.8))
plot(life_exp_GRC$Year,life_exp_GRC$Male,type="l",lwd=1.5,col="blue",ylim=c(73.5,84),
      ylab="Προσδόκιμο ζωής",xlab="Έτος",font.axis=3,font.lab=4)
lines(life_exp_GRC$Year,life_exp_GRC$Female,col="red",lwd=1.5)
legend("topleft", legend=c("Άνδρες","Γυναίκες"),col=c("blue","red"),
      lty=c(1,1),lwd=1.5,cex=0.7,text.width=10,text.font =10,text.col = 1,
      x.intersp = 0.3, y.intersp = 1.3,bty="n")

births_GRC<-read.table("HMD_GRC_birth.txt",header=TRUE)
par(mar=c(4.2,4,0.6,0.8))
plot(births_GRC$Year,births_GRC$Total,type="l",lwd=1.5,
      ylab="Αριθμός Γεννήσεων",xlab="Έτος",col="blue",font.axis=3,font.lab=4)

Dx_GRC<-read.table("HMD_GRC_Deaths.txt",header=TRUE)
Exp_GRC<-read.table("HMD_GRC_Exposure_to_risk.txt",header=TRUE)
Pop_GRC<-read.table("HMD_GRC_Population.txt",header=TRUE)

Years=rep(c(1981:2019),each=86)
Ages=c(0:85)
Dx_GRC_new<-filter(Dx_GRC,Dx_GRC$Age %in% c(0:85))
Exp_GRC_new<-filter(Exp_GRC,Exp_GRC$Age %in% c(0:85))
Pop_GRC_new<-filter(Pop_GRC,Pop_GRC$Age %in% c(0:85),Pop_GRC$Year <= 2019)

Dx_GRC_new$Total <- round(Dx_GRC_new$Total,0)
Dx_GRC_new$Male <- round(Dx_GRC_new$Male,0)
Dx_GRC_new$Female <- round(Dx_GRC_new$Female,0)

GRC.data.T=data.frame(Dx_GRC_new$Total,
                      Exp_GRC_new$Total,
                      Pop_GRC_new$Total,
                      Years,
                      Ages)

GRC.data.M=data.frame(Dx_GRC_new$Male,
                      Exp_GRC_new$Male,
                      Pop_GRC_new$Male,
                      Years,
                      Ages)

GRC.data.F=data.frame(Dx_GRC_new$Female,
                      Exp_GRC_new$Female,
                      Pop_GRC_new$Female,
```


Years,
Ages)

```
GRC.data.F[2155,1]<-(GRC.data.F[2069,1]+GRC.data.F[2241,1])/2  
GRC.data.F[3276,1]<-GRC.data.F[3190,1]
```

```
Deaths_Male<-matrix(GRC.data.M$Dx_GRC_new.Male,86,39,byrow=F)  
rownames(Deaths_Male)<-0:85  
colnames(Deaths_Male)<-1981:2019  
par(mar=c(1.8,1,0.6,0.2))  
persp(seq(0,85),seq(1981,2019),Deaths_Male,ticktype="detailed",theta=28,  
      xlab="Ηλικία",ylab="Έτος",zlab="Αριθμός θανάτων",phi=3,cex.lab=1.2,cex.axis=0.8,  
      col="springgreen",shade=0.4,box=T,axes=T,nticks=5,r=4,font.axis=3,font.lab=4)
```

```
Deaths_Female<-matrix(GRC.data.F$Dx_GRC_new.Female,86,39,byrow=F)  
rownames(Deaths_Female)<-0:85  
colnames(Deaths_Female)<-1981:2019  
par(mar=c(1.8,1,0.6,0.2))  
persp(seq(0,85),seq(1981,2019),Deaths_Female,ticktype="detailed",theta=28,  
      xlab="Ηλικία",ylab="Έτος",zlab="Αριθμός  
θανάτων",phi=3,cex.lab=1.2,cex.axis=0.8,col="springgreen",  
      shade=0.4,box=T,axes=T,nticks=5,r=4,font.axis=3,font.lab=4)
```

```
par(mar=c(4.2,4,0.8,0.8))  
plot(seq(1981,2019),Deaths_Male[68,],xlab="Έτος",ylab="Αριθμός θανάτων",  
     cex.lab=1.2,cex.axis=1,col="blue",pch=20,font.axis=3,font.lab=4)  
lines(smooth.spline(seq(1981,2019),Deaths_Male[68,],df=25),col="red")
```

```
par(mar=c(4.2,4,2,0.8,0.8))  
plot(seq(0,85),Deaths_Male[,39],xlab="Ηλικία",ylab="Αριθμός θανάτων",  
     cex.lab=1.2,cex.axis=1,type="l",col="blue",lwd=1.5,font.axis=3,font.lab=4)  
lines(seq(0,85),Deaths_Male[,1],col="red",lwd=1.5)  
legend("topleft", legend=c("Αριθμός θανάτων 1981","Αριθμός θανάτων  
2019"),col=c("red","blue"),  
      lty=c(1,1),lwd=1.5,cex=0.7,text.width=12,text.font =13,text.col = 1,  
      x.intersp = 0.3, y.intersp = 1.3,bty="n")
```

```
par(mar=c(4.2,4,0.8,0.8))  
plot(seq(1981,2019),Deaths_Female[68,],xlab="Έτος",ylab="Αριθμός θανάτων",  
     cex.lab=1.2,cex.axis=1,col="blue",pch=20,font.axis=3,font.lab=4)  
lines(smooth.spline(seq(1981,2019),Deaths_Female[68,],df=25),col="red")
```

```
par(mar=c(4.2,4,0.8,0.8))  
plot(seq(0,85),Deaths_Female[,39],xlab="Ηλικία",ylab="Αριθμός θανάτων",  
     cex.lab=1.2,cex.axis=1,type="l",col="blue",lwd=1.5,ylim=c(0,3200),font.axis=3,font.lab=4)  
lines(seq(0,85),Deaths_Female[,1],col="red",lwd=1.5)
```

```

legend("topleft", legend=c("Αριθμός θανάτων 1981", "Αριθμός θανάτων
2019"),col=c("red", "blue"),
      lty=c(1,1),lwd=1.5,cex=0.7,text.width=12,text.font =13,text.col = 1,
      x.intersp = 0.3, y.intersp = 1.3,bty="n")

D_Female <-GRC.data.F$Dx_GRC_new.Female/100000
GRC.deaths.F.per.100000<-matrix(D_Female,86,39,byrow=F)
colnames(GRC.deaths.F.per.100000)<-1981:2019

D_Male <-GRC.data.M$Dx_GRC_new.Male/100000
GRC.deaths.M.per.100000<-matrix(D_Male,86,39,byrow=F)
colnames(GRC.deaths.M.per.100000)<-1981:2019

par(mar=c(4.2,4,0.8,0.8))
plot(seq(0:85),GRC.deaths.F.per.100000[,39],xlab="Ηλικία",ylab="Αριθμός θανάτων",
      cex.lab=1.2,cex.axis=1,type="l",col="red",lwd=1.5,font.axis=3,font.lab=4)
lines(seq(0,85),GRC.deaths.M.per.100000[,39],col="blue",lwd=1.5)

legend("topleft", legend=c("Γυναίκες", "Άνδρες"),col=c("red", "blue"),
      lty=c(1,1),lwd=1.5,cex=0.7,text.width=12,text.font =13,text.col = 1,
      x.intersp = 0.3, y.intersp = 1.3,bty="n")

Expo_Male<-matrix(GRC.data.M$Exp_GRC_new.Male,86,39,byrow=F)
Expo_Female<-matrix(GRC.data.F$Exp_GRC_new.Female,86,39,byrow=F)
par(mar=c(4.2,4,0.8,0.6))
plot(seq(0,85),Expo_Male[,39],xlab="Ηλικία",ylab="Πληθυσμός",
      cex.lab=1.2,cex.axis=1,type="l",col="blue",lwd=1.5,ylim=c(5000,90000),font.axis=3,font.lab
=4)
lines(seq(0,85),Expo_Male[,1],col="red",lwd=1.5)
legend("bottomleft", legend=c("1981", "2019"),col=c("red", "blue"),
      lty=c(1,1),lwd=1.5,cex=0.7,text.width=10,text.font =13,text.col = 1,
      x.intersp = 0.3, y.intersp = 1.3,bty="n")

par(mar=c(4.2,4,0.8,0.6))
plot(seq(0,85),Expo_Female[,39],xlab="Ηλικία",ylab="Πληθυσμός",
      cex.lab=1.2,cex.axis=1,type="l",col="blue",lwd=1.5,ylim=c(10000,90000),font.axis=3,font.la
b=4)
lines(seq(0,85),Expo_Female[,1],col="red",lwd=1.5)
legend("bottomleft", legend=c("1981", "2019"),col=c("red", "blue"),
      lty=c(1,1),lwd=1.5,cex=0.7,text.width=10,text.font =13,text.col = 1,
      x.intersp = 0.3, y.intersp = 1.3,bty="n")

mort_total1 <-
matrix(GRC.data.T$Dx_GRC_new.Total/GRC.data.T$Exp_GRC_new.Total,86,39,byrow=F)

```

```

mort_total <-
matrix(log(GRC.data.T$Dx_GRC_new.Total/GRC.data.T$Exp_GRC_new.Total),86,39,byrow=F)
rownames(mort_total)<- 0:85
colnames(mort_total)<-1981:2019

par(mar=c(4.2,4,0.8,0.6))
plot(seq(0,85),mort_total[,39],xlab="Ηλικία",ylab="θνησιμότητα",
      cex.lab=1.2,cex.axis=1,type="l",col="blue",lwd=1.5,ylim=c(0,0.14),font.axis=3,font.lab=4)
lines(seq(0,85),mort_total[,1],col="red",lwd=1.5)

legend("topleft", legend=c("1981","2019"),col=c("red","blue"),
      lty=c(1,1),lwd=1.5,cex=0.7,text.width=6,text.font =13,text.col = 1,
      x.intersp = 0.3, y.intersp = 1.3, bty="n")

par(mar=c(4.2,4,0.8,0.6))
plot(seq(0,85),mort_total[,39],xlab="Ηλικία",ylab="Λογάριθμος θνησιμότητας",
      cex.lab=1.2,cex.axis=1,type="l",col="blue",lwd=1.5,ylim=c(-10,-2),font.axis=3,font.lab=4)
lines(seq(0,85),mort_total[,1],col="red",lwd=1.5)

legend("topleft", legend=c("1981","2019"),col=c("red","blue"),
      lty=c(1,1),lwd=1.5,cex=0.7,text.width=6,text.font =13,text.col=1,
      x.intersp = 0.3, y.intersp = 1.5,bty="n")

mort_male <-
matrix(log(GRC.data.M$Dx_GRC_new.Male/GRC.data.M$Exp_GRC_new.Male),86,39,byrow=F)
rownames(mort_male)<- 0:85
colnames(mort_male)<-1981:2019

mort_female<-
matrix(log(GRC.data.F$Dx_GRC_new.Female/GRC.data.F$Exp_GRC_new.Female),86,39,byrow=F)
rownames(mort_female)<- 0:85
colnames(mort_female)<-1981:2019

par(mar=c(4.2,4,0.8,0.8))
plot(seq(1981,2019),mort_male[68,],xlab="Έτος",ylab="Λογάριθμος θνησιμότητας",
      cex.lab=1.2,cex.axis=1,col="blue",pch=20,font.axis=3,font.lab=4)
lines(smooth.spline(seq(1981,2019),mort_male[68,],df=25),col="red")

plot(seq(0,85),mort_male[,39],xlab="Ηλικία",ylab="Λογάριθμος θνησιμότητας",
      cex.lab=1.2,cex.axis=1,type="l",col="blue",lwd=1.5,ylim=c(-11,-1.9),font.axis=3,font.lab=4)
lines(seq(0,85),mort_male[,1],col="red",lwd=1.5)
legend("topleft", legend=c("1981","2019"),col=c("red","blue"),

```

```
lty=c(1,1),lwd=1.5,cex=0.7,text.width=6,text.font =13,text.col=1,  
x.intersp = 0.3, y.intersp = 1.5,bty="n")
```

```
plot(seq(1981,2019),mort_female[68,],xlab="Έτος",ylab="Θνησιμότητα",  
cex.lab=1.2,cex.axis=1,col="blue",pch=20,font.axis=3,font.lab=4)  
lines(smooth.spline(seq(1981,2019),mort_female[68,],df=25),col="red")
```

```
plot(seq(0,85),mort_female[,39],xlab="Ηλικία",ylab="Λογάριθμος θνησιμότητας",  
cex.lab=1.2,cex.axis=1,type="l",col="blue",lwd=1.5,ylim=c(-11,-  
1.9),font.axis=3,font.lab=4)  
lines(seq(0,85),mort_female[,1],col="red",lwd=1.5)  
legend("topleft", legend=c("1981","2019"),col=c("red","blue"),  
lty=c(1,1),lwd=1.5,cex=0.7,text.width=6,text.font =13,text.col = 1,  
x.intersp = 0.3, y.intersp = 1.3,bty="n")
```

Πίνακας Α.1: Παρατηρούμενος αριθμός θανάτων ελληνικού Πληθυσμού ηλικίας 67 ετών.

Έτος	Παρατηρούμενος αριθμός θανάτων		
	Άνδρες	Γυναίκες	Συνολικός Πληθυσμός
1981	1018	616	1634
1982	1083	679	1762
1983	960	651	1611
1984	831	593	1424
1985	851	634	1485
1986	764	519	1283
1987	952	562	1514
1988	900	601	1501
1989	903	546	1449
1990	867	590	1457
1991	959	592	1551
1992	1015	670	1685
1993	1198	670	1868
1994	1173	716	1889
1995	1232	693	1925
1996	1272	662	1934
1997	1224	700	1924
1998	1303	698	2001
1999	1153	602	1755
2000	1087	655	1742
2001	1186	595	1781
2002	1187	604	1791
2003	1083	573	1656
2004	1065	532	1597
2005	1027	499	1526
2006	934	500	1434
2007	994	493	1487
2008	869	429	1298
2009	719	372	1091
2010	684	343	1027
2011	887	388	1275
2012	978	447	1425
2013	1044	501	1545
2014	1085	463	1548
2015	987	494	1481
2016	820	456	1276

2017	926	461	1387
2018	942	445	1387
2019	1005	464	1469

Πίνακας Α.2: Παρατηρούμενες τιμές λογαρίθμου δείκτη θνησιμότητας για τα έτη 1981 και 2019.

Παρατηρούμενες τιμές λογαρίθμου δείκτη θνησιμότητας						
Ηλικία	Ανδρες		Γυναίκες		Συνολικός Πληθυσμός	
	1981	2019	1981	2019	1981	2019
0	-4.182	-5.604	-4.391	-5.626	-4.277	-5.615
1	-7.347	-7.969	-7.32	-8.199	-7.333	-8.074
2	-7.518	-8.855	-7.998	-8.667	-7.723	-8.759
3	-7.589	-9.205	-8.224	-8.363	-7.848	-8.708
4	-7.632	-9.2	-8.1	-9.371	-7.833	-9.279
5	-8.371	-9.422	-8.429	-9.369	-8.399	-9.396
6	-8.079	-9.442	-8.492	-9.386	-8.258	-9.415
7	-8.409	-9.284	-8.361	-9.448	-8.386	-9.948
8	-8.065	-10.23	-8.888	-9.481	-8.384	-9.797
9	-8.194	-9.353	-8.246	-10.915	-8.219	-9.84
10	-7.914	-9.875	-8.427	-10.929	-8.132	-10.259
11	-8.389	-9.55	-8.586	-9.784	-8.48	-9.657
12	-8.126	-9.329	-8.762	-9.269	-8.386	-9.299
13	-7.743	-9.134	-8.385	-9.257	-8.005	-9.192
14	-7.964	-8.981	-8.298	-9.066	-8.113	-9.021
15	-7.361	-9.825	-8.169	-8.77	-7.674	-9.183
16	-7.313	-8.107	-7.792	-8.661	-7.518	-8.335
17	-7.107	-8.066	-7.975	-8.92	-7.443	-8.387
18	-6.865	-8.194	-7.775	-8.801	-7.217	-8.439
19	-7.077	-7.792	-8.264	-8.805	-7.495	-8.156
20	-6.833	-7.589	-7.592	-8.392	-7.135	-7.897
21	-6.67	-7.525	-8.01	-8.679	-7.126	-7.923
22	-6.837	-7.335	-8.03	-8.805	-7.265	-7.794
23	-6.737	-7.754	-7.519	-8.688	-7.056	-8.097
24	-6.655	-7.21	-7.943	-8.804	-7.113	-7.689
25	-6.8	-7.395	-7.652	-8.679	-7.145	-7.823
26	-6.767	-7.397	-7.794	-8.689	-7.158	-7.829
27	-6.782	-7.459	-7.683	-8.219	-7.134	-7.758
28	-6.9	-7.314	-7.448	-9.003	-7.137	-7.828
29	-6.917	-7.004	-7.547	-7.972	-7.182	-7.373
30	-6.966	-7.555	-7.474	-8.682	-7.187	-7.967
31	-6.704	-7.245	-7.511	-8.047	-7.028	-7.57

32	-6.856	-7.082	-7.387	-8.136	-7.092	-7.48
33	-6.861	-7.092	-7.467	-7.887	-7.126	-7.413
34	-6.835	-6.837	-7.293	-7.843	-7.042	-7.216
35	-6.535	-7.162	-7.012	-8.174	-6.75	-7.542
36	-6.502	-6.991	-7.713	-8.151	-6.946	-7.408
37	-6.46	-6.838	-6.911	-8.051	-6.663	-7.27
38	-6.71	-6.721	-7.258	-7.589	-6.95	-7.065
39	-6.591	-6.661	-7.443	-7.44	-6.94	-6.977
40	-6.309	-6.465	-7.037	-7.376	-6.622	-6.821
41	-6.179	-6.43	-6.767	-7.457	-6.44	-6.818
42	-6.333	-6.432	-6.781	-7.188	-6.539	-6.742
43	-6.035	-6.353	-6.696	-7.134	-6.324	-6.673
44	-6.135	-6.385	-6.585	-7.1	-6.345	-6.687
45	-5.952	-6.152	-6.594	-6.93	-6.235	-6.474
46	-5.841	-6.158	-6.185	-6.797	-6.002	-6.432
47	-5.718	-5.994	-6.168	-6.75	-5.921	-6.31
48	-5.676	-5.821	-6.158	-6.726	-5.89	-6.185
49	-5.531	-5.819	-6.295	-6.562	-5.837	-6.134
50	-5.365	-5.654	-5.81	-6.288	-5.564	-5.932
51	-5.231	-5.463	-5.902	-6.245	-5.518	-5.795
52	-5.187	-5.302	-5.691	-6.066	-5.411	-5.627
53	-5.134	-5.321	-5.752	-6.09	-5.403	-5.646
54	-4.959	-5.242	-5.638	-5.908	-5.251	-5.533
55	-4.776	-5.115	-5.467	-5.992	-5.076	-5.481
56	-4.712	-4.925	-5.51	-5.92	-5.05	-5.331
57	-4.602	-4.884	-5.303	-5.734	-4.906	-5.246
58	-4.484	-4.811	-5.248	-5.66	-4.819	-5.173
59	-4.491	-4.696	-5.256	-5.547	-4.819	-5.058
60	-4.271	-4.585	-4.967	-5.466	-4.573	-4.957
61	-4.139	-4.509	-4.758	-5.356	-4.42	-4.87
62	-4.153	-4.473	-4.751	-5.428	-4.431	-4.869
63	-4.026	-4.415	-4.538	-5.404	-4.272	-4.819
64	-3.899	-4.295	-4.526	-5.157	-4.187	-4.659
65	-3.831	-4.25	-4.409	-5.112	-4.101	-4.615
66	-3.729	-4.187	-4.205	-4.929	-3.954	-4.511
67	-3.65	-4.044	-4.255	-4.933	-3.922	-4.417
68	-3.5	-4.006	-4.154	-4.811	-3.805	-4.353
69	-3.458	-4.011	-4.092	-4.854	-3.76	-4.372
70	-3.364	-3.98	-3.858	-4.72	-3.612	-4.309
71	-3.297	-3.782	-3.639	-4.502	-3.47	-4.107

72	-3.129	-3.613	-3.486	-4.308	-3.3	-3.931
73	-3.061	-3.523	-3.423	-4.257	-3.238	-3.856
74	-2.985	-3.521	-3.374	-4.151	-3.179	-3.815
75	-2.924	-3.423	-3.169	-4.058	-3.057	-3.718
76	-2.76	-3.488	-3.016	-4.085	-2.895	-3.769
77	-2.677	-3.357	-2.955	-3.908	-2.816	-3.622
78	-2.593	-3.116	-2.745	-3.621	-2.674	-3.367
79	-2.494	-2.979	-2.866	-3.441	-2.694	-3.216
80	-2.352	-2.893	-2.474	-3.422	-2.425	-3.163
81	-2.235	-2.846	-2.415	-3.178	-2.342	-3.025
82	-2.254	-2.656	-2.51	-3.045	-2.396	-2.865
83	-2.087	-2.556	-2.199	-2.832	-2.151	-2.709
84	-2.044	-2.366	-2.229	-2.653	-2.15	-2.527
85	-1.922	-2.246	-2.017	-2.451	-1.98	-2.363

Παράρτημα Β: Κώδικας R και πίνακες αποτελεσμάτων για το Κεφάλαιο 4.

```

Year.new=unique(Years)
nC=length(Year.new)
nL=length(Ages)

deathrate_male=matrix(GRC.data.M$Dx_GRC_new.Male/GRC.data.M$Exp_GRC_new.Male,nL,nC)
deathrate_female=matrix(GRC.data.F$Dx_GRC_new.Female/GRC.data.F$Exp_GRC_new.Female,nL,nC)
deathrate_total=matrix(GRC.data.T$Dx_GRC_new.Total/GRC.data.T$Exp_GRC_new.Total,nL,nC)

expos_male=matrix(GRC.data.M$Exp_GRC_new.Male,nL,nC)
expos_female=matrix(GRC.data.F$Exp_GRC_new.Female,nL,nC)
expos_total=matrix(GRC.data.T$Exp_GRC_new.Total,nL,nC)

demo_male<-demogdata(data=deathrate_male,pop=expos_male,ages=Ages,years=Year.new,
  type="mortality",label="Greece",name="Male",lambda=1)
demo_female<-
demogdata(data=deathrate_female,pop=expos_female,ages=Ages,years=Year.new,
  type="mortality",label="Greece",name="Female",lambda=1)
demo_total<-demogdata(data=deathrate_total,pop=expos_total,ages=Ages,years=Year.new,
  type="mortality",label="Greece",name="Total",lambda=1)

Lca_male<- lca(demo_male,max.age=85,interpolate=F,adjust="none")
Lca_female<- lca(demo_female,max.age=85,interpolate=T,adjust="none")
Lca_total<- lca(demo_total,max.age=85,interpolate=T,adjust="none")

par(mar=c(4.2,4,1.8,1.2))

```



```

plot(Lca_male$fit,cex.lab=1.2,cex.axis=1,
     ylab="Δείκτης θνησιμότητας",xlab="Ηλικία",font.axis=3,font.lab=4)
plot(Lca_female$fit,cex.lab=1.2,cex.axis=1,
     ylab="Δείκτης θνησιμότητας",xlab="Ηλικία",font.axis=3,font.lab=4)
plot(Lca_total$fit,cex.lab=1.2,cex.axis=1,
     ylab="Δείκτης θνησιμότητας",xlab="Ηλικία",font.axis=3,font.lab=4)

ax_male<- Lca_male$ax
bx_male<- Lca_male$bx
kt_male<- Lca_male$kt

ax_female<- Lca_female$ax
bx_female<- Lca_female$bx
kt_female<- Lca_female$kt

ax_total<- Lca_total$ax
bx_total<- Lca_total$bx
kt_total<- Lca_total$kt

ax_values=matrix(c(ax_female,ax_male,ax_total),nL)
colnames(ax_values)<-c("ax_female","ax_male","ax_total")
rownames(ax_values)<-0:85
format(round(ax_values,5))
write.csv(format(round(ax_values,5)),"ax_values.csv",row.names=T)

bx_values=matrix(c(bx_female,bx_male,bx_total),nL)
colnames(bx_values)<-c("bx_female","bx_male","bx_total")
rownames(bx_values)<-0:85
format(round(bx_values,5))
write.csv(format(round(bx_values,5)),"bx_values.csv",row.names=T)

kt_values=matrix(c(kt_female,kt_male,kt_total),nC)
colnames(kt_values)<-c("kt_female","kt_male","kt_total")
rownames(kt_values)<-1981:2019
format(round(kt_values,5))
write.csv(round(kt_values,5),"kt_values.csv",row.names=T)

par(mar=c(4,4.2,0.6,0.4))
plot(ax_male,type="l",xlab="Ηλικία",ylab=expression("α"[x]),lty=1,lwd=1.5,
     col="blue",cex.lab=1.2,cex.axis=1,font.axis=3,font.lab=4)
lines(ax_female,type="l",lty=1,lwd=1.5,col="red")
lines(ax_total,type="l",lty=2,lwd=1.5,col="green")
legend("topleft",legend=c("Άνδρες","Γυναίκες","Σύνολο"),text.font=13,text.width=7,
     lty=c(1,1,2),col=c("blue","red","green"),bg="white",lwd=1.5,cex=0.7,pt.cex=1,bty="n",
     text.col=1.5,x.intersp = 0.3, y.intersp = 1.3)

par(mar=c(4,4.2,0.6,0.4))
plot(bx_male,type="l",xlab="Ηλικία",ylab=expression("b"[x]),lty=1,lwd=1.5,

```

```

col="blue",cex.lab=1.2,cex.axis=1,font.axis=3,font.lab=4)
lines(bx_female,type="l",lty=1,lwd=1.5,col="red")
lines(bx_total,type="l",lty=2,lwd=1.5,col="green")
legend("top",legend=c("Ανδρες", "Γυναίκες", "Σύνολο"),text.font=13,text.width=7,
      lty=c(1,1,2),col=c("blue","red","green"),bg="white",lwd=1.5,cex=0.7,pt.cex=1,bty="n",
      text.col=1.5,x.intersp = 0.3, y.intersp = 1.3)

par(mar=c(4,4.2,0.6,0.8))
plot(kt_male,type="l",xlab="Ηλικία",ylab=expression("k"[t]),lty=1,lwd=1.5,
     col="blue",cex.lab=1.2,cex.axis=1,ylim=c(-33,33),font.axis=3,font.lab=4)
lines(kt_female,type="l",lty=1,lwd=1.5,col="red")
lines(kt_total,type="l",lty=2,lwd=1.5,col="green")
legend("top",legend=c("Ανδρες", "Γυναίκες", "Σύνολο"),text.font=13,text.width=7,
      lty=c(1,1,2),col=c("blue","red","green"),bg="white",lwd=1.5,cex=0.7,pt.cex=1,bty="n",
      text.col=1.5,x.intersp = 0.3, y.intersp = 1.3)

tsDATA_total = ts(kt_total,start=min(Year.new),end=max(Year.new),frequency=1)
par(mfrow=c(1,2))
acf(tsDATA_total)
pacf(tsDATA_total)
adf.test(tsDATA_total)

fit.ARIMA.total=Arima(kt_total,order=c(0,1,0),include.drift=T)

checkresiduals(fit.ARIMA.total)
forecast_total<-forecast(fit.ARIMA.total,h=31)
par(mfrow=c(1,1))
par(mar=c(4.2,4.3,0.8,0.4))
plot(forecast_total,lwd=1.5,xlab="Ετη",ylab=expression(hat(kt)),
     main="",font.axis=3,font.lab=4)

tsDATA_male = ts(kt_male,start=min(Year.new),end=max(Year.new),frequency=1)
par(mfrow=c(1,2))
acf(tsDATA_male)
pacf(tsDATA_male)
adf.test(tsDATA_male)

fit.ARIMA.male=Arima(kt_male,order=c(0,1,0),include.drift=T)

checkresiduals(fit.ARIMA.male)
forecast_male<-forecast(fit.ARIMA.male,h=31)
par(mfrow=c(1,1))
par(mar=c(4.2,4.3,0.8,0.4))
plot(forecast_male,lwd=1.5,xlab="Ετη",ylab=expression(hat(kt)),
     main="",font.axis=3,font.lab=4)

tsDATA_female = ts(kt_female,start=min(Year.new),end=max(Year.new),frequency=1)

```

```

par(mfrow=c(1,2))
acf(tsDATA_female)
pacf(tsDATA_female)
adf.test(tsDATA_female)
fit.ARIMA.female=Arima(kt_female,order=c(0,1,0),include.drift=T)

checkresiduals(fit.ARIMA.female)
forecast_female<-forecast(fit.ARIMA.female,h=31)
par(mfrow=c(1,1))
par(mar=c(4.2,4.3,0.6,0.4))
plot(forecast_female,lwd=1.5,xlab="Έτη",ylab=expression(hat(kt)),
     main="",font.axis=3,font.lab=4)

LC.poisson.GRC.T <- gnm(round(GRC.data.T$Dx_GRC_new.Total,0)~-
1+as.factor(GRC.data.T$Ages)+
      Mult(as.factor(GRC.data.T$Ages),as.factor(GRC.data.T$Years))+
      offset(log(GRC.data.T$Exp_GRC_new.Total)),
      family=poisson(link="log"))

LC.poisson.GRC.T.stdres <- rstandard(LC.poisson.GRC.T)

par(mar=c(4.2,5,0.6,0.4))
plot(LC.poisson.GRC.T$fitted.values,LC.poisson.GRC.T.stdres,ylim=c(-30,30),
     ylab="Κανονικοποιημένα \nΚατάλοιπα",xlab="Εκτιμήσεις",cex.lab=1,cex.axis=1,
     font.axis=3,font.lab=4,pch=20)
abline(0,0,col="red",lwd=1.5)

par(mar=c(4.2,4,0.6,0.4))
qqnorm(LC.poisson.GRC.T.stdres,main="",ylab="Δειγματικά ποσοστημόρια",
       xlab="Θεωρητικά ποσοστημόρια",cex.lab=1,cex.axis=1,
       font.axis=3,font.lab=4,pch=20)
qqline(LC.poisson.GRC.T.stdres,lwd=1.5,lty=1,col="red")

Dx_fitted_values_T=matrix(LC.poisson.GRC.T$fitted.values,nL,nC)
colnames(Dx_fitted_values_T)<-1981:2019
rownames(Dx_fitted_values_T)<-0:85

LC.poisson.GRC.M <- gnm(GRC.data.M$Dx_GRC_new.Male~-
1+as.factor(GRC.data.M$Ages)+
      Mult(as.factor(GRC.data.M$Ages),as.factor(GRC.data.M$Years))+
      offset(log(GRC.data.M$Exp_GRC_new.Male)),
      family=poisson(link="log"))

LC.poisson.GRC.M.stdres <- rstandard(LC.poisson.GRC.M)

par(mar=c(4.2,5,0.6,0.4))
plot(LC.poisson.GRC.M$fitted.values,LC.poisson.GRC.M.stdres,ylim=c(-30,30),
     ylab="Κανονικοποιημένα \nΚατάλοιπα",xlab="Εκτιμήσεις",cex.lab=1,cex.axis=1,

```

```

font.axis=3,font.lab=4,pch=20)
abline(0,0,col="red",lwd=1.5)

par(mar=c(4.2,4,0.6,0.4))
qqnorm(LC.poisson.GRC.M.stdres,main="",ylab="Δειγματικά ποσοστημόρια",
  xlab="Θεωρητικά ποσοστημόρια",cex.lab=1,cex.axis=1,
  font.axis=3,font.lab=4,pch=20)
qqline(LC.poisson.GRC.M.stdres,lwd=1.5,lty=1,col="red")

LC.poisson.GRC.F <- gnm(GRC.data.F$Dx_GRC_new.Female~
1+as.factor(GRC.data.F$Ages)+
  Mult(as.factor(GRC.data.F$Ages),as.factor(GRC.data.F$Years))+
  offset(log(GRC.data.F$Exp_GRC_new.Female)),
  family=poisson(link="log"))

LC.poisson.GRC.F.stdres <- rstandard(LC.poisson.GRC.F)
par(mar=c(4.2,5,0.6,0.4))
plot(LC.poisson.GRC.F$fitted.values,LC.poisson.GRC.F.stdres,ylim=c(-30,30),
  ylab="Κανονικοποιημένα\ηΚατάλοιπα",xlab="Εκτιμήσεις",cex.lab=1,cex.axis=1,
  font.axis=3,font.lab=4,pch=20)
abline(0,0,col="red",lwd=1.5)

par(mar=c(4.2,4,0.6,0.4))
qqnorm(LC.poisson.GRC.F.stdres,main="",ylab="Δειγματικά ποσοστημόρια",
  xlab="Θεωρητικά ποσοστημόρια",cex.lab=1,cex.axis=1,
  font.axis=3,font.lab=4,pch=20)
qqline(LC.poisson.GRC.T.stdres,lwd=1.5,lty=1,col="red")

theta<-theta.ml(GRC.data.T$Dx_GRC_new.Total,fitted(LC.poisson.GRC.T))
LC.neg.binom.GRC.T <- gnm(GRC.data.T$Dx_GRC_new.Total~
1+as.factor(GRC.data.T$Ages)+
  Mult(as.factor(GRC.data.T$Ages),as.factor(GRC.data.T$Years))+
  offset(log(GRC.data.T$Exp_GRC_new.Total)),
  family=negative.binomial(theta))

LC.neg.binom.GRC.T.stdres <- rstandard(LC.neg.binom.GRC.T)
par(mar=c(4.2,5,0.6,0.4))
plot(LC.neg.binom.GRC.T$fitted.values,LC.neg.binom.GRC.T.stdres,ylim=c(-30,30),
  ylab="Κανονικοποιημένα\ηΚατάλοιπα",xlab="Εκτιμήσεις",cex.lab=1,cex.axis=1,
  font.axis=3,font.lab=4,pch=20)
abline(0,0,col="red",lwd=1.5)

par(mar=c(4.2,4,0.6,0.4))
qqnorm(LC.neg.binom.GRC.T.stdres,main="",ylab="Δειγματικά ποσοστημόρια",
  xlab="Θεωρητικά ποσοστημόρια",cex.lab=1,cex.axis=1,
  font.axis=3,font.lab=4,pch=20)
qqline(LC.neg.binom.GRC.T.stdres,lwd=1.5,lty=1,col="red")

```

```
theta<-theta.ml(GRC.data.M$Dx_GRC_new.Male,fitted(LC.poisson.GRC.M))
LC.neg.binom.GRC.M <- gnm(GRC.data.M$Dx_GRC_new.Male~-
1+as.factor(GRC.data.M$Ages)+
      Mult(as.factor(GRC.data.M$Ages),as.factor(GRC.data.M$Years))+
      offset(log(GRC.data.M$Exp_GRC_new.Male)),
      family=negative.binomial(theta))
```

```
LC.neg.binom.GRC.M.stdres <- rstandard(LC.neg.binom.GRC.M)
par(mar=c(4.2,5,0.6,0.4))
plot(LC.neg.binom.GRC.M$fitted.values,LC.neg.binom.GRC.M.stdres,ylim=c(-30,30),
      ylab="Κανονικοποιημένα\ηΚατάλοιπα",xlab="Εκτιμήσεις",cex.lab=1,cex.axis=1,
      font.axis=3,font.lab=4,pch=20)
abline(0,0,col="red",lwd=1.5)
```

```
par(mar=c(4.2,4,0.6,0.4))
qqnorm(LC.neg.binom.GRC.M.stdres,main="",ylab="Δειγματικά ποσοστημόρια",
      xlab="Θεωρητικά ποσοστημόρια",cex.lab=1,cex.axis=1,
      font.axis=3,font.lab=4,pch=20)
qqline(LC.neg.binom.GRC.M.stdres,lwd=1.5,lty=1,col="red")
```

```
theta<-theta.ml(GRC.data.F$Dx_GRC_new.Female,fitted(LC.poisson.GRC.F))
LC.neg.binom.GRC.F <- gnm(GRC.data.F$Dx_GRC_new.Female~-
1+as.factor(GRC.data.F$Ages)+
      Mult(as.factor(GRC.data.F$Ages),as.factor(GRC.data.F$Years))+
      offset(log(GRC.data.F$Exp_GRC_new.Female)),
      family=negative.binomial(theta))
```

```
LC.neg.binom.GRC.F.stdres <- rstandard(LC.neg.binom.GRC.F)
par(mar=c(4.2,5,0.6,0.4))
plot(LC.neg.binom.GRC.F$fitted.values,LC.neg.binom.GRC.F.stdres,ylim=c(-30,30),
      ylab="Κανονικοποιημένα\ηΚατάλοιπα",xlab="Εκτιμήσεις",cex.lab=1,cex.axis=1,
      font.axis=3,font.lab=4,pch=20)
abline(0,0,col="red",lwd=1.5)
```

```
par(mar=c(4.2,4,0.6,0.4))
qqnorm(LC.neg.binom.GRC.F.stdres,main="",ylab="Δειγματικά ποσοστημόρια",
      xlab="Θεωρητικά ποσοστημόρια",cex.lab=1,cex.axis=1,
      font.axis=3,font.lab=4,pch=20)
qqline(LC.neg.binom.GRC.F.stdres,lwd=1.5,lty=1,col="red")
```

```
LC.binom.GRC.T <-
gnm(GRC.data.T$Dx_GRC_new.Total/GRC.data.T$Pop_GRC_new.Total~-
1+as.factor(GRC.data.T$Ages)+
      Mult(as.factor(GRC.data.T$Ages),as.factor(GRC.data.T$Years)),
      family=binomial(link="logit"),
      weights=GRC.data.T$Pop_GRC_new.Total)
```

```

LC.binom.GRC.T.stdres <- rstandard(LC.binom.GRC.T)
par(mar=c(4.2,5,0.6,0.4))
plot(LC.binom.GRC.T$fitted.values,LC.binom.GRC.T.stdres,ylim=c(-30,30),
     ylab="Κανονικοποιημένα\ηΚατάλοιπα",xlab="Εκτιμήσεις",cex.lab=1,cex.axis=1,
     font.axis=3,font.lab=4,pch=20)
abline(0,0,col="red",lwd=1.5)

par(mar=c(4.2,4,0.6,0.4))
qqnorm(LC.binom.GRC.T.stdres,main="",ylab="Δειγματικά ποσοστημόρια",
       xlab="Θεωρητικά ποσοστημόρια",cex.lab=1,cex.axis=1,
       font.axis=3,font.lab=4,pch=20)
qqline(LC.binom.GRC.T.stdres,lwd=1.5,lty=1,col="red")
LC.binom.GRC.M <-
gnm(GRC.data.M$Dx_GRC_new.Male/GRC.data.M$Pop_GRC_new.Male~
1+as.factor(GRC.data.M$Ages)+
     Mult(as.factor(GRC.data.M$Ages),as.factor(GRC.data.M$Years)),
     family=binomial(link="logit"),
     weights=GRC.data.M$Pop_GRC_new.Male)

LC.binom.GRC.M.stdres <- rstandard(LC.binom.GRC.M)
par(mar=c(4.2,5,0.6,0.4))
plot(LC.binom.GRC.M$fitted.values,LC.binom.GRC.M.stdres,ylim=c(-30,30),
     ylab="Κανονικοποιημένα\ηΚατάλοιπα",xlab="Εκτιμήσεις",cex.lab=1,cex.axis=1,
     font.axis=3,font.lab=4,pch=20)
abline(0,0,col="red",lwd=1.5)

par(mar=c(4.2,4,0.6,0.4))
qqnorm(LC.binom.GRC.M.stdres,main="",ylab="Δειγματικά ποσοστημόρια",
       xlab="Θεωρητικά ποσοστημόρια",cex.lab=1,cex.axis=1,
       font.axis=3,font.lab=4,pch=20)
qqline(LC.binom.GRC.M.stdres,lwd=1.5,lty=1,col="red")

LC.binom.GRC.F <-
gnm(GRC.data.F$Dx_GRC_new.Female/GRC.data.F$Pop_GRC_new.Female~
1+as.factor(GRC.data.F$Ages)+
     Mult(as.factor(GRC.data.F$Ages),as.factor(GRC.data.F$Years)),
     family=binomial(link="logit"),
     weights=GRC.data.F$Pop_GRC_new.Female)

LC.binom.GRC.F.stdres <- rstandard(LC.binom.GRC.F)
par(mar=c(4.2,5,0.6,0.4))
plot(LC.binom.GRC.F$fitted.values,LC.binom.GRC.F.stdres,ylim=c(-30,30),
     ylab="Κανονικοποιημένα\ηΚατάλοιπα",xlab="Εκτιμήσεις",cex.lab=1,cex.axis=1,
     font.axis=3,font.lab=4,pch=20)
abline(0,0,col="red",lwd=1.5)

par(mar=c(4.2,4,0.6,0.4))
qqnorm(LC.binom.GRC.F.stdres,main="",ylab="Δειγματικά ποσοστημόρια",

```

```

xlab="Θεωρητικά ποσοστημόρια",cex.lab=1,cex.axis=1,
font.axis=3,font.lab=4,pch=20)
qqline(LC.binom.GRC.F.stdres,lwd=1.5,lty=1,col="red")

phi_total_poisson = sum(residuals(LC.poisson.GRC.T,
type="pearson")^2)/LC.poisson.GRC.T$df.residual
phi_male_poisson = sum(residuals(LC.poisson.GRC.M,
type="pearson")^2)/LC.poisson.GRC.M$df.residual
phi_female_poisson = sum(residuals(LC.poisson.GRC.F,
type="pearson")^2)/LC.poisson.GRC.F$df.residual

phi_total_neg_bin = sum(residuals(LC.neg.binom.GRC.T,
type="pearson")^2)/LC.neg.binom.GRC.T$df.residual
phi_male_neg_bin = sum(residuals(LC.neg.binom.GRC.M,
type="pearson")^2)/LC.neg.binom.GRC.M$df.residual
phi_female_neg_bin = sum(residuals(LC.neg.binom.GRC.F,
type="pearson")^2)/LC.neg.binom.GRC.F$df.residual

phi_total_binom = sum(residuals(LC.binom.GRC.T,
type="pearson")^2)/LC.binom.GRC.T$df.residual
phi_male_binom = sum(residuals(LC.binom.GRC.M,
type="pearson")^2)/LC.binom.GRC.M$df.residual
phi_female_binom = sum(residuals(LC.binom.GRC.F,
type="pearson")^2)/LC.binom.GRC.F$df.residual

deviances_GRC<-
matrix(c(LC.poisson.GRC.T$deviance,LC.neg.binom.GRC.T$deviance,LC.binom.GRC.T$de
viance,
LC.poisson.GRC.M$deviance,LC.neg.binom.GRC.M$deviance,LC.binom.GRC.M$deviance,
LC.poisson.GRC.F$deviance,LC.neg.binom.GRC.F$deviance,LC.binom.GRC.F$deviance),1,
9)
colnames(deviances_GRC)<-rep(c("Poisson","Negative Binomial","Binomial"),3)

AIC_GRC<-
matrix(c(AIC(LC.poisson.GRC.T),AIC(LC.neg.binom.GRC.T),AIC(LC.binom.GRC.T),
AIC(LC.poisson.GRC.M),AIC(LC.neg.binom.GRC.M),AIC(LC.binom.GRC.M),
AIC(LC.poisson.GRC.F),AIC(LC.neg.binom.GRC.F),AIC(LC.binom.GRC.F)),1,9)
colnames(AIC_GRC)<-rep(c("Poisson","Negative Binomial","Binomial"),3)

BIC_GRC<-
matrix(c(BIC(LC.poisson.GRC.T),BIC(LC.neg.binom.GRC.T),BIC(LC.binom.GRC.T),
BIC(LC.poisson.GRC.M),BIC(LC.neg.binom.GRC.M),BIC(LC.binom.GRC.M),
BIC(LC.poisson.GRC.F),BIC(LC.neg.binom.GRC.F),BIC(LC.binom.GRC.F)),1,9)
colnames(BIC_GRC)<-rep(c("Poisson","Negative Binomial","Binomial"),3)

```

```

mod_nul_poisson_males<-gnm(GRC.data.M$Dx_GRC_new.Male~1,
                             family=poisson(link="log"))
mod_nul_poisson_females<-gnm(GRC.data.F$Dx_GRC_new.Female~1,
                              family=poisson(link="log"))
mod_nul_poisson_total<-gnm(GRC.data.T$Dx_GRC_new.Total~-1,
                           family=poisson(link="log"))

theta_null<-theta.ml(GRC.data.M$Dx_GRC_new.Male,fitted(LC.poisson.GRC.M))
mod_nul_neg.bin_males<- gnm(GRC.data.M$Dx_GRC_new.Male~1,
                             family=negative.binomial(theta_null))

theta_null<-theta.ml(GRC.data.F$Dx_GRC_new.Female,fitted(LC.poisson.GRC.F))
mod_nul_neg.bin_females<- gnm(GRC.data.F$Dx_GRC_new.Female~1,
                              family=negative.binomial(theta_null))

theta_null<-theta.ml(GRC.data.T$Dx_GRC_new.Total,fitted(LC.poisson.GRC.T))
mod_nul_neg.bin_total<- gnm(GRC.data.T$Dx_GRC_new.Total~1,
                            family=negative.binomial(theta_null))

mod_nul_binom_males<-
gnm(GRC.data.M$Dx_GRC_new.Male/GRC.data.M$Pop_GRC_new.Male~1,
     family=binomial(link="logit"),
     weights=GRC.data.M$Pop_GRC_new.Male)

mod_nul_binom_females<-
gnm(GRC.data.F$Dx_GRC_new.Female/GRC.data.F$Pop_GRC_new.Female~1,
     family=binomial(link="logit"),
     weights=GRC.data.F$Pop_GRC_new.Female)

mod_nul_binom_total<-
gnm(GRC.data.T$Dx_GRC_new.Total/GRC.data.T$Pop_GRC_new.Total~1,
     family=binomial(link="logit"),
     weights=GRC.data.T$Pop_GRC_new.Total)

R2_mcf_poisson_m = 1- (logLik(LC.poisson.GRC.M)/logLik(mod_nul_poisson_males))
R2_mcf_poisson_f = 1- (logLik(LC.poisson.GRC.F)/logLik(mod_nul_poisson_females))
R2_mcf_poisson_t = 1- (logLik(LC.poisson.GRC.M)/logLik(mod_nul_poisson_males))

R2_mcf_neg.bin_m = 1- (logLik(LC.neg.binom.GRC.M)/logLik(mod_nul_neg.bin_males))
R2_mcf_neg.bin_f = 1- (logLik(LC.neg.binom.GRC.F)/logLik(mod_nul_neg.bin_females))
R2_mcf_neg.bin_t = 1- (logLik(LC.neg.binom.GRC.M)/logLik(mod_nul_neg.bin_males))

R2_mcf_binom_m = 1- (logLik(LC.binom.GRC.M)/logLik(mod_nul_binom_males))
R2_mcf_binom_f = 1- (logLik(LC.neg.binom.GRC.F)/logLik(mod_nul_binom_females))
R2_mcf_binom_t = 1- (logLik(LC.binom.GRC.M)/logLik(mod_nul_binom_males))

```


alpha_total_poisson=coefficients(LC.poisson.GRC.T)[1:86]
 alpha_male_poisson=coefficients(LC.poisson.GRC.M)[1:86]
 alpha_female_poisson=coefficients(LC.poisson.GRC.F)[1:86]

alpha_total_neg.binomial=coefficients(LC.neg.binom.GRC.T)[1:86]
 alpha_male_neg.binomial=coefficients(LC.neg.binom.GRC.M)[1:86]
 alpha_female_neg.binomial=coefficients(LC.neg.binom.GRC.F)[1:86]

alpha_total_binomial=coefficients(LC.binom.GRC.T)[1:86]
 alpha_male_binomial=coefficients(LC.binom.GRC.M)[1:86]
 alpha_female_binomial=coefficients(LC.binom.GRC.F)[1:86]

beta_total_poisson=coefficients(LC.poisson.GRC.T)[87:172]
 beta_male_poisson=coefficients(LC.poisson.GRC.M)[87:172]
 beta_female_poisson=coefficients(LC.poisson.GRC.F)[87:172]

beta_total_neg.binomial=coefficients(LC.neg.binom.GRC.T)[87:172]
 beta_male_neg.binomial=coefficients(LC.neg.binom.GRC.M)[87:172]
 beta_female_neg.binomial=coefficients(LC.neg.binom.GRC.F)[87:172]

beta_total_binomial=coefficients(LC.binom.GRC.T)[87:172]
 beta_male_binomial=coefficients(LC.binom.GRC.M)[87:172]
 beta_female_binomial=coefficients(LC.binom.GRC.F)[87:172]

kappa_total_poisson=coefficients(LC.poisson.GRC.T)[173:211]
 kappa_male_poisson=coefficients(LC.poisson.GRC.M)[173:211]
 kappa_female_poisson=coefficients(LC.poisson.GRC.F)[173:211]

kappa_total_neg.binomial=coefficients(LC.neg.binom.GRC.T)[173:211]
 kappa_male_neg.binomial=coefficients(LC.neg.binom.GRC.M)[173:211]
 kappa_female_neg.binomial=coefficients(LC.neg.binom.GRC.F)[173:211]

kappa_total_binomial=coefficients(LC.binom.GRC.T)[173:211]
 kappa_male_binomial=coefficients(LC.binom.GRC.M)[173:211]
 kappa_female_binomial=coefficients(LC.binom.GRC.F)[173:211]

Πίνακας Β.1: Εκτιμήσεις παραμέτρων α_x και β_x μέσω του κλασικού μοντέλου Lee – Carter.

Εκτιμήσεις κλασικού μοντέλου Lee - Carter						
Ηλικία	α_x			β_x		
	Άνδρες	Γυναίκες	Σύνολο	Άνδρες	Γυναίκες	Σύνολο
0	-5.0389	-5.2085	-5.1168	0.0308	0.0254	0.0292
1	-7.9908	-8.0398	-7.9982	0.0193	0.0153	0.0173
2	-8.3218	-8.4651	-8.3617	0.0253	0.0158	0.0211
3	-8.5056	-8.8103	-8.6046	0.0275	0.0220	0.0234

4	-8.6360	-8.8013	-8.7329	0.0263	0.0188	0.0251
5	-8.7016	-9.0026	-8.8044	0.0159	0.0165	0.0165
6	-8.7753	-9.0634	-8.8649	0.0276	0.0212	0.0241
7	-8.8628	-9.0812	-8.9607	0.0261	0.0184	0.0246
8	-8.9133	-9.2947	-9.0397	0.0253	0.0206	0.0219
9	-8.9500	-9.2218	-9.0082	0.0279	0.0191	0.0216
10	-8.9520	-9.1677	-9.0238	0.0264	0.0224	0.0250
11	-9.0760	-9.1269	-9.0635	0.0240	0.0110	0.0179
12	-8.8152	-9.2171	-8.9537	0.0259	0.0162	0.0209
13	-8.5971	-9.0679	-8.7688	0.0174	0.0125	0.0153
14	-8.2988	-8.9411	-8.5342	0.0186	0.0195	0.0179
15	-8.0274	-8.6473	-8.2571	0.0233	0.0131	0.0184
16	-7.6699	-8.5718	-7.9975	0.0174	0.0183	0.0171
17	-7.4094	-8.4773	-7.7823	0.0209	0.0175	0.0199
18	-7.1898	-8.2604	-7.5656	0.0196	0.0143	0.0178
19	-7.0879	-8.2462	-7.4875	0.0163	0.0132	0.0151
20	-6.9671	-8.2423	-7.3906	0.0159	0.0155	0.0153
21	-6.9459	-8.1127	-7.3476	0.0162	0.0123	0.0148
22	-6.8935	-8.1904	-7.3230	0.0138	0.0147	0.0134
23	-6.8745	-8.1075	-7.2928	0.0155	0.0113	0.0139
24	-6.8750	-8.1637	-7.3059	0.0141	0.0091	0.0122
25	-6.8758	-8.1388	-7.2997	0.0113	0.0170	0.0122
26	-6.9344	-8.1432	-7.3514	0.0100	0.0125	0.0105
27	-6.8992	-8.0096	-7.2932	0.0110	0.0098	0.0101
28	-6.9289	-8.0572	-7.3242	0.0111	0.0136	0.0115
29	-6.9087	-7.9785	-7.2954	0.0070	0.0126	0.0086
30	-6.8651	-7.8955	-7.2441	0.0094	0.0118	0.0101
31	-6.8771	-7.8894	-7.2530	0.0068	0.0087	0.0070
32	-6.8700	-7.7522	-7.2076	0.0054	0.0083	0.0066
33	-6.8180	-7.7076	-7.1614	0.0073	0.0063	0.0068
34	-6.7553	-7.6107	-7.0886	0.0054	0.0105	0.0071
35	-6.7313	-7.5508	-7.0528	0.0068	0.0111	0.0082
36	-6.6747	-7.5268	-7.0076	0.0073	0.0078	0.0076
37	-6.6268	-7.4597	-6.9541	0.0066	0.0086	0.0072
38	-6.5751	-7.3636	-6.8906	0.0051	0.0093	0.0069
39	-6.4850	-7.2560	-6.7976	0.0069	0.0068	0.0069
40	-6.3937	-7.1698	-6.7078	0.0077	0.0083	0.0080
41	-6.3245	-7.0641	-6.6255	0.0045	0.0080	0.0056
42	-6.2585	-6.9857	-6.5572	0.0058	0.0061	0.0060

43	-6.1178	-6.8418	-6.4175	0.0059	0.0058	0.0059
44	-6.0365	-6.7847	-6.3444	0.0052	0.0076	0.0061
45	-5.9207	-6.6740	-6.2320	0.0056	0.0054	0.0056
46	-5.8412	-6.5727	-6.1448	0.0046	0.0066	0.0055
47	-5.7561	-6.5008	-6.0650	0.0037	0.0048	0.0042
48	-5.6638	-6.4261	-5.9781	0.0032	0.0045	0.0039
49	-5.5508	-6.3169	-5.8671	0.0036	0.0048	0.0043
50	-5.4418	-6.1833	-5.7508	0.0039	0.0048	0.0045
51	-5.3559	-6.1320	-5.6762	0.0033	0.0036	0.0036
52	-5.2627	-6.0507	-5.5872	0.0026	0.0053	0.0038
53	-5.1667	-5.9728	-5.4978	0.0029	0.0049	0.0038
54	-5.0711	-5.8780	-5.4034	0.0027	0.0053	0.0039
55	-4.9708	-5.7773	-5.3036	0.0041	0.0063	0.0052
56	-4.8740	-5.6963	-5.2125	0.0040	0.0059	0.0049
57	-4.7756	-5.6159	-5.1209	0.0034	0.0060	0.0045
58	-4.6890	-5.5410	-5.0393	0.0048	0.0064	0.0054
59	-4.5970	-5.4421	-4.9456	0.0045	0.0072	0.0055
60	-4.5128	-5.3372	-4.8559	0.0055	0.0077	0.0064
61	-4.4218	-5.2646	-4.7717	0.0061	0.0084	0.0070
62	-4.3348	-5.1480	-4.6766	0.0068	0.0092	0.0078
63	-4.2539	-5.0665	-4.5970	0.0079	0.0104	0.0089
64	-4.1719	-4.9632	-4.5092	0.0081	0.0108	0.0091
65	-4.0747	-4.8413	-4.4052	0.0093	0.0117	0.0103
66	-3.9900	-4.7397	-4.3156	0.0085	0.0122	0.0101
67	-3.9012	-4.6251	-4.2194	0.0095	0.0130	0.0110
68	-3.8106	-4.5068	-4.1209	0.0100	0.0132	0.0115
69	-3.7035	-4.3817	-4.0096	0.0114	0.0138	0.0126
70	-3.6001	-4.2439	-3.8951	0.0121	0.0147	0.0135
71	-3.5189	-4.1205	-3.7985	0.0108	0.0149	0.0129
72	-3.4200	-3.9887	-3.6886	0.0115	0.0146	0.0132
73	-3.3238	-3.8686	-3.5841	0.0122	0.0151	0.0140
74	-3.2321	-3.7298	-3.4745	0.0124	0.0151	0.0142
75	-3.1255	-3.5954	-3.3575	0.0122	0.0157	0.0145
76	-3.0229	-3.4567	-3.2412	0.0127	0.0149	0.0145
77	-2.9218	-3.3046	-3.1182	0.0119	0.0141	0.0137
78	-2.8193	-3.1667	-3.0003	0.0111	0.0138	0.0132
79	-2.7004	-3.0317	-2.8760	0.0114	0.0125	0.0127
80	-2.5943	-2.8883	-2.7536	0.0113	0.0127	0.0128
81	-2.5024	-2.7657	-2.6477	0.0104	0.0108	0.0113

82	-2.3973	-2.6376	-2.5318	0.0096	0.0098	0.0103
83	-2.2968	-2.4982	-2.4114	0.0096	0.0089	0.0098
84	-2.1889	-2.3759	-2.2967	0.0087	0.0079	0.0088
85	-2.0892	-2.2545	-2.1856	0.0085	0.0075	0.0084

Πίνακας Β.2: Εκτιμήσεις παραμέτρου k_t μέσω του κλασικού μοντέλου Lee – Carter.

Έτος	Εκτιμήσεις παραμέτρου k_t		
	Άνδρες	Γυναίκες	Σύνολο
1981	24.122	32.239	25.722
1982	22.009	31.118	24.525
1983	24.243	31.552	25.99
1984	20.19	25.213	21.168
1985	21.349	24.602	21.879
1986	14.428	20.716	15.82
1987	18.549	22.83	19.426
1988	12.996	20.377	15.408
1989	15.492	21.448	17.185
1990	14.043	16.622	14.261
1991	10.78	8.046	9.736
1992	10.071	14.609	11.802
1993	8.349	13.276	10.514
1994	9.037	9.467	9.646
1995	8.203	8.904	8.504
1996	9.759	10.68	10.562
1997	9.343	9.439	8.965
1998	3.542	7.325	4.72
1999	5.143	5.346	6.04
2000	-1.185	-1.139	-0.122
2001	-0.287	-4.308	-1.014
2002	-3.67	-4.879	-3.848
2003	-5.157	-8.67	-6.134
2004	-0.499	-6.189	-3.04
2005	-2.082	-1.343	-1.59
2006	-10.156	-15.401	-12.467
2007	-7.315	-14.91	-8.73
2008	-12.454	-18.508	-12.737
2009	-7.749	-13.544	-9.802
2010	-12.824	-14.862	-12.662
2011	-10.377	-16.279	-12.793

2012	-18.324	-20.566	-19.518
2013	-23.406	-23.527	-23.046
2014	-19.373	-28.809	-23.378
2015	-30.053	-28.038	-27.426
2016	-19.702	-21.59	-21.558
2017	-22.374	-29.316	-24.061
2018	-25.002	-33.189	-28.058
2019	-29.661	-28.742	-29.888

Πίνακας Β.3: Εκτιμήσεις αριθμού θανάτων ανδρών ηλικίας 65 ετών.

Έτη	Εκτιμήσεις αριθμού θανάτων ανδρών ηλικίας 65 ετών			
	Παρατηρούμενος αριθμός	Poisson	Αρνητική Διωνυμική	Διωνυμική
1981	1018.05	827.38	831.59	886.36
1982	1083.02	753.2	755.51	768.56
1983	960.15	727.67	731.94	752.34
1984	831	681.55	684.78	688.57
1985	851.01	763.95	766.79	692.35
1986	764	814.7	812.43	811.56
1987	952	820.88	820.32	828.05
1988	900.04	790.98	789.23	796.22
1989	903.01	812.44	816.46	773.53
1990	867	942.08	941.49	851.99
1991	959	1075.16	1079.8	1030.27
1992	1015	1092.51	1091.02	1129.86
1993	1198	1080.02	1077.39	1046.39
1994	1173	1088.7	1092	1105.23
1995	1232	1104.41	1106.66	1089.54
1996	1272	1106.62	1113.35	1114.57
1997	1224	1058.01	1063.68	1080.33
1998	1303	1040.08	1040.97	1039.81

1999	1153	1047.3	1053.26	1045.13
2000	1087	1048.09	1045.02	1062.44
2001	1186	990.14	994.36	1004.89
2002	1187	974.67	968.36	984.89
2003	1083	953.33	945.4	970.54
2004	1065	919.54	918.03	940.02
2005	1027	876.73	876.67	909.77
2006	934	799.77	800.42	833.7
2007	994	769.4	773.09	790.21
2008	869	734.38	734.59	731.23
2009	719	764	773.26	739.76
2010	684	801.74	810.08	774.74
2011	887	843.73	850.07	833.65
2012	978	865.52	863.56	867.2
2013	1044	816.72	817.34	820.13
2014	1085	808.07	812.2	806.84
2015	987	836.38	832.58	829.81
2016	820	819.53	821.91	819.26
2017	926	819.64	812.7	822.36
2018	942	777.61	776.7	785.51
2019	1005	788.63	782.79	777.63

Πίνακας Β.4: Εκτιμήσεις αριθμού θανάτων γυναικών ηλικίας 65 ετών.

Έτη	Εκτιμήσεις αριθμού θανάτων γυναικών ηλικίας 65 ετών			
	Παρατηρούμενος αριθμός	Poisson Lee - Carter	Αρνητικό Διωνομικό Lee - Carter	Διωνομικό Lee - Carter
1981	567.04	511.3	526.21	541.37
1982	481	455.38	466.93	475.69
1983	557.09	462.88	472.88	450.06

1984	478	420.01	429.34	459.24
1985	530	456.02	463.17	391.84
1986	453	470	474.18	505.02
1987	515	489.85	494.27	446.66
1988	467.01	468.78	472.21	505.57
1989	475	464.85	469.63	425.86
1990	508	556.55	558.98	500.56
1991	590	608.99	607.84	606.95
1992	533	592.04	592.94	615.1
1993	575	584.99	583.52	560.14
1994	571	578.22	576.81	597.74
1995	536	573.41	571.01	562.16
1996	531	562.66	561.83	576.74
1997	534	536.42	537.11	533.25
1998	525	546.36	546.27	548.5
1999	497	537.37	536.75	536.73
2000	486	538.86	531.12	540.09
2001	524	516.65	509.72	522.31
2002	439	504.21	496.26	510.89
2003	429	490.45	480.61	498.86
2004	460	467.29	463.53	479.36
2005	409	430.91	426.93	450.24
2006	380	386.45	382.17	406.54
2007	319	366.51	362.27	377.68
2008	302	347.68	341.47	345.15
2009	341	357.1	354.32	343.11
2010	398	383.66	384.22	368.32
2011	425	399.04	398.9	392.38
2012	469	414.1	409.78	413.39
2013	408	388.3	389.16	389.73
2014	341	371.22	374.51	371.39
2015	360	383.35	384.5	381.66
2016	442	373.89	382.39	373.42
2017	407	372.49	379.18	372.36
2018	419	350.1	359.45	350.97
2019	396	358.81	368.37	350.13

Πίνακας Β.5: Εκτιμήσεις παραμέτρου α_x για τις γυναίκες.

Ηλικία	Εκτιμήσεις παραμέτρου α_x			
	LC model	Poisson	Negative Binomial	Binomial
1	-5.20852	-5.24635	-5.15332	-5.25178
2	-8.03983	-8.03393	-7.98276	-8.04262
3	-8.46507	-8.43016	-8.378	-8.43828
4	-8.81032	-8.70701	-8.65113	-8.71414
5	-8.80128	-8.77417	-8.7218	-8.78281
6	-9.00255	-8.92103	-8.87335	-8.9294
7	-9.06344	-8.98382	-8.92744	-8.99228
8	-9.08117	-9.06678	-9.00767	-9.07556
9	-9.29467	-9.19038	-9.1352	-9.19872
10	-9.22184	-9.12128	-9.06847	-9.12887
11	-9.16765	-9.11574	-9.05565	-9.12379
12	-9.12693	-9.0836	-9.0485	-9.08996
13	-9.21711	-9.13462	-9.09344	-9.14247
14	-9.06786	-9.01015	-8.97436	-9.01751
15	-8.94106	-8.8718	-8.82014	-8.88056
16	-8.64727	-8.6057	-8.56545	-8.61415
17	-8.57177	-8.54027	-8.48824	-8.54825
18	-8.47731	-8.44483	-8.39261	-8.45295
19	-8.26037	-8.24508	-8.20141	-8.25254
20	-8.2462	-8.225	-8.18483	-8.23225
21	-8.24225	-8.21875	-8.17459	-8.22621
22	-8.11266	-8.10327	-8.06516	-8.11017
23	-8.19035	-8.17168	-8.12703	-8.17897
24	-8.10751	-8.09948	-8.06357	-8.1059
25	-8.16364	-8.13672	-8.10973	-8.14282
26	-8.13877	-8.10149	-8.05408	-8.10851
27	-8.14322	-8.13297	-8.09352	-8.13912
28	-8.00959	-8.00144	-7.97229	-8.00566
29	-8.05717	-8.02855	-7.98843	-8.03387
30	-7.97846	-7.97336	-7.93411	-7.97755
31	-7.89548	-7.88783	-7.85337	-7.89102
32	-7.88942	-7.88531	-7.85778	-7.88675
33	-7.7522	-7.75093	-7.72318	-7.75416
34	-7.70755	-7.70376	-7.68353	-7.70601

35	-7.6107	-7.61243	-7.57934	-7.61517
36	-7.5508	-7.54917	-7.51461	-7.55063
37	-7.52684	-7.52322	-7.49619	-7.52179
38	-7.45968	-7.45645	-7.42935	-7.45285
39	-7.36355	-7.35859	-7.33025	-7.35476
40	-7.25597	-7.25159	-7.23063	-7.24764
41	-7.16976	-7.17585	-7.14836	-7.17465
42	-7.0641	-7.06765	-7.04235	-7.06484
43	-6.98572	-6.98247	-6.96176	-6.9804
44	-6.84176	-6.84266	-6.82498	-6.83937
45	-6.78468	-6.78787	-6.76412	-6.7864
46	-6.67397	-6.67472	-6.65792	-6.67212
47	-6.57267	-6.57559	-6.55469	-6.57301
48	-6.50083	-6.50217	-6.48709	-6.49852
49	-6.4261	-6.42637	-6.41175	-6.42181
50	-6.31685	-6.31704	-6.30177	-6.31166
51	-6.18326	-6.18381	-6.16871	-6.17997
52	-6.13197	-6.131	-6.11961	-6.12472
53	-6.0507	-6.04967	-6.03383	-6.04449
54	-5.97281	-5.97396	-5.9581	-5.96837
55	-5.87804	-5.87996	-5.86402	-5.87538
56	-5.77726	-5.7794	-5.75939	-5.7737
57	-5.69627	-5.69643	-5.67864	-5.68861
58	-5.61591	-5.61934	-5.6006	-5.61031
59	-5.54103	-5.54351	-5.52408	-5.5347
60	-5.44207	-5.44581	-5.42335	-5.43512
61	-5.33717	-5.3437	-5.31881	-5.33448
62	-5.26457	-5.27222	-5.24455	-5.25983
63	-5.14797	-5.15943	-5.1279	-5.14887
64	-5.06648	-5.07811	-5.04325	-5.06627
65	-4.96324	-4.97456	-4.93845	-4.9635
66	-4.84125	-4.85619	-4.81628	-4.84603
67	-4.73971	-4.75329	-4.71182	-4.7417
68	-4.62514	-4.64098	-4.59741	-4.62828
69	-4.50676	-4.52132	-4.47773	-4.50998
70	-4.38171	-4.39798	-4.35236	-4.38246
71	-4.24386	-4.25931	-4.21135	-4.24475
72	-4.12052	-4.13734	-4.08806	-4.11578

73	-3.98869	-4.00387	-3.95693	-3.98049
74	-3.86859	-3.8864	-3.83645	-3.85969
75	-3.72979	-3.74713	-3.69785	-3.71865
76	-3.59538	-3.6111	-3.5613	-3.58017
77	-3.45665	-3.47095	-3.42429	-3.43338
78	-3.30455	-3.32012	-3.27421	-3.27604
79	-3.16671	-3.18193	-3.1373	-3.13116
80	-3.03165	-3.04557	-3.00481	-2.98881
81	-2.8883	-2.90196	-2.86126	-2.83626
82	-2.76566	-2.77725	-2.74256	-2.70032
83	-2.63755	-2.64722	-2.61594	-2.55783
84	-2.49815	-2.50738	-2.47854	-2.40564
85	-2.37592	-2.38368	-2.35842	-2.26982
86	-2.25447	-2.26146	-2.23775	-2.13358

Πίνακας Β.6: Εκτιμήσεις παραμέτρου α_x για τους άνδρες

Ηλικία	Εκτιμήσεις παραμέτρου α_x			
	Κλασικό μοντέλο Lee-Carter	Poisson Lee- Carter	Αρνητικό Διωνυμικό Lee- Carter	Διωνυμικό Lee- Carter
1	-5.03893	-5.07698	-4.98041	-5.08719
2	-7.99082	-7.96003	-7.91076	-7.97162
3	-8.32182	-8.30215	-8.23564	-8.31539
4	-8.50557	-8.49836	-8.42846	-8.5111
5	-8.63597	-8.60863	-8.54298	-8.62107
6	-8.70157	-8.6662	-8.62831	-8.67652
7	-8.77529	-8.73818	-8.67703	-8.7521
8	-8.86279	-8.8345	-8.77158	-8.84807
9	-8.91329	-8.8741	-8.81629	-8.88621
10	-8.95002	-8.89351	-8.83267	-8.90565
11	-8.95196	-8.89613	-8.83629	-8.90835
12	-9.07601	-9.00732	-8.94059	-9.01901
13	-8.81522	-8.78163	-8.71997	-8.79503
14	-8.59705	-8.57257	-8.5265	-8.58351
15	-8.29874	-8.29312	-8.24518	-8.3048
16	-8.02734	-8.00052	-7.95202	-8.01276
17	-7.66992	-7.66836	-7.62782	-7.67792
18	-7.4094	-7.40335	-7.35565	-7.41426
19	-7.18977	-7.17401	-7.13166	-7.18459
20	-7.08786	-7.07748	-7.04208	-7.08676
21	-6.9671	-6.96158	-6.92654	-6.97075
22	-6.94586	-6.93671	-6.90258	-6.9454
23	-6.89345	-6.87906	-6.84972	-6.88632
24	-6.87445	-6.85687	-6.82659	-6.86424
25	-6.87499	-6.85295	-6.82807	-6.85976
26	-6.8758	-6.85026	-6.82996	-6.85569
27	-6.93442	-6.90856	-6.89169	-6.91385
28	-6.89914	-6.8701	-6.85184	-6.87458
29	-6.92887	-6.9024	-6.8827	-6.90704
30	-6.90865	-6.88677	-6.87359	-6.89034
31	-6.86514	-6.84781	-6.8301	-6.85115
32	-6.87714	-6.86538	-6.85199	-6.86662
33	-6.87004	-6.8509	-6.8387	-6.85321

34	-6.81803	-6.81202	-6.79549	-6.81449
35	-6.75528	-6.74704	-6.73489	-6.7485
36	-6.73129	-6.72359	-6.70809	-6.72396
37	-6.67472	-6.67124	-6.65361	-6.66924
38	-6.62675	-6.62159	-6.60543	-6.61785
39	-6.57509	-6.56592	-6.55263	-6.56102
40	-6.48497	-6.48479	-6.46752	-6.48046
41	-6.39368	-6.39209	-6.37291	-6.38987
42	-6.32451	-6.31756	-6.30841	-6.31319
43	-6.25848	-6.2519	-6.23774	-6.24855
44	-6.11779	-6.11713	-6.10233	-6.11298
45	-6.03651	-6.03225	-6.02049	-6.02882
46	-5.92071	-5.92052	-5.90723	-5.91697
47	-5.84122	-5.83936	-5.82835	-5.83547
48	-5.7561	-5.75674	-5.74765	-5.75197
49	-5.66374	-5.66356	-5.65468	-5.65804
50	-5.55078	-5.55036	-5.54103	-5.54464
51	-5.44176	-5.44496	-5.43497	-5.43908
52	-5.35586	-5.35649	-5.34836	-5.34897
53	-5.26271	-5.26311	-5.25638	-5.25562
54	-5.16668	-5.16818	-5.161	-5.15993
55	-5.07111	-5.0722	-5.06534	-5.06402
56	-4.9708	-4.97283	-4.96251	-4.96373
57	-4.87397	-4.8767	-4.86666	-4.86558
58	-4.77562	-4.77731	-4.76884	-4.7643
59	-4.68895	-4.69221	-4.67951	-4.67836
60	-4.59699	-4.59909	-4.58777	-4.5844
61	-4.51283	-4.51728	-4.50257	-4.50223
62	-4.42182	-4.42724	-4.4104	-4.40926
63	-4.33483	-4.34127	-4.3226	-4.32307
64	-4.25386	-4.26036	-4.23952	-4.24147
65	-4.17185	-4.17939	-4.1577	-4.15992
66	-4.0747	-4.08385	-4.05913	-4.06514
67	-3.98995	-3.99723	-3.97448	-3.97656
68	-3.90119	-3.90817	-3.88357	-3.88586
69	-3.81056	-3.81921	-3.79211	-3.7959
70	-3.7035	-3.71411	-3.68404	-3.687
71	-3.60008	-3.61176	-3.57977	-3.58302

72	-3.51886	-3.52845	-3.50029	-3.49531
73	-3.42004	-3.42889	-3.40004	-3.3925
74	-3.32376	-3.33457	-3.30328	-3.29411
75	-3.23207	-3.2411	-3.21069	-3.19695
76	-3.12553	-3.13515	-3.10452	-3.08676
77	-3.02291	-3.03175	-3.00138	-2.97824
78	-2.92176	-2.93	-2.90148	-2.8698
79	-2.81927	-2.82829	-2.80076	-2.76047
80	-2.70037	-2.71071	-2.68152	-2.63371
81	-2.59434	-2.60528	-2.57548	-2.51819
82	-2.50239	-2.51192	-2.48449	-2.41495
83	-2.39734	-2.40594	-2.3807	-2.29739
84	-2.29684	-2.30685	-2.28083	-2.18647
85	-2.1889	-2.19793	-2.17406	-2.0636
86	-2.08919	-2.09758	-2.07391	-1.94845

Πίνακας Β.7: Εκτιμήσεις παραμέτρου β_x για τους άνδρες

Ηλικία	Εκτιμήσεις παραμέτρου β_x			
	Κλασικό Lee - Carter	Poisson Lee - Carter	Αρνητικό Διωνυμικό Lee - Carter	Διωνυμικό Lee - Carter
0	0.031	1.107	-0.753	-0.581
1	0.019	0.612	-0.446	-0.321
2	0.025	0.805	-0.585	-0.425
3	0.027	0.852	-0.622	-0.451
4	0.026	0.807	-0.59	-0.426
5	0.016	0.456	-0.332	-0.241
6	0.028	0.734	-0.534	-0.392
7	0.026	0.772	-0.566	-0.414
8	0.025	0.703	-0.513	-0.375
9	0.028	0.745	-0.546	-0.398
10	0.026	0.735	-0.54	-0.392
11	0.024	0.799	-0.58	-0.423
12	0.026	0.756	-0.554	-0.403
13	0.017	0.554	-0.403	-0.296
14	0.019	0.582	-0.426	-0.313
15	0.023	0.588	-0.43	-0.319
16	0.017	0.495	-0.363	-0.269
17	0.021	0.583	-0.429	-0.32

18	0.02	0.519	-0.383	-0.287
19	0.016	0.435	-0.321	-0.24
20	0.016	0.435	-0.321	-0.239
21	0.016	0.424	-0.315	-0.236
22	0.014	0.366	-0.272	-0.206
23	0.016	0.385	-0.287	-0.217
24	0.014	0.321	-0.241	-0.184
25	0.011	0.272	-0.204	-0.158
26	0.01	0.227	-0.172	-0.136
27	0.011	0.246	-0.188	-0.147
28	0.011	0.263	-0.199	-0.154
29	0.007	0.177	-0.133	-0.108
30	0.009	0.231	-0.174	-0.135
31	0.007	0.169	-0.128	-0.104
32	0.005	0.156	-0.116	-0.096
33	0.007	0.206	-0.153	-0.12
34	0.005	0.155	-0.115	-0.092
35	0.007	0.193	-0.142	-0.112
36	0.007	0.217	-0.159	-0.125
37	0.007	0.202	-0.148	-0.118
38	0.005	0.17	-0.124	-0.1
39	0.007	0.214	-0.157	-0.122
40	0.008	0.237	-0.174	-0.129
41	0.004	0.118	-0.086	-0.068
42	0.006	0.182	-0.133	-0.1
43	0.006	0.184	-0.134	-0.101
44	0.005	0.148	-0.109	-0.08
45	0.006	0.158	-0.117	-0.085
46	0.005	0.133	-0.098	-0.071
47	0.004	0.105	-0.077	-0.057
48	0.003	0.107	-0.077	-0.056
49	0.004	0.11	-0.081	-0.057
50	0.004	0.116	-0.084	-0.058
51	0.003	0.098	-0.071	-0.05
52	0.003	0.083	-0.06	-0.041
53	0.003	0.084	-0.061	-0.043
54	0.003	0.082	-0.06	-0.041
55	0.004	0.124	-0.091	-0.067
56	0.004	0.12	-0.088	-0.067
57	0.003	0.098	-0.072	-0.056

58	0.005	0.142	-0.105	-0.079
59	0.004	0.135	-0.1	-0.074
60	0.005	0.167	-0.123	-0.091
61	0.006	0.188	-0.139	-0.107
62	0.007	0.207	-0.153	-0.117
63	0.008	0.241	-0.177	-0.133
64	0.008	0.243	-0.179	-0.133
65	0.009	0.281	-0.206	-0.153
66	0.009	0.267	-0.194	-0.146
67	0.009	0.294	-0.214	-0.16
68	0.01	0.323	-0.234	-0.174
69	0.011	0.353	-0.257	-0.193
70	0.012	0.371	-0.271	-0.203
71	0.011	0.336	-0.245	-0.186
72	0.011	0.355	-0.259	-0.197
73	0.012	0.376	-0.274	-0.211
74	0.012	0.382	-0.28	-0.213
75	0.012	0.378	-0.277	-0.215
76	0.013	0.381	-0.28	-0.218
77	0.012	0.361	-0.264	-0.208
78	0.011	0.342	-0.249	-0.199
79	0.011	0.352	-0.256	-0.205
80	0.011	0.354	-0.257	-0.205
81	0.01	0.331	-0.238	-0.193
82	0.01	0.307	-0.22	-0.18
83	0.01	0.301	-0.217	-0.178
84	0.009	0.28	-0.2	-0.167
85	0.009	0.281	-0.196	-0.169

Πίνακας Β.8: Εκτιμήσεις παραμέτρου β_x για τις γυναίκες

Ηλικία	Εκτιμήσεις παραμέτρου β_x			
	Κλασικό Lee - Carter	Poisson Lee - Carter	Αρνητικό Διωνυμικό Lee - Carter	Διωνυμικό Lee - Carter
0	0.025	1.107	-0.833	0.334
1	0.015	0.612	-0.514	0.198
2	0.016	0.805	-0.524	0.202
3	0.022	0.852	-0.557	0.215
4	0.019	0.807	-0.506	0.197
5	0.017	0.456	-0.502	0.191

6	0.021	0.734	-0.55	0.214
7	0.018	0.772	-0.581	0.228
8	0.021	0.703	-0.544	0.213
9	0.019	0.745	-0.526	0.204
10	0.022	0.735	-0.568	0.224
11	0.011	0.799	-0.34	0.132
12	0.016	0.756	-0.384	0.152
13	0.013	0.554	-0.371	0.141
14	0.02	0.582	-0.496	0.196
15	0.013	0.588	-0.384	0.154
16	0.018	0.495	-0.495	0.197
17	0.018	0.583	-0.486	0.196
18	0.014	0.519	-0.407	0.164
19	0.013	0.435	-0.369	0.151
20	0.015	0.435	-0.416	0.166
21	0.012	0.424	-0.359	0.144
22	0.015	0.366	-0.41	0.167
23	0.011	0.385	-0.334	0.136
24	0.009	0.321	-0.252	0.104
25	0.017	0.272	-0.43	0.178
26	0.013	0.227	-0.364	0.15
27	0.01	0.246	-0.275	0.114
28	0.014	0.263	-0.373	0.154
29	0.013	0.177	-0.367	0.15
30	0.012	0.231	-0.327	0.132
31	0.009	0.169	-0.259	0.107
32	0.008	0.156	-0.261	0.108
33	0.006	0.206	-0.191	0.079
34	0.011	0.155	-0.313	0.126
35	0.011	0.193	-0.332	0.133
36	0.008	0.217	-0.26	0.106
37	0.009	0.202	-0.26	0.106
38	0.009	0.17	-0.273	0.112
39	0.007	0.214	-0.202	0.083
40	0.008	0.237	-0.263	0.101
41	0.008	0.118	-0.243	0.096
42	0.006	0.182	-0.202	0.08
43	0.006	0.184	-0.17	0.067
44	0.008	0.148	-0.23	0.087
45	0.005	0.158	-0.163	0.064

46	0.007	0.133	-0.204	0.077
47	0.005	0.105	-0.149	0.058
48	0.005	0.107	-0.143	0.055
49	0.005	0.11	-0.154	0.059
50	0.005	0.116	-0.152	0.054
51	0.004	0.098	-0.118	0.043
52	0.005	0.083	-0.166	0.06
53	0.005	0.084	-0.163	0.061
54	0.005	0.082	-0.163	0.06
55	0.006	0.124	-0.203	0.078
56	0.006	0.12	-0.189	0.074
57	0.006	0.098	-0.188	0.073
58	0.006	0.142	-0.196	0.075
59	0.007	0.135	-0.228	0.089
60	0.008	0.167	-0.244	0.092
61	0.008	0.188	-0.266	0.104
62	0.009	0.207	-0.293	0.113
63	0.01	0.241	-0.33	0.13
64	0.011	0.243	-0.334	0.128
65	0.012	0.281	-0.363	0.141
66	0.012	0.267	-0.387	0.152
67	0.013	0.294	-0.413	0.163
68	0.013	0.323	-0.416	0.161
69	0.014	0.353	-0.434	0.171
70	0.015	0.371	-0.464	0.178
71	0.015	0.336	-0.473	0.19
72	0.015	0.355	-0.457	0.18
73	0.015	0.376	-0.476	0.192
74	0.015	0.382	-0.473	0.187
75	0.016	0.378	-0.49	0.196
76	0.015	0.381	-0.465	0.189
77	0.014	0.361	-0.444	0.184
78	0.014	0.342	-0.433	0.18
79	0.012	0.352	-0.391	0.161
80	0.013	0.354	-0.398	0.164
81	0.011	0.331	-0.338	0.141
82	0.01	0.307	-0.312	0.133
83	0.009	0.301	-0.285	0.123
84	0.008	0.28	-0.252	0.109
85	0.007	0.281	-0.238	0.103

Πίνακας Β.9: Εκτιμήσεις παραμέτρου k_t για τους άνδρες

Έτος	Εκτιμήσεις παραμέτρου k_t			
	Κλασικό Lee - Carter	Poisson Lee - Carter	Αρνητικό Διωνυμικό Lee - Carter	Διωνυμικό Lee - Carter
1	24.122	0.761	-0.932	-1.411
2	22.009	0.747	-0.904	-1.397
3	24.243	0.787	-0.972	-1.489
4	20.19	0.681	-0.822	-1.278
5	21.349	0.754	-0.917	-1.406
6	14.428	0.635	-0.722	-1.186
7	18.549	0.661	-0.768	-1.232
8	12.996	0.563	-0.626	-1.056
9	15.492	0.488	-0.559	-0.903
10	14.043	0.474	-0.512	-0.885
11	10.78	0.444	-0.495	-0.835
12	10.071	0.454	-0.482	-0.841
13	8.349	0.399	-0.402	-0.758
14	9.037	0.338	-0.344	-0.643
15	8.203	0.365	-0.376	-0.681
16	9.759	0.321	-0.337	-0.611
17	9.343	0.228	-0.206	-0.433
18	3.542	0.21	-0.159	-0.398
19	5.143	0.193	-0.16	-0.37
20	-1.185	0.196	-0.122	-0.388
21	-0.287	0.042	0.053	-0.109
22	-3.67	0.015	0.142	-0.064
23	-5.157	-0.031	0.214	0.015
24	-0.499	-0.093	0.266	0.124
25	-2.082	-0.131	0.311	0.197
26	-10.156	-0.253	0.474	0.428
27	-7.315	-0.22	0.409	0.367
28	-12.454	-0.384	0.654	0.667
29	-7.749	-0.451	0.688	0.798
30	-12.824	-0.571	0.861	1.049
31	-10.377	-0.58	0.888	1.095
32	-18.324	-0.53	0.866	1.023
33	-23.406	-0.706	1.092	1.355
34	-19.373	-0.723	1.094	1.39

35	-30.053	-0.612	0.989	1.198
36	-19.702	-0.694	1.066	1.332
37	-22.374	-0.676	1.096	1.295
38	-25.002	-0.815	1.25	1.546
39	-29.661	-0.78	1.232	1.486

Πίνακας Β.10: Εκτιμήσεις παραμέτρου k_t για τις γυναίκες

Έτος	Εκτιμήσεις παραμέτρου k_t			
	Κλασικό Lee - Carter	Poisson Lee - Carter	Αρνητικό Διωνυμικό Lee - Carter	Διωνυμικό Lee - Carter
1	32.236	1.537	-0.917	2.469
2	31.116	1.455	-0.856	2.348
3	31.551	1.547	-0.902	2.529
4	25.211	1.324	-0.767	2.154
5	24.601	1.412	-0.803	2.259
6	20.714	1.272	-0.699	2.028
7	22.829	1.367	-0.758	2.211
8	20.374	1.095	-0.585	1.77
9	21.448	0.992	-0.53	1.604
10	16.621	0.934	-0.478	1.502
11	8.045	0.857	-0.413	1.398
12	14.606	0.868	-0.429	1.364
13	13.273	0.763	-0.354	1.274
14	9.463	0.616	-0.263	1.002
15	8.905	0.609	-0.254	0.981
16	10.677	0.517	-0.205	0.84
17	9.438	0.355	-0.112	0.579
18	7.322	0.399	-0.136	0.641
19	5.341	0.305	-0.075	0.505
20	-1.142	0.271	-0.017	0.466
21	-4.307	0.078	0.1	0.161
22	-4.88	-0.003	0.156	0.05
23	-8.673	-0.087	0.22	-0.077
24	-6.192	-0.217	0.266	-0.281
25	-1.346	-0.373	0.366	-0.533
26	-15.313	-0.526	0.465	-0.794
27	-14.913	-0.481	0.439	-0.716
28	-18.511	-0.701	0.592	-1.072
29	-13.546	-0.872	0.669	-1.357

30	-14.867	-0.96	0.698	-1.522
31	-16.279	-1.058	0.763	-1.706
32	-20.567	-0.97	0.737	-1.585
33	-23.532	-1.232	0.864	-2.004
34	-28.816	-1.401	0.95	-2.27
35	-28.037	-1.258	0.877	-2.051
36	-21.593	-1.377	0.897	-2.233
37	-29.318	-1.381	0.913	-2.251
38	-33.193	-1.622	1.038	-2.636
39	-28.744	-1.581	1.013	-2.575

ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΕΙΡΑΙΩΣ

