



ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ
ΣΧΟΛΗ ΕΦΑΡΜΟΣΜΕΝΩΝ ΜΑΘΗΜΑΤΙΚΩΝ ΚΑΙ
ΦΥΣΙΚΩΝ ΕΠΙΣΤΗΜΩΝ
ΤΟΜΕΑΣ ΜΑΘΗΜΑΤΙΚΩΝ

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

ΤΟ ΑΝΑΛΟΓΙΣΤΙΚΟ ΜΟΝΤΕΛΟ LEE-CARTER

ΟΝΟΜΑΤΕΠΩΝΥΜΟ : ΑΝΑΣΤΑΣΙΑΔΗΣ ΚΩΣΤΑΣ

ΑΘΗΝΑ, ΝΟΕΜΒΡΙΟΣ 2011

Εισηγήτρια και επιβλέπων καθηγήτρια

Καρώνη Χρυσής

Αναπληρώτρια Καθηγήτρια σχολής Ε.Μ.Φ.Ε

Εθνικό Μετσόβιο Πολυτεχνείο

Εξεταστική Επιτροπή

<p>Καρώνη Χ. (επιβλέπων) Αναπληρώτρια Καθηγήτρια Σχολής Εφαρμοσμένων Μαθηματικών και Φυσικών Επιστημών Εθνικό Μετσόβιο Πολυτεχνείο Ηρώων Πολυτεχνείου 9, ΤΚ 15780, Πολυτεχνειούπολη Ζωγράφου, Αθήνα, Ελλάδα</p>	<p>Κουκουβίνος Χ. (μέλος) Καθηγητής Σχολής Εφαρμοσμένων Μαθηματικών και Φυσικών Επιστημών Εθνικό Μετσόβιο Πολυτεχνείο Ηρώων Πολυτεχνείου 9, ΤΚ 15780, Πολυτεχνειούπολη Ζωγράφου, Αθήνα, Ελλάδα</p>	<p>Λουλάκης Μ. (μέλος) Επίκουρος Καθηγητής Σχολής Εφαρμοσμένων Μαθηματικών και Φυσικών Επιστημών Εθνικό Μετσόβιο Πολυτεχνείο Ηρώων Πολυτεχνείου 9, ΣΚ 15780, Πολυτεχνειούπολη Ζωγράφου, Αθήνα, Ελλάδα</p>
--	---	--

Πρόλογος

Το αναλογιστικό μοντέλο Lee-Carter αποτελεί μια σύγχρονη μέθοδο πρόβλεψης ρυθμού θνησιμότητας και έχει εφαρμοστεί σε διάφορους πληθυσμούς. Χρησιμοποιείται σε αναλογιστικές μελέτες για να ερευνηθεί πως επηρεάζει η κοινωνική και οικονομική κατάσταση ενός πληθυσμού το ρυθμό θνησιμότητας ενώ ακόμη εφαρμόζεται σε διάφορες μελέτες ασφαλιστικών εταιρειών για προβλέψεις ατυχημάτων. Αυτή η διπλωματική εργασία παρουσιάζει τα βασικά στοιχεία που χαρακτηρίζουν το μοντέλο καθώς και εφαρμογές του για προβλέψεις ρυθμών θνησιμότητας.

Στο πρώτο κεφάλαιο παρουσιάζονται τα διάφορα μοντέλα πληθυσμών που εφαρμόζεται το μοντέλο, ανάλυση των τάσεων θνησιμότητας σε πληθυσμούς καθώς και τη μορφή των δεδομένων που επεξεργαζόμαστε με το μοντέλο Lee-Carter. Στο δεύτερο κεφάλαιο αναφέρεται η θεωρία των γενικευμένων γραμμικών μοντέλων και των μοντέλων παλινδρόμησης *Poisson* και αρνητικής Διωνυμικής. Η αναφορά αυτή γίνεται για τη καλύτερη κατανόηση της λειτουργίας του μοντέλου το οποίο εντάσσεται στο πλαίσιο λειτουργίας των γενικευμένων γραμμικών μοντέλων αλλά δεν αποτελεί γενικευμένο γραμμικό μοντέλο.

Στο επόμενο κεφάλαιο αναφέρεται εκτενώς η δομή του μοντέλου Lee-Carter, οι παράμετροι που το χαρακτηρίζουν καθώς και τα διάφορα σφάλματα που προκύπτουν από την εφαρμογή του. Ακόμη αναφέρεται η μεθοδολογία πρόβλεψης με τη χρήση χρονοσειρών και μοντέλων ARIMA και εφαρμογή του μοντέλου σε ομάδα πληθυσμού για πρόβλεψη του ρυθμού θνησιμότητας. Στο τέταρτο κεφάλαιο αναφέρονται οι διάφορες επεκτάσεις που μπορούν να γίνουν στο μοντέλο Lee-Carter με ιδιαίτερη αναφορά στα μοντέλα της αρνητικής Διωνυμικής και της Διωνυμικής. Στο τελευταίο κεφάλαιο παρουσιάζεται μια ολοκληρωμένη εφαρμογή του μοντέλου σε δεδομένα από ασφαλιστική εταιρεία για προβλέψεις με το ρυθμό ατυχημάτων την επόμενη χρόνια.

Θα ήθελα να ευχαριστήσω θερμά την Αναπληρώτρια Καθηγήτρια κυρία Χρυσήδα Καρώνη για τη συνεχή ενθάρρυνση της καθώς και την αμέριστη βοήθεια που μου πρόσφερε για την εκπόνηση της διπλωματικής εργασίας. Τέλος ευχαριστώ την οικογένεια μου για την υποστήριξη και την πίστη που δείχνει σε εμένα.

Περιεχόμενα

1	Εισαγωγή	5
1.1	Μοντέλα πληθυσμών	5
1.2	Τάσεις θνησιμότητας	5
1.3	Δεδομένα	9
2	Γενικευμένα Γραμμικά Μοντέλα	11
2.1	Εισαγωγή	11
2.2	Μοντέλο παλινδρόμησης <i>Poisson</i>	12
2.2.1	Μοντέλο παλινδρόμησης αρνητικής Διωνυμικής	13
2.2.2	Έλεγχος μοντέλων	14
3	Αναλογιστικό Μοντέλο <i>Lee Carter</i>	17
3.1	Δομή	17
3.2	Προσαρμογή μοντέλου	18
3.2.1	Μοντελοποίηση	18
3.2.2	Προσέγγιση SVD	20
3.2.3	Προσέγγιση <i>Poisson</i>	21
3.2.4	Προβλήματα με τη μέθοδο	22
3.3	Υπόλοιπα	23
3.3.1	Είδη υπολοίπων	23
3.3.2	Η δομή και το μέγεθος των υπολοίπων στην πρόβλεψη θνησιμότητας	23
3.4	Προβλέψεις του μοντέλου	24
3.4.1	Μοντέλα ARIMA	25
3.4.2	Παράδειγμα μοντέλου ARIMA	25
3.5	Παράδειγμα μοντελοποίησης και πρόβλεψης θνησιμότητας	30
3.5.1	Προσαρμογή του μοντέλου	30
3.5.2	Εκτιμήσεις παραμέτρων	30
3.5.3	Μοντελοποίηση και πρόβλεψη της παραμέτρου κ	32
3.5.4	Υπολογισμός αναμενόμενης διάρκειας ζωής και ρυθμών θνησιμότητας	38
4	Επεκτάσεις βασικού μοντέλου <i>Lee – Carter</i>	41
4.1	Διωνυμικό μοντέλο <i>Lee – Carter</i>	41
4.2	Αρνητικό Διωνυμικό μοντέλο <i>Lee – Carter</i>	42
4.2.1	Παράδειγμα αρνητικού διωνυμικού μοντέλου	42

5	Εφαρμογή μοντέλου <i>Lee – Carter</i>	55
5.1	Δεδομένα	55
5.2	Μοντελοποίηση και εφαρμογή του μοντέλου <i>Lee – Carter</i> . . .	57
5.2.1	Μοντέλο	57
5.2.2	Υπολογισμός παραμέτρων	57
5.2.3	Πρόβλεψη του μοντέλου	59
5.2.4	Συμπεράσματα	68

Κεφάλαιο 1

Εισαγωγή

1.1 Μοντέλα πληθυσμών

Με το πέρασμα των χρόνων και τη ραγδαία ανάπτυξη της ανθρωπότητας σε διάφορους τομείς άρχισε να παρατηρείται η ανάγκη για την ανάλυση δεδομένων όπως προκύπτουν για τους διάφορους ανθρώπινους πληθυσμούς. Από τη στιγμή που δεδομένα για τους διάφορους πληθυσμούς έγιναν διαθέσιμα δημιουργήθηκε έντονα η ανάγκη για σύγκριση των διάφορων πληθυσμών όσον αφορά το ρυθμό θνησιμότητας και κατά πόσον η κοινωνικοοικονομική κατάσταση μιας χώρας επηρεάζει αυτό το ρυθμό.

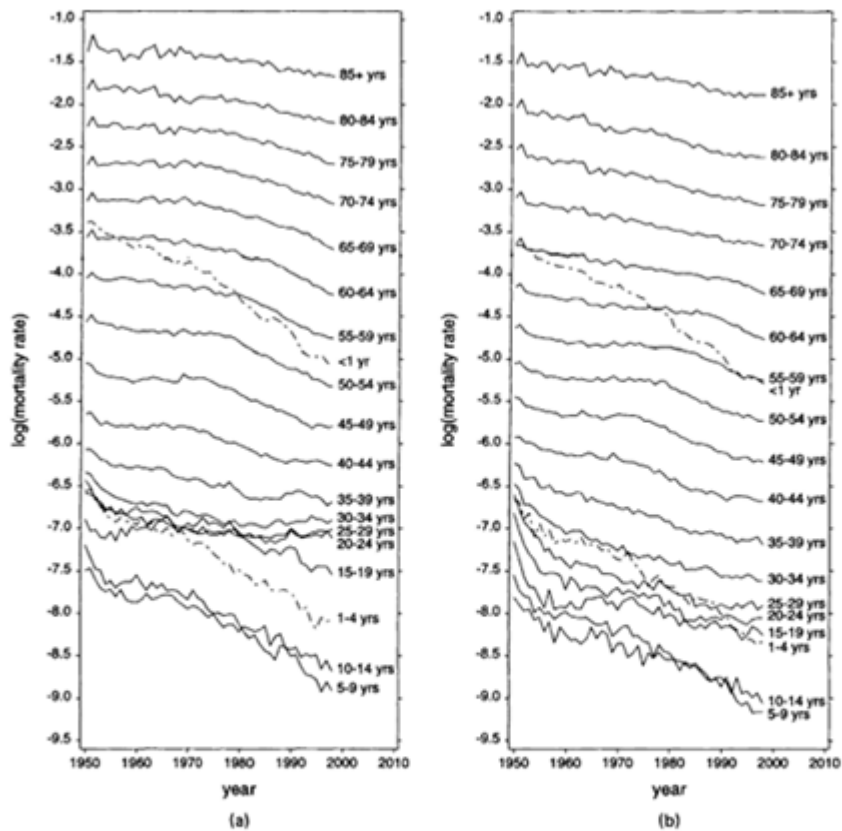
Έπειτα από συνεχείς μελέτες και έρευνες παρατηρήθηκε έντονα το φαινόμενο μειωμένης θνησιμότητας σε πολλές χώρες και αυτό είχε ως αποτέλεσμα το θέμα της θνησιμότητας να γίνει τεράστιας σημασίας για τους αναλογιστές και ειδικά για ασφαλιστικές εταιρείες και διάφορα άλλα ιδρύματα που προσφέρουν συντάξεις και άλλα οφέλη κατά τη διάρκεια της ζωής ενός ατόμου.

1.2 Τάσεις θνησιμότητας

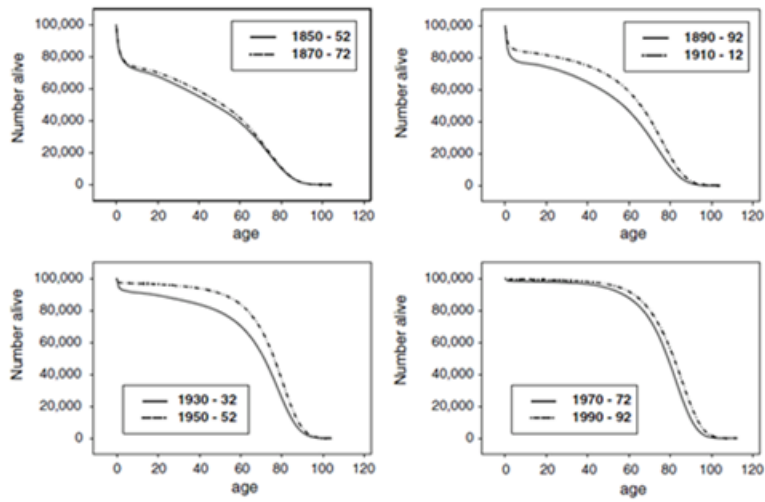
Σε αυτήν την ενότητα θα αναφερθούμε στις τάσεις θνησιμότητας που προκύπτουν για διάφορες αναπτυγμένες χώρες. Οι βασικές αρχές που επηρεάζουν τις καμπύλες θνησιμότητας συναρτήσει της ηλικίας είναι :

1. Μειώσεις στους ρυθμούς θνησιμότητας στο χρόνο εκτός από νεαρές ηλικίες σε μερικές χώρες όπου τα ναρκωτικά και το AIDS δημιούργησαν καταστροφικά αποτελέσματα.
2. Η καμπύλη της τυχαίας συνάρτησης ζωής (σχήματα 1.2,1.3) η οποία οδηγεί στην αντίστοιχη συνάρτηση επιβίωσης η οποία με την πάροδο του χρόνου τετραγωνοποιείται.
3. Αυξησή του μέσου της συνάρτησης πυκνότητας του (2) το οποίο ονομάζεται επέκταση της συνάρτησης επιβίωσης.

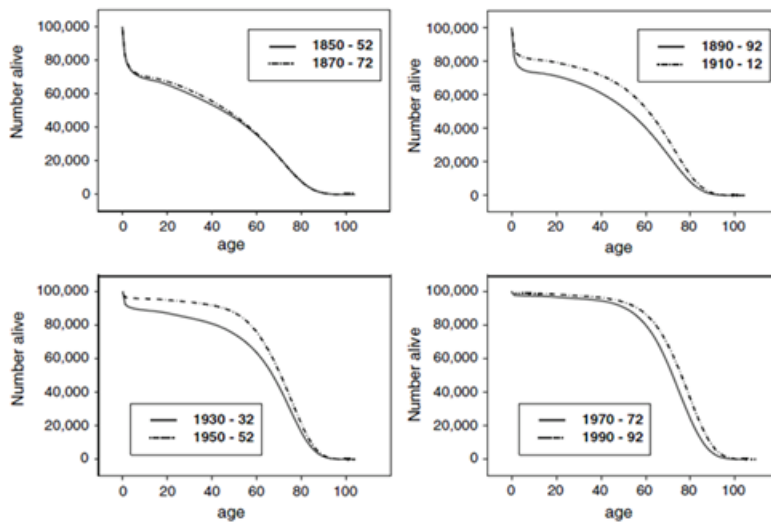
Μπορούμε να παρουσιάσουμε οποιαδήποτε από τις πιο πάνω αρχές χρησιμοποιώντας δεδομένα από το ρυθμό θνησιμότητας στην Αγγλία και την Ουαλία (Haberman και Renshaw, 2003). Το πρώτο σχήμα δείχνει το ρυθμό θνησιμότητας για ομάδες ηλικιών σε σχέση με το χρόνο ξεχωριστά για κάθε φύλο. Παρατηρούμε ότι οι τάσεις θνησιμότητας επιβεβαιώνουν την πρώτη βασική αρχή όπως αναφέρθηκε πιο πάνω καθώς και ότι για το αρσενικό φύλο στις ηλικίες 25-34 παρατηρείται αύξηση στο ρυθμό θνησιμότητας. Το δεύτερο και τρίτο σχήμα δείχνουν την 'καμπύλη των θανάτων' όπως προκύπτει από τους πίνακες ζωής που βασίζονται σε απογραφές για τη θνησιμότητα στις δύο χώρες σε διαστήματα 20 χρόνων και επιβεβαιώνουν τις άλλες δύο βασικές αρχές. Παρατηρούμε την αύξηση του αριθμού των επιζώντων γύρω από τον δειγματικό μέσο. Οι αντίστοιχες συναρτήσεις επιβίωσης που παρουσιάζονται στα σχήματα 1.4 και 1.5 επιβεβαιώνουν την υπόθεση της τετραγωνοποίησης.



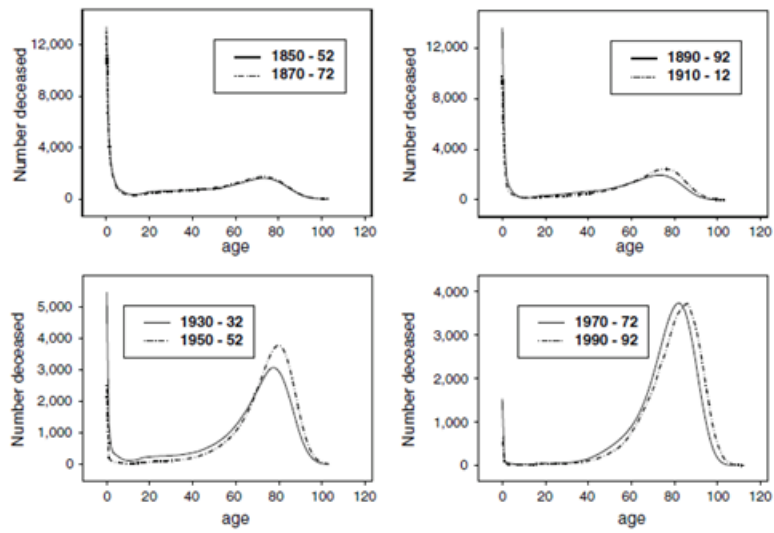
Σχήμα.1.1 Λογαριθμοποιημένοι ρυθμοί θνησιμότητας για ομαδοποιημένες ηλικίες από (α) την Αγγλία και (β) την Ουαλία (Haberman & Renshaw 2003).



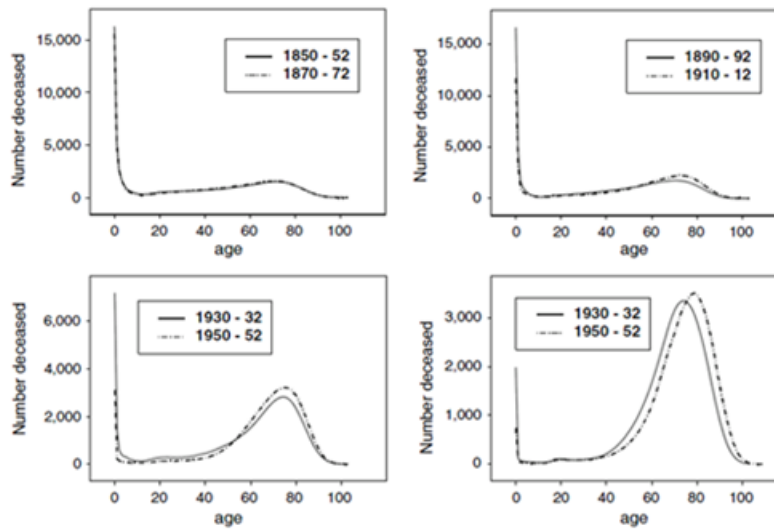
Σχημά.1.2 Γραφικές παραστάσεις διάρκειας ζωής για γυναίκες σε δείγμα 100,000 (Haberman & Renshaw 2003).



Σχημά.1.3 Γραφικές παραστάσεις διάρκειας ζωής για άνδρες σε δείγμα 100,000 (Haberman & Renshaw 2003).



Σχήμα.1.4 Γραφικές παραστάσεις διάρκειας ζωής για γυναίκες σε δείγμα 100,000 (Haberman & Renshaw 2003).



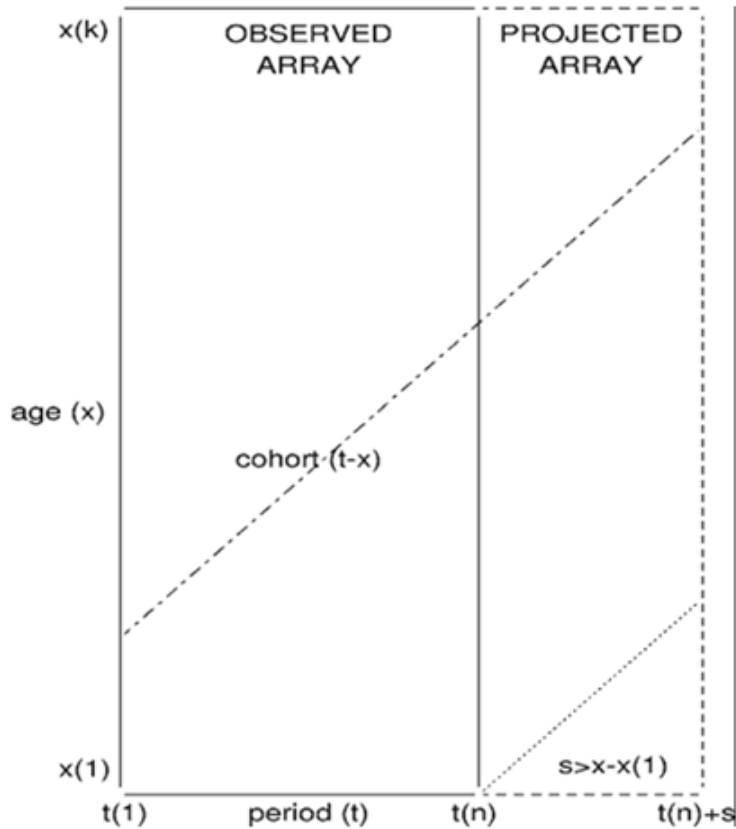
Σχήμα.1.5 Γραφικές παραστάσεις διάρκειας ζωής για άνδρες σε δείγμα 100,000 (Haberman & Renshaw 2003).

1.3 Δεδομένα

Η τυχαία μεταβλητή D_{xt} ορίζει τον αριθμό των θανάτων σε ένα πληθυσμό σε ηλικία x και χρόνο t . Παρουσιάζουμε τα δεδομένα σε τετραγωνική μορφή (2 πίνακες όπου για κάθε ηλικία έχουμε τον αριθμό θανάτων και τον αντίστοιχο πληθυσμό για κάθε χρονιά) το d_{xt} να ορίζει τον παρατηρούμενο αριθμό θανάτων και το e_{xt} τον αριθμό (άτομα-χρόνια) που εκτίθενται σε ρίσκο θανάτου. Οι εμπειρικοί ρυθμοί θνησιμότητας δίνονται από τον τύπο :

$$\hat{m}_{xt} = \frac{d_{xt}}{e_{xt}} \quad (1.1)$$

Έχουμε τότε με βάση τους Haberman και Renshaw (2008) διασταυρωμένη ταξινόμηση (σχήμα 1.6) των δεδομένων ανά ημερολογιακό χρόνο t στο διάστημα $[t_1, t_n]$ και ανά ηλικία x στο διάστημα $[x_1, x_n]$ είτε ομαδοποιημένα σε k αριθμό κατηγοριών είτε σε ατομικό χρόνο. Το μ_{xt} ορίζεται ως ο ρυθμός κινδύνου για ηλικία x και χρόνο t και εκτιμάται από το \hat{m}_{xt} .



Σχήμα.1.6 Παρουσίαση της διαμόρφωσης των δεδομένων (Haberman & Renshaw 2008).

Κεφάλαιο 2

Γενικευμένα Γραμμικά Μοντέλα

Στο κεφάλαιο αυτό θα ασχοληθούμε με τη λειτουργία των γενικευμένων γραμμικών μοντέλων και ειδικότερα του μοντέλου της παλινδρόμησης *Poisson* και της αρνητικής Διωνυμικής (Mc Cullagh και Nelder, 1989) για να κατανοήσουμε το πλαίσιο εργασίας του αναλογιστικού μοντέλου Lee-Carter όταν το προσεγγίζουμε με την κατανομή *Poisson*.

2.1 Εισαγωγή

Όπως αναφέρεται από τους Οικονόμου και Καρώνη (2010) θεωρούμε το μοντέλο της γραμμικής παλινδρόμησης το οποίο εκφράζεται ως :

$$y_i = x_i' \beta + \varepsilon_i \quad (2.1)$$

όπου το $i = 1, \dots, n$ y_i εξαρτημένη τυχαία μεταβλητή, x_i διάνυσμα k ανεξάρτητων τυχαίων μεταβλητών, β διάνυσμα $k \times 1$ άγνωστων παραμέτρων και τα ε_i υπόλοιπα με μηδενικό μέσο και σταθερή διασπορά σ^2 . Το γραμμικό μοντέλο παλινδρόμησης χαρακτηρίζεται από τις πιο κάτω αρχές :

- τα y_i ακολουθούν κανονική κατανομή με μέση τιμή $E(y_i) = \mu_i$ σταθερή διασπορά σ^2
- οι συμμεταβλητές x_i συνδυάζονται γραμμικά με τους συντελεστές για να δημιουργήσουν την γραμμική πρόβλεψη $\eta_i = x_i' \beta$
- η γραμμική πρόβλεψη $\eta_i = x_i' \beta$ είναι συνάρτηση του μέσου μ_i μέσω μιας συνάρτησης σύνδεσης $g(\mu_i)$ που στη περίπτωση αυτή είναι η ταυτότητα $g(\mu_i) = x_i' \beta$

Από το μοντέλο της γραμμικής παλινδρόμησης και τις αρχές που το χαρακτηρίζουν δημιουργούνται ερωτήματα όσον αφορά την εξαρτημένη μεταβλητή y όταν είναι συνεχής αλλά δεν ακολουθεί την κανονική κατανομή και όταν είναι διακριτή τυχαία μεταβλητή.

Για τη πρώτη περίπτωση χρησιμοποιούμε κατάλληλο μετασχηματισμό της y όπως για παράδειγμα το λογαριθμικό όμως επειδή δεν μπορούμε πάντα να βρούμε κατάλληλο μετασχηματισμό χρειαζόμαστε ένα μοντέλο όμοιο με τη γραμμική παλινδρόμηση χρησιμοποιώντας όμως μια κατανομή διαφορετική από την Κανονική.

Για τη δεύτερη περίπτωση, που σχετίζεται και με την εφαρμογή του μοντέλου Lee-Carter, πολλές φορές αγνοούμε το γεγονός ότι οι τιμές είναι διακριτές. Όμως υπάρχουν περιπτώσεις όπως ο αριθμός αυτοκινητιστικών δυστυχημάτων και ο αριθμός θανάτων σε μια χώρα (παράδειγματα με την εφαρμογή του μοντέλου Lee-Carter) που δεν μπορούμε να αγνοήσουμε το γεγονός ότι οι τιμές που καταγράφονται είναι διακριτές και για αυτό το λόγο ψάχνουμε μοντέλο παλινδρόμησης αφού οι τιμές εξαρτώνται και από άλλες μεταβλητές.

2.2 Μοντέλο παλινδρόμησης *Poisson*

Το αναλογιστικό μοντέλο Lee-Carter εφαρμόζεται στο πλαίσιο λειτουργίας των γενικευμένων γραμμικών μοντέλων και στη θεωρία του μοντέλου παλινδρόμησης *Poisson* για αυτό θα αναφερθούμε πιο λεπτομερώς (Οικονόμου και Καρώνη, 2010) στο συγκεκριμένο μοντέλο για καλύτερη κατανόηση της εφαρμογής του αναλογιστικού μοντέλου. Θεωρούμε τον αριθμό των δυστυχημάτων που πραγματοποιούνται ως την τυχαία μεταβλητή y η οποία είναι διακριτή και χρησιμοποιούμε για την περιγραφή του γεγονότος την κατανομή *Poisson* με συνάρτηση πιθανότητας :

$$f(y) = \frac{e^{-\mu} \mu^y}{y!}, \mu > 0, y = 0, 1, 2, \dots \quad (2.2)$$

και αναμενόμενη μέση τιμή $E(y) = \mu$ και διασπορά $V(y) = \mu$. Παρατηρούμε ότι πρέπει να αναπτύξουμε μοντέλο τύπου παλινδρόμησης που να είναι προσαρμοσμένο στη διακριτή μεταβλητή y .

Το μοντέλο παλινδρόμησης *Poisson* υποθέτει ότι η μεταβλητή y ακολουθεί κατανομή *Poisson* και ότι ο λογάριθμος της αναμενόμενης τιμής μοντελοποιείται από ένα γραμμικό συνδυασμό αγνώστων παραμέτρων. Με βάση τους Cullagh και Nelder (1989) η μοντελοποίηση αυτή γίνεται μέσω ενός κατάλληλου μετασχηματισμού $g(\cdot)$ της αναμενόμενης μέσης τιμής της y έτσι ώστε να ισχύει μια σχέση της μορφής

$$g(\mu_i) = x_i' \beta \quad (2.3)$$

για κάθε καταγεγραμμένη παρατήρηση y_i . Η συνάρτηση $g(\cdot)$ καλείται συνάρτηση σύνδεσης και ο περιορισμός $\mu > 0$ που επιβάλλεται από τον ορισμό της κατανομής *Poisson* σημαίνει ότι δεν μπορεί να ισχύει $g(\mu_i) = \mu_i$ γιατί τότε θα πρέπει $\mu_i = x_i' \beta$ η οποία δεν ικανοποιεί τον περιορισμό $\mu > 0$ για οποιοδήποτε x . Για να τηρείται ο περιορισμός πρέπει η μ_i να είναι μη αρνητική συνάρτηση του $x_i' \beta$ και μπορούμε να επιλέξουμε την $\mu_i = e^{x_i' \beta}$ όπου η συνάρτηση σύνδεσης

είναι η $g(\mu) = \ln \mu$

Προκύπτουν οι ακόλουθες προϋποθέσεις για το μοντέλο παλινδρόμησης *Poisson*

- $y_i \sim \text{Poisson}(\mu_i)$
- $\mu_i = e^{x_i\beta}$
- ανεξαρτησία μεταξύ των y_x παρατηρήσεων
- οι μετρήσεις τόσο της εξαρτημένης όσο και των ανεξάρτητων μεταβλητών δεν υπόκεινται σε σφάλματα μέτρησης

Για τον υπολογισμό των άγνωστων παραμέτρων β_i χρησιμοποιούμε τη μέθοδο της μέγιστης πιθανοφάνειας. Η συνάρτηση πιθανοφάνειας L για τις τιμές y_1, y_2, \dots, y_n με αντίστοιχες συμμεταβλητές $x'_i = (x_{i0}, x_{i1}, \dots, x_{ik})$ δίνεται από τη σχέση

$$L = \prod_{i=1}^n \frac{e^{-\mu_i} \mu_i^{y_i}}{y_i!} \quad (2.4)$$

και η λογαριθμοποιημένη συνάρτηση πιθανοφάνειας δίνεται από τη σχέση

$$l = \sum_{i=1}^n [-\mu_i + y_i \ln \mu_i - \ln(y_i!)]$$

Η πιο πάνω σχέση χρησιμοποιώντας τη σχέση (2.3) παίρνει τη μορφή

$$l = \sum_{i=1}^n [-e^{x'_i\beta} + y_i x'_i\beta - \ln(y_i!)]$$

και συνεπώς

$$\begin{aligned} \frac{\partial l}{\partial \beta_j} &= \sum_{i=1}^n (-x_{ij} e^{x'_i\beta} + y_i x_{ij}) \\ &= \sum_{i=1}^n [x_{ij}(y_i - e^{x'_i\beta})] \end{aligned}$$

Οι εκτιμήτριες μέγιστης πιθανοφάνειας των β_j προκύπτουν από την επίλυση των εξισώσεων

$$\sum_{i=1}^n [x_{ij}(y_i - e^{x'_i\hat{\beta}})] = 0, j = 0, 1, \dots, k \quad (2.5)$$

η οποία λύνεται μόνο με επαναληπτικές μεθόδους αφού είναι μη γραμμική ως προς τα $\hat{\beta}$.

2.2.1 Μοντέλο παλινδρόμησης αρνητικής Διωνυμικής

Σε αυτή την ενότητα θα αναφερθούμε στη θεωρία του μοντέλου παλινδρόμησης της αρνητικής Διωνυμικής για να κατανοήσουμε τη λειτουργία της επέκτασης του βασικού μοντέλου Lee-Carter που θα αναφερθούμε σε πιο κάτω κεφάλαιο. Όπως είδαμε και στην προηγούμενη ενότητα στη περίπτωση της *Poisson* η μέση τιμή μ_i θεωρείται σταθερή για τις διάφορες τιμές του i . Θεωρούμε τώρα (Ismail και Jemain, 2007) ότι η παράμετρος μ_i δεν είναι σταθερή αλλά ακολουθεί κατανομή Γάμμα με αναμενόμενη μέση τιμή $E(\mu_i) = \rho_i$ και

διασπορά $V(\mu_i) = \rho_i^2 \nu_i^{-1}$ όπου ρ και ν οι παράμετροι κλίμακας και σχήματος της κατανομής Γάμμα. Ισχύει τότε ότι $y_i | \mu_i$ ακολουθεί κατανομή *Poisson* με δεσμευμένη μέση τιμή $E(y_i | \mu_i) = \mu_i$ και η περιθώρια κατανομή του y_i ακολουθεί αρνητική Διωνυμική κατανομή με αναμενόμενη μέση τιμή $E(y_i) = \rho_i$ και διασπορά $V(y_i) = \rho_i + \rho_i^2 \nu_i^{-1}$. Παρατηρούμε ότι για διάφορες παραμετρικοποιήσεις θα λαμβάνουμε διάφορους τύπους αρνητικής Διωνυμικής κατανομής. Για παράδειγμα αν ορίσουμε $\nu_i = \alpha^{-1}$ η τυχαία μεταβλητή y ακολουθεί αρνητική Διωνυμική κατανομή με μέση τιμή $E(y_i) = \rho_i$ και διασπορά $V(y_i) = \rho_i + \alpha \rho_i^2$ όπου το α δηλώνει την παράμετρο μεταβλητότητας. Αν η παράμετρος α είναι ίση με μηδέν τότε θα έχουμε ίση μέση τιμή και διασπορά και οδηγούμαστε στην κατανομή *Poisson*.

Η περίπτωση που μας ενδιαφέρει άμεσα και η εφαρμογή της χρησιμοποιείται για διάφορες αναλογιστικές μελέτες είναι όταν η παράμετρος α λαμβάνει τιμές μεγαλύτερες του μηδέν. Τότε η τιμή της διασποράς θα υπερβαίνει την αναμενόμενη μέση τιμή γεγονός που δηλώνει μεγαλύτερη μεταβλητότητα από αυτή που αναμέναμε από την εφαρμογή του στατιστικού μοντέλου (*overdispersion*). Για τη κατασκευή του μοντέλου παλινδρόμησης της αρνητικής Διωνυμικής στη συγκεκριμένη περίπτωση, όπου πρέπει να τηρούνται οι προϋποθέσεις μη αρνητικότητας των παραμέτρων όπως αναφέραμε και στη προηγούμενη ενότητα, θα θεωρήσουμε ότι η αναμενόμενη μέση τιμή είναι πολλαπλασιαστική και δίνεται από τη πιο κάτω σχέση:

$$\mu_i = e_i \exp(x_i' \beta) \quad (2.6)$$

όπου e_i δηλώνει τιμή από τα άτομα που εκτίθενται στο κίνδυνο (σε εφαρμογή σε επομένο κεφάλαιο είναι ο αριθμός του πληθυσμού για διάφορες χρονικές περιόδους), x_i διάνυσμα k ανεξάρτητων τυχαίων μεταβλητών και β διάνυσμα $k \times 1$ άγνωστων παραμέτρων. Για τον υπολογισμό των παραμέτρων β θα χρησιμοποιήσουμε τη μέθοδο της μέγιστης πιθανοφάνειας. Η λογαριθμοποιημένη συνάρτηση πιθανοφάνειας δίνεται από τη σχέση :

$$l(\beta, \alpha) = \sum_i \{ \sum_{r=1}^{y_i-1} \log(1+r\alpha) \} - y_i \log(\alpha) - \log(y_i!) + y_i \log(\alpha \rho_i) - (y_i + \alpha^{-1}) \log(1 + \alpha \rho_i) \quad (2.7)$$

Οι εκτιμήτριες μέγιστης πιθανοφάνειας ($\hat{\beta}, \hat{\alpha}$) τότε προκύπτουν μεγιστοποιώντας τη πιο πάνω σχέση ως προς α και β . Προκύπτουν οι ακόλουθες σχέσεις:

$$\frac{\partial l(\beta, \alpha)}{\partial \beta_j} = \sum_i \frac{(y_i - \rho_i) x_{ij}}{1 + \alpha \rho_i} = 0, j = 1, 2, \dots, k \quad (2.8)$$

και

$$\frac{\partial l(\beta, \alpha)}{\partial \alpha} = \sum_i \{ \sum_{r=1}^{y_i-1} \left(\frac{r}{1+r\alpha} \right) \} + \alpha^{-2} \log(1 + \alpha \rho_i) - \frac{(y_i + \alpha^{-1}) \rho_i}{(1 + \alpha \rho_i)} = 0 \quad (2.9)$$

Η επίλυση των εξισώσεων (2.8), (2.9) γίνεται ταυτόχρονα χρησιμοποιώντας επαναληπτικές μεθόδους (Ismail και Jemain, 2007).

2.2.2 Έλεγχος μοντέλων

Στην ενότητα αυτή θα αναφερθούμε στις διάφορες μεθόδους ελέγχου της προσαρμογής του κατάλληλου μοντέλου στα δεδομένα μας.

Λόγος πιθανοφάνειων

Με τη χρησιμοποίηση της μεθόδου μέγιστης πιθανοφάνειας μπορούμε να χρησιμοποιήσουμε τις τιμές που λαμβάνουμε για να ελέγξουμε τη καταλληλότητα του μοντέλου που εφαρμόσαμε. Εκτελούμε τον έλεγχο για τις υποθέσεις $H_0: y = y_0$ έναντι $H_1: y \neq y_0$ με τον λόγο των πιθανοφάνειων να δίνεται από τη πιο κάτω σχέση :

$$\lambda = \frac{\hat{L}_0}{\hat{L}_1} = \frac{L(y_0, x)}{L(\hat{y}, x)}, 0 \leq \lambda \leq 1 \quad (2.10)$$

όπου $\hat{y} = \operatorname{argmax} L(y, x)$ και $L(y, x)$ η συνάρτηση πιθανοφάνειας. Ο έλεγχος καταλληλότητας γίνεται από τη σχέση

$$-2 \ln \lambda = -2(\hat{l}_0 - \hat{l}_1) \sim \chi_q^2 \quad (2.11)$$

όπου οι βαθμοί ελευθερίας q είναι $p_1 - p_0$ με p τον αριθμό των παραμέτρων, l_1 , l_0 οι μεγιστοποιημένες λογαριθμοποιημένες συναρτήσεις πιθανοφάνειας και για μεγάλη τιμή του ελέγχου το μοντέλο κρίνεται ακατάλληλο.

Ελεγχοςυνάρτηση *Deviance*

Η συνάρτηση *Deviance* δίνεται από τη πιο κάτω σχέση:

$$D = 2(l(y; y) - l(\mu; y)) \quad (2.12)$$

όπου $l(y; y)$ και $l(\mu; y)$ οι λογαριθμοποιημένες πιθανοφάνειες του μοντέλου. Χρησιμοποιούμε τη συνάρτηση για τη σύγκριση δύο εμφωλευμένων μοντέλων, δηλαδή το ένα μοντέλο είναι απλούστερη μορφή του άλλου. Για παράδειγμα έστω ένα μοντέλο A_1 με β εκτιμήτρια μέγιστης πιθανοφάνειας και ένα μοντέλο A_2 πιο απλή μορφή του A_1 με βαθμούς ελευθερίας β^* . Εκτελούμε τον έλεγχο με μηδενική υπόθεση H_0 να ισχύει το μοντέλο A_2 και εναλλακτική να ισχύει το μοντέλο A_1 . Με βάση το λόγο μεγιστοποιημένων πιθανοφάνειων προκύπτει η σχέση :

$$D(\hat{\beta}^*) - D(\hat{\beta}) \sim \chi_q^2 \quad (2.13)$$

όπου οι βαθμοί ελευθερίας q είναι $p_1 - p_0$ με p τον αριθμό των παραμέτρων στις δύο περιπτώσεις. Αν η τιμή της ελεγχοςυνάρτησης είναι μεγάλη τότε απορρίπτουμε τη μηδενική υπόθεση.

AIC και *BIC*

Όταν έχουμε διαθέσιμα αρκετά μοντέλα μέγιστης πιθανοφάνειας και θέλουμε να συγκρίνουμε δύο ή και περισσότερα μοντέλα χρησιμοποιούμε τα κριτήρια *AIC* και *BIC* που ορίζονται από τις σχέσεις:

$$AIC = -2l + 2p \quad (2.14)$$

$$BIC = -2l + p \ln n \quad (2.15)$$

όπου p είναι ο αριθμός των παραμέτρων, l η τιμή της πιθανοφάνειας και n ο αριθμός των παρατηρήσεων.

Κεφάλαιο 3

Αναλογιστικό Μοντέλο *Lee Carter*

Στο κεφάλαιο αυτό θα εστιάσουμε στην ανάπτυξη και την επέκταση του βασικού μοντέλου Lee-Carter, όπως προκύπτει μέσα από τους Haberman και Renshaw (2008) καθώς και τους Lee και Carter (1992) το οποίο αξίζει να σημειώσουμε πώς έχει γίνει μια από τις πιο διαδεδομένες τεχνικές που χρησιμοποιείται από πολλές υπηρεσίες και είναι πλέον από τα πιο σημαντικά στατιστικά μοντέλα θνησιμότητας στην δημογραφική βιβλιογραφία.

3.1 Δομή

Το πρόβλημα που έχουμε να αντιμετωπίσουμε είναι η μοντελοποίηση και η πρόβλεψη του μ_{xt} , το οποίο έχουμε ορίσει ως το ρυθμό κινδύνου, σαν συνάρτηση του χρόνου και της ηλικίας. Το πλαίσιο μοντελοποίησης του Lee-Carter στοχεύει στη μείωση της διάστασης του προβλήματος καθορίζοντας ένα δείκτη χρόνου ο οποίος συνοψίζει παλιότερες τάσεις και που επηρεάζει μ_{xt} σε χρόνο t σε όλες τις ηλικίες ταυτόχρονα και μπορεί να μοντελοποιηθεί κατά προσέγγιση. Για αυτό το σκοπό προτάθηκε από τους Lee και Carter (1992) το πιο κάτω μοντέλο :

$$\ln \mu_{xt} = \alpha_x + \beta_x \kappa_t + \varepsilon_{xt} \quad (3.1)$$

όπου ε_{xt} ανεξάρτητες τυχαίες μεταβλητές $\sim N(0, \sigma^2)$, α_x, β_x παράμετροι ηλικίας και κ_t παράμετρος χρόνου για τον υπολογισμό του οποίου χρησιμοποιούνται μέθοδοι χρονοσειρών στις οποίες θα αναφερθούμε πιο κάτω.

Το μοντέλο Lee-Carter έχει περισσότερες παραμέτρους (overparameterised) υπό την έννοια ότι η δομή παραμένει αναλλοίωτη κάτω από οποιαδήποτε μετατροπή παραμέτρου από τις ακόλουθες :

$$\{\alpha_x, \beta_x, \kappa_t\} \mapsto \{\alpha_x, \frac{\beta_x}{c}, c\kappa_t\}$$

$$\{\alpha_x, \beta_x, \kappa_t\} \mapsto \{\alpha_x - c\beta_x, \beta_x, \kappa_t + c\}$$

για οποιαδήποτε σταθερά C .

Για να εξασφαλίσουμε την αναγνωσιμότητα του μοντέλου επιβάλλουμε δύο περιορισμούς οι οποίοι έχουν εισηγηθεί από τους Lee και Carter και εφαρμόζονται στο πλαίσιο της τετραγωνικής διάταξης των δεδομένων.

$$\sum_{t=t_1}^{t_n} \kappa_t = 0 \quad (3.2)$$

$$\sum_{x=x_1}^{x_k} \beta_x = 0 \quad (3.3)$$

Αποτέλεσμα των εξισώσεων (3.1),(3.2),(3.3) είναι η εκτίμηση του α_x που δίνεται από τον πιο κάτω τύπο :

$$\hat{\alpha}_x = \ln \prod_{t=t_1}^{t_n} \mu_{xt}^{\frac{1}{h}} \quad (3.4)$$

όπου $h = t_n - t_1 + 1$ είναι η εκτιμήτρια ελαχίστων τετραγώνων του $\hat{\alpha}_x$. Οι παράμετροι του βασικού μοντέλου Lee-Carter μπορούν να ερμηνευθούν ως εξής : α_x : αποτελεί παράμετρο ηλικίας που περιγράφει τη μορφή κάθε ομάδας ηλικίας.

κ_t : αντιπροσωπεύει το δείκτη θνησιμότητας. Είναι η παράμετρος του χρόνου όπου θα γίνει η πρόβλεψη για τον υπολογισμό μελλοντικών τιμών.

β_x : αντιπροσωπεύει την παράμετρο ηλικίας που δείχνει την τάση θνησιμότητας σε σχέση με το χρόνο και κατά πόσο η θνησιμότητα σε κάποια ομάδα ηλικίας τείνει να αυξηθεί ή να μειωθεί.

ε_{xt} : αντιπροσωπεύει την ηλικία και ειδικά αποτελέσματα χρόνου που δεν συλλαμβάνονται από το μοντέλο.

3.2 Προσαρμογή μοντέλου

3.2.1 Μοντελοποίηση

Για την προσαρμογή του μοντέλου (Lee και Carter, 1992) οι Lee και Carter πρότειναν τα παρακάτω στάδια :

- υπολογισμός του α σαν $\hat{\alpha}_x$ χρησιμοποιώντας τον τύπο (3.4)
- υπολογισμός του πίνακα $[Z_{xt}] = [\ln \mu_{xt} - \hat{\alpha}_{xt}]$ και μετά υπολογίζουμε το κ_t και το β_x ως το πρώτο δεξιά και πρώτο αριστερά διάνυσμα στο SVD (Singular Value Decomposition) του πίνακα Z_{xt}
- προσαρμόζουμε τα υπολογισμένα κ_t έτσι ώστε για κάθε t ο παρατηρούμενος αριθμός θανάτων και ο αναμενόμενος αριθμός να είναι ίσοι όπως για παράδειγμα :

$$\sum_{x=x_1}^{x_k} d_{xt} = \sum_{x=x_1}^{x_k} e_{xt} \exp(\hat{\alpha}_x + \hat{\beta}_x \kappa_t) \quad (3.5)$$

για όλα τα t .

Η ισότητα του παρατηρούμενου αριθμού θανάτου και του αναμενόμενου αριθμού θανάτων δεν ικανοποιείται αυτόματα από το μοντέλο Lee-Carter για αυτό το λόγο χρειάζεται δεύτερο στάδιο προσαρμογής το οποίο περιγράφεται από την πιο πάνω εξίσωση (3.5). Ο λόγος που δεν ισχύει αυτή η ισότητα είναι ότι όλες οι ηλικίες λαμβάνουν το ίδιο βάρος στο SVD ανεξάρτητα από το μέγεθος του m_{xt} και του υποκείμενου αριθμού θανάτων. Άρα οι τιμές του ε_{xt} που αντιστοιχούν σε νεότερες ηλικίες είναι ίσες με αυτές που αντιστοιχούν σε μεγαλύτερες ηλικίες παρόλο που ο αριθμός θανάτων θα είναι ψηλότερος σε πιο μεγάλες ηλικίες. Παρατηρούμε ότι η υπόθεση τα σφάλματα ε_{xt} να είναι ομοσκεδαστικά δεν ισχύει. Αναμένουμε ο λογάριθμος του ρυθμού κινδύνου να είναι πιο αξιόπιστος στις μεγαλύτερες ηλικίες σε σχέση με τις μικρότερες ηλικίες λόγω του μικρότερου αριθμού θανάτων.

Για τη διαδικασία προσαρμογής και μοντελοποίησης έχουν προταθεί δύο επίπεδα βελτίωσης.

1. Το πρώτο επίπεδο είναι μια μορφή ελαχίστων τετραγώνων με βάρη της βασικής προσέγγισης στην προσαρμογή του μοντέλου

$$D(y_{xt}, \hat{y}_{xt}) = \sum_{x,t} \omega_{xt} (y_{xt} - \hat{y}_{xt})^2 \quad (3.6)$$

όπου $y_{xt} = \ln m_{xt}$ και $\hat{y}_{xt} = \alpha_x + \beta_x \kappa_t$ και τα βάρη ω_{xt} επιλέχθηκαν να είναι ίσα με τον παρατηρούμενο αριθμό θανάτων d_{xt}

2. Στο δεύτερο επίπεδο θα περιγράψουμε πως μπορούμε να χρησιμοποιήσουμε την κατανομή *Poisson* για την εύρεση προσεγγιστικής λύσης με την εφαρμογή του μοντέλου. Με βήμα n μπορούμε να αναβαθμίσουμε μια παράμετρο με τις υπόλοιπες να προσαρμοσμένες στους εκτιμητές τους χρησιμοποιώντας τον πιο κάτω αναδρομικό τύπο :

$$\hat{\theta}^{(n+1)} = \hat{\theta}^{(n)} - \frac{\frac{\partial D^{(n)}}{\partial \theta}}{\frac{\partial^2 D^{(n)}}{\partial \theta^2}} \quad (3.7)$$

Έτσι τα D_{xt} μοντελοποιούνται ως ανεξάρτητες μεταβλητές της *Poisson* και ισχύει ότι:

$$E(D_{xt}) = e_{xt} \exp \eta_{xt} \quad (3.8)$$

και

$$Var(D_{xt}) = E(D_{xt}) \quad (3.9)$$

όπου

$$\eta_{xt} = \alpha_x + \beta_x \kappa_t \quad (3.10)$$

Επειδή όμως το η_{xt} είναι μη γραμμικό στις παραμέτρους του δεν μπορεί να εφαρμοστεί σαν γενικευμένο γραμμικό μοντέλο. Όμως προσαρμόζοντας κατάλληλη μέθοδο μπορούμε να χρησιμοποιήσουμε την πιθανοφάνεια *Poisson* παρακολουθώντας τη συσχετισμένη απόκλιση :

$$D(d_{xt}, \hat{d}_{xt}) = \sum_{all x,t} dev(d_{xt}, \hat{d}_{xt}) = \sum_{x,t} 2\omega_{xt} \left\{ d_{xt} \log\left(\frac{d_{xt}}{\hat{d}_{xt}}\right) - (d_{xt} - \hat{d}_{xt}) \right\} \quad (3.11)$$

όπου $\hat{d}_{xt} = e_{xt} \exp(\hat{\alpha}_x + \hat{\beta}_x \hat{\kappa}_t)$, $\omega_{xt} = 1$ αν $e_{xt} > 0$ και 0 διαφορετικά.

Για την λειτουργία του πιο πάνω ακολουθούμε την παρακάτω διαδικασία : Θέτουμε αρχικές τιμές για τα $\alpha_q, \beta_q, \kappa_t$ και υπολογίζουμε τα d_{xt} και $D(d_{xt}, \hat{d}_{xt})$

↓

Ενημερώνουμε τη τιμή του α_x , υπολογίζουμε \hat{d}_{xt} , ενημερώνουμε το κ_t προσαρμόζοντας $\sum_{t=t_1}^{t_n} \kappa_t = 0$, υπολογίζουμε \hat{d}_{xt} , ενημερώνουμε το β_x , υπολογίζουμε \hat{d}_{xt} και τέλος υπολογίζουμε το $D(d_{xt}, \hat{d}_{xt})$

↓

Επαναλαμβάνουμε την πιο πάνω διαδικασία και σταματάμε όταν το $D(d_{xt}, \hat{d}_{xt})$ συγκλίνει. Το μοντέλο της *Poisson* έχει το πλεονέκτημα ότι δεν υποθέτει ομοσκεδαστικότητα στη δομή σφάλματος του μοντέλου Lee-Carter, αναγνωρίζει την αθέρα φύση του D_{xt} και αποφεύγει την ανάγκη για το κ_t να επαναυπολογίζεται μέσα από διαδικασία δύο στάδιων. Αυτό ακολουθείται επειδή οι εξισώσεις πιθανοτήτων επαληθεύουν αυτόματα την εξίσωση (3.5) Από διάφορες έρευνες και μελέτες παρατηρούμε ότι υπάρχει συμφωνία στον υπολογισμό των παραμέτρων μέσα από τις μεθόδους :

(α') προσέγγιση Wilmoth με την επιλογή βάρους $\omega_{xt} = d_{xt}$

(β') προσέγγιση γενικευμένου γραμμικού μοντέλου με σφάλματα *Poisson*

Παρόλο που υπάρχει συμφωνία στις δύο μεθόδους η δεύτερη είναι προτιμότερη επειδή βασίζεται στις αρχές της μέγιστης πιθανοφάνειας και τα αποτελέσματα των εκτιμήσεων θα έχουν καλύτερες στατιστικές ιδιότητες. Πρέπει να επισημάνουμε ότι σε πολλές εφαρμογές συμπεριλαμβάνουμε μια παράμετρο ϕ στην εξίσωση (3.9) έτσι ώστε να παίρνει την πιο κάτω μορφή:

$$\text{Var}(D_{xt}) = \phi E(D_{xt}) \quad (3.12)$$

Αυτό είναι σχετικό με αρκετές αναλογιστικές εφαρμογές όπου τα δεδομένα βασίζονται σε διαδικασίες ασφάλισης αντί σε άτομα έτσι ώστε να μπορεί να αντιπροσωπευθεί περισσότερες από μια φορές στα δεδομένα.

3.2.2 Προσέγγιση SVD

Η μέθοδος προσέγγισης με SVD μπορεί να περιγραφεί ως εξής (Haberman και Renshaw, 2008) : Θέτουμε $M = (M_{xt})$ και ο πίνακας $N \times T$ των ρυθμών που έχουν παρατηρηθεί για N ηλικίες κατά τη διάρκεια T χρόνων. Χρησιμοποιώντας το Singular Value Decomposition γράφουμε $M = BK^T$ όπου οι στήλες του B και του K είναι ορθογώνιες. Αυτή η παρουσίαση μπορεί να χρησιμοποιηθεί για να προσεγγίσουμε το M υπό την έννοια των ελαχίστων τετραγώνων. Η καλύτερη προσέγγιση που μπορούμε να πάρουμε είναι $M = B_1 K_1^T$ όπου τα K_1 και B_1 είναι οι στήλες του K και B που ανταποκρίνονται στη μεγαλύτερη ιδιοτιμή. Στην μέθοδο Lee-Carter η προσέγγιση εφαρμόζεται στους ρυθμούς θνησιμότητας που έχουν καταγραφεί. Είναι αξιοσημείωτο το γεγονός

ότι η μονοδιάστατη προσέγγιση είναι εμπειρικά επαρκής για τη θνησιμότητα σε αρκετές χώρες. Για δεδομένα γονιμότητας χρειάζεται προσέγγιση μεγαλύτερης τάξης.

Το δεύτερο βήμα είναι να αναπτύξουμε ένα στοχαστικό μοντέλο για τις παραμέτρους κ με συχνότερη επιλογή ένα απλό τυχαίο περίπατο. Για την πρόβλεψη το B_1 διατηρείται σταθερό και η βέλτιστη πρόβλεψη για το $\kappa_{\tau+t}$ παράγεται από μια ευθεία γραμμή που περνά ανάμεσα από τις τιμές κ_1 και κ_τ . Η πρόβλεψη για το διάστημα των μετασχηματισμένων ρυθμών είναι $\hat{\kappa}_{\tau+t}B_1$.

3.2.3 Προσέγγιση *Poisson*

Παρόλο που η μέθοδος των ελαχίστων τετραγώνων είναι λογική δεν αποτελεί και τη βέλτιστη μέθοδο. Παίρνουμε τα δεδομένα όπως τα έχουμε προαναφέρει με D_{xt} τον αριθμό θανάτων σε ηλικία x κατά τη διάρκεια του χρόνου t και V_{xt} αντίστοιχα τα άτομα που εκτίθενται σε κίνδυνο ανά χρόνο. Χρησιμοποιούμε μοντέλο *Poisson*

$$D_{xt} \sim Po(\mu_{xt}E_{xt})$$

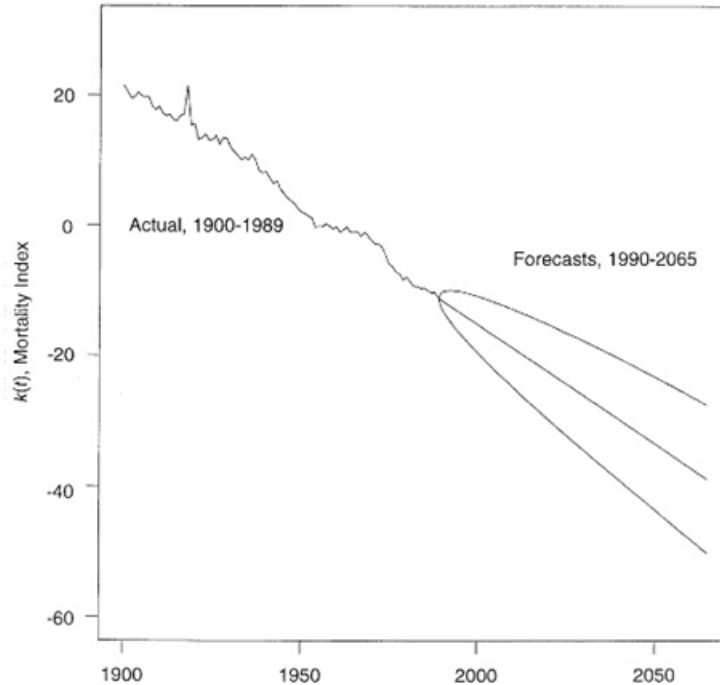
όπου η ένταση είναι λογαριθμο-διγγραμμικής μορφής.

$$\log(\mu_{xt}) = \alpha_x + \beta_x \kappa_t$$

Χρησιμοποιώντας το πρόγραμμα LEM υπολογίζουμε τις εκτιμήσεις των παραμέτρων. Το επόμενο βήμα είναι να μοντελοποιήσουμε το κ σαν στοχαστική χρονοσειρά. Σε πολλές εφαρμογές το κ_t μοντελοποιείται σαν τυχαίος περίπατος : $\kappa_t = c + \kappa_{t-1} + u_t$

Σε αυτή την περίπτωση η πρόβλεψη του κ αλλάζει γραμμικά και κάθε αντίστοιχος ρυθμός θανάτων αλλάζει με σταθερό εκθετικό ρυθμό. Έπειτα προσαρμόζουμε κατάλληλο μοντέλο χρονοσειρών για την πρόβλεψη του κ στην επιθυμητή χρονική περίοδο.

Σε εφαρμογή του μοντέλου για την πρόβλεψη του ρυθμού θνησιμότητας στην Χιλή και στις ΗΠΑ πήραμε προβλέψεις για το δείκτη χρόνου κ . Στη γραφική παράσταση 3.1 βλέπουμε τις τιμές του κ για την περίοδο 1900-1989 όπου έχουμε πάρει τα δεδομένα και τις τιμές για την περίοδο πρόβλεψης από το 1990-2065. Από την γραφική παράσταση παρατηρούμε τη γραμμικότητα που ακολουθούν οι τιμές του κ . Στη συνέχεια μετατρέπουμε τις προβλέψεις για το κ σε προβλέψεις για ρυθμούς θνησιμότητας χρησιμοποιώντας την εξίσωση του $\ln(m_{x,t})$.



Σχήμα.3.1 Πρόβλεψη δείκτη θνησιμότητας για δεδομένα από το 1900-1989 μέχρι το 2065 με 95% διαστήματα εμπιστοσύνης (Lee 2000).

3.2.4 Προβλήματα με τη μέθοδο

Το μοντέλο Lee-Carter είναι χρήσιμο και κατάλληλος τρόπος για να εξάγουμε ιστορικές τάσεις για τη θνησιμότητα με βάση την ηλικία. Ωστόσο υπάρχουν κάποια προβλήματα (Lee 2000) που έχουμε να αντιμετωπίσουμε χρησιμοποιώντας τη μέθοδο.

- Η μέθοδος υποθέτει συγκεκριμένο σχήμα αλλαγής στην κατανομή της ηλικίας στη θνησιμότητα έτσι ώστε οι ρυθμοί στις διάφορες ηλικίες (b_x (dk_t/dt)) να διατηρούν τον ίδιο λόγο κατά τη διάρκεια του χρόνου. Στην πραγματικότητα όμως οι ρυθμοί στις διάφορες ηλικίες μπορούν να ποικίλουν. Για παράδειγμα στη Σουηδία ο ρυθμός θνησιμότητας στις πιο μεγάλες ηλικίες ήταν πιο αργός σε σχέση με άλλες ηλικίες αλλά πιο πρόσφατα κινείται πιο γρήγορα σε σχέση με άλλες ηλικίες.
- Η μέθοδος παρέχει ένα είδος πρόβλεψης για το τι θα συνέβαινε αν οι σημερινές τάσεις θνησιμότητας συνεχιζόνταν. Αν υπήρχε σημαντικός λόγος ούτως ώστε να αναμένουμε οι μακροπρόθεσμες μελλοντικές τάσεις να είναι περισσότερο ή λιγότερο ταχείς η πρόβλεψη θα μπορούσε να τροποποιηθεί κατάλληλα με βάση τις τάσεις θνησιμότητας. Ωστόσο τέτοιοι λόγοι που να επηρεάζουν τις μελλοντικές τάσεις είναι σπάνια διαθέσιμοι.
- Η μέθοδος παρουσιάζει αβεβαιότητα στην πρόβλεψη που προκύπτει κυρίως από την πρόβλεψη του παράγοντα k καθώς και στην εκτιμώμενη

τάση στην θνησιμότητα. Ακόμη αβεβαιότητα προκύπτει και από σφάλματα στον υπολογισμό των συντελεστών b_x αλλά και από την υπόθεση ότι τα προσαρμοσμένα σφάλματα στο βασικό μοντέλο είναι ασυσχέτιστα με βάση την ηλικία.

- Επειδή τα διαστήματα πιθανότητας δεν αντανακλούν αβεβαιότητα για το αν η προδιαγραφή του μοντέλου είναι σωστή ούτε αβεβαιότητα για το αν το μέλλον θα μοιάζει με το παρελθόν υπάρχει διχασμός απόψεων όσον αφορά τα μελλοντικά επίπεδα επέκτασης ζωής. Κάποιοι περιμένουν μεγάλες ανακαλύψεις στη φαρμακοβιομηχανία που θα επιταχύνει την παρακμή της θνησιμότητας ενώ κάποιοι άλλοι ότι θα πολλαπλασιαστούν οι ασθένειες και θα προκαλέσει μείωση της παρακμής.
- Ακόμη ένα πρόβλημα που παρουσιάζει η μέθοδος είναι οι αβάσιμες διαφορές ανάμεσα των φυλών. Σε εφαρμογή δεδομένων για την πρόβλεψη του ρυθμού θνησιμότητας παρατηρήθηκαν αποκλίνουσες τάσεις στα ιστορικά δεδομένα είτε όταν η πρόβλεψη για τα δύο φύλα έγινε ξεχωριστά είτε με κοινό παράγοντα κ .

3.3 Υπόλοιπα

Σε αυτήν την ενότητα θα ασχοληθούμε με τη συμπεριφορά και τα σχήματα των υπολοίπων για να παρατηρήσουμε το μέγεθος των σφαλμάτων πρόβλεψης πιο λεπτομερώς (Lee και Carter, 1992).

3.3.1 Είδη υπολοίπων

- Αν χρησιμοποιήσουμε τη προσέγγιση SVD τα υπόλοιπα μπορούν να ορισθούν ως εξής :

$$e_{xt} = \ln \hat{m}_{xt} - \hat{\alpha}_x - \hat{\beta}_x \hat{\kappa}_t \quad (3.13)$$

και κανονικοποιούνται διαιρώντας με

$$\sum_{x,t} \frac{e_{xt}^2}{(h-2)(k-1)}$$

όπου $h = t_n - t_1 + 1$ και κ_t αντιπροσωπεύει τον αριθμό των ομάδων ηλικίας.

- Μπορούμε να χρησιμοποιήσουμε υπόλοιπα Pearson

$$\frac{d_{xt} - \hat{d}_{xt}}{\sqrt{\hat{d}_{xt}}} \text{ όπου}$$

$\hat{d}_{xt} = e_{xt} \exp(\hat{\alpha}_x + \hat{\beta}_x \hat{\kappa}_t)$ τα οποία έχουν το ίδιο πρόσημο λόγω της ιδιότητας

$$e_{xt} = \ln \frac{d_{xt}}{\hat{d}_{xt}}$$

3.3.2 Η δομή και το μέγεθος των υπολοίπων στην πρόβλεψη θνησιμότητας

Στο βασικό μοντέλο Lee-Carter και στις περισσότερες εφαρμογές του γίνεται περισσότερη αναφορά για τη συμπεριφορά των σφαλμάτων που προέρχονται

από την πρόβλεψη του παράγοντα κ και αγνοούνται τα σφάλματα που προκύπτουν από τους υπόλοιπους παράγοντες. Σε αυτή την ενότητα θα αναφερθούμε πιο λεπτομερώς στη συμπεριφορά των υπολοίπων.

Ορίζουμε το λογάριθμο για τη πρόβλεψη του ρυθμού θνησιμότητας, σε κάθε ηλικία, για s περιόδους πιο πάνω από τη βασική περίοδο t σύμφωνα με την πιο κάτω εξίσωση :

$$\ln m_{x,t+s} = \hat{\alpha}_x + \hat{\kappa}_{t+s}\hat{\beta}_x \quad (3.14)$$

όπου το "" αντιπροσωπεύει μια εκτίμηση των παραμέτρων α_x και β_x για μια πρόβλεψη του παράγοντα κ_t . Η πραγματική τιμή του λογάριθμου $\ln m_{x,t+s}$ θεωρώντας τις προδιαγραφές και τα δεδομένα του μοντέλου σωστά δίνεται από τον πιο κάτω τύπο :

$$\ln m_{x,t+s} = (\hat{\alpha}_x + \alpha_x) + (\hat{\kappa}_{t+s} + u_{t+s})(\hat{\beta}_x + \beta_x) + \varepsilon_{x,t+s} \quad (3.15)$$

όπου τα α_x και β_x είναι τα σφάλματα που προκύπτουν από την εκτίμηση των παραμέτρων a_x και b_x και το u_{t+s} το σφάλμα της πρόβλεψης του παράγοντα κ . Το σφάλμα πρόβλεψης που προκύπτει είναι η διαφορά των δύο πιο πάνω εξισώσεων :

$$E_{x,t+s} = \alpha_x + \varepsilon_{x,t+s} + (\hat{\beta}_x + \beta_x)u_{t+s} + \beta_x\hat{\kappa}_{t+s} \quad (3.16)$$

Στην πιο πάνω εξίσωση υπάρχουν τέσσερις όροι σφαλμάτων οι οποίοι δεν είναι ξεκάθαρο πώς συσχετίζονται μεταξύ τους. Για να αποκτήσουμε μια ιδέα θα υποθέσουμε ότι τα σφάλματα είναι ανεξάρτητα. Υπό αυτή την υπόθεση η διασπορά του σφάλματος $E_{x,t+s}$ δίνεται από τον τύπο :

$$\sigma_{E_{x,t+s}}^2 = \sigma_{\varepsilon_{x,t+s}}^2 + \sigma_{\alpha_x}^2 + \hat{\beta}_x^2 \sigma_{u_{t+s}}^2 + \sigma_{\beta_x}^2 (\hat{\kappa}_{t+s}^2 + \sigma_{u_{t+s}}^2) \quad (3.17)$$

Η διασπορά του όρου $\varepsilon_{x,t+s}$ υπολογίζεται από τη διασπορά του σφάλματος στην προσαρμοσμένη ομάδα ηλικίας x για την αντίστοιχη περίοδο. Το $\hat{\alpha}_x$ είναι ο μέσος όρος του λογάριθμου του ρυθμού θανάτων για κάθε ηλικία x στο χρόνο και η διασπορά του είναι η διασπορά του $\ln(m_{x,t})$ διαιρεμένη με T που είναι ο αριθμός των παρατηρήσεων του m_x .

3.4 Προβλέψεις του μοντέλου

Σε αυτήν την ενότητα θα ασχοληθούμε με τη διαδικασία πρόβλεψης του μοντέλου για μελλοντικούς χρόνους. Για να επιτύχουμε κατάλληλες προβλέψεις μοντελοποιούμε τις τιμές του κ_t χρησιμοποιώντας συνήθεις μεθόδους όπως τη διαδικασία ARIMA (0,1,0). Δηλώνουμε τα αποτελέσματα των προβλέψεων για την παράμετρο του χρόνου ως $\{\hat{\kappa}_{t_n+s} : s > 0\}$ και οι προβλέψεις για το ρυθμό θνησιμότητας υπολογίζονται ως εξής:

$$\hat{m}_{x,t_n+s} = \exp(\hat{\alpha}_x + \hat{\beta}_x \hat{\kappa}_{t_n+s}), \quad s > 0 \quad (3.18)$$

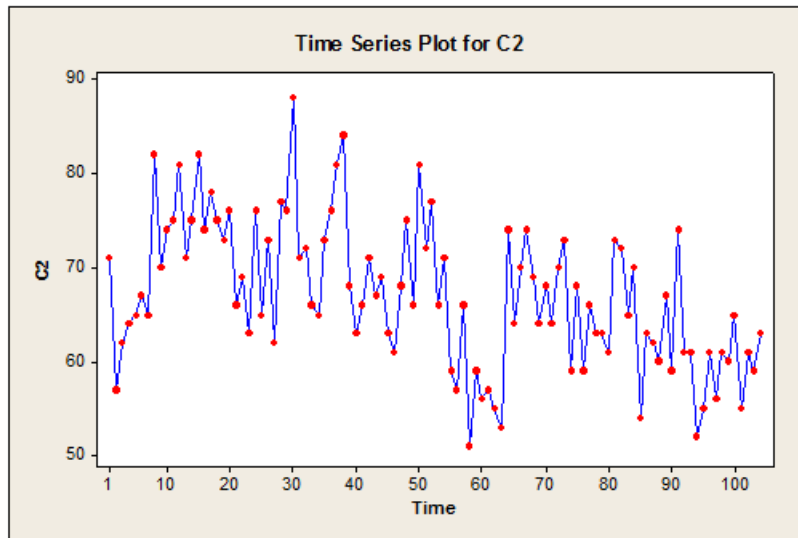
3.4.1 Μοντέλα ARIMA

Η μέθοδος πρόβλεψης χρονοσειρών ARIMA χρησιμοποιείται στις περιπτώσεις όπου τα δεδομένα μας δεν εκδηλώνουν σταθερή συμπεριφορά στο χρόνο (Montgomery, Jennings, Kulahci, 2008) αλλά με τις κατάλληλες μετατροπές, προσθέτοντας διαστήματα που προκύπτουν από τη χρονοσειρά και τα σφάλματα πρόβλεψης για να αφαιρεθούν οποιαδήποτε ίχνη αυτοσυσχέτισης υπάρχουν στο μοντέλο χρονοσειρών που θα εφαρμόσουμε και να γίνει στάσιμο. Η ονομασία προκύπτει από το Auto-Regressive Integrated Moving Average όπου ο όρος Integrated χρησιμοποιείται για την περίπτωση όπου μια μη στάσιμη χρονοσειρά αποτελεί ολοκληρώσιμη μορφή μιας στάσιμης χρονοσειράς.

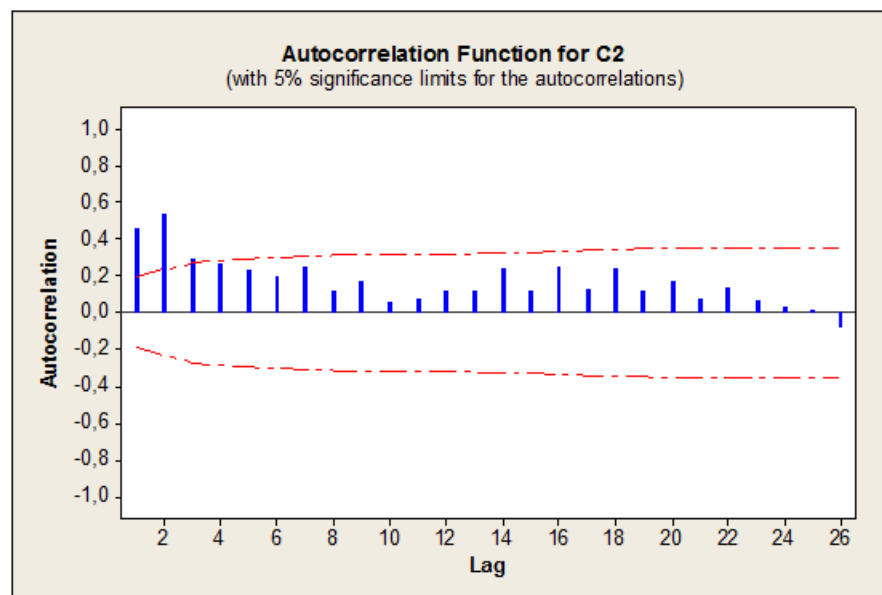
Για να καθορίσουμε το κατάλληλο μοντέλο ARIMA για μια χρονική σειρά αρχίζουμε καθορίζοντας τις τάξεις αφαίρεσης που χρειάζονται για να γίνει στάσιμη η χρονοσειρά και να αφαιρεθούν όλα τα γνωρίσματα εποχικότητας που παρουσιάζει. Αν σταματήσουμε εδώ και θεωρήσουμε ότι η σειρά είναι σταθερή δημιουργούμε ένα απλό τυχαίο περίπατο (για παράδειγμα το μοντέλο ARIMA (0,1,0) αποτελεί τη πιο απλή μορφή τυχαίου περιπάτου), διαφορετικά προσθέτουμε παράγοντες στην εξίσωση πρόβλεψης για να διορθώσουμε τυχόν λάθη που προκύπτουν. Αν τα σφάλματα που προκύπτουν από το τυχαίο περίπατο αυτοσυσχετίζονται το πρόβλημα μπορεί να διορθωθεί προσθέτοντας μια καθυστέρηση στην εξαρτημένη μεταβλητή της εξίσωσης πρόβλεψης και θα είχαμε το μοντέλο ARIMA (1,1,0). Ακόμη σε διάφορες περιπτώσεις ο τυχαίος περίπατος δεν λειτουργεί καλύτερα από τη διαδικασία του κινητού μέσου για πιο παλιές τιμές και εφαρμόζουμε μοντέλο ARIMA (0,1,1) για να διορθώσουμε τυχόν σφάλματα στο μοντέλο πρόβλεψης. Αυτές είναι οι πιο απλές μορφές μοντέλων ARIMA που μπορούν να εφαρμοστούν σε μια σειρά από δεδομένα.

3.4.2 Παράδειγμα μοντέλου ARIMA

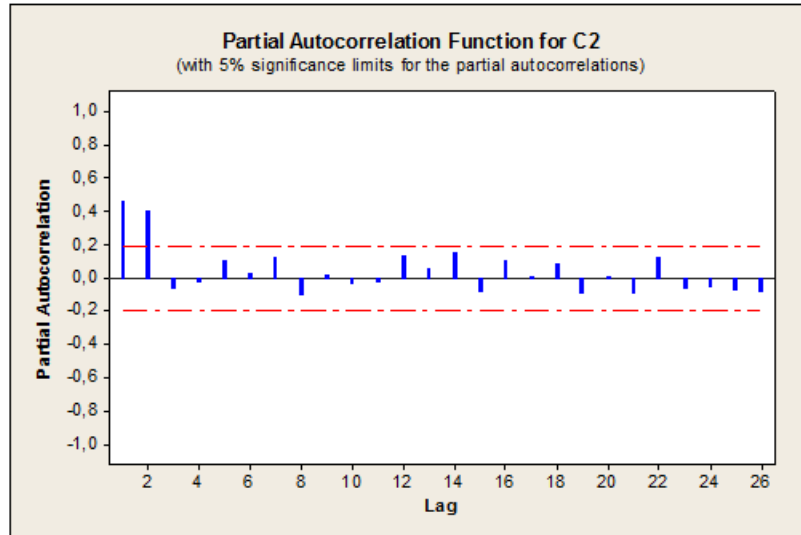
Για την εφαρμογή του μοντέλου ARIMA που θα χρησιμοποιήσουμε στην πρόβλεψη χρησιμοποιούμε το στατιστικό πακέτο MINITAB με το οποίο θα ελέγξουμε πως συσχετίζονται οι τιμές που θέλουμε να προβλέψουμε και πως μπορούμε βρούμε το κατάλληλο μοντέλο που προσαρμόζεται στα δεδομένα μας από τις γραφικές παραστάσεις αυτοσυσχέτισης και μερικής αυτοσυσχέτισης των δεδομένων. Η διαδικασία για την επιλογή του κατάλληλου μοντέλου χρονοσειρών περιγράφεται στο πιο κάτω παράδειγμα για τον εβδομαδιαίο αριθμό αιτήσεων δανείων μιας τράπεζας τα τελευταία δύο χρόνια. Δίνουμε πρώτα τον πίνακα με τον αριθμό των δανείων και έπειτα κατασκευάζουμε τις γραφικές παραστάσεις του αριθμού δανείων(3.2) ανά εβδομάδα, της αυτοσυσχέτισης (3.3) καθώς και της μερικής αυτοσυσχέτισης(3.4) των δεδομένων μας.



Σχήμα.3.2 Γραφική παράσταση για τον αριθμό δανείων ανά εβδομάδα που πραγματοποιούνται.



Σχήμα.3.3 Γραφική παράσταση αυτοσυσχέτισης των δεδομένων.

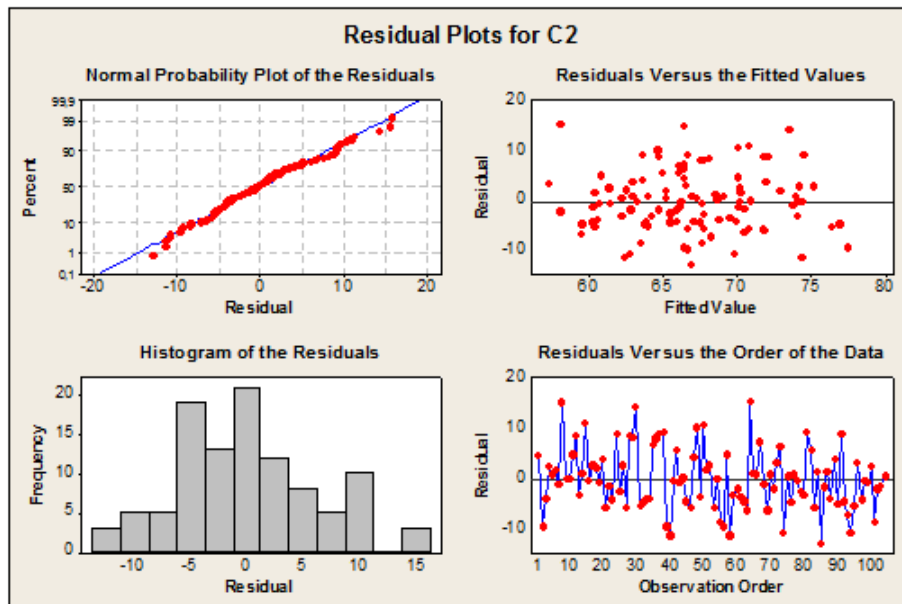


Σχήμα.3.4 Γραφική παράσταση μερικής αυτοσυσχέτισης των δεδομένων.

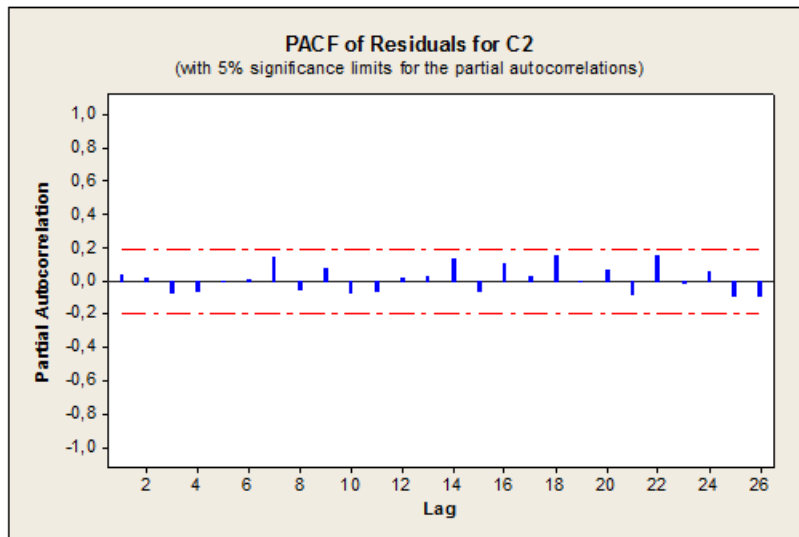
Type	Coef	SE Coef	T	P
AR 1	0,2682	0,0903	2,97	0,004
AR 2	04212	0,0908	4,64	0,000
Constant	20,7642	0,6157	33,73	0,000

Πίνακας.3.7 Αποτελέσματα παραμέτρων μοντέλου ARIMA (2,0,0).

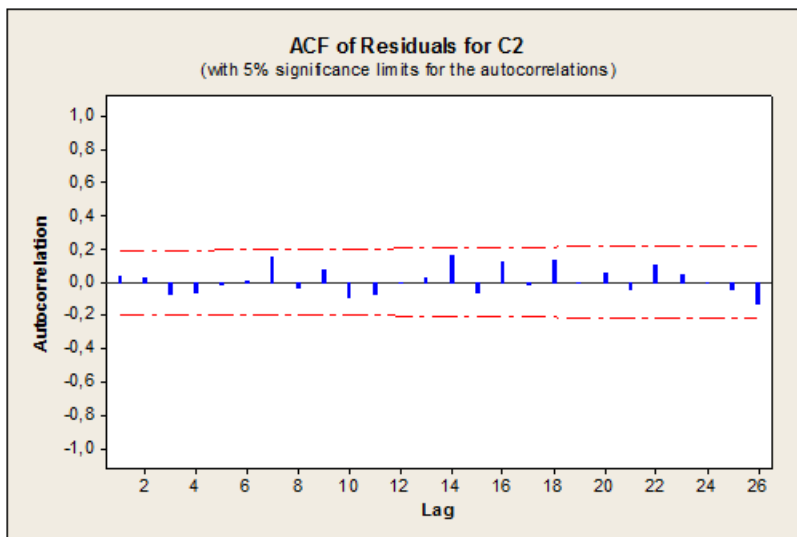
Παρατηρούμε από τις γραφικές παραστάσεις της αυτοσυσχέτισης και μερικής αυτοσυσχέτισης ότι αποκόπτονται μετά τη δεύτερη τιμή. Η ερμηνεία των δύο γραφικών παραστάσεων μας οδηγεί στην χρησιμοποίηση μοντέλου $ARIMA(2,0,0)$. Εκτελούμε την εντολή για το συγκεκριμένο μοντέλο και παίρνουμε τον πίνακα αποτελεσμάτων (3.7) για τις τιμές των παραμέτρων του μοντέλου. Παρατηρούμε ότι και οι δύο παράμετροι καθώς και η σταθερά που υποθέσαμε για την εφαρμογή του μοντέλου είναι στατιστικά σημαντικές με βάση τις p -τιμές. Μπορούμε μετά τον έλεγχο καταλληλότητας να πραγματοποιήσουμε προβλέψεις με για μελλοντικούς αριθμούς δανείων στη τράπεζα.



Σχήμα.3.5 Γραφική παράσταση των υπολοίπων για έλεγχο καταλληλότητας του μοντέλου.



Σχήμα 3.6 Γραφική παράσταση μερικής αυτοσυσχέτισης των υπολοίπων.



Σχήμα 3.7 Γραφική παράσταση αυτοσυσχέτισης των υπολοίπων.

3.5 Παράδειγμα μοντελοποίησης και πρόβλεψης θνησιμότητας

Στην ενότητα αυτή θα ασχοληθούμε με τον ρυθμό θνησιμότητας στις ΗΠΑ και τις προβλέψεις που προκύπτουν χρησιμοποιώντας το αναλογιστικό μοντέλο Lee-Carter. Προσαρμόζουμε τα δεδομένα για την διάρκεια ζωής στο μοντέλο μας ακολουθώντας την διαδικασία για την προσαρμογή του βασικού μοντέλου Lee-Carter και έπειτα χρησιμοποιούμε μεθόδους χρονοσειρών για την πρόβλεψη της παραμέτρου του χρόνου.

3.5.1 Προσαρμογή του μοντέλου

Προσαρμόζουμε τα δεδομένα μας στο βασικό μοντέλο

$$\ln \mu_{xt} = \alpha_x + \beta_x \kappa_t + \varepsilon_{xt}.$$

Η παράμετρος β_x δηλώνει την τάση που ακολουθεί η θνησιμότητα για τις διάφορες ηλικίες με βάση τις αλλαγές της παραμέτρου κ_t . Πιο συγκεκριμένα αν παρατηρηθεί αρνητική τιμή για κάποιες ηλικίες τότε η θνησιμότητα τείνει να αυξηθεί ενώ θετική τιμή προδιαθέτει πτώση της θνησιμότητας. Αξίζει να σημειώσουμε πως όταν το κ είναι γραμμικό ως προς τον χρόνο η θνησιμότητα μεταβάλλεται με εκθετικό ρυθμό και καθώς το κ τείνει αρνητικά στο άπειρο τότε λόγω της εκθετικής μεταβολής ο ρυθμός θα τείνει στο 0. Αυτό συνεπάγεται ότι δεν προσαρμόζονται αρνητικοί ρυθμοί θανάτων στο μοντέλο το οποίο είναι σημαντικό πλεονέκτημα για τις προβλέψεις των επόμενων χρόνων. Στο παράδειγμα μας (Lee και Carter, 1992) έχουν χρησιμοποιηθεί δεδομένα θνησιμότητας από το 1900 ως το 1989 και υπολογιστήκαν προβλέψεις έως το 2065.

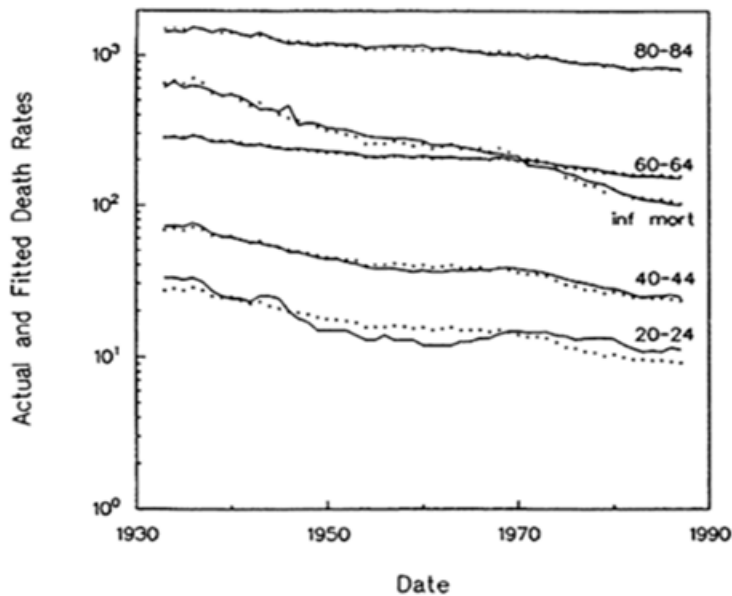
3.5.2 Εκτιμήσεις παραμέτρων

Χωρίζουμε τις ηλικίες σε ομάδες και χρησιμοποιούμε τα δεδομένα ρυθμού θνησιμότητας από το 1933-1987. Στη συνέχεια υπολογίζουμε τις τιμές των παραμέτρων α_x, β_x χρησιμοποιώντας τη μέθοδο *SVD* οι οποίες δίνονται στον πίνακα (3.1).

3.5. ΠΑΡΑΔΕΙΓΜΑ ΜΟΝΤΕΛΟΠΟΙΗΣΗΣ ΚΑΙ ΠΡΟΒΛΕΨΗΣ ΘΝΗΣΙΜΟΤΗΤΑΣ³¹

Age group	α_x	β_x
0	-3.64109	.09064
1-4	-6.70581	.11049
5-9	-7.51064	.09179
10-14	-7.55717	.08358
15-19	-6.76012	.04744
20-24	-6.44334	.05351
25-29	-6.40062	.05966
30-34	-6.22909	.06173
35-39	-5.91325	.05899
40-44	-5.51323	.05279
45-49	-5.09024	.04458
50-54	-4.65680	.03830
55-59	-4.25497	.03382
60-64	-3.85608	.02949
65-69	-3.47313	.02880
70-74	-3.06117	.02908
75-79	-2.63023	.03240
80-84	-2.20498	.03091
85-89	-1.79960	.03091
90-94	-1.40963	.03091
95-99	-1.03655	.03091
100-104	-.68035	.03091
105-109	-.34105	.03091

Πίνακας.3.1 Εκτιμήσεις των παραμέτρων ηλικίας για το μοντέλο (Lee & Carter 1992).



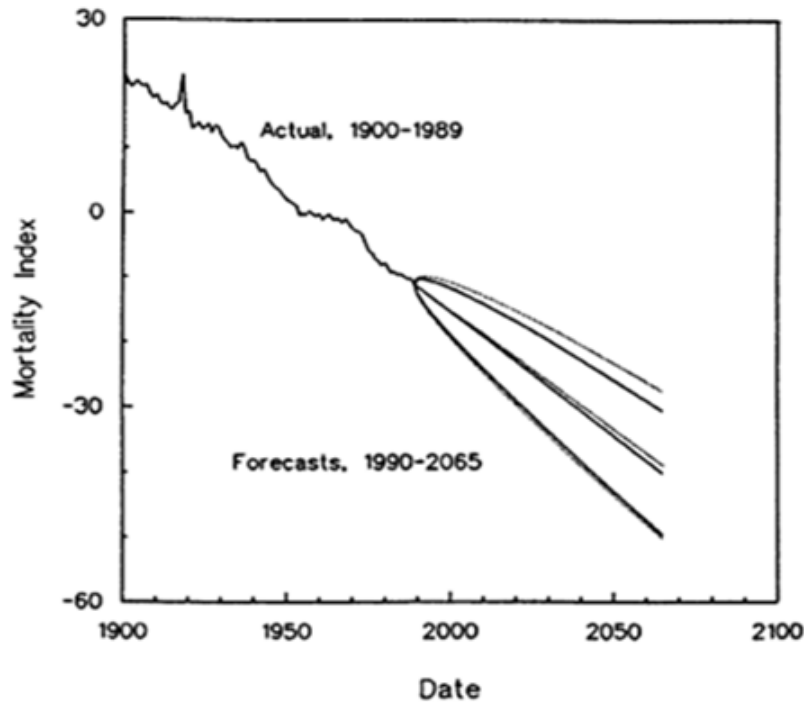
Σχήμα.3.7 Σύγκριση πραγματικού ρυθμού θανάτων και προσαρμοσμένου στο μοντέλο Lee-Carter για συγκεκριμένες ομάδες ηλικιών (Lee & Carter 1992).

Στη γραφική παράσταση (3.7) βλέπουμε το ρυθμό θανάτων για τις ηλικίες από 1933-1987 όπως έχουν καταγραφεί και όπως έχουν υπολογιστεί από το προσαρμοσμένο μοντέλο μας. Παρατηρούμε ότι η προσαρμογή των δεδομένων είναι πολύ καλή με τις μεγαλύτερες αποκλίσεις από την κανονική τιμή να είναι στις περιπτώσεις που οι ρυθμοί θνησιμότητας είναι χαμηλοί.

3.5.3 Μοντελοποίηση και πρόβλεψη της παραμέτρου κ

Μετά την ανάπτυξη και την προσαρμογή του μοντέλου προχωράμε στο πρόβλημα της πρόβλεψης. Αρχικά πρέπει να βρούμε κατάλληλο μοντέλο χρονοσειρών ARIMA για το δείκτη θνησιμότητας κ . Για τον υπολογισμό των προβλέψεων του δείκτη θνησιμότητας χρησιμοποιούμε ARIMA(1,0,0). Το πρόβλημα που παρουσιάζεται στο συγκεκριμένο παράδειγμα είναι πως θα αντιμετωπίσουμε στην προσαρμογή του μοντέλου την επιδημία του 1918 όπου επιφέρει μεγάλο αριθμό θανάτων. Σε αυτή την περίπτωση μπορούμε να χρησιμοποιήσουμε μια μεταβλητή στο μοντέλο πρόβλεψης του δείκτη θνησιμότητας ο οποίος θα αφαιρεί την επίδραση του συγκεκριμένου γεγονότος στην πρόβλεψη ή μπορούμε να εκλάβουμε το γεγονός σαν ένα περιστατικό το οποίο μπορεί να ξανασυμβεί στο μέλλον και να το εκλάβουμε σαν οποιαδήποτε άλλη παρατήρηση. Σε οποιαδήποτε περίπτωση η πρόβλεψη της θνησιμότητας δεν αλλάζει αλλά μεταβάλλεται το διάστημα εμπιστοσύνης της πρόβλεψης.

3.5. ΠΑΡΑΔΕΙΓΜΑ ΜΟΝΤΕΛΟΠΟΙΗΣΗΣ ΚΑΙ ΠΡΟΒΛΕΨΗΣ ΘΝΗΣΙΜΟΤΗΤΑΣ33



Σχήμα.3.8 Σύγκριση προβλέψεων του δείκτη θνησιμότητας μέχρι το 2065 για τα διαστήματα 1900-1989 (τελείες) και 1933-1989 με 95% διαστήματα εμπιστοσύνης (Lee & Carter 1992).

Στο σχήμα (3.8) παρουσιάζουμε τις τιμές του δείκτη θνησιμότητας για τα δεδομένα από το 1900-1989 μαζί με τις προβλέψεις που προκύπτουν εφαρμόζοντας το μοντέλο $ARIMA(0,1,0)$. Η προσαρμογή του μοντέλου έγινε δύο φορές. Η μία πρόβλεψη έγινε για τα δεδομένα από το 1900-1989 χρησιμοποιώντας παράμετρο για αντιμετώπιση της επιδημίας όπως αναφέρθηκε πιο πάνω και η άλλη για δεδομένα από το 1933-1989. Παρατηρούμε οι προβλέψεις είναι παρόμοιες όσον αφορά τις αναμενόμενες τιμές καθώς και τα διαστήματα εμπιστοσύνης. Από τη σύγκριση μπορούμε ακόμη να ελέγξουμε εάν η πρόβλεψη μεταβάλλεται χρησιμοποιώντας πιο σύγχρονα δεδομένα. Στον πιο κάτω πίνακα (3.2) δίνονται οι τιμές των προβλέψεων μέχρι το 2065 μαζί με τα τυπικά σφάλματα.

Date	k	Standard Deviation	Date	K	Standard deviation
1990	-11.41	.65	2028	-25.29	4.07
1991	-11.78	.92	2029	-25.65	4.12
1992	-12.14	1.13	2030	-26.02	4.17
1993	-12.51	1.30	2031	-26.38	4.22
1994	-12.87	1.46	2032	-26.75	4.27
1995	-13.24	1.60	2033	-27.11	4.32
1996	-13.60	1.72	2034	-27.48	4.37
1997	-13.97	1.84	2035	-27.84	4.42
1998	-14.33	1.95	2036	-28.21	4.46
1999	-14.70	2.06	2037	-28.57	4.51
2000	-15.06	2.16	2038	-28.94	4.56
2001	-15.43	2.26	2039	-29.30	4.61
2002	-15.79	2.35	2040	-29.67	4.65
2003	-16.16	2.44	2041	-30.03	4.70
2004	-16.52	2.52	2042	-30.40	4.74
2005	-16.89	2.61	2043	-30.76	4.79
2006	-17.25	2.69	2044	-31.13	4.83
2007	-17.62	2.76	2045	-31.49	4.87
2008	-17.98	2.84	2046	-31.86	4.92
2009	-18.35	2.91	2047	-32.22	4.96
2010	-18.71	2.98	2048	-32.59	5.00
2011	-19.08	3.05	2049	-32.95	5.04
2012	-19.44	3.12	2050	-33.32	5.09
2013	-19.81	3.19	2051	-33.68	5.13
2015	-20.54	3.32	2053	-34.41	5.21
2016	-20.91	3.38	2054	-34.78	5.25
2017	-21.27	3.45	2055	-35.14	5.29
2018	-21.64	3.51	2056	-35.51	5.33
2019	-22.00	3.57	2057	-35.87	5.37
2020	-22.37	3.63	2058	-36.24	5.41
2021	-22.73	3.68	2059	-36.61	5.45
2022	-23.10	3.74	2060	-36.97	5.49
2023	-23.46	3.80	2061	-37.34	5.53
2024	-23.83	3.85	2062	-37.70	5.56
2025	-24.19	3.91	2063	-38.07	5.60
2026	-24.56	3.96	2064	-38.43	5.64
2027	-24.92	4.01	2065	-38.80	5.68

Πίνακας.3.2 Προβλέψεις για το δείκτη θνησιμότητας (Lee & Carter 1992).

Επιπλέον για να ελέγξουμε καλύτερα την επίδραση της περιόδου των δεδομένων στις προβλέψεις έγινε προσαρμογή του μοντέλου χρονοσειρών για διάφορες χρονικές περιόδους και τα αποτελέσματα δίνονται στον πιο κάτω πίνακα (3.3).

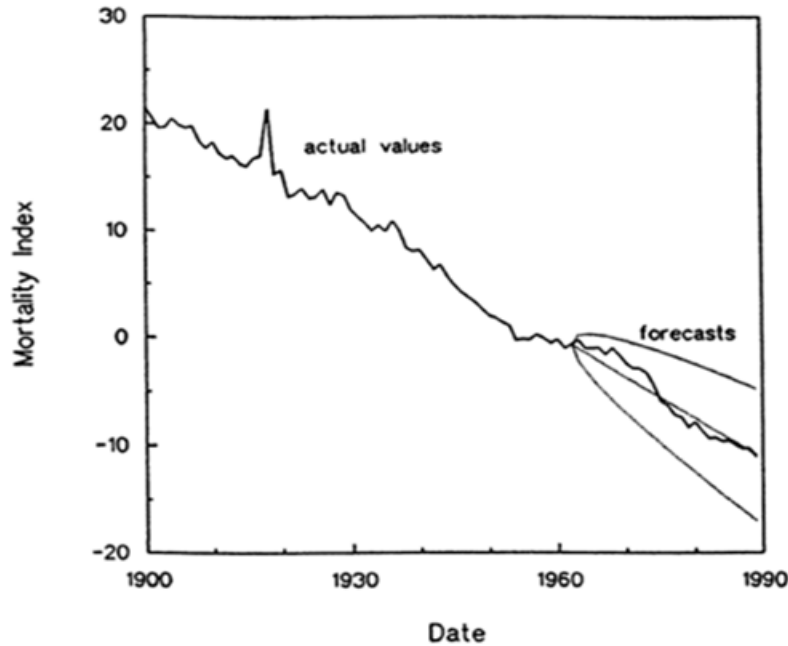
3.5. ΠΑΡΑΔΕΙΓΜΑ ΜΟΝΤΕΛΟΠΟΙΗΣΗΣ ΚΑΙ ΠΡΟΒΛΕΨΗΣ ΘΝΗΣΙΜΟΤΗΤΑΣ³⁵

Starting date K(2065)	Base per length	Constant (drift)	Standard error of constant	Standard error of equation	Forecast K(2065)	Standard error of forecast
1930	60	-.406	.071	.550	-41.9	4.80
1940	50	-.382	.071	.500	-40.1	4.36
1950	40	-.342	.079	.500	-37.1	4.36
1960	30	-.351	.094	.516	-37.7	4.50
1970	20	-.473	.104	.465	-47.0	4.06
1980	10	-.278	.137	.434	-32.2	3.78

Πίνακας 3.3 Προβλέψεις του δείκτη θνησιμότητας για διάφορες περιόδους οι οποίες φτάνουν μέχρι το 1989 (Lee & Carter 1992).

Παρατηρούμε ότι καθώς μεταβάλλουμε το διάστημα των χρόνων που χρησιμοποιούμε για τα δεδομένα μας οι προβλέψεις του κ μεταβάλλονται παρουσιάζοντας αστάθεια ειδικά όταν το διάστημα μειώνεται στα 20 και 10 χρόνια. Πιο συγκεκριμένα στο παράδειγμα παρατηρούμε ότι για το διάστημα 1970-1989 η τιμή του κ μειώνεται κατά μέσο όρο 0.548 κάθε χρόνο σε σχέση με 0.365 για το διάστημα 1900-1989 ενώ στο διάστημα 1980-1989 επανέρχεται στα ίδια επίπεδα και μειώνεται κατά 0.363 κάθε χρόνο.

Τέλος ελέγξαμε τη διαδικασία προσαρμόζοντας το μοντέλο για να προβλέψουμε τιμές τις οποίες γνωρίζαμε. Στο σχήμα 3.9 φαίνονται οι προβλέψεις από μοντέλο που προσαρμόστηκε σε δεδομένα από το 1933-1962 για πρόβλεψη του διαστήματος 1963-1989. Παρατηρούμε ότι η πρόβλεψη είναι πάρα πολύ καλή αφού κυμαίνεται πολύ κοντά στην αναμενόμενη τιμή. Ο πίνακας 3.3 περιλαμβάνει προβλέψεις για ρυθμούς θανάτων. Οι τιμές του πιο πάνω πίνακα προκύπτουν εύκολα με αντικατάσταση των τιμών για τις παραμέτρους α_x και β_x (πίνακας 3.1) καθώς και τις προβλέψεις του δείκτη θνησιμότητας κ_t (πίνακας 3.2).



Σχήμα.3.9 Πρόβλεψη δείκτη θνησιμότητας με προσαρμοσμένο μοντέλο σε δεδομένα από 1933-1962 και πρόβλεψη για 1963-1989 (Lee & Carter 1992).

Age group	1990	1995	2000	2010	2020	2030	2040	2050	2065
0-1	932	790	669	481	345	248	178	128	78
1-4	35	28	23	15	10	7	5	3	2
5-9	19	16	14	10	7	5	4	3	2
10-14	20	17	15	11	8	6	4	3	2
15-19	67	62	57	48	40	34	28	24	18
20-24	86	78	71	58	48	40	33	27	20
25-29	84	75	68	54	44	35	28	23	16
30-34	97	87	78	62	50	40	32	25	18
35-39	138	124	111	90	72	58	47	38	27
40-44	221	201	182	150	124	102	84	69	52
45-49	370	341	315	267	227	193	164	139	109
50-54	613	572	533	464	403	351	305	265	215
55-59	965	907	853	754	666	589	520	460	382
60-64	1,511	1,432	1,357	1,218	1,094	982	882	792	674
65-69	2,233	2,119	2,010	1,810	1,629	1,466	1,320	1,188	1,015
70-74	3,361	3,187	3,022	2,718	2,444	2,198	1,976	1,777	1,515
75-79	4,979	4,693	4,423	3,930	3,491	3,102	2,756	2,448	2,050
80-84	7,748	7,323	6,921	6,182	5,523	4,933	4,407	3,936	3,323
85-89	12,267	11,687	10,609	10,108	9,177	8,331	7,564	6,868	5,942
90-94	19,099	18,341	16,915	16,246	14,987	13,827	12,758	11,774	10,439
95-99	29,744	28,864	27,188	26,390	24,869	23,442	22,102	20,844	19,095
100-04	46,334	45,554	44,053	43,329	41,933	40,600	39,325	38,104	36,364
105-09	72,195	72,100	71,956	71,906	71,845	71,831	71,861	71,930	72,097

Πίνακας.3.4 Προβλέψεις για ρυθμούς θανάτων ανά 100,000 άτομα σε διαστήματα 5 χρόνων (Lee & Carter 1992).

3.5. ΠΑΡΑΔΕΙΓΜΑ ΜΟΝΤΕΛΟΠΟΙΗΣΗΣ ΚΑΙ ΠΡΟΒΛΕΨΗΣ ΘΝΗΣΙΜΟΤΗΤΑΣ37

Age	Date									
	1990	1995	2000	2010	2020	2030	2040	2050	2060	2065
0	100,000	100,000	100,000	100,000	100,000	100,000	100,000	100,000	100,000	100,000
1	99,120	99,253	99,366	99,544	99,672	99,764	99,830	99,878	99,912	99,926
5	98,985	99,143	99,276	99,483	99,631	99,737	99,812	99,866	99,904	99,919
10	98,890	99,062	99,208	99,434	99,596	99,712	99,794	99,853	99,895	99,911
15	98,791	98,977	99,134	99,380	99,556	99,682	99,773	99,837	99,883	99,901
20	98,458	98,671	98,853	99,143	99,357	99,514	99,631	99,718	99,783	99,809
25	98,033	98,285	98,503	98,854	99,118	99,318	99,469	99,585	99,673	99,710
30	97,622	97,915	98,170	98,586	98,902	99,143	99,329	99,472	99,582	99,628
35	97,148	97,490	97,789	98,280	98,657	98,947	99,172	99,346	99,482	99,539
40	96,480	96,888	97,247	97,840	98,301	98,659	98,939	99,158	99,330	99,402
45	95,420	95,921	96,365	97,108	97,694	98,157	98,523	98,814	99,046	99,144
50	93,669	94,297	94,860	95,818	96,590	97,214	97,718	98,128	98,461	98,604
55	90,837	91,636	92,362	93,620	94,661	95,523	96,239	96,835	97,333	97,550
60	86,550	87,565	88,500	90,153	91,556	92,749	93,765	94,632	95,373	95,702
65	80,235	81,500	82,681	84,813	86,673	88,296	89,714	90,953	92,037	92,528
70	71,723	73,275	74,745	77,451	79,872	82,036	83,969	85,694	87,234	87,940
75	60,561	62,419	64,204	67,561	70,643	73,464	76,039	78,384	80,514	81,504
80	47,098	49,256	51,364	55,423	59,256	62,850	66,201	69,311	72,183	73,532
85	31,780	33,972	36,166	40,531	44,821	48,992	53,007	56,840	60,471	62,206
90	16,953	18,681	20,470	24,202	28,091	32,078	36,108	40,131	44,101	46,055
95	6,290	7,220	8,226	10,462	12,983	15,763	18,771	21,968	25,314	27,030
100	1,301	1,569	1,874	2,608	3,520	4,621	5,920	7,418	9,113	10,031
105	104	131	163	248	363	515	710	955	1,254	1,427
110	0	0	0	0	0	0	0	0	0	0

Πίνακας.3.5 Προβλέψεις αριθμού ατόμων που επιζούν σε συγκεκριμένες ηλικίες ανά 100,000 άτομα (Lee & Carter 1992).

Age	Date									
	1990	1995	2000	2010	2020	2030	2040	2050	2060	2065
0	75.83	76.68	77.49	79.04	80.48	81.84	83.13	84.34	85.50	86.05
1	75.50	76.25	76.99	78.40	79.75	81.04	82.27	83.44	84.57	85.12
5	71.60	72.33	73.05	74.45	75.78	77.06	78.28	79.45	80.58	81.12
10	66.66	67.39	68.10	69.48	70.81	72.08	73.30	74.46	75.59	76.13
15	61.73	62.45	63.15	64.52	65.83	67.10	68.31	69.48	70.59	71.13
20	56.93	57.63	58.32	59.67	60.96	62.21	63.40	64.56	65.66	66.20
25	52.16	52.85	53.52	54.83	56.10	57.32	58.50	59.64	60.73	61.26
30	47.37	48.04	48.69	49.98	51.22	52.42	53.58	54.70	55.78	56.31
35	42.59	43.24	43.87	45.12	46.34	47.52	48.66	49.77	50.84	51.36
40	37.87	38.49	39.10	40.31	41.50	42.65	43.77	44.86	45.91	46.42
45	33.26	33.85	34.44	35.60	36.74	37.85	38.94	40.00	41.04	41.54
50	28.83	29.39	29.94	31.04	32.13	33.19	34.24	35.27	36.26	36.75
55	24.64	25.16	25.68	26.71	27.73	28.73	29.73	30.70	31.65	32.12
60	20.73	21.21	21.68	22.63	23.57	24.51	25.44	26.35	27.25	27.68
65	17.16	17.59	18.02	18.89	19.75	20.61	21.47	22.31	23.14	23.54
70	13.88	14.27	14.65	15.43	16.21	16.99	17.75	18.52	19.26	19.63
75	10.96	11.30	11.63	12.31	12.99	13.66	14.33	15.00	15.65	15.98
80	8.36	8.63	8.90	9.44	9.98	10.53	11.08	11.62	12.16	12.42
85	6.18	6.37	6.57	6.97	7.37	7.78	8.18	8.59	9.00	9.20
90	4.46	4.59	4.72	4.99	5.27	5.54	5.82	6.10	6.38	6.52
95	3.10	3.18	3.27	3.43	3.60	3.77	3.93	4.10	4.26	4.34
100	1.91	1.98	2.04	2.16	2.27	2.37	2.46	2.55	2.64	2.68
105	.00	.00	.11	.31	.48	.63	.76	.88	.99	1.04
110	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00

Πίνακας.3.6 Προβλέψεις υπολειπόμενης διάρκειας ζωής για συγκεκριμένες ηλικίες σε διαστήματα 5 χρόνων (Lee & Carter 1992)

Από τις προβλέψεις του ρυθμού θανάτων στον πίνακα (3.4) μπορούμε να υπολογίσουμε πίνακες διάρκειας ζωής. Ο πίνακας 3.5 περιλαμβάνει προβλέψεις για την διάρκεια ζωής έως κάποια ηλικία ανά 100,000 άτομα που γεννιούνται κάθε χρονιά που γίνεται η πρόβλεψη. Για παράδειγμα προβλέπεται ότι το 2060 το 44/100 θα επιζήσει μέχρι την ηλικία των 90 χρόνων. Ο πίνακας 3.6 περιλαμβάνει προβλέψεις για την υπολειπόμενη διάρκεια ζωής σε διαφορετικές ηλικίες και περιόδους. Για παράδειγμα το 1995 ο μέσος όρος υπολειπόμενης διάρκειας ζωής για άτομα 20 χρονών είναι 57,63 χρόνια.

3.5.4 Υπολογισμός αναμενόμενης διάρκειας ζωής και ρυθμών θνησιμότητας

Στο παράδειγμα που αναφέραμε πιο πάνω χρησιμοποιώντας τις προβλέψεις του δείκτη θνησιμότητας δημιουργήσαμε πίνακα προβλέψεων για το ρυθμό θανάτων και ακολούθως πίνακες για την αναμενόμενη διάρκεια ζωής και το ρυθμό θνησιμότητας για ομάδες ηλικιών και χρόνων. Σε αυτή την ενότητα θα αναφερθούμε πιο λεπτομερώς στον υπολογισμό αυτών των πινάκων και ειδικότερα στο σχηματισμό του πίνακα για την αναμενόμενη διάρκεια ζωής. Μπορούμε να ορίσουμε την αναμενόμενη διάρκεια ζωής τον αριθμό των χρόνων που αναμένεται να ζήσει ένα άτομο που γεννιέται σε μια περιοχή ή πληθυσμό υπό την προϋπόθεση ότι οι παρούσες συνθήκες για τον συγκεκριμένο πληθυσμό θα παραμείνουν σταθερές και στο μέλλον.

Για τον υπολογισμό του πίνακα της αναμενόμενης διάρκειας ζωής χρειαζόμαστε το ρυθμό θανάτων ανά ομάδα ηλικίας (στο παράδειγμα ο πίνακας 3.3) και ακολουθούμε την πιο κάτω διαδικασία :

- Ορίζουμε n τον αριθμό των χρόνων για κάθε ηλικία. Για παράδειγμα η ομάδα < 1 έχει $n = 1$ η ομάδα 1-4 έχει $n = 4$ και οι υπόλοιπες $n = 5$
- Η μέση αναλογία της χρονιάς από τα άτομα που πέθαναν την συγκεκριμένη χρονιά. Κατά μέσο όρο υπολογίζεται ότι τα άτομα ζουν περίπου το 0.5 του διαστήματος πριν πεθάνουν με κάποιες εξαιρέσεις όπως την ομάδα < 1 όπου ο μέσος όρος είναι 0.1. Η μέση αναλογία είναι το na_x
- Η πιθανότητα θανάτου (nq_x) όπου το q_x ο ρυθμός θνησιμότητας για κάθε ηλικία και η αντίστοιχη πιθανότητα επιβίωσης (np_x)
- Ο αριθμός ατόμων που είναι ζωντανός στην αρχή κάθε διαστήματος (I_x). Για παράδειγμα στην ηλικία 0 έχουμε ένα δείγμα από 100,000 ζωντανούς, αυτοί που θα είναι ζωντανοί και 1-4 η πιθανότητα επιβίωσης στο προηγούμενο διάστημα πολλαπλασιασμένη με το πληθυσμό που είναι ζωντανός στο προηγούμενο διάστημα ($I_{x-n} * np_{x-n}$)
- Ο αριθμός θανάτων (d_x) κατά τη διάρκεια του διαστήματος που είναι ο αριθμός των ζωντανών στο διάστημα -πληθυσμό στην αρχή του επόμενου διαστήματος ($I_x - I_{x-n}$)

3.5. ΠΑΡΑΔΕΙΓΜΑ ΜΟΝΤΕΛΟΠΟΙΗΣΗΣ ΚΑΙ ΠΡΟΒΛΕΨΗΣ ΘΝΗΣΙΜΟΤΗΤΑΣ39

- Αριθμός ατόμων που έζησαν το διάστημα αυτό (nL_x) και υπολογίζεται από το $n(I_{x+n} + n\alpha_x * nd_x)$
- Συνολικός αριθμός ατόμων που έζησαν μετά το διάστημα (nT_x) δηλαδή το $T_{x+n} + nL_x$
- Τέλος υπολογίζουμε την αναμενόμενη διάρκεια ζωής (e_x) από την έκφραση T_x/I_x

Κεφάλαιο 4

Επεκτάσεις βασικού μοντέλου *Lee – Carter*

Στο κεφάλαιο αυτό θα αναφερθούμε σε μεθόδους επέκτασης του βασικού μοντέλου *Lee-Carter*. Υπάρχουν δύο βασικές επεκτάσεις του βασικού μοντέλου όπου μπορούμε να χρησιμοποιήσουμε διωνυμικό πλαίσιο εφαρμογής του μοντέλου η αρνητικό διωνυμικό μοντέλο.

4.1 Διωνυμικό μοντέλο *Lee – Carter*

Μια εναλλακτική πρόταση στη δομή με σφάλματα *Poisson* που περιγράψαμε προηγουμένως είναι να χρησιμοποιήσουμε διωνυμικό πλαίσιο (Brouhns, Denuit, Vermunt, 2002). Έτσι υποθέτουμε ότι έχουμε διαθέσιμο ένα τετραγωνικό πλαίσιο δεδομένων (d_{xt}, e^i_{xt}) . Τα κενά κελιά δεδομένων καθορίζονται από τον καταμερισμό 0-1 ως εξής : $\omega_{xt}=1$ αν $e^i_{xt} > 0$ $\omega_{xt}=0$ αν $e^i_{xt} = 0$ Στόχος μας είναι η πιθανότητα του θανάτου q_{xt} μοντελοποιώντας τον αριθμό θανάτων σαν ανεξάρτητες διωνυμικές μεταβλητές $D_{xt}:\text{bin}(e^2_{xt}, q_{xt})$ και γράφουμε

$$E(D_{xt}) = e^i_{xt}q_{xt} \quad (4.1)$$

$$\text{Var}(D_{xt}) = \varphi \frac{V(E(D_{xt}))}{\omega_{xt}} \quad (4.2)$$

$$V(u) = u(1 - \frac{u}{e_{xt}}) \quad (4.3)$$

με παράμετρο κλίμακας $\varphi=1$ και χαρακτηριστική συνάρτηση διακύμανσης V . Ενδιαφερόμαστε τότε στη δομή πρόβλεψης παραμέτρων $\eta_{xt}=\alpha_x+\beta_x\kappa_t$, $\sum_x \beta_x=1$, κ_{t_n} η οποία σχετίζεται με την $E(D_{xt})$ με μία από τις παρακάτω 1-1 συναρτήσεις :

1. $\eta_{xt} = \log\{-\log(1 - q_{xt})\} \Leftrightarrow q_{xt} = 1 - \exp(-\exp \eta_{xt})$
2. $\eta_{xt} = \log(\frac{q_{xt}}{1-q_{xt}}) \Leftrightarrow q_{xt} = \frac{\exp \eta_{xt}}{1+\exp \eta_{xt}}$
3. $\eta_{xt} = \log \Phi^{-1}(q_{xt}) \Leftrightarrow q_{xt} \approx 1 - \exp(-\mu_{xt})$

Χρησιμοποιώντας το μη γραμμικό εκτιμητή

$$\mu_{xt} \approx -\log(1-q_{xt}) \Leftrightarrow q_{xt} \approx 1-\exp(-\mu_{xt}) \text{ και τη συνάρτηση (1) προκύπτει}$$

η

$\mu_{xt} = \exp(\alpha_x + \beta_x \kappa_t)$ η οποία είναι η δομή του βασικού Lee-Carter Poisson μοντέλου

4.2 Αρνητικό Διωνυμικό μοντέλο Lee – Carter

Σαν επέκταση του Poisson Lee-Carter μοντέλου ερευνούμε την επίδραση αλλαγής από *Poisson* σε αρνητικά Διωνυμικό μοντέλο χωρίς παράμετρο κλίμακας. Αυτό σημαίνει ότι αντικαθιστούμε τη συνάρτηση διακύμανσης $V(u) = u$ με

$$V(u) = u + \lambda_x u^2 \quad (4.4)$$

Τότε οι μη γραμμικοί Lee-Carter παράμετροι $\{\alpha_x, \beta_x, \kappa_t\}$ και η παράμετρος $\{\lambda_x\}$ υπολογίζονται σε ένα στάδιο μεγιστοποιώντας την κατάλληλη συνάρτηση πιθανότητας.

4.2.1 Παράδειγμα αρνητικού διωνυμικού μοντέλου

Υπολογισμός Παραμέτρων

Παίρνουμε τη συνάρτηση πιθανοφάνειας για παρατηρούμενο αριθμό θανάτων d_{xt} από σύνολο e_{xt} για ηλικίες x και ημερολογιακό έτος t $\prod_{x,t} \frac{\Gamma(d_{xt} + \frac{1}{\kappa})}{\Gamma(\frac{1}{\kappa}) d_{xt}!} \left(\frac{\kappa \delta_{xt}}{1 + \kappa \delta_{xt}}\right)^{d_{xt}} \left(\frac{1}{1 + \kappa \delta_{xt}}\right)^{1/\kappa}$

με $\delta_{xt} = e_{xt} \exp(\alpha_x + \beta_x \kappa_t)$ όπου το δ_{xt} είναι ο αναμενόμενος αριθμός θανά-

των. Η λογαριθμοποιημένη συνάρτηση πιθανοφάνειας δίνεται από τον τύπο

$$L(\alpha, \beta, \kappa, k) = \sum_{x,t} \left(\sum_{i=1}^{d_{xt}} \ln\left(\frac{1}{k} + d_{xt} - i\right) - \ln(d_{xt}!) - \left(d_{xt} + \frac{1}{k}\right) \ln(1 + k\delta_{xt}) + d_{xt} \ln(k\delta_{xt}) \right) \quad (4.5)$$

όπου το $k = \text{Var}[\exp(\varepsilon_{xt})]$

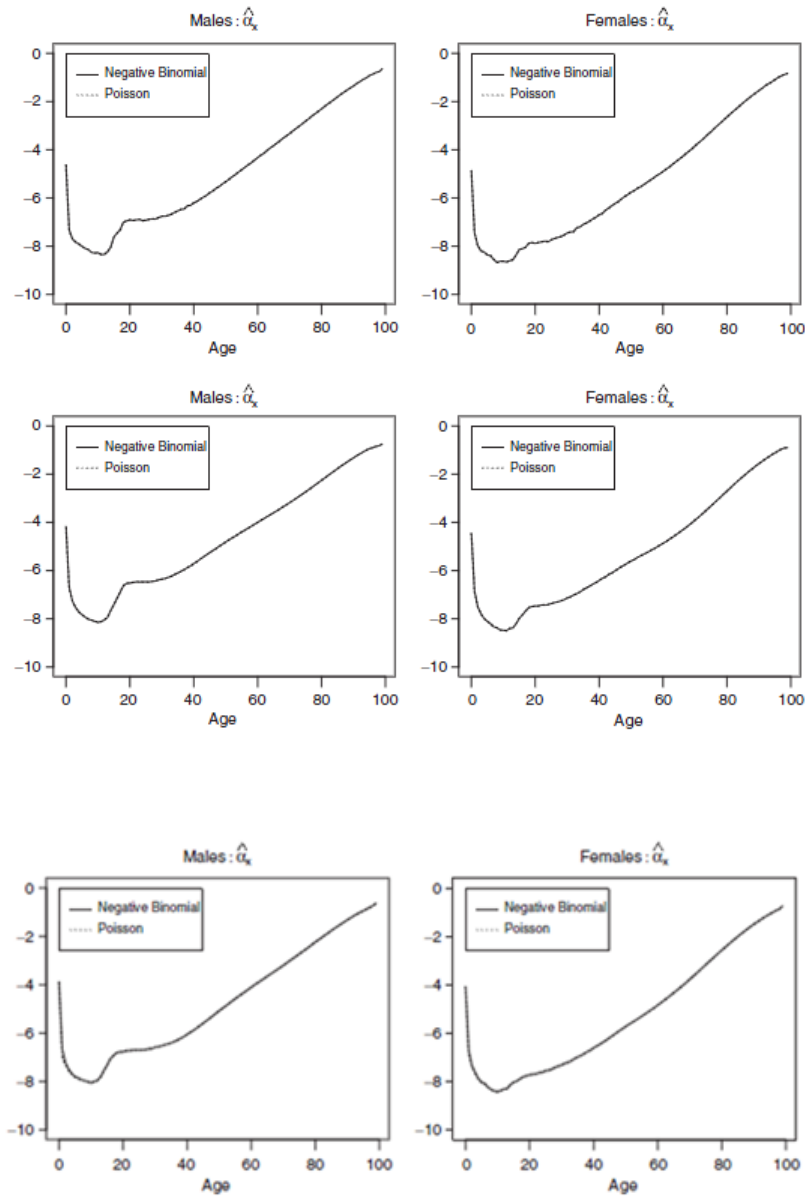
Όπως αναφέραμε και προηγουμένως λόγω του διγραμμικού όρου $\beta_x \kappa_t$ δεν μπορούμε να εκτιμήσουμε το μοντέλο χρησιμοποιώντας συνηθισμένα στατιστικά που εφαρμόζουν την αρνητική Διωνυμική κατανομή. Ο αλγόριθμος που εφαρμόζεται για την επίλυση της εξίσωσης πιθανοφάνειας είναι της μεθόδου *Newton* και χρησιμοποιούμε τον ίδιο αναδρομικό τύπο που αναφέραμε στο κεφάλαιο 3.

Αριθμητικό παράδειγμα

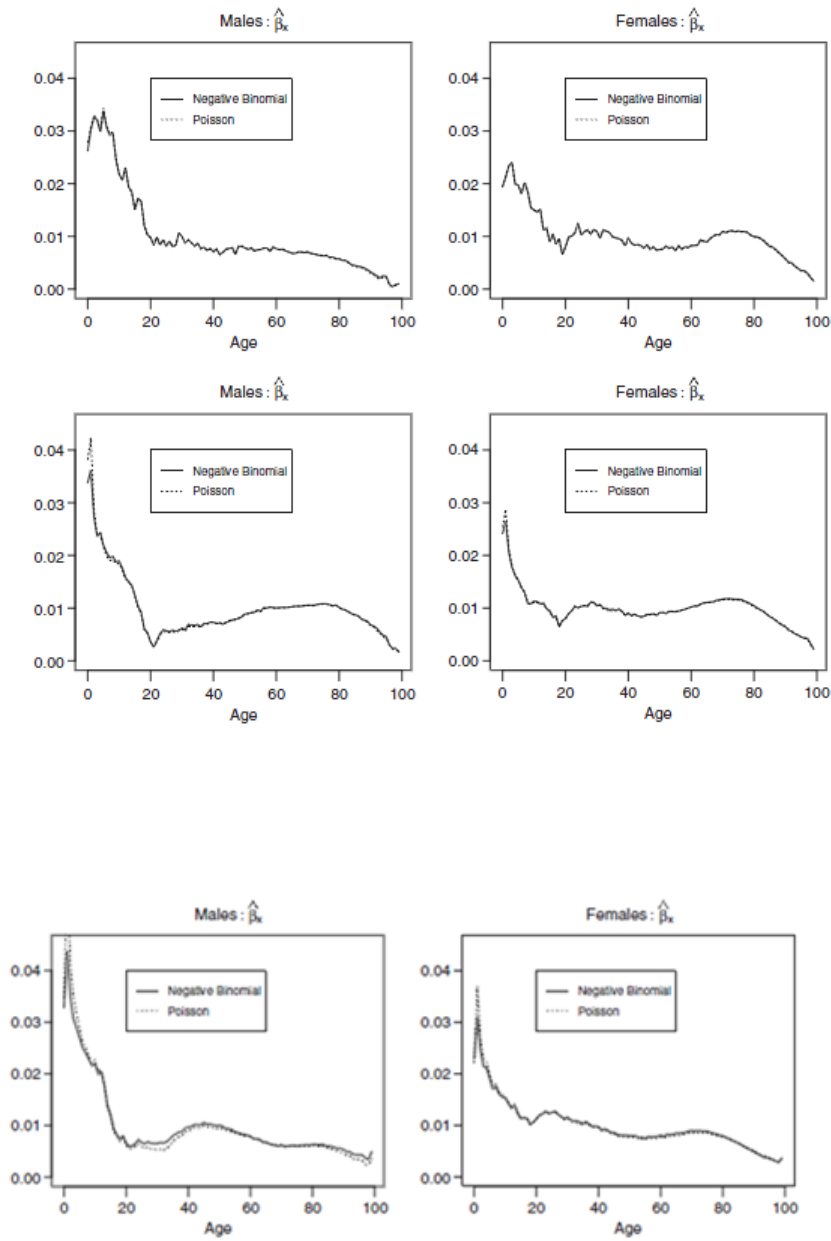
Για την εφαρμογή του αρνητικού διωνυμικού μοντέλου χρησιμοποιήθηκαν δεδομένα (Delwarde, Denuit, Partrat, 2007), (Brouhns, Denuit, Vermunt, 2002) από τους πληθυσμούς της Σουηδίας, της Γαλλίας και της Ιταλίας. Ο λόγος που επιλέγηκαν αυτές οι χώρες για την εφαρμογή είναι για να συγκρίνουμε το ρυθμό θνησιμότητας σε χώρες με διαφορετική οικονομική και κοινωνική κατάσταση. Στα δεδομένα εφαρμόστηκε και η κατανομή *Poisson* για σύγκριση

της εφαρμογής με τις δύο μεθόδους. Στις γραφικές παραστάσεις (4.1, 4.2, 4.3) δίνονται οι εκτιμήσεις των παραμέτρων σε σχέση με τα χρόνια διαχωρίζοντας το φύλο σε κάθε χώρα.

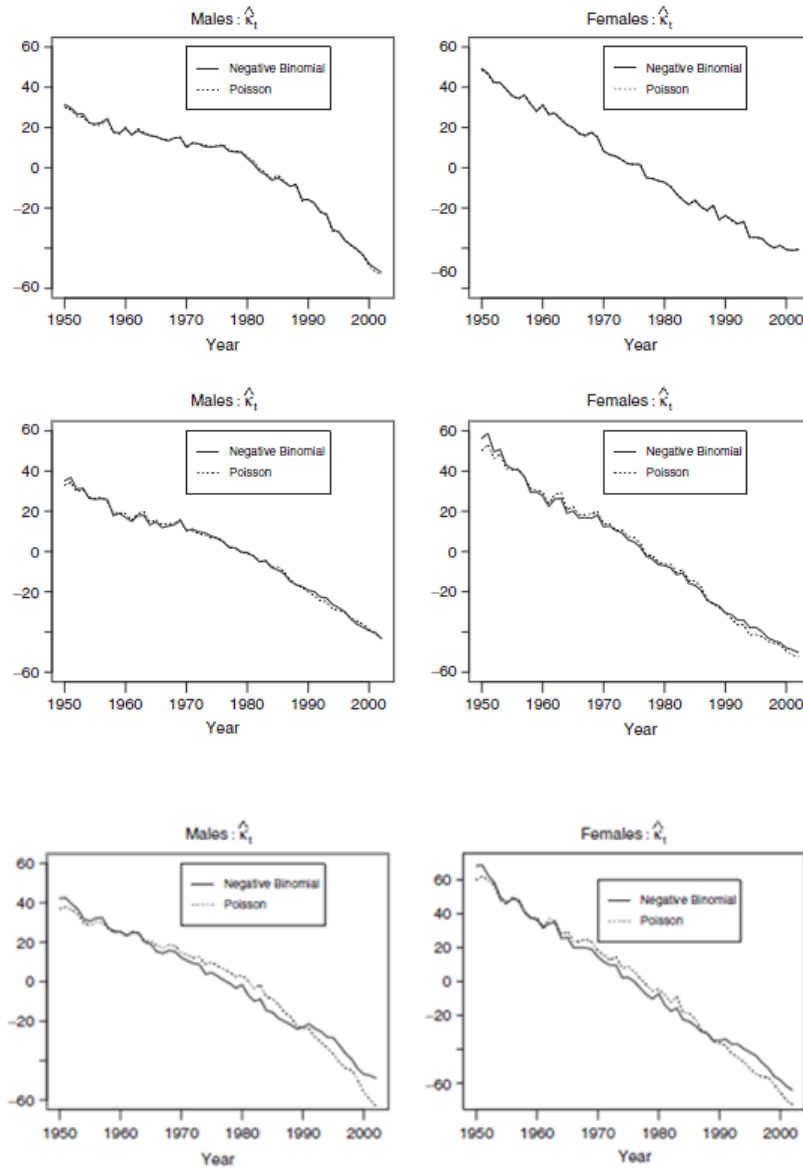
44ΚΕΦΑΛΑΙΟ 4. ΕΠΕΚΤΑΣΕΙΣ ΒΑΣΙΚΟΥ ΜΟΝΤΕΛΟΥ LEE-CARTER



Σχήμα.4.1 Εκτιμήσεις των παραμέτρων α_x εφαρμόζοντας τα μοντέλα της Poisson και της αρνητικής Διωνυμικής διαχωρίζοντας τα φύλα (άνδρες αριστερά και γυναίκες δεξιά) για τη Σουηδία (πάνω), Γαλλία (μέση) και Ιταλία (Delwarde, Denuit, Partrat, 2007).



Σχήμα.4.2 Εκτιμήσεις των παραμέτρων β_x εφαρμόζοντας τα μοντέλα της Poisson και της αρνητικής Διωνυμικής διαχωρίζοντας τα φύλα (άνδρες αριστερά και γυναίκες δεξιά) για τη Σουηδία (πάνω), Γαλλία (μέση) και Ιταλία (Delwarde, Denuit, Partrat, 2007).



Σχήμα.4.3 Εκτιμήσεις των παραμέτρων κ_t εφαρμόζοντας τα μοντέλα της Poisson και της αρνητικής Διωνυμικής διαχωρίζοντας τα φύλα (άνδρες αριστερά και γυναίκες δεξιά) για τη Σουηδία (πάνω), Γαλλία (μέση) και Ιταλία (Delwarde, Denuit, Partrat, 2007).

Παρατηρούμε ότι οι εκτιμήσεις της παραμέτρου α_x είναι σχεδόν οι ίδιες και για τις δύο μεθόδους κάτι που αναμενόταν αφού τα α_x αντιπροσωπεύουν ένα μέσο επίπεδο θνησιμότητας σε ηλικία x για μια περίοδο που μελετάμε. Για τα β_x παρατηρούμε ότι υπάρχουν κάποιες διαφορές ανάμεσα στην *Poisson* και την αρνητική Διωνυμική ειδικά σε νεαρές ηλικίες. Στα κ_t παρατηρούμε μεγαλύτερες διαφοροποιήσεις οι οποίες αξίζει να επισημάνουμε είναι πιο έντονες στην Ιταλία σε σχέση με την Γαλλία και στην Γαλλία σε σχέση με την Σουηδία. Στη συνέχεια χρησιμοποιήθηκε κατάλληλο μοντέλο ARIMA για την εκτίμηση του δείκτη χρόνου κ_t και χρησιμοποιούμε τις τιμές αυτές για να προσδιορίσουμε μελλοντικούς ρυθμούς θνησιμότητας και επεκτάσεις ζωής σε συγκεκριμένες ηλικίες. Το μοντέλο που χρησιμοποιήθηκε δίνεται από την πιο κάτω εξίσωση.

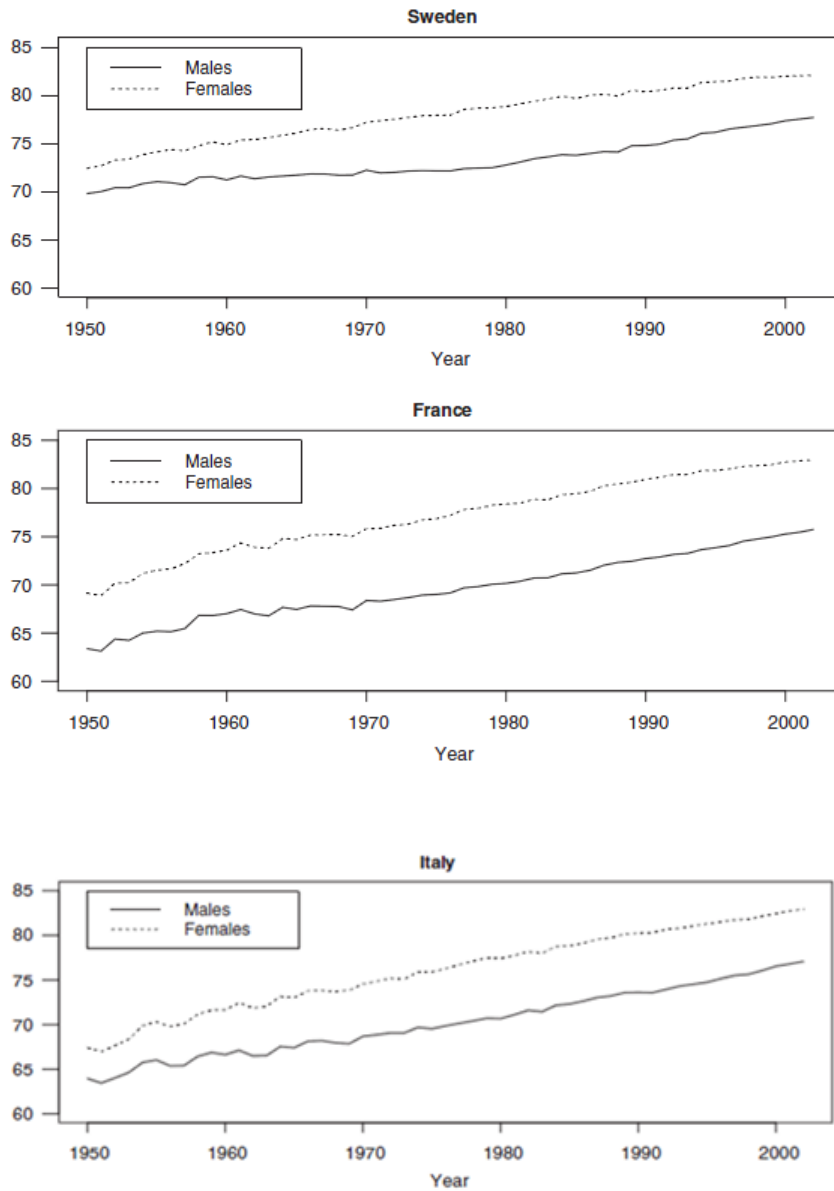
$$\hat{\kappa}_t = \hat{\kappa}_{t-1} + \theta + \varepsilon_t \quad (4.6)$$

όπου $\varepsilon_t \sim N(0, s_\varepsilon^2)$. Στον πιο κάτω πίνακα (4.1) δίνονται οι τιμές των παραμέτρων του μοντέλου ARIMA

	Males		Females	
	$\hat{\theta}$	$\hat{\sigma}_\varepsilon^2$	$\hat{\theta}$	$\hat{\sigma}_\varepsilon^2$
Sweden				
Poisson	-1.605	2.528	-1.710	2.471
Negative binomial	-1.607	2.206	-1.731	2.396
France				
Poisson	-1.456	1.928	-1.968	2.851
Negative binomial	-1.510	2.017	-2.049	2.575
Italy				
Poisson	-1.921	2.071	-2.552	3.187
Negative binomial	-1.753	1.978	-2.551	2.799

Πίνακας 4.1 Εκτιμήσεις παραμέτρων μοντέλου ARIMA (Delwarde, Denuit, Partrat, 2007).

Στο σχήμα 4.4 δείχνουμε τη διάρκεια ζωής για τις παρατηρούμενες τιμές που χρησιμοποιήθηκαν στην εφαρμογή του μοντέλου.



Σχήμα.4.4 Παρατηρούμενες αναμενόμενες τιμές διάρκειας ζωής στις 3 χώρες (Delwarde, Denuit, Partrat, 2007).

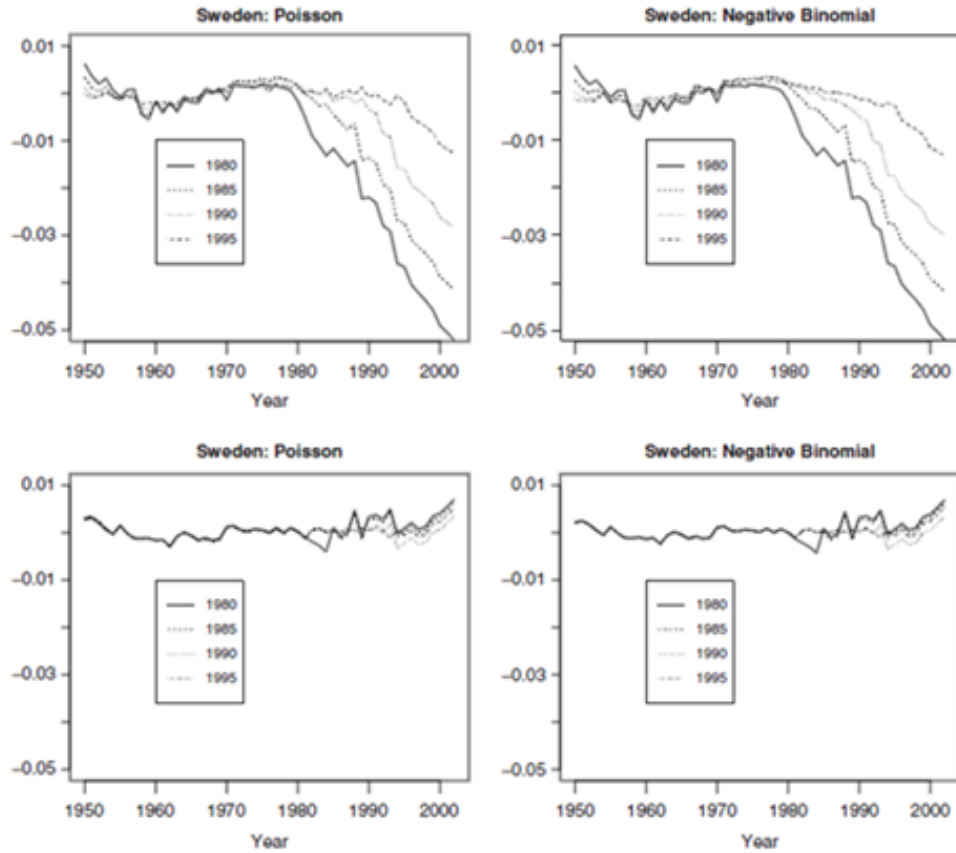
Για να επιβεβαιωθεί η καταλληλότητα του μοντέλου έγινε πρόβλεψη για τις

πιο πρόσφατες τιμές χρησιμοποιώντας τις προηγούμενες για την εφαρμογή του μοντέλου. Πιο συγκεκριμένα χρησιμοποιήσαμε τα δεδομένα για διάφορα διαστήματα και προβλέψαμε την αναμενόμενη διάρκεια ζωής (διαστήματα δεδομένων 1950-1980, 1950-1985, 1950-1990, 1950-1995). Στον πίνακα (4.2) δείχνουμε τις τιμές των παραμέτρων από την εφαρμογή του μοντέλου ARIMA για τον υπολογισμό προβλέψεων του δείκτη θνησιμότητας και κατ' επέκταση των αναμενόμενων τιμών διάρκειας ζωής.

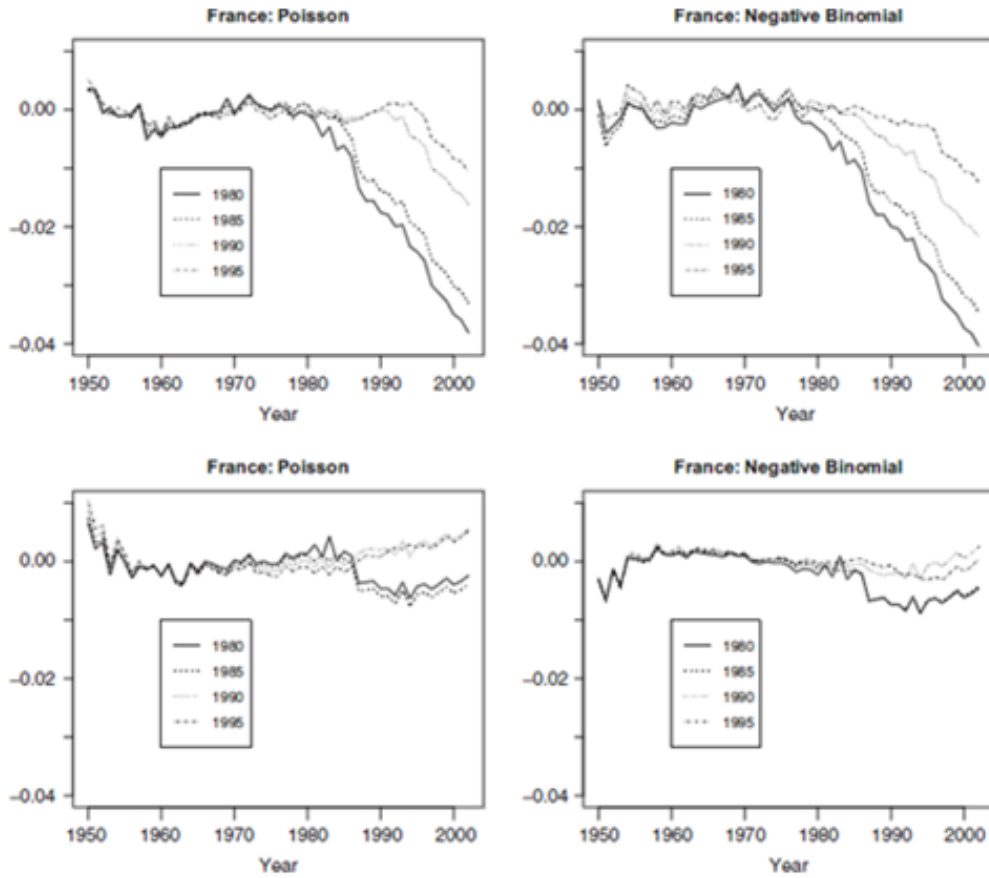
	Period to	Males		Females	
		$\hat{\theta}$	$\hat{\sigma}_\varepsilon^2$	$\hat{\theta}$	$\hat{\sigma}_\varepsilon^2$
Sweden					
Poisson	1980	-0.802940608	1.367189192	-1.796568864	2.189390714
	1985	-1.010694671	1.867350562	-1.830799681	2.233097477
	1990	-1.204292934	2.509530103	-1.766260812	2.416446267
	1995	-1.465140515	2.783170993	-1.809227715	2.536918384
Negative binomial	1980	-0.810631331	1.325141916	-1.801872654	2.150541349
	1985	-1.010079681	1.713852004	-1.838709576	2.191361542
	1990	-1.185704288	2.166841863	-1.775673181	2.381662615
	1995	-1.472297135	2.417020616	-1.82323332	2.474396276
France					
Poisson	1980	-1.037818681	1.590034651	-2.063347037	3.340662721
	1985	-1.072585893	1.674204881	-1.967351717	3.186544518
	1990	-1.269064816	1.830303772	-2.099930388	3.102537187
	1995	-1.311002364	1.861464538	-2.038829553	3.00303587
Negative binomial	1980	-1.01685612	1.726567609	-2.096128912	2.980753758
	1985	-1.072747022	1.818628627	-2.037485736	2.846374143
	1990	-1.233529291	1.958799363	-2.133116425	2.767190514
	1995	-1.331058086	1.990521416	-2.062518593	2.698158916
Italy					
Poisson	1980	-1.382736469	1.280518922	-2.528822941	3.655572918
	1985	-1.614653254	1.818817761	-2.613841952	3.864578567
	1990	-1.746072609	2.035600494	-2.691513437	3.711897958
	1995	-1.739056115	1.993346043	-2.591434619	3.432491285
Negative binomial	1980	-1.349392349	1.324787543	-2.516885908	2.993310291
	1985	-1.557005717	1.641850436	-2.649765828	3.094455158
	1990	-1.52100111	1.821675111	-2.616276795	2.989201355
	1995	-1.509925366	1.872936373	-2.455823827	2.925189661

Πίνακας.4.2 Εκτιμήσεις των παραμέτρων του μοντέλου ARIMA για τα διαστήματα των παρατηρούμενων τιμών (Delwarde, Denuit, Partrat, 2007).

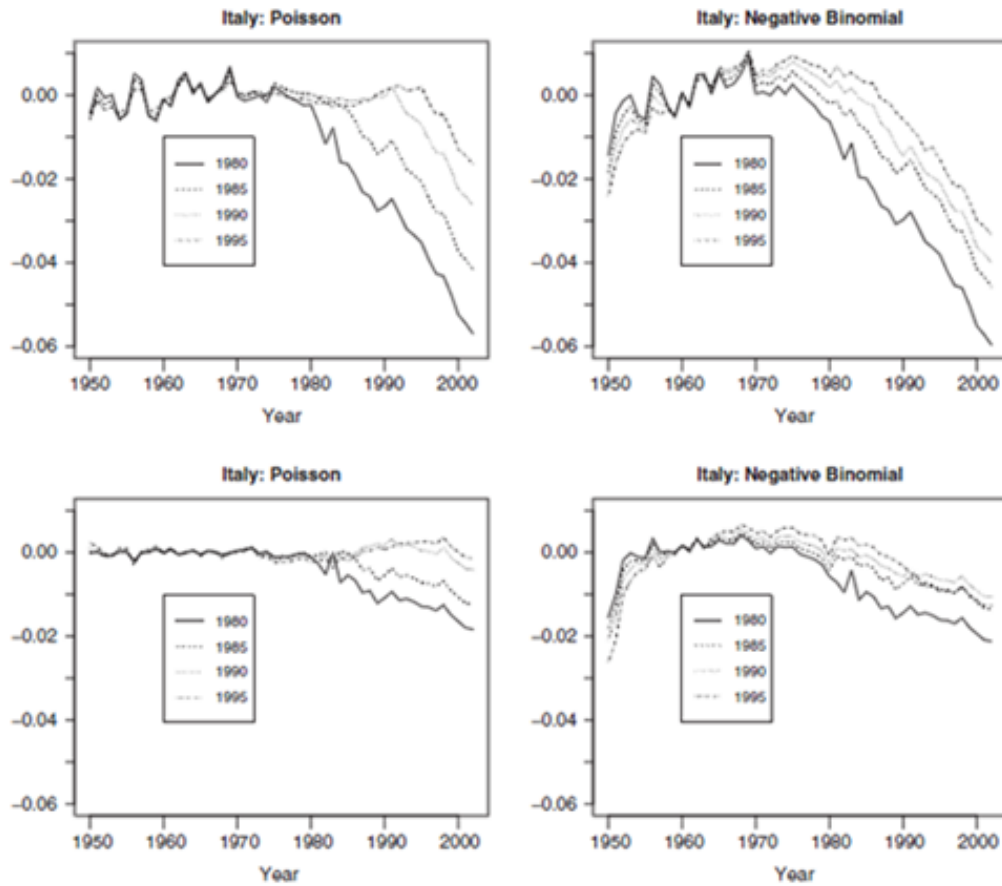
Για να ελέγξουμε την ικανότητα πρόβλεψης του μοντέλου δημιουργούμε γραφικές παραστάσεις και για τις τέσσερις περιόδους που χρησιμοποιήσαμε για να συγκρίνουμε τις τιμές που προκύπτουν. Παρατηρούμε ότι η προσαρμογή του μοντέλου είναι καλύτερη όταν χρησιμοποιούμε το μοντέλο της αρνητικής Διωνυμικής ωστόσο οι ικανότητες πρόβλεψης και των δύο μοντέλων είναι κατά προσέγγιση οι ίδιες. Ακόμη η ακρίβεια στην πρόβλεψη είναι ικανοποιητική αφού οι διαφορές ανάμεσα στις τιμές πρόβλεψης και στις παρατηρούμενες τιμές είναι μικρότερες της τάξης του 0.05.



Σχήμα.4.5 Σχετικές διαφορές ανάμεσα στη παρατηρούμενη διάρκεια ζωής και στη πρόβλεψη που προκύπτει από την εφαρμογή του μοντέλου Poisson (αριστερά) και αρνητικής διωνυμικής (δεξιά) για τους άνδρες (πάνω) και τις γυναίκες (κάτω) (Delwarde, Denuit, Partrat, 2007).



Σχήμα.4.6 Σχετικές διαφορές ανάμεσα στη παρατηρούμενη διάρκεια ζωής και στη πρόβλεψη που προκύπτει από την εφαρμογή του μοντέλου Poisson (αριστερά) και αρνητικής διωνυμικής (δεξιά) για τους άνδρες (πάνω) και τις γυναίκες (κάτω) (Delwarde, Denuit, Partrat, 2007).



Σχήμα.4.7 Σχετικές διαφορές ανάμεσα στη παρατηρούμενη διάρκεια ζωής και στη πρόβλεψη που προκύπτει από την εφαρμογή του μοντέλου Poisson (αριστερά) και αρνητικής διωνυμικής (δεξιά) για τους άνδρες (πάνω) και τις γυναίκες (κάτω) (Delwarde, Denuit, Partrat, 2007).

Υπάρχουν αρκετές επεκτάσεις του βασικού μοντέλου για το διαχωρισμό διάφορων παραμέτρων που θέλουμε να αναλύσουμε:

- Διαχωρισμός των φύλων για τον οποίο υπάρχουν διάφορες προσεγγίσεις (Lee και Carter, 1992). Μπορούμε να θεωρήσουμε τις προβλέψεις για τους άνδρες και τις γυναίκες ενός πληθυσμού σαν δύο ξεχωριστές εφαρμογές της βασικής προσέγγισης του μοντέλου. Ωστόσο μπορούμε να εκμεταλλευτούμε τις όποιες ομοιότητες παρουσιάζουν τα δύο φύλα για να μειώσουμε τη διαστατικότητα του μοντέλου. Για να το πετύχουμε αυτό μπορούμε να συμπεριφερθούμε και στα δύο φύλα σαν να είναι διαφορετικές ομάδες του ίδιου πληθυσμού με $2N$ ομάδες ηλικίας αν ο πραγματικός αριθμός ομάδων ηλικίας είναι N . Με αυτό τον τρόπο εκτιμούμαι και προβλέπουμε μια σειρά για το δείκτη θνησιμότητας k και υπολογίζουμε την θνησιμότητα h οποία θα διαφέρει μεταξύ των δύο φύλων λόγω των διαφορετικών τιμών των παραμέτρων α_x και β_x .
- Μπορούμε να χρησιμοποιήσουμε την πιο πάνω διαδικασία για να κάνουμε και διαχωρισμό με βάση τη γεωγραφική θέση των διάφορων πληθυσμών (Lee και Nault, 1993).
- Μια άλλη επέκταση του βασικού μοντέλου είναι ο διαχωρισμός με βάση την αιτία θανάτου (Wilmoth, 1995). Η πρόβλεψη του μοντέλου με βάση αυτό το διαχωρισμό αποδίδει ψηλότερες θνησιμότητες στο μέλλον σε σχέση με τη συνολική πρόβλεψη επειδή η αιτία θανάτου που φθίνει πιο αργά θα υπερισχύσει μακροπρόθεσμα και αυξάνει τον αριθμό των θανάτων.
- Για τον υπολογισμό των προβλέψεων του δείκτη θνησιμότητας k μπορούμε να θεωρήσουμε τον όρο της τάσης σαν τυχαία μεταβλητή (Carter 1996).

Κεφάλαιο 5

Εφαρμογή μοντέλου *Lee – Carter*

Στο κεφάλαιο αυτό θα ασχοληθούμε με τη λεπτομερή εφαρμογή του αναλογιστικού μοντέλου Lee-Carter για να μελετήσουμε το ρυθμό αυτοκινητιστικών δυστυχημάτων και να προβλέψουμε μελλοντικές τιμές ρυθμών χρησιμοποιώντας μεθόδους χρονοσειρών όπως σε προηγούμενα παραδείγματα και να καταλήξουμε σε συμπεράσματα για τη λειτουργία του μοντέλου.

5.1 Δεδομένα

Στην εφαρμογή όπως αναφέραμε θα χρησιμοποιήσουμε δεδομένα που αφορούν το ρυθμό αυτοκινητιστικών δυστυχημάτων. Έχουμε πάρει δεδομένα από ασφαλιστική εταιρεία για τον αριθμό των δυστυχημάτων που έχουν πραγματοποιηθεί από το 2006 έως το 2010. Για καλύτερη και πιο ακριβή εφαρμογή του μοντέλου χρησιμοποιήσαμε τα δεδομένα μας με βάση τον αριθμό των δυστυχημάτων ανά εξάμηνο και όχι ανά χρόνο όπως χρησιμοποιήθηκε για τον ρυθμό θνησιμότητας στα προηγούμενα παραδείγματα. Πιο κάτω δίνονται οι πίνακες με τα δεδομένα που θα χρησιμοποιήσουμε για την εφαρμογή. Ο πίνακας (5.1) περιλαμβάνει τον αριθμό των δυστυχημάτων (αντίστοιχα είναι ο αριθμός θανάτων για ρυθμό θανάτων) ανά εξάμηνο που καταγράφηκε στην εταιρεία και ο πίνακας (5.2) ο αριθμός των ασφαλιζομένων ανά ομάδα ηλικίας σε κάθε εξάμηνο στην εταιρεία.

ΑΡΙΘΜΟΣ ΑΤΥΧΗΜΑΤΩΝ ΑΝΑ ΕΞΑΜΗΝΟ										
ΗΛΙΚΙΑ\ΕΤΟΣ	A'2006	B'2006	A'2007	B'2007	A'2008	B'2008	A'2009	B'2009	A'2010	B'2010
17-22	315	375	323	389	297	379	287	381	263	360
23-28	414	523	456	542	476	556	451	561	395	487
29-34	324	402	323	407	391	485	380	456	383	452
35-40	318	411	320	398	355	459	325	415	292	373
41-46	342	402	327	388	378	467	344	423	322	386
47-52	278	343	299	355	314	383	334	398	301	357
53-58	222	264	200	239	244	285	243	292	255	303
59-64	132	168	157	201	182	234	202	262	165	205
65-70	86	122	88	116	97	127	98	128	117	152
71-76	38	58	47	71	51	79	63	83	69	90
77-82	14	25	15	26	16	29	21	33	19	31
83-88	3	5	5	9	5	5	2	7	4	11

Πίνακας.5.1 Αριθμός ατυχημάτων ανά εξάμηνο για τα έτη 2006-2010.

ΑΡΙΘΜΟΣ ΑΣΦΑΛΙΖΟΜΕΝΩΝ ΑΝΑ ΕΞΑΜΗΝΟ										
ΗΛΙΚΙΑ\ΕΤΟΣ	A'2006	B'2006	A'2007	B'2007	A'2008	B'2008	A'2009	B'2009	A'2010	B'2010
17-22	10235	10235	10204	10204	10218	10218	10114	10114	10081	10081
23-28	21195	21195	21249	21249	21279	21279	21365	21365	21451	21451
29-34	20790	20790	20830	20830	20930	20930	21044	21044	21105	21105
35-40	17779	17779	17813	17813	17873	17873	17895	17895	17948	17948
41-46	17628	17628	17675	17675	17716	17716	17758	17758	17794	17794
47-52	18587	18587	18634	18634	18683	18683	18728	18728	18769	18769
53-58	15621	15621	15657	15657	15697	15697	15729	15729	15753	15753
59-64	12433	12433	12463	12463	12476	12476	12501	12501	12513	12513
65-70	7833	7833	7850	7850	7856	7856	7869	7869	7878	7878
71-76	4946	4946	4963	4963	4965	4965	4980	4980	4984	4984
77-82	1495	1495	1503	1503	1500	1500	1506	1506	1500	1500
83-88	238	238	292	292	298	298	299	299	302	302

Πίνακας.5.2 Αριθμός ασφαλιζομένων ανά εξάμηνο για τα έτη 2006-2010.

5.2 Μοντελοποίηση και εφαρμογή του μοντέλου Lee – Carter

5.2.1 Μοντέλο

Όπως αναφέραμε σε προηγούμενο κεφάλαιο ο αριθμός ατυχημάτων αποτελεί διακριτή τυχαία μεταβλητή και για την εφαρμογή του μοντέλου Lee-Carter θα χρησιμοποιήσουμε προσέγγιση *Poisson* με βάση το πλαίσιο εργασίας των γενικευμένων γραμμικών μοντέλων. Έτσι έχουμε ότι

$$D_{xt} \sim \text{Poisson}(E_{xt}\mu_x(t), \mu_x(t)) = \exp(\alpha_x + \beta_x \kappa_t) \quad (5.1)$$

όπου το $\mu_x(t)$ αντιπροσωπεύει το ρυθμό των ατυχημάτων και οι παράμετροι $\alpha_x, \beta_x, \kappa_t$ όπως ορίζονται από το βασικό μοντέλο Lee-Carter.

5.2.2 Υπολογισμός παραμέτρων

Για τον υπολογισμό των παραμέτρων του πιο πάνω μοντέλου μεγιστοποιούμε την λογαριθμοποιημένη συνάρτηση πιθανοφάνειας όπως προκύπτει για το μοντέλο που δίνεται στην εξίσωση. Το πρόβλημα που παρουσιάζεται σε αυτή την περίπτωση είναι ο διγραμμικός όρος $\beta_x \kappa_t$ που δεν μας επιτρέπει να χρησιμοποιήσουμε στατιστικά πακέτα που εφαρμόζουν την παλινδρόμηση *Poisson*. Αυτός είναι και ο λόγος που το μοντέλο Lee-Carter εντάσσεται στο πλαίσιο εργασίας των γενικευμένων γραμμικών μοντέλων αλλά δεν αποτελεί γενικευμένο γραμμικό μοντέλο. Για το σκοπό αυτό χρησιμοποιούμε το στατιστικό πακέτο LEM (Vermunt, 1997). Ο αλγόριθμος που εφαρμόζεται είναι μια μέθοδος *Newton* και χρησιμοποιείται για τον υπολογισμό λογαριθμοποιημένων γραμμικών μοντέλων με διγραμμικούς όρους. Το LEM χρησιμοποιεί τον περιορισμό $\sum_t \hat{\kappa}_t = 0$ ο οποίος είναι ο ίδιος που χρησιμοποιείται στον υπολογισμό των παραμέτρων του μοντέλου Lee-Carter. Ο περιορισμός αυτός καθορίζεται χρησιμοποιώντας την εντολή *spe()* στον κώδικα που θα εφαρμόσουμε τα δεδομένα μας. Ο περιορισμός κλίμακας που χρησιμοποιείται από το LEM είναι $\hat{\beta}_1 = 1$ ο οποίος είναι διαφορετικός από αυτό που υποθέτει το μοντέλο Lee-Carter ($\sum_x \hat{\beta}_x = 1$). Για να πετύχουμε τον περιορισμό αυτό διαιρούμε τις εκτιμήσεις των παραμέτρων β_x με $\sum_x \hat{\beta}_x$ και πολλαπλασιάζουμε τις εκτιμήσεις του κ_t με τον ίδιο αριθμό.

Ο κώδικας που χρησιμοποιούμε στο στατιστικό πακέτο LEM για τον υπολογισμό των παραμέτρων του μοντέλου Lee-Carter για την προσέγγιση *Poisson* είναι ο εξής:

```
man 2
dim 12 10
lab O D
mod {wei(OD), O, spe(D, 1a, O, b)}
sta (OD) exposures.dat
dat deaths.dat
```

Η εντολή *man* δηλώνει τον αριθμό των μεταβλητών που στην περίπτωση μας είναι δύο. Με την εντολή *dim* καθορίζουμε τη διάσταση των μεταβλητών

όπως προκύπτει από τα δεδομένα μας. Στο παράδειγμα που εκτελούμε έχουμε δώδεκα ομάδες ηλικίας και δέκα χρονικές περιόδους. Η εντολή *lab* δίνει όνομα στις δύο μεταβλητές και η εντολή *mod* χρησιμοποιείται για να καθορίσουμε τους τρεις όρους του μοντέλου, τους ασφαλιζομένους $wei(OD)$, την επίδραση της ηλικίας O και το διγραμμικό όρο $spe(D, 1a, O, b)$. Οι εντολές *sta* και *dat* χρησιμοποιούνται για να καθορίσουμε τα αρχεία δεδομένων όπου το *exposures.dat* περιέχει τον πίνακα με τον αριθμό των ασφαλιζομένων και το *deaths.dat* το πίνακα με τον αριθμό των ατυχημάτων όπως τους αναφέραμε στην προηγούμενη ενότητα. Στους πιο κάτω πίνακες (5.3, 5.4) δίνονται οι τιμές για τις εκτιμήσεις των παραμέτρων που προκύπτουν για τις διάφορες ομάδες ηλικιών και για τις διάφορες χρονικές στιγμές.

Ηλικία	α_x	β_x
17-22	-3.413	1.0000
23-28	-3.7859	0.9793
29-34	-3.9642	1.1053
35-40	-3.8945	1.2109
41-46	-3.8531	0.9932
47-52	-4.0229	0.9763
53-58	-4.1254	0.9020
59-64	-4.1934	1.5138
65-70	-4.2504	1.2950
71-76	-4.3568	1.8303
77-82	-4.4526	1.8996
83-88	-3.9722	2.4686

Πίνακας 5.3 Εκτιμήσεις των παραμέτρων για κάθε ηλικία

Έτος	κ_t
A'2006	-0.1525
B'2006	0.0495
A'2007	-0.1223
B'2007	0.0665
A'2008	-0.0412
B'2008	0.1571
A'2009	-0.0664
B'2009	0.1339
A'2010	-0.1102
B'2010	0.0856

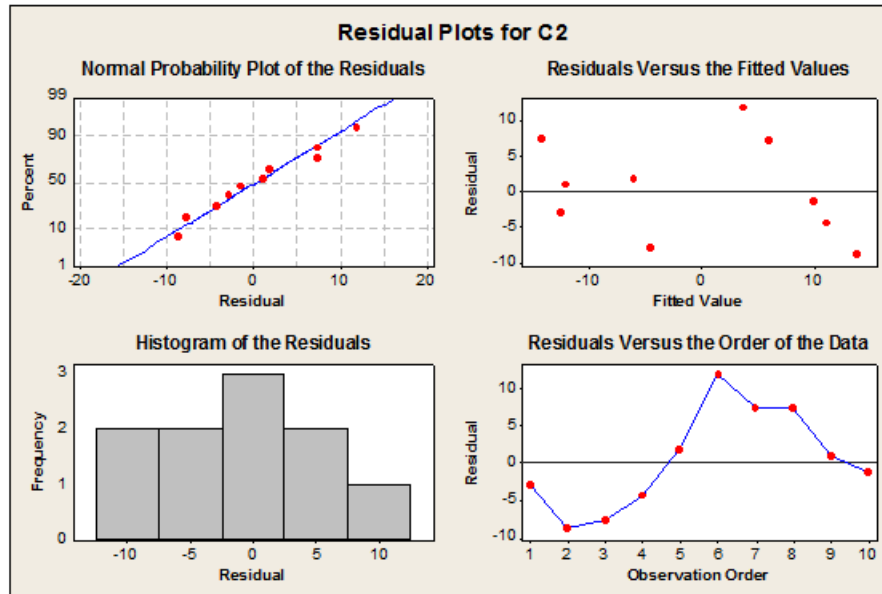
Πίνακας.5.4 Εκτιμήσεις του δείκτη θνησιμότητας κ_t για κάθε εξάμηνο

5.2.3 Πρόβλεψη του μοντέλου

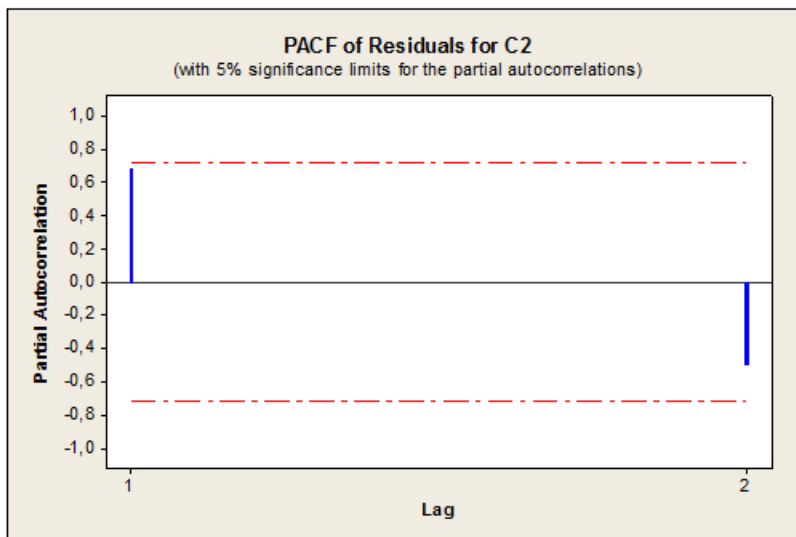
Σε αυτή την ενότητα θα αναφερθούμε στην πρόβλεψη του δείκτη θνησιμότητας και κατ' επέκταση του αριθμού ατυχημάτων για μελλοντικούς χρόνους χρησιμοποιώντας μεθόδους χρονοσειρών. Χρησιμοποιώντας το στατιστικό πακέτο MINITAB και εφαρμόζοντας τους κατάλληλους ελέγχους επιλέγουμε να χρησιμοποιήσουμε το μοντέλο ARIMA (1,0,0) το οποίο δίνεται από την πιο κάτω εξίσωση:

$$\kappa_t = -0.90\kappa_{t-1} + \varepsilon_t \quad (5.2)$$

Αρχικά δοκιμάσαμε το πιο πάνω μοντέλο χρονοσειρών προσθέτοντας και ένα σταθερό όρο όμως η p -τιμή του στο μοντέλο ήταν πολύ μεγάλη και γι' αυτό το λόγο αφαιρέθηκε. Στα πιο κάτω σχήματα επιβεβαιώνουμε την καλή προσαρμογή του μοντέλου στα δεδομένα για την πρόβλεψη του δείκτη "θνησιμότητα".

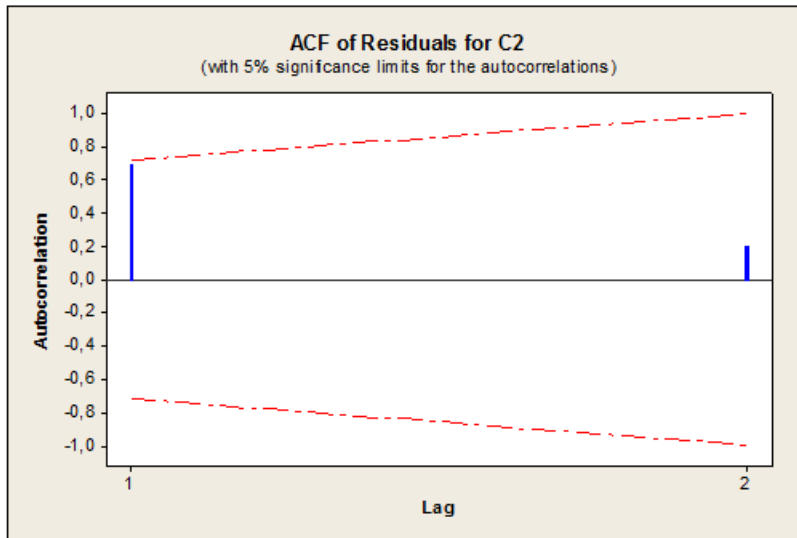


Σχήμα.5.1 Γραφικές παραστάσεις των υπολοίπων για το μοντέλο ARIMA(1,0,0) για το δείκτη θνησιμότητας κτ.

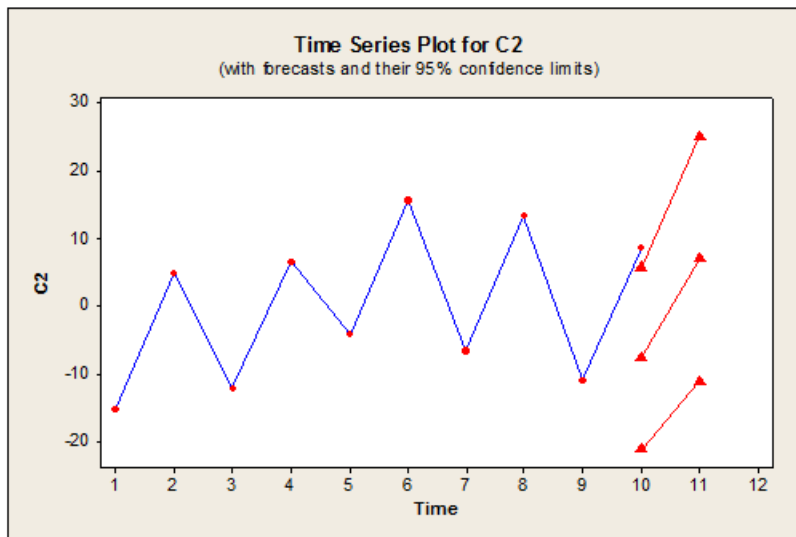


Σχήμα.5.2 Γραφική παράσταση μερικής αυτοσυσχέτισης των υπολοίπων για τις τιμές του δείκτη θνησιμότητας κτ.

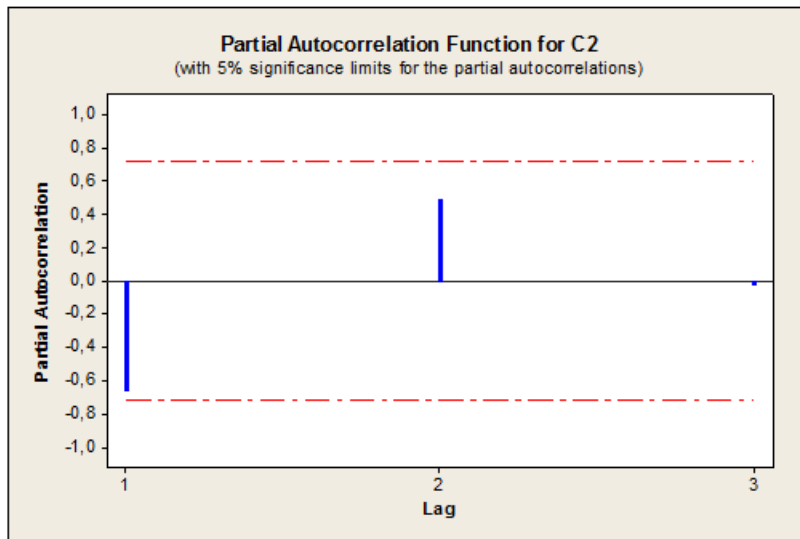
5.2. ΜΟΝΤΕΛΟΠΟΙΗΣΗ ΚΑΙ ΕΦΑΡΜΟΓΗ ΤΟΥ ΜΟΝΤΕΛΟΥ LEE-CARTER61



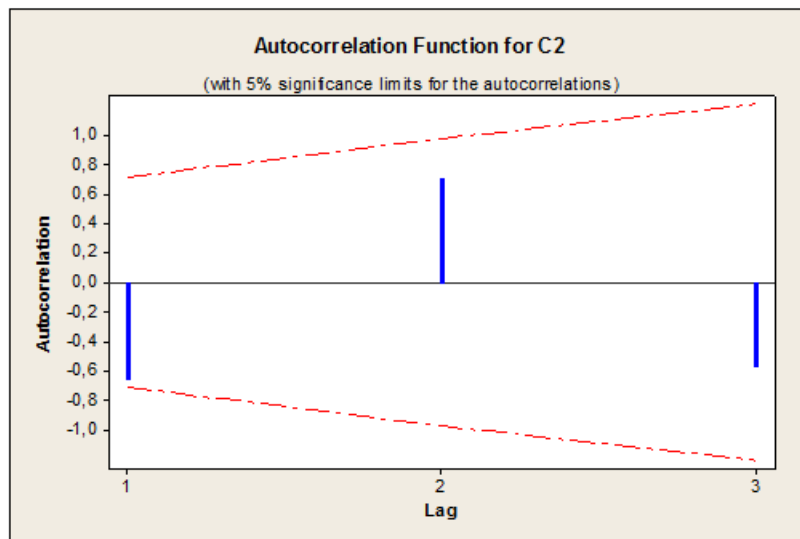
Σχήμα.5.3 Γραφική παράσταση αυτοσυσχέτισης των υπολοίπων για τις τιμές του δείκτη θνησιμότητας κτ.



Σχήμα.5.4 Γραφική παράσταση για τις τιμές του δείκτη θνησιμότητας κτ μαζί με τις προβλέψεις και 95% διαστήματα εμπιστοσύνης.



Σχήμα 5.5 Γραφική παράσταση μερικής αυτοσυσχέτισης των δεδομένων.



Σχήμα 5.6 Γραφική παράσταση αυτοσυσχέτισης των δεδομένων.

Από τις γραφικές παραστάσεις της αυτοσυσχέτισης και της μερικής αυ-

5.2. ΜΟΝΤΕΛΟΠΟΙΗΣΗ ΚΑΙ ΕΦΑΡΜΟΓΗ ΤΟΥ ΜΟΝΤΕΛΟΥ LEE-CARTER63

το συσχετίσις των δεδομένων (σχήμα 5.5, σχήμα 5.6) παρατηρούμε ότι αποκόπτονται μετά τη πρώτη τιμή και οδηγούμαστε στην υπόθεση ότι ένα κατάλληλο μοντέλο είναι το ARIMA (1,0,0). Από τις υπόλοιπες γραφικές παραστάσεις των υπολοίπων παρατηρούμε ότι το μοντέλο προσαρμόζεται αρκετά καλά στα δεδομένα μας.

Πιο κάτω δίνουμε τον πίνακα (5.6) με τα αποτελέσματα όπως προκύπτουν από το MINITAB καθώς και το πίνακα με τις προβλέψεις (5.5) για το δείκτη θνησιμότητας που θα χρησιμοποιήσουμε για να προβλέψουμε τον αριθμό των ατυχημάτων την επόμενη χρονιά.

Type	Coef	SE Coef	T	P
AR 1	-0,9000	0,1949	-4,62	0,001

Πίνακας 5.6 Αποτελέσματα παραμέτρων μοντέλου ARIMA (1,0,0).

Ηλικία	Πρόβλεψη (Α' Εξάμηνο)	Πρόβλεψη (Β' Εξάμηνο)
17-22	305	353
23-28	210	243
29-34	174	205
35-40	185	221
41-46	196	227
47-52	166	192
53-58	150	172
59-64	134	167
65-70	129	156
71-76	111	145
77-82	100	133
83-88	155	223

Πίνακας 5.5 Προβλέψεις αριθμού ατυχημάτων ανά εξάμηνο και ανά ηλικία για ανά 10,000 ασφαλιζομένους σε κάθε ομάδα ηλικίας.

Για να ελέγξουμε την καταλληλότητα του μοντέλου εφαρμόζουμε την ίδια διαδικασία για εκτίμηση παραμέτρων ηλικίας και δείκτη θνησιμότητας (πίνακας 5.7, 5.8) για δεδομένα από το 2006-2009. Στη συνέχεια εφαρμόζουμε το μοντέλο ARIMA (1,0,0) για τα δεδομένα από 2006-2009 και προβλέπουμε τις τιμές για την επόμενη χρονιά. Οι τιμές που προκύπτουν για τη χρονιά 2010 ανά 10,000 ασφαλιζομένους σε κάθε ομάδα ηλικίας δίνονται στον πιο κάτω πίνακα (5.9). Επειδή γνωρίζουμε τον ακριβή αριθμό ασφαλιζομένων για το 2010 (πίνακας 5.2) κατασκευάζουμε ακόμη ένα πίνακα για τις προβλέψεις του αριθμού των ατυχημάτων σε σχέση με τον γνωστό αριθμό των ασφαλιζομένων (5.10). Από την εφαρμογή του μοντέλου παρατηρούμε ότι η πρόβλεψη για τις είδη γνωστές τιμές είναι αρκετά καλή ειδικά για το Α' εξάμηνο του 2010. Πιο κάτω δίνονται οι γραφικές παραστάσεις (5.8, 5.9) για σύγκριση των τιμών των μοντέλων που χρησιμοποιήθηκαν και για τις δύο περιπτώσεις όπου παρατηρούμε ότι η προσαρμογή του μοντέλου στις μεγαλύτερες ηλικίες είναι πολύ καλή και για τα δύο εξάμηνα ενώ στις πιο μικρές ηλικίες η προσαρμογή για το Α' εξάμηνο του 2010 είναι καλύτερη από αυτή του Β' εξαμήνου.

5.2. ΜΟΝΤΕΛΟΠΟΙΗΣΗ ΚΑΙ ΕΦΑΡΜΟΓΗ ΤΟΥ ΜΟΝΤΕΛΟΥ LEE-CARTER65

Ηλικία	α_x	β_x
17-22	-3.3951	1.0000
23-28	-3.7609	1.0903
29-34	-3.9745	1.4028
35-40	-3.8698	1.3602
41-46	-3.8363	1.1493
47-52	-4.0171	1.1921
53-58	-4.1495	1.1374
59-64	-4.1882	1.9492
65-70	-4.2997	1.6148
71-76	-4.4211	2.4995
77-82	-4.2465	3.1182
83-88	-4.0319	2.1378

Πίνακας.5.7 Εκτιμήσεις των παραμέτρων για κάθε ηλικία για δεδομένα από το 2006-2009.

Έτος	K_t
A'2006	-0.1344
B'2006	0.0340
A'2007	-0.1089
B'2007	0.0487
A'2008	-0.0386
B'2008	0.1259
A'2009	-0.0459
B'2009	0.1192

Πίνακας.5.8 Εκτιμήσεις για το δείκτη θνησιμότητας k_t για δεδομένα από το 2006-2009.

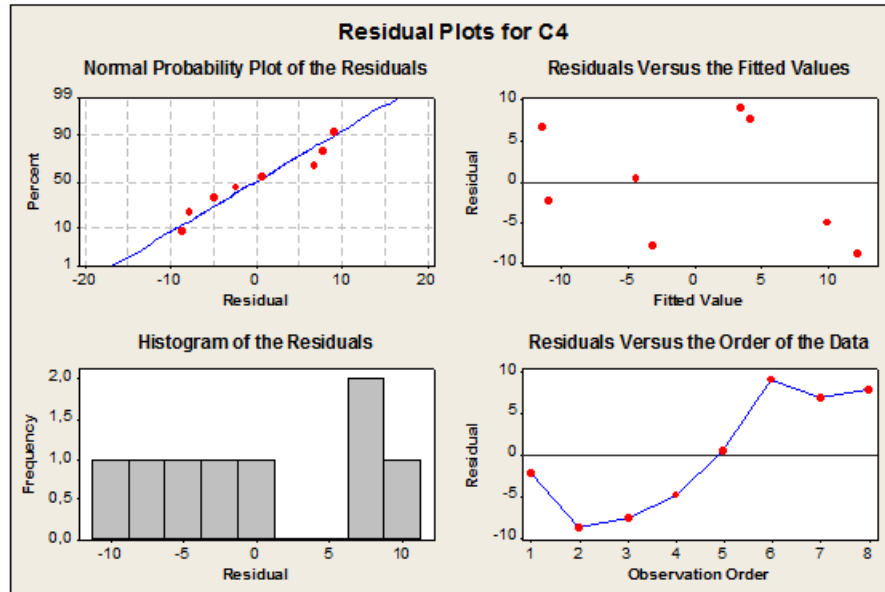
Ηλικία	Πρόβλεψη (Α' Εξάμηνο)	Πρόβλεψη (Β' Εξάμηνο)
17-22	301	370
23-28	206	259
29-34	161	215
35-40	180	238
41-46	190	241
47-52	158	202
53-58	139	176
59-64	123	183
65-70	114	159
71-76	92	153
77-82	102	194
83-88	140	218

Πίνακας.5.9 Προβλέψεις αριθμού ατυχημάτων ανά εξάμηνο και ανά ηλικία για ανά 10,000 ασφαλιζόμενους σε κάθε ομάδα ηλικίας για το έτος 2010.

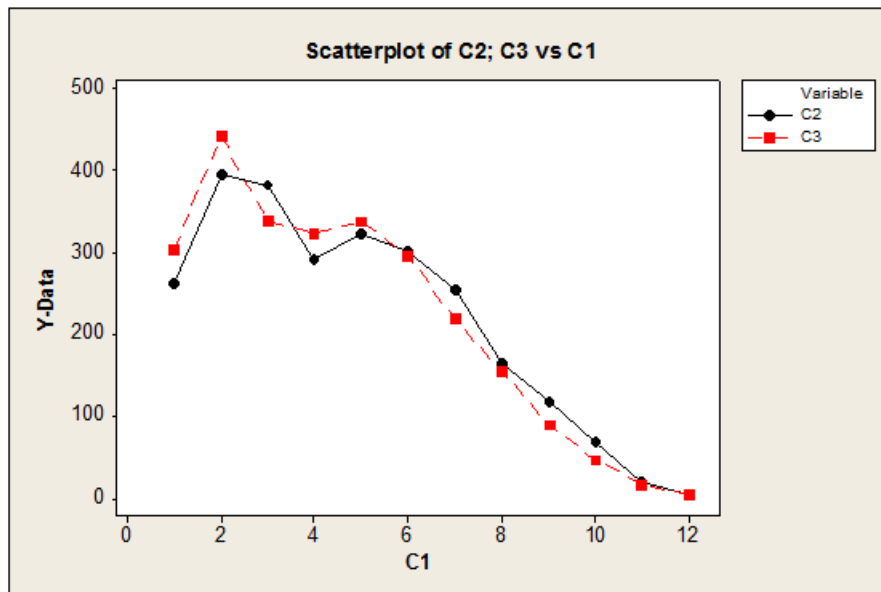
Ηλικία	Πρόβλεψη (Α' Εξάμηνο)	Πρόβλεψη (Β' Εξάμηνο)
17-22	303	373
23-28	442	555
29-34	339	454
35-40	323	427
41-46	338	429
47-52	296	379
53-58	219	277
59-64	154	229
65-70	89	125
71-76	46	76
77-82	15	29
83-88	4	6

Πίνακας.5.10 Προβλέψεις αριθμού ατυχημάτων ανά εξάμηνο και ανά ηλικία με βάση τον αριθμό ασφαλιζόμενων του πίνακα 5.2 σε κάθε ομάδα ηλικίας για το έτος 2010.

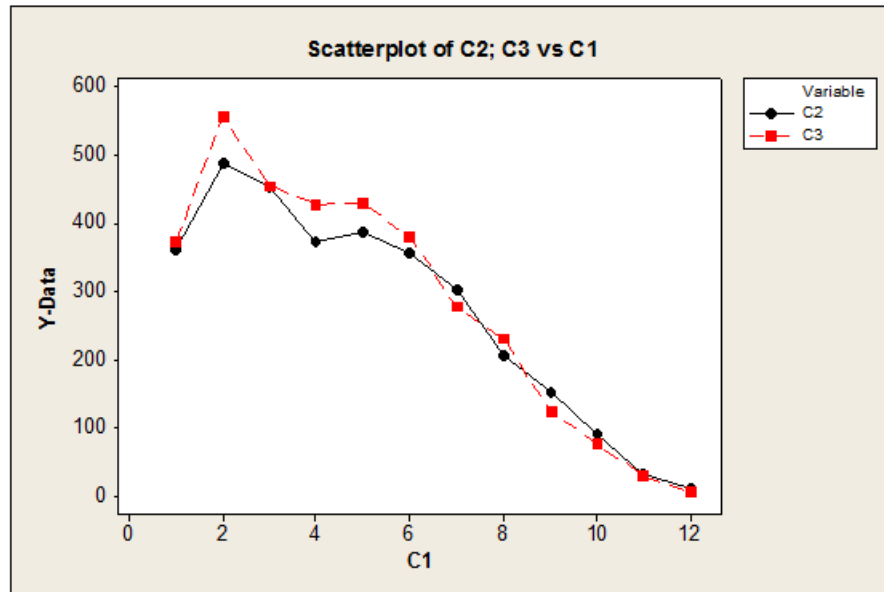
5.2. ΜΟΝΤΕΛΟΠΟΙΗΣΗ ΚΑΙ ΕΦΑΡΜΟΓΗ ΤΟΥ ΜΟΝΤΕΛΟΥ LEE-CARTER67



Σχήμα.5.7 Γραφικές παραστάσεις των υπολοίπων για το μοντέλο που προσαρμόσαμε στο δείκτη θνησιμότητας για δεδομένα από 2006-2009.



Σχήμα.5.8 Γραφική παράσταση σύγκρισης των παρατηρούμενων τιμών και των προβλέψεων για τον αριθμό ατυχημάτων για κάθε ομάδα ηλικίας (C1) το Α' εξάμηνο του 2010.



Σχήμα.5.9 Γραφική παράσταση σύγκρισης των παρατηρούμενων τιμών και των προβλέψεων για τον αριθμό ατυχημάτων για κάθε ομάδα ηλικίας (C1) το Β' εξάμηνο του 2010.

5.2.4 Συμπεράσματα

- Η προσαρμογή του μοντέλου που εφαρμόσαμε στα δεδομένα μας για πρόβλεψη του αριθμού αυτοκινητιστικών ατυχημάτων ήταν αρκετά καλή ωστόσο θα μπορούσε να είναι πιο ακριβής και με καλύτερα αποτελέσματα αν είχαμε περισσότερα δεδομένα για την εφαρμογή. Αυτό θα είχε ως αποτέλεσμα να γίνει καλύτερη μελλοντική πρόβλεψη τιμών αφού οι προβλέψεις που παίρνουμε με την εφαρμογή του μοντέλου Lee-Carter προϋποθέτουν ότι θα διατηρηθεί η ίδια συμπεριφορά για τα επόμενα χρόνια με βάση το διάστημα που μελετούμε. Θα μπορούσαμε να επιχειρήσουμε την πρόβλεψη για πιο μεγάλο μελλοντικό διάστημα.
- Παρατηρούμε ότι τα δεδομένα στην πιο πάνω εφαρμογή συμπεριφέρονται σαν δύο διαφορετικά παραδείγματα με βάση το εξάμηνο που πραγματοποιήθηκε το ατύχημα και θα μπορούσαμε να εφαρμόσουμε το μοντέλο μας για δύο ομάδες δεδομένων αναλύοντας ξεχωριστά κάθε εξάμηνο όμως το γεγονός ότι τα δεδομένα μας είναι λίγα δεν μας επιτρέπει(η πρόβλεψη για το δείκτη θνησιμότητας δεν θα μπορούσε να πραγματοποιηθεί ή δεν θα ήταν καθόλου καλή) να κάνουμε την πιο πάνω εφαρμογή όπου θα μπορούσαμε να συγκρίνουμε σε ποια περίπτωση το μοντέλο μας εφαρμόζεται καλύτερα.
- Στην εφαρμογή του μοντέλου χρονοσειρών για την πρόβλεψη του δείκτη θνησιμότητας προσθέσαμε και σταθερό όρο ο οποίος όμως δεν συνέβαλλε

5.2. ΜΟΝΤΕΛΟΠΟΙΗΣΗ ΚΑΙ ΕΦΑΡΜΟΓΗ ΤΟΥ ΜΟΝΤΕΛΟΥ LEE–CARTER⁶⁹

σημαντικά στην εφαρμογή και για αυτό το λόγο αφαιρέθηκε.

- Όπως αναφέραμε και πιο πάνω το μοντέλο προϋποθέτει ότι η συμπεριφορά των δεδομένων θα διατηρηθεί και τα επόμενα χρόνια. Αυτός είναι και ο λόγος που στις προβλέψεις για τους ρυθμούς θνησιμότητας οι τιμές για τα πρώτα χρόνια της πρόβλεψης είναι πιο ακριβής σε σχέση με τις υπόλοιπες προβλέψεις.

Βιβλιογραφία

- [1] Brouhns N., Denuit M., Vermunt J.K. (2002). A Poisson log-bilinear regression approach to the construction of projected lifetables, *Mathematics and Economics*, 31, 373-393.
- [2] 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, 385–401.
- [3] Haberman S., Renshaw A. (2008). Mortality, longevity and experiments with the Lee-Carter model. *Lifetime Data Analysis*, 14, 286-315.
- [4] Lee R.D., Carter L.R. (1992). Modeling and Forecasting U. S. Mortality. *Journal of the American Statistical Association*, 87, 659-671.
- [5] Ismail N., Jemain A.A.H. (2007). Overdispersion with Negative Binomial and Generalized Poisson Regression Models. *Casualty Actuarial Society Forum*, 103-158.
- [6] Lee R. (2000). The Lee-Carter method for forecasting mortality, with various extensions and applications. *North American Actuarial Journal*, 4, 80-93
- [7] Renshaw A., Haberman S. (2003). Lee-Carter Mortality Forecasting: A Parallel Generalized Linear Modelling Approach for England and Wales Mortality Projections. *Applied Statistics*, 52, 119-137.
- [8] Οικονόμου Π., Καρώνη Χ. (2010). *Στατιστικά μοντέλα παλινδρόμησης*. Εκδόσεις Συμεών, Αθήνα.
- [9] Montgomery D.C., Jennings C.L., Kulahci M. (2008). *Introduction to Time Series Analysis and Forecasting*, Wiley, Hoboken, N.J.
- [10] Mc Cullagh P., Nelder J.A. (1989). *Generalized linear models*. Chapman and Hall (2nd edition), London.
- [11] Wilmoth J.R. (1995). Are Mortality Projections Always More Pessimistic When Disaggregated by Cause of Death? *Mathematical Population Studies* 5, 4, 293-319.

- [12] Carter L. (1996). Forecasting U.S. Mortality: A Comparison of Box-Jenkins ARIMA and Structural Time Series Models. *The Sociological Quarterly*, 37, 127-44.
- [13] Vermunt J.K. (1997). LEM: A general program for the analysis of categorical data. Tilburg University
- [14] Lee R.D., Nault F. (1993). Modeling and Forecasting Provincial Mortality in Canada. *World Congress of the International Union for the Scientific Study of Population*, Montreal.

Ιστοσελίδες

<http://www.camri.uqam.ca/camri/app/statistics/LCA/LEM/manual.pdf>

<http://www.duke.edu/rnau/411arim.htm>