ΠΟΛΥΤΕΧΝΕΙΟ ΚΡΗΤΗΣ
ΓΕΝΙΚΟ ΤΜΗΜΑ

ΠΡΟΓΡΑΜΜΑ ΜΕΤΑΠΤΥΧΙΑΚΩΝ ΣΠΟΥΔΩΝ
ΕΦΑΡΜΟΣΜΕΝΕΣ ΕΠΙΣΤΗΜΕΣ ΚΑΙ ΤΕΧΝΟΛΟΓΙΑ

ΔΙΠΛΩΜΑΤΙΚΗ ΔΙΑΤΡΙΒΗ ΜΕΤΑΠΤΥΧΙΑΚΟΥ ΔΙΠΛΩΜΑΤΟΣ ΕΙΔΙΚΕΥΣΗΣ
ΚΑΤΕΥΘΥΝΣΗ : «ΕΦΑΡΜΟΣΜΕΝΑ ΚΑΙ ΥΠΟΛΟΓΙΣΤΙΚΑ ΜΑΘΗΜΑΤΙΚΑ»

ΜΕΘΟΔΟΙ ΠΡΟΒΛΕΨΗΣ

ΠΑΠΑΔΟΠΟΥΛΟΥ ΓΕΩΡΓΙΑ

Επιβλέπων: Επίκουρος Καθηγητής **Δάρας Τρύφων**

ΧΑΝΙΑ, 2016

Ευχαριστίες

Η παρούσα μεταπτυχιακή διπλωματική εργασία εκπονήθηκε στα πλαίσια του Π.Μ.Σ.: "Εφαρμοσμένες Επιστήμες & Τεχνολογία" του Γενικού Τμήματος του Πολυτεχνείου Κρήτης, με επιβλέποντα καθηγητή, τον κ. Τρύφωνα Δάρα. Πρωτίστως λοιπόν, θα ήθελα να ευχαριστήσω θερμά τον κ. Δάρα για την πολύτιμη βοήθεια, αλλά και για την εμπιστοσύνη που μου έδειξε, σε όλη αυτή τη προσπάθεια διεκπαιρέωσης της εργασίας. Οφείλω επίσης, ένα μεγάλο ευχαριστώ στον φίλο και σύντεκνο μου, Δρόσο, ο οποίος με βοήθησε πολύ με τις πολύτιμες συμβουλές του. Και τέλος, θα ήθελα να ευχαριστήσω θερμά τους γονείς μου Ιωάννη και Φωτεινή, καθώς και τον σύζυγό μου Σπύρο, οι οποίοι ήταν, είναι και θα είναι αρωγοί σε κάθε προσπάθεια της ζωής μου.

Περιεχόμενα

ΕΙΣΑΓΩΓΗ

Σκοπός αυτής της εργασίας είναι η παρουσίαση κάποιων πολύ χρήσιμων μεθόδων πρόβλεψης, οι οποίες χρησιμοποιούνται για τη μελέτη υπαρκτών προβλημάτων. Πιο αναλυτικά, στο πρώτο μέρος της εργασίας περιγράφονται οι μέθοδοι ανάλυσης παλινδρόμησης, καθώς και μέθοδοι ανάλυσης χρονοσειρών. Πιο συγκεκριμένα, γίνεται αναφορά στην απλή γραμμική παλινδρόμηση, την πολλαπλή γραμμική παλινδρόμηση, καθώς και σε κάποιες χρήσιμες τεχνικές υπολογισμών. Επίσης, γίνεται αναφορά στην έννοια της χρονοσειράς, στα κριτήρια αξιολόγησης των μεθόδων πρόβλεψης, σε μεθόδους εξομάλυνσης και στη μέθοδο διάσπασης χρονοσειρών. Στο δεύτερο μέρος της εργασίας, παρουσιάζεται εφαρμογή των μεθόδων εξομάλυνσης, διάσπασης χρονοσειρών και ανάλυσης παλινδρόμησης.

Γ. Παπαδοπούλου, Χανιά 2016.

Μέρος Ι

ΜΕΘΟΔΟΙ ΠΡΟΒΛΕΨΕΩΝ

Κεφάλαιο 1

ΜΕΘΟΔΟΙ ΑΝΑΛΥΣΗΣ ΠΑΛΙΝΔΡΟΜΗΣΗΣ

1.1 ΕΙΣΑΓΩΓΗ ΣΤΗΝ ΠΡΟΒΛΕΨΗ

Ο άνθρωπος από τα αρχαία χρόνια ασχολήθηκε με την πρόβλεψη του μέλλοντος. Η επιθυμία για την αποκάλυψη μελλοντικών γεγονότων και καταστάσεων έχει τις ρίζες της πολλούς αιώνες πριν. Ως γνωστόν, στην Αρχαία Ελλάδα υπήρχε το περίφημο «Μαντείο των Δελφών», το οποίο έδινε χρησμούς για το μέλλον. Η τάση αυτή λοιπόν, για τη γνώση του μέλλοντος, με την πάροδο του χρόνου κάθε άλλο παρά αποδυναμώθηκε. Ο άνθρωπος πάντα είχε, έχει και θα έχει την ανάγκη να γνωρίζει τι του επιφυλάσσει το μέλλον. Βέβαια αξίζει να σημειωθεί ότι δεν είναι δυνατόν να γνωρίζει με βεβαιότητα εκ των προτέρων τη μελλοντική εξέλιξη των γεγονότων και των καταστάσεων που τον αφορούν. Αυτό το οποίο επιθυμεί να πετύχει με την διαμόρφωση προβλέψεων για την πρόβλεψη του μέλλοντος, είναι να περιορίσει όσο είναι δυνατόν το στοιχείο της αβεβαιότητας έτσι ώστε να λάβει τις καλύτερες δυνατές αποφάσεις.

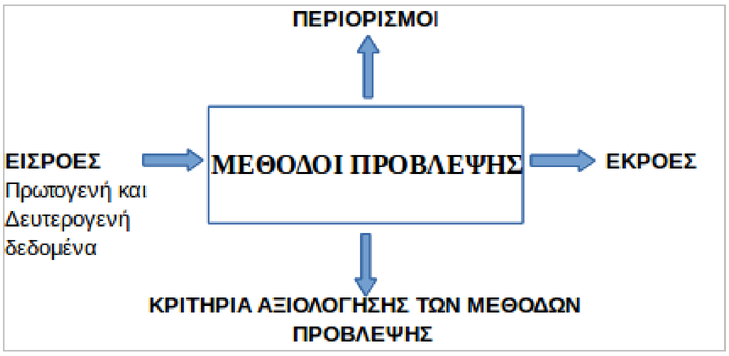
Για τη διαμόρφωση των προβλέψεων, "πηγή τροφοδότησης" αποτελούν τα δεδομένα. Η συλλογή τους σε αρκετές περιπτώσεις είναι κάθε άλλο παρά απλή διαδικασία. Δεν είναι λίγες οι περιπτώσεις εκείνες που η συλλογή των απαραίτητων δεδομένων – πληροφοριών είναι δύσκολη, χρονοβόρα και υψηλού κόστους. Επιπροσθέτως, για να διαμορφώσουμε σωστές και όσο γίνεται περισσότερο αξιόπιστες προβλέψεις, θα πρέπει τα δεδομένα που συλλέγουμε να είναι αξιόπιστα και ακριβή. Διότι ακόμα και αν έχουμε επιλέξει την πλέον κατάλληλη μέθοδο προβλέψεων είναι δυνατό να αποτύχει και να μας δώσει παραπλανητικά στοιχεία, αν τα δεδομένα που έχουμε χρησιμοποιήσει στερούνται ακρίβειας και αξιοπιστίας.

Τα δεδομένα που χρησιμοποιούνται για τη δημιουργία προβλέψεων μπορούμε να τα διακρίνουμε σε **διαστρωματικά** (cross section) και σε **χρονοσειρές** (time series). Διαστρωματικά είναι εκείνα τα δεδομένα που αφορούν τη συμπεριφορά μιας συγκεκριμένης μεταβλητής σε μια δεδομένη χρονική περίοδο ή στιγμή. Για παράδειγμα οι πωλήσεις ενός ή περισσοτέρων προϊόντων μιας επιχείρησης σε διάφορες γεωγραφικές περιοχές για ένα συγκεκριμένο μήνα. Ενώ με τον όρο χρονοσειρά εννοούμε ένα σύνολο τιμών για κάποια μεταβλητή, από παρατηρήσεις σε ίσα χρονικά διαστήματα. Για παράδειγμα οι μηνιαίες πωλήσεις ενός προϊόντος μιας επιχείρησης για τα τελευταία πέντε χρόνια.Γίνεται κατανοητό ότι η ουσιώδης διαφορά μεταξύ των χρονοσειρών και των διαστρωματικών δεδομένων είναι το χρονικό σημείο αναφοράς τους. Οι μεν χρονοσειρές παρέχουν πληροφόρηση για τη διαχρονική εξέλιξη των τιμών μιας μεταβλητής που μας ενδιαφέρει, τα δε διαστρωματικά δεδομένα παρέχουν πληροφόρηση για τις τιμές της μεταβλητής στο ίδιο χρονικό σημείο αναφοράς.

Αξίζει να σημειωθεί ότι τα δεδομένα ανάλογα με την πηγή προελεύσεως τους, διακρίνονται σε **πρωτογενή** (primary data source) και σε **δευτερογενή** (secondary data source). Πρωτογενή είναι τα δεδομένα εκείνα που συλλέγονται για πρώτη φορά αποκλειστικά για το συγκεκριμένο σκοπό της έρευνας και περιέχουν, ως επί το πλείστον, πρωτότυπη πληροφόρηση. Η συλλογή τους γίνεται κυρίως με διάφορες μεθόδους δειγματοληψίας. Ενώ δευτερογενή δεδομένα είναι εκείνα τα στοιχεία που ήδη υπάρχουν και έχουν συλλεγεί για κάποιο άλλο σκοπό. Για παράδειγμα ισολογισμοί και καταστάσεις αποτελεσμάτων χρήσης μίας επιχείρησης.

Όσον αφορά στη διαμόρφωση «σωστών» προβλέψεων, βασική προϋπόθεση είναι η επιλογή εκείνης της μεθόδου πρόβλεψης που ερμηνεύει με τον καλύτερο δυνατό τρόπο την συμπεριφορά των τιμών της μεταβλητής που μας ενδιαφέρει. Η συμπεριφορά των τιμών μιας μεταβλητής, σύμφωνα με την στατιστική θεωρία και ανάλυση, προσδιορίζεται από δύο μέρη. Τα μέρη αυτά είναι το **συστηματικό** (systematic) και το **στοχαστικό** ή τυχαίο (random). Το συστηματικό μέρος αναφέρεται στους παράγοντες εκείνους που μπορούν να προσδιοριστούν και οι οποίοι ερμηνεύουν τον τρόπο συμπεριφοράς της μεταβλητής.Το στοχαστικό ή τυχαίο μέρος έχει ως σκοπό να λάβει υπόψιν του, όλους εκείνους τους μη – συστηματικούς παράγοντες (άρα μη – ελέγξιμους, μη – παρατηρήσιμους και συνεπώς τυχαίους και στοχαστικούς), οι οποίοι αν και επηρεάζουν τις τιμές της μεταβλητής δεν μπορούν να προσδιοριστούν. Συνεπώς, γίνεται κατανοητό ότι όταν προσπαθούμε να προσδιορίσουμε τον τρόπο συμπεριφοράς των τιμών μιας μεταβλητής στην ουσία προσπαθούμε να προσδιορίσουμε το συστηματικό μέρος. Το τυχαίο μέρος δεν μπορεί να προσδιοριστεί.

Στο διαγραμμα που ακολουθεί μπορούμε να δούμε τον τρόπο με τον οποίο διαρθρώνεται και λειτουργεί ένα σύστημα πρόβλεψης.



Σχήμα 1.1: Διάρθρωση ενός συστήματος πρόβλεψης.

Ένα σύστημα προβλέψεων λειτουργεί ως εξής : Αρχικά έχουμε τη συλλογή των δεδομένων που αποτελούν τις εισροές του συστήματος. Εν συνεχεία, θα πρέπει να γίνει επιλογή της μεθόδου προβλέψεως που θα χρησιμοποιηθεί για την επεξεργασία των δεδομένων μας. Για να επιλέξουμε την καταλληλότερη μέθοδο πρόβλεψης θα πρέπει να λάβουμε υπόψη μας τα εξής:

- Την ακρίβεια των προβλέψεων
- Την ευστάθεια της μεθόδου πρόβλεψης
- Την αντικειμενικότητα στην επεξεργασία των δεδομένων
- Τον απαιτούμενο χρόνο για τη διαμόρφωση προβλέψεων
- Το κόστος εφαρμογής της μεθόδου

Επιπροσθέτως εκτός από τα ανωτέρω κριτήρια αξιολόγησης των μεθόδων πρόβλεψης θα πρέπει να λάβουμε υπόψη μας και τυχόν περιορισμούς που υπάρχουν στη λειτουργία ενός συστήματος προβλέψεων. Οι σπουδαιότεροι από αυτούς είναι οι ακόλουθοι:

- Ο διαθέσιμος χρόνος για την προετοιμασία μιας πρόβλεψης
- Η έλλειψη στοιχείων
- Η ποιότητα και η αξιοπιστία των διαθέσιμων στοιχείων

- Η έλλειψη εξειδικευμένου προσωπικού
- Τα διαθέσιμα μέσα για την επεξεργασία των στοιχείων (υπολογιστές, ειδικά προγράμματα κ.τ.λ)

Και τέλος, εφόσον έχουμε επιλέξει την καταλληλότερη μέθοδο προβλέψεων, την εφαρμόζουμε και προβαίνουμε στη διαμόρφωση προβλέψεων που αφορούν τη μελλοντική εξέλιξη των τιμών της υπό εξέταση μεταβλητής.

Κάνοντας αναφορά στις κατηγορίες προβλέψεων, οι προβλέψεις δύναται να κατηγοριοποιηθούν βάσει διαφόρων κριτηρίων. Χρησιμοποιώντας ως κριτήριο το χρονικό ορίζοντα προγραμματισμού στον οποίο αναφέρονται οι προβλέψεις μπορούμε να τις διακρίνουμε σε βραχυπρόθεσμες, μεσοπρόθεσμες και μακροπρόθεσμες. Βραχυπρόθεσμες είναι εκείνες οι προβλέψεις που αναφέρονται στο εγγύς μέλλον. Συνήθως οι προβλέψεις αυτές αφορούν περίοδο μικρότερη των τριων μηνών. Οι μεσοπρόθεσμες προβλέψεις αφορούν ένα χρονικό ορίζοντα προγραμματισμού που είναι ενδιάμεσος του ορίζοντα των βραχυπρόθεσμων και των μακροπρόθεσμων προβλέψεων. Οι προβλέψεις αυτές αφορούν περίοδο από τρεις μήνες έως τρία χρόνια. Τέλος, οι μακροπρόθεσμες προβλέψεις αφορούν το απώτερο μέλλον. Στις περισσότερες περιπτώσεις αφορούν χρονικές περιόδους άνω των τριών ετών. Βέβαια να σημειωθεί ότι οι ανωτέρω χρονικοί προσδιορισμοί είναι ενδεικτικοί και εξαρτώνται κυρίως από την μεταβλητή για την οποία γίνεται η πρόβλεψη.

Λαμβάνοντας ως κριτήριο το σκοπό για τον οποίο διενεργούμε την πρόβλεψη μπορούμε να διακρίνουμε τις προβλέψεις σε οικονομικές, τεχνολογικές και κοινωνικές. Οι οικονομικές προβλέψεις διακρίνονται σε προβλέψεις μικροοικονομικού επιπέδου, για παράδειγμα οι προβλέψεις για τη ζήτηση ενός προϊόντος μιας επιχείρησης, οι προβλέψεις για το ανθρώπινο δυναμικό μιας επιχείρησης κ.τ.λ και σε προβλέψεις μακροοικονομικού επιπέδου, όπως για παράδειγμα πρόβλεψη του ακαθάριστου εθνικού προϊόντος, της προσφοράς χρήματος, της απασχόλησης του εργατικού δυναμικού μιας χώρας κ.τ.λ. Οι τεχνολογικές προβλέψεις ενδιαφέρονται για την πρόβλεψη μελλοντικών τεχνολογικών αναπτύξεων, καθώς και για την επίδραση που θα έχουν οι αναπτύξεις αυτές σε μια επιχείρηση.Και τέλος, με τις κοινωνικές προβλέψεις προσπαθούμε να αντιληφθούμε τη φύση της μεταβαλλόμενης κοινωνίας μας και να διαπιστώσουμε την επιρροή των μεταβαλλόμενων αξιών και της συμπεριφοράς των μελών της κοινωνίας στην επιχείρηση.

Μια άλλη διάκριση είναι αυτή μεταξύ **ποιοτικών** και **ποσοτικών** μεθόδων προβλέψεων. Οι ποιοτικές μέθοδοι προβλέψεων στηρίζονται στην ποιοτική ανάλυση των δεδομένων. Στην περίπτωση αυτή που δεν έχουμε στη διάθεση μας στατιστικά στοιχεία αναγκαζόμαστε να στηριχθούμε σε οποιαδήποτε ποιοτική πληροφορία που μπορούμε να συγκεντρώσουμε. Χαρακτηριστικό παράδειγμα αποτελεί η προσπάθεια πρόβλεψης του όγκου πωλήσεων ενός νέου προϊόντος. Προσπαθούμε δηλαδή να διαπιστώσουμε τις προθέσεις των πιθανών καταναλωτών και ζητούμε τη γνώμη των πωλητών και άλλων προσώπων που έχουν σχέση με την αγορά κ.τ.λ. Εν συνεχεία όλες αυτές τις ποιοτικές πληροφορίες προσπαθούμε να τις μετατρέψουμε σε κάποια ποσοτική εκτίμηση. Αντιθέτως, οι ποσοτικές μέθοδοι προβλέψεων στηρίζονται στην ποσοτική ανάλυση αριθμητικών δεδομένων για την πρόβλεψη της υπό εξέτασης μεταβλητής. Εδώ να αναφέρουμε ότι στην παρούσα εργασία θα μελετήσουμε τις ποσοτικές μεθόδους πρόβλεψης.

Με τη σειρά τους, οι ποσοτικές μέθοδοι, μπορούν να διακριθούν σε αυτές που βασίζονται σε **μοντέλα χρονοσειρών** (time series models) και σε αυτές που βασίζονται σε **αιτιακά μοντέλα** (causal models). Τα μοντέλα χρονοσειρών προϋποθέτουν ότι η απαραίτητη πληροφορία για την πρόβλεψη περιέχεται στη χρονοσειρά των στοιχείων.Όπου χρονοσειρά είναι μια σειρά παρατηρήσεων που λαμβάνονται σε κανονικά διαστήματα μέσα σε ένα καθορισμένο χρονικό διάστημα. Για παράδειγμα, αν καταγράφουμε τις μηνιαίες τιμές του χρυσού στο χρηματιστήριο μετάλλων του Λονδίνου για χρονικό διάστημα 5 ετών, τότε έχουμε στη διάθεσή μας μια χρονοσειρά των μηνιαίων τιμών του χρυσού. Η ανάλυση χρονοσειράς κάνει την υπόθεση ότι μπορεί να γίνει πρόβλεψη με βάση τα μοτίβα (patterns) των διαθέσιμων δεδομένων. Έτσι, η ανάλυση αυτή αναζητάει τάσεις, κυκλικότητα, περιοδικότητα κτλ στα δεδομένα προκειμένου να δημιουργήσει ένα μοντέλο πρόβλεψης. Από την άλλη μεριά, τα αιτιακά μοντέλα χρησιμοποιούν μια αρκετά διαφορετική προσέγγιση για την δημιουργία πρόβλεψης. θεωρούν ότι η μεταβλητή για την οποία θέλουμε να κάνουμε πρόβλεψη είναι εξαρτημένη με κάποιο τρόπο από μία ή περισσότερες παραμέτρους. Η δυσκολία έγκειται στην εύρεση της μαθηματικής σχέσης με την οποία επηρεάζεται η ζητούμενη μεταβλητή από τις παραμέτρους αυτές.

Οι κατηγορίες των αιτιακών μοντέλων είναι αρκετές, όπως είναι τα οικονομικά και τα μοντέλα δυναμικής συστημάτων. Ωστόσο εμείς στην παρούσα εργασία θα παρουσιάσουμε το αιτιακό μοντέλο που βασίζεται στην ανάλυση παλινδρόμησης. Η ανάλυση παλινδρόμησης (regression analysis), διερευνά ένα σύστημα στο οποίο εμπλέκονται τουλάχιστον δύο μεταβλητές. Αποτελεί μία μεθοδολογία που αποσκοπεί στον προσδιορισμό των ποσοτικών σχέσεων μεταξύ των εμπλεκόμενων μεταβλητών και στη δημιουργία προβλέψεων. Οι προβλέψεις αυτές, βάσει της ανάλυσης παλινδρομήσεως, μπορούν να προκύψουν απ' τα υποδείγματα της απλής και πολλαπλής Παλινδρόμησης. Όσο για την ανάλυση χρονοσειρών (time series analysis), προσπαθεί να διερευνήσει τη διαχρονική συμπεριφορά των τιμών μιας μεταβλητής, οι παρατηρήσεις της οποίας προέρχονται από χρονοσειρά. Η πρόβλεψη των μελλοντικών τιμών μιας μεταβλητής βάσει της ανάλυσης χρονοσειρών μπορεί να προκύψει από τις ακόλουθες τρεις κατηγορίες μεθόδων προβλέψεων:

- **Μέθοδοι Εξομάλυνσης** (smoothing methods). Η δημιουργία προβλέψεων προέρχεται από την εξομάλυνση της διαχρονικής εξέλιξης των τιμών της υπό εξέτασης μεταβλητής, έτσι ώστε να αναγνωριστεί όσο το δυνατό καλύτερα ο τρόπος συμπεριφοράς της.

- **Διάσπαση Χρονοσειρών** (time series decomposition). Η βασική υπόθεση πάνω στην οποία στηρίζεται η συγκεκριμένη μέθοδος είναι ότι οι τιμές μιας χρονοσειράς σχηματίζονται από τα συνθετικά της στοιχεία που είναι η τάση, η κυκλικότητα, η εποχικότητα και η μη – κανονικότητα. Προκειμένου να δημιουργήσουμε προβλέψεις με βάση τη μέθοδο αυτή διασπούμε τη χρονοσειρά στα ανωτέρω τέσσερα συνθετικά της στοιχεία και προσδιορίζουμε την επιρροή που έχει κάθε ένα από αυτά στη διαμόρφωση των τιμών της μεταβλητής που μας ενδιαφέρει.

- **Μέθοδοι Box- Jenkins**. Η μεθοδολογία των Box – Jenkins αποτελεί μια στατιστικά εξειδικευμένη προσέγγιση στην ανάλυση και την κατασκευή ενός υποδείγματος πρόβλεψης, με στόχο την όσο το δυνατόν καλύτερη αναπαράσταση μιας χρονοσειράς. Ωστόσο, εδώ θα πρέπει να επισημάνουμε ότι στην παρούσα εγασία, απ'την ανάλυση χρονοσειρών θα ασχοληθούμε μόνο με τις μεθόδους εξομάλυνσης και τη μέθοδο διάσπασης χρονοσειρών.

Η ανάλυση παλινδρόμησης είναι μια μεθοδολογία που βασίζεται σε στατιστικές και μαθηματικές μεθόδους, οι οποίες χρησιμοποιούνται για την εκτίμηση των σχέσεων μεταξύ των μεταβλητών που εξετάζονται στα πλαίσια της διερεύνησης ενός, κατά βάσην, οικονομικού φαινομένου. Αναλυτικότερα, με την ανάλυση παλινδρόμησης προσδιορίζεται η ποσοτική σχέση μεταξύ της εξαρτημένης και μιας ή περισσοτέρων ανεξάρτητων μεταβλητών με σκοπό την πρόβλεψη των τιμών της εξαρτημένης μεταβλητής με βάση συγκεκριμένες τιμές των ανεξάρτητων μεταβλητών. Η ανάλυση παλινδρόμησης μπορεί να πραγματοποιηθεί σε πέντε στάδια. Τα στάδια αυτά είναι ιεραρχημένα ως προς τη σειρά εφαρμογής τους.

(i) Στο πρώτο στάδιο κατασκευάζεται το θεωρητικό υπόδειγμα που περιλαμβάνει το σύνολο των ανεξάρτητων μεταβλητών οι οποίες αναμένεται να ερμηνεύσουν τη συμπεριφορά της εξαρτημένης μεταβλητής.

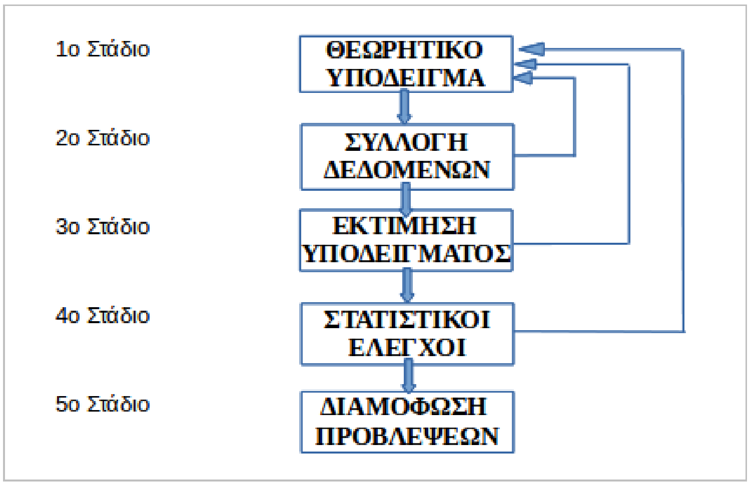
(ii) Στο δεύτερο στάδιο ακολουθεί η συλλογή των δεδομένων. Στο στάδιο αυτό είναι πολύ πιθανό να προκύψουν δυσκολίες αναφορικά με τη διαθεσιμότητα και την ποιότητα των δεδομένων. Έτσι ίσως χρειαστεί να αναθεωρήσουμε το αρχικό θεωρητικό μας υπόδειγμα με μεταβλητές για τις οποίες υπάρχουν τα κατάλληλα δεδομένα.

(iii) Στο τρίτο στάδιο γίνεται η εκτίμηση του υποδείγματος. Λέγοντας εκτίμηση του θεωρητικού υποδείγματος εννοούμε το προσδιορισμό της ποσοτικής σχέσης μεταξύ της εξαρτημένης και της ανεξάρτητης ή των ανεξάρτητων μεταβλητών.

(iv) Στο τέταρτο στάδιο εφαρμόζονται οι διάφοροι στατιστικοί έλεγχοι που υπάρχουν, προκειμένου να διαπιστώσουμε την αξιοπιστία της εκτιμηθείσας μορφής του θεωρητικού υποδείγματος. Αξίζει να σημειωθεί ότι και στο στάδιο αυτό όπως και στο στάδιο της εκτίμησης του υποδείγματος είναι δυνατή η αναθεώρηση του θεωρητικού υποδείγματος, αν βέβαια έχουν προκύψει προβλήματα σχετικά με την εκτίμηση, τον στατιστικό έλεγχο ή και τα δύο.

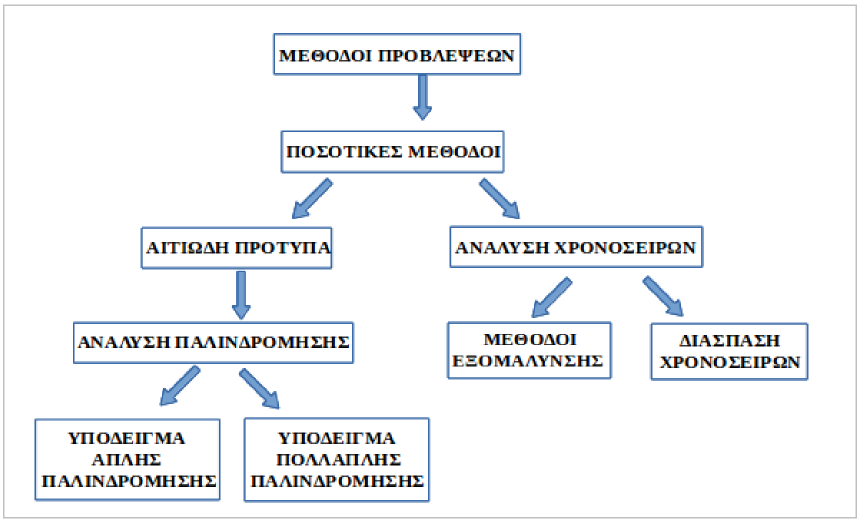
(v) Στο πέμπτο και τελευταίο στάδιο χρησιμοποιούμε το εκτιμηθέν υπόδειγμα για τη διαμόρφωση προβλέψεων των τιμών της εξαρτημένης μεταβλητής με βάση δεδομένες τιμές των ανεξάρτητων μεταβλητών.

Τα παραπάνω στάδια απεικονίζονται στο ακόλουθο διάγραμμα.



Σχήμα 1.2: Τα στάδια της Ανάλυσης Παλινδρόμησης.

Εν κατακλείδι, στο διάγραμμα που ακολουθεί μπορούμε να δούμε τις μεθόδους που θα παρουσιαστούν αναλυτικά στη συνέχεια της εργασίας.



Σχήμα 1.3: Οι Μέθοδοι προβλέψεων που θα μελετήσουμε στην παρούσα εργασία.

1.2 ΑΠΛΗ ΓΡΑΜΜΙΚΗ ΠΑΛΙΝΔΡΟΜΗΣΗ

Η απλή γραμμική παλινδρόμηση είναι μια μέθοδος, που μπορεί να προσδιορίσει ποσοτικά τη συγκεκριμένη γραμμική σχέση μεταξύ μιας εξαρτημένης μεταβλητής Y και μιας ανεξάρτητης μεταβλητής X. Η μαθηματική μορφή που λαμβάνει το υπόδειγμα της απλής γραμμικής παλινδρόμησης για τις μεταβλητές X και Y είναι η ακόλουθη:

$$Y_t = \beta_0 + \beta_1 X_t + \varepsilon_t \tag{1.1}$$

’Οπου:
Y_t = η εξαρτημένη μεταβλητή
X_t = η ανεξάρτητη μεταβλητή
β_0 = ο σταθερός όρος
β_1 = η κλίση του γραμμικού υποδείγματος
ε_t = το τυχαίο σφάλμα ή διαταρακτικός όρος

Τα β_0 και β_1 είναι γνωστά ως συντελεστές παλινδρόμησης (regression coefficients), τις τιμές των οποίων επιθυμούμε να προσδιορίσουμε με βάση τις παρατηρήσεις των μεταβλητών X και Y ενός συγκεκριμένου δείγματος. Η σχέση (1.1) αποτελείται από δύο μέρη. Το $\beta_0 + \beta_1 X_t$ καλείται συστηματικό ή προσδιοριστικό μέρος της εξίσωσης της απλής γραμμικής παλινδρόμησης και το ε_t καλείται τυχαίο ή στοχαστικό μέρος.

Οι υποθέσεις που συνιστούν το κλασικό γραμμικό υπόδειγμα παλινδρόμησης είναι οι εξής:

(i) $Y_t = \beta_0 + \beta_1 X_t + \varepsilon_t$
Η πρώτη υπόθεση αναφέρεται στη γραμμική σχέση που συνδέει τις μεταβλητές X και Y. Δηλαδή, πως κάθε τιμή Y_t είναι γραμμική συνάρτηση της τιμής X_t συν το διαταρακτικό όρο ε_t.

(ii) $E(\varepsilon_t) = 0$
Η μεταβλητή ε_t είναι τυχαία μεταβλητή που παίρνει τιμές θετικές και αρνητικές, αλλά κατά μέσο όρο η τιμή της είναι μηδεν.

(iii) $Var(\varepsilon_t) = E\varepsilon_t^2 = \sigma^2$
Η διακύμανση της τυχαίας μεταβλητής ε_t είναι σταθερή για όλες τις τιμές της X_t. Δηλαδή η διασπορά των τιμών της ε_t από τον μέσο δεν αλλάζει όταν μεταβάλλεται η τιμή της X_t, αλλά παραμένει η ίδια. Η υπόθεση αυτή είναι γνωστή ως υπόθεση της ομοσκεδαστικότητας και η παραβίαση της προκαλεί το πρόβλημα της ετεροσκεδαστικότητας.

(iv) $Cov(\varepsilon_t, \varepsilon_s) = 0$
Η υπόθεση αυτή σημαίνει ότι οι διαταρακτικοί όροι είναι ασυσχέτιστοι μεταξύ τους και επομένως, η συνδιακύμανση του διαταρακτικού όρου της παρατήρησης t με το διαταρακτικό όρο οποιασδήποτε άλλης παρατήρησης s, είναι μηδέν. Δηλαδή, $Cov(\varepsilon_t, \varepsilon_s) = E(\varepsilon_t - E\varepsilon_t)(\varepsilon_s - E\varepsilon_s) = E\varepsilon_t\varepsilon_s = 0$. Η παραβίαση της υπόθεσης αυτής δημιουργεί το πρόβλημα της αυτοσυσχέτισης.

(v) $Cov(X_t, \varepsilon_t) = 0$
Η μεταβλητή X δεν είναι στοχαστική. Οι τιμές της παραμένουν σταθερές και δεν είναι όλες ίσες μεταξύ τους. Με άλλα λόγια, η υπόθεση αυτή μας λέει ότι εφόσον η μεταβλητή X δεν είναι στοχαστική, αυτό αυτόματα συνεπάγεται ότι δεν συσχετίζεται με τον διαταρακτικό όρο και ότι επομένως η συνδιακύμανση τους είναι μηδέν. Δηλαδή, $Cov(X_t, \varepsilon_t) = 0$ ή $E(X_t, \varepsilon_t) = 0$, καθώς $E\varepsilon_t = 0$.

(vi) $\varepsilon_t \sim N(0, \sigma^2)$
Δηλαδή, οι όροι σφάλματος ή διαταρακτικοί όροι ακολουθούν την κανονική κατανομή. Και όπως ήδη αναφέραμε, με μέσο όρο 0 και διασπορά ίση με σ^2.

Η μεταβλητή Y στη σχέση (1.1) είναι συνάρτηση της τυχαίας μεταβλητής ε και επομένως η Y είναι επίσης τυχαία μεταβλητή. Επιπλέον, η κατανομή της Y είναι κατανομή υπό συνθήκη, δεδομένης της τιμής της X. Δηλαδή, για κάθε τιμή της X, δεν υπάρχει μία μόνο τιμή της Y, αλλά ολόκληρη κατανομή. Μπορεί να δειχτεί εύκολα ότι:

$$E(Y_t) = \beta_0 + \beta_1 X_t \tag{1.2}$$
$$V(Y_t) = \sigma^2$$

Επομένως,η μεταβλητή Y_t, όπως ορίσαμε στο υπόδειγμα, ακολουθεί μία κατανομή με μέσο $E(Y_t) = \beta_0 + \beta_1 X_t$ και διασπορά $V(Y_t) = \sigma^2$. Δηλαδή, $Y_t \sim (\beta_0 + \beta_1 X_t, \sigma^2)$.
Η σχέση (1.2) ονομάζεται γραμμή παλινδρομήσεως στον πληθυσμό (population regres-sion line). Η γραμμή παλινδρομήσεως όμως στον πληθυσμό είναι άγνωστη, εφόσον δεν γνωρίζουμε τις τιμές των παραμέτρων β_0 και β_1. Αν γνωρίζαμε όλες τις δυνατές τιμές που παίρνει η Y για δύο τουλάχιστον τιμές της X, θα μπορούσαμε να υπολογίσουμε τις τιμές των παραμέτρων β_0 και β_1, αφού σ' αυτήν την περίπτωση θα γνωρίζαμε δύο σημεία από τα οποία διέρχεται η γραμμή παλινδρομήσεως. Εφόσον όμως αυτό είναι αδύνατο, εκτιμάμε τις τιμές των συντελεστών β_0 και β_1 από δείγμα παρατηρήσεων για τις μεταβλητές Y και X. Έστω ότι $\widehat{\beta}_0$ και $\widehat{\beta}_1$ είναι οι εκτιμήσεις για τους συντελεστές β_0 και β_1, οπότε:

$$\widehat{Y}_t = \widehat{\beta}_0 + \widehat{\beta}_1 X_t \tag{1.3}$$

Η σχέση (1.3) ονομάζεται γραμμή παλινδρομήσεως στο δείγμα (sample regression line) και $\widehat{Y}_t$ είναι η τιμή της Y που υπολογίζουμε από τη γραμμή παλινδρομήσεως του δείγματος. Είναι φανερό ότι οι υπολογισμένες τιμές $\widehat{Y}_t$ δεν θα είναι όλες ίσες με τις πραγματικές τιμές της Y_t, πράγμα που σημαίνει ότι η γραμμή παλινδρόμησης του δείγματος, δεν περνάει από όλα τα σημεία που ορίζουν τα ζεύγη των παρατηρήσεων (Y_t, X_t). Η διαφορά μεταξύ των πραγματικών τιμών Y_t και των υπολογισμένων τιμών $\widehat{Y}_t$, δηλαδή:

$$\widehat{\varepsilon}_t = Y_t - \widehat{Y}_t$$

ονομάζεται κατάλοιπο (residual) ή απόκλιση και μπορεί να θεωρηθεί εκτίμηση της άγνωστης τιμής του διαταρακτικού όρου ε_t. Επομένως, η σχέση $Y_t = \beta_0 + \beta_1 X_t + \varepsilon_t$ αναφέρεται στον πληθυσμό και η σχέση $\widehat{Y}_t = \widehat{\beta}_0 + \widehat{\beta}_1 X_t + \widehat{\varepsilon}_t$ αναφέρεται στο δείγμα.

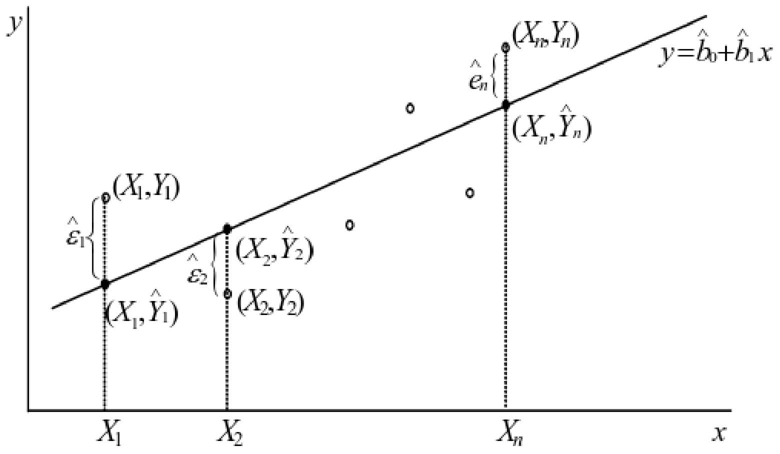
Η εκτίμηση των παραμέτρων του υποδείγματος της γραμμικής παλινδρόμησης μπορεί να γίνει με διάφορους τρόπους, όπως η μέθοδος των ελαχίστων τετραγώνων και η μέθοδος της μεγίστης πιθανοφάνειας. Στην παρούσα εργασία όμως, θα χρησιμοποιήσουμε αποκλειστικά τη μέθοδο των ελαχίστων τετραγώνων (least squares method).

1.2.1 ΜΕΘΟΔΟΣ ΕΛΑΧΙΣΤΩΝ ΤΕΤΡΑΓΩΝΩΝ

Με τη μέθοδο των ελαχίστων τετραγώνων, όπως το όνομα φανερώνει, επιλέγουμε εκείνη τη γραμμή για την οποία το άθροισμα των τετραγώνων των αποκλίσεων (καταλοίπων) των παρατηρήσεων της Y από τη γραμμη παλινδρομήσεως του δείγματος είναι ελάχιστο. Με άλλα λόγια, αυτό σημαίνει ότι οι εκτιμητές $\widehat{\beta}_0$ και $\widehat{\beta}_1$ που προκύπτουν από τη μέθοδο των ελαχίστων τετραγώνων είναι οι τιμές για τις οποίες ελαχιστοποιείται η συνάρτηση:

$$\Phi = \sum_{t=1}^{n} \left(Y_t - \widehat{\beta}_0 - \widehat{\beta}_1 X_t \right)^2 = \sum_{t=1}^{n} \widehat{\varepsilon}_t^2 \tag{1.4}$$

Τα παραπάνω γίνονται περισσότερο κατανοητά με τη βοήθεια του ακόλουθου σχήματος. Από την ελαχιστοποίηση της συναρτήσεως (1.4), δηλαδή $\frac{\partial \Phi}{\partial \widehat{\beta}_0} = 0$



Σχήμα 1.4: Η εκτιμώμενη ευθεία γραμμικής παλινδρόμησης.

και $\frac{\partial \Phi}{\partial \widehat{\beta}_1} = 0$ προκύπτει ένα σύστημα με δύο εξισώσεις και δύο αγνώστους, τους $\widehat{\beta}_0$ και $\widehat{\beta}_1$. Οι εξισώσεις αυτές ονομάζονται κανονικές εξισώσεις και είναι οι ακόλουθες:

$$\sum_{t=1}^{n} Y_t = n\widehat{\beta}_0 + \widehat{\beta}_1 \sum_{t=1}^{n} X_t \tag{1.5}$$

$$\sum_{t=1}^{n} Y_t X_t = \widehat{\beta}_0 \sum_{t=1}^{n} X_t + \widehat{\beta}_1 \sum_{t=1}^{n} X_t^2 \tag{1.6}$$

Απ' τη λύση των εξισώσεων (1.5) και (1.6) παίρνουμε τους εκτιμητές για τους συντελεστές β_0 και β_1, που είναι οι ακόλουθοι:

$$\widehat{\beta}_1 = \frac{n\sum_{t=1}^{n} XY - \sum_{t=1}^{n} X \sum_{t=1}^{n} Y}{n\sum_{t=1}^{n} X^2 - \left(\sum_{t=1}^{n} X\right)^2} \tag{1.7}$$

$$\widehat{\beta}_0 = \frac{\sum_{t=1}^{n} Y}{n} - \widehat{\beta}_1 \frac{\sum_{t=1}^{n} X}{n} \tag{1.8}$$

Αν και από τη μέθοδο των ελαχίστων τετραγώνων προκύπτουν εκτιμητές, που όπως θα δούμε έχουν ορισμένες επιθυμητές ιδιότητες, η γραμμή παλινδρομήσεως του δείγματος δεν παύει να υπόκειται σε σφάλματα. Το ερώτημα, επομένως, είναι κατά πόσο "καλή" είναι η εκτίμηση της γραμμής παλινδρόμησης, δηλαδή πόσο καλοί είναι οι συντελεστές που βρίκαμε πιο πάνω, δηλαδη $\widehat{\beta}_0$ και $\widehat{\beta}_1$ και τι κριτήρια μπορούμε να χρησιμοποιήσουμε για να αξιολογήσουμε τα αποτελέσματα της εκτιμήσεως. Τα κριτήρια αυτά είναι οικονομικά, οικονομετρικά και στατιστικά. Στην παρούσα εργασία θα αναφερθούμε σε ορισμένα κριτήρια, που αποτελούν απόρροια του κλασικού υποδείγματος της κλασσικής γραμμικής παλινδρόμησης, καθώς και σε ορισμένα στατιστικά.

1.2.2 ΣΥΝΤΕΛΕΣΤΗΣ ΠΡΟΣΔΙΟΡΙΣΜΟΥ

Ένα από αυτά τα κριτήρια λοιπόν, είναι και ο συντελεστής προσδιορισμού (Coef-ficient of determination), που συμβολίζεται με R^2. Ο συντελεστής αυτός λαμβάνει τιμές από μηδέν έως και ένα. Όσο πιο κοντά στην μονάδα είναι η τιμή που λαμβάνει, τόσο πιο μεγάλη είναι η ερμηνευτική ικανότητα του υποδείγματος, δηλαδή η παλινδρόμηση ερμηνεύει ένα πολύ μεγάλο μέρος της μεταβλητότητας της Y. Με άλλα λόγια, με τη γραμμή παλινδρομήσεως του δείγματος $\widehat{Y}_t = \widehat{\beta}_0 + \widehat{\beta}_1 X_t$, προσπαθούμε να ερμηνεύσουμε τη μεταβλητότητα της Y που εξηγείται από τις μεταβολές στην τιμή της X, δηλαδή από τη γραμμή παλινδρομήσεως. Το ερώτημα είναι πόση είναι η μεταβλητότητα της Y που εξηγείται από την παλινδρόμηση και πόση μένει ανεξήγητη, δηλαδή οφείλεται στους τυχαίους παράγοντες.

Θα πρέπει να επισημάνουμε ότι η μεταβλητότητα της Y, γενικά ορίζεται πάντοτε σε σχέση με κάποιο σημείο αναφοράς. Δηλαδή σε σχέση με μία τιμή προς την οποία συγκρίνονται όλες οι άλλες τιμές που παίρνει η μεταβλητή Y. Η τιμή αυτή, δηλαδή το σημείο αναφοράς, μπορεί να είναι το μηδέν, ο μέσος της μεταβλητής ή οποιαδήποτε άλλη στατιστική από το δείγμα. Συνήθως, όμως χρησιμοποιείται ο μέσος, όπως θα κάνουμε και στην παρούσα εργασία. Οπότε, η μεταβλητότητα της Y ορίζεται ως το άθροισμα των τετραγώνων των των αποκλίσεων των παρατηρήσεων της Y από το μέσο τους. Δηλαδή:

$$\text{Μεταβλητότητα της Y} = \sum \left(Y - \overline{Y}\right)^2$$

Ο συντελεστής προσδιρισμού μπορεί να υπολογιστεί ως εξής:

$$R^2 = \frac{\sum \left(\widehat{Y} - \overline{Y}\right)^2}{\sum \left(Y - \overline{Y}\right)^2} \tag{1.9}$$

Η σχεση (1.9) μας λέει ότι ο συντελεστής προσδιορισμού ορίζεται ως το πηλικό της μεταβλητότητας που ερμηνεύεται από την παλινδρόμηση προς την συνολική μεταβλητότητα. Βέβαια, επεξεργάζοντας περισσότερο τον τύπο στη σχέση (1.9), παρατηρούμε ότι ο συντελεστής προσδιορισμού μπορεί να πάρει και τις ακόλουθες εκφράσεις:

$$R^2 = \widehat{\beta}_1 \frac{\sum_{t=1}^{n} XY - \frac{\sum_{t=1}^{n} X \sum_{t=1}^{n} Y}{n}}{\sum_{t=1}^{n} Y^2 - \frac{\left(\sum_{t=1}^{n} Y\right)^2}{n}} \tag{1.10}$$

$$R^2 = \frac{\widehat{\beta}_1 \sum \left(X - \overline{X}\right)\left(Y - \overline{Y}\right)}{\sum \left(Y - \overline{Y}\right)^2} \tag{1.11}$$

$$R^2 = \frac{\left(\sum \left(X - \overline{X}\right)\left(Y - \overline{Y}\right)\right)^2}{\sum \left(Y - \overline{Y}\right)^2 \sum \left(X - \overline{X}\right)^2} \tag{1.12}$$

Για χάριν απλότητας και απλούστευσης στην παρουσίαση των τύπων, αντί για $\sum_{t=1}^{n}$, όπου n το μέγεθος του δείγματος, χρησιμοποιήσαμε απλά το σύμβολο $\sum$. Η εύρεση του συντελεστη προσδιορισμου, R^2 μπορεί να υπολογιστεί με έναν από τους ανωτέρω τύπους. Εδώ να επισημάνουμε ότι για τους σκοπούς της εργασίας, δεν μας ενδιαφέρει ο τρόπος απόδειξης των τύπων, αλλά να κατανοήσουμε την σημαντικότητα αυτού του δείκτη στην μετέπειτα πρόβλεψη που θα ακολουθήσει σε επόμενο κεφάλαιο της εργασίας μας.

Θα πρέπει να γίνει κατανοητό ότι ο συντελεστής προσδιορισμού αποτελεί ένα δείκτη, που μας δείχνει πόσο ''καλό'' είναι το υπόδειγμα που έχουμε εκτιμήσει. Όσο μεγαλύτερη είναι η τιμή του συντελεστή , τόσο καλύτερη είναι η προσαρμογή του υποδείγματος στα δεδομένα του δείγματος και αντίστροφα. Δηλαδή, όσο μικρότερη είναι η τιμή του R^2, τόσο φτωχότερη είναι η ποσαρμογή της γραμμής παλινδρομήσεως στα δεδομένα του δείγματος.

1.2.3 ΙΔΙΟΤΗΤΕΣ ΤΩΝ ΕΚΤΙΜΗΤΩΝ

Στο σημείο αυτό, θα πρέπει να ορίσουμε τις έννοιες της αμεροληψίας και της αποτελεσματικότητας.

- **Αμεροληψία**:
  Ένας εκτιμητής $\widehat{\theta}$ μίας παραμέτρου του πληθυσμού θ είναι αμερόληπτος εάν η προσδοκώμενη τιμή του είναι ίση με την παράμετρο, δηλαδη,

$$E\widehat{\theta} = \theta$$

- **Αποτελεσματικότητα**:
  Ένας εκτιμητής $\widehat{\theta}$ είναι αποτελεσματικός εκτιμητής της παραμέτρου θ εάν
  (i) $\widehat{\theta}$ είναι αμερόληπτος
  (ii) $V(\widehat{\theta}) < V(\theta^*)$, όπου θ^* οποιοσδήποτε άλλος αμερόληπτος εκτιμητής της παραμέτρου θ.

Όπως, ήδη έχουμε αναφέρει, οι συντελεστές που προκύπτουν από τη μέθοδο των ελαχίστων τετραγώνων, έχουν ορισμένες επιθυμητές ιδιότητες. Σύμφωνα με το θεώρημα των Gauss-Markov, για το κλασσικό γραμμικό υπόδειγμα, οι εκτιμητές που προκύπτουν από τη μέθοδο των ελαχίστων τετραγώνων είναι άριστοι γραμμικοί αμερόληπτοι εκτιμητές των παραμέτρων τους. Αυτό σημαίνει ότι:

(i) Είναι γραμμικές συναρτήσεις των παρατηρήσεων της εξαρτημένης μεταβλητής Y

(ii) Είναι αμερόληπτοι, δηλαδή

$$E\left(\widehat{\beta}_0\right) = \beta_0$$
$$E\left(\widehat{\beta}_1\right) = \beta_1$$

(iii) Μεταξύ όλων των γραμμικών αμερόληπτων εκτιμητών έχουν τη μικρότερη διακύμανση,δηλ είναι αποτελεσματικοί. Και δίνεται από τις ακόλουθες σχέσεις:

$$V\left(\widehat{\beta}_0\right) = \sigma^2 \frac{\sum X^2}{n \sum \left(X - \overline{X}\right)^2} \tag{1.13}$$

$$V\left(\widehat{\beta}_1\right) = \sigma^2 \frac{1}{\sum \left(X - \overline{X}\right)^2} \tag{1.14}$$

Επιπλέον, η συνδιακύμανση των συντελεστών $\widehat{\beta}_0$ και $\widehat{\beta}_1$ δίνεται από την ακόλουθη σχέση:

$$\mathrm{Cov}\left(\widehat{\beta}_0, \widehat{\beta}_1\right) = -\sigma^2 \frac{\overline{X}}{\sum \left(X - \overline{X}\right)^2}$$

Εκτός από τους εκτιμητές των παραμέτρων β_0 και β_1, ένας άλλος εκτιμητής που θα πρέπει να εξεταστεί είναι αυτός της διακυμάνσεως σ^2. Προηγουμένως, είδαμε ότι οι διακυμάνσεις των εκτιμητών $\widehat{\beta}_0$ και $\widehat{\beta}_1$ είναι συναρτήσεις της άγνωστης διακυμάνσεως (σ^2) του διαταρακτικού όρου ε. Επομένως, για να εκτιμήσουμε τις διακυμάνσεις των συντελεστών, που είναι απαραίτητες για την εφαρμογή των στατιστικών κριτηρίων, θα πρέπει να έχουμε μία εκτίμηση για τη σ^2. Είναι λογικό η εκτίμηση της σ^2 να βασίζεται στα κατάλοιπα $\widehat{\varepsilon}_t$ της γραμμής παλινδρομήσεως. Πράγματι, ένας αμερόληπτος εκτιμητής της σ^2 δίνεται από τη σχέση:

$$S^2 = \frac{\sum \widehat{\varepsilon}_t^2}{n-2}$$

Μία άλλη έκφραση για την S^2, περισσότερο κατάλληλη για τον υπολογισμό της είναι η ακόλουθη:

$$S^2 = \frac{\sum \left(Y - \overline{Y}\right)^2 - \widehat{\beta}_1 \sum \left(X - \overline{X}\right)\left(Y - \overline{Y}\right)}{n-2}$$

Η θετική τετραγωνική ρίζα της διακυμάνσεως S^2 ονομάζεται τυπικό σφάλμα εκτιμήσεως της Y.

1.2.4 ΔΙΑΣΤΗΜΑΤΑ ΕΜΠΙΣΤΟΣΥΝΗΣ

Όπως έχει ήδη ειπωθεί, από το υπόδειγμα της απλής γραμμικής παλινδρόμησης, οι διαταρακτικοί όροι ε_t, κατανέμονται κανονικά και ανεξάρτητα με μέσο το μηδέν και σταθερή διακύμανση, δηλαδή $\varepsilon_t \sim N\left(0, \sigma^2\right)$. Εφόσον, λοιπόν οι εκτιμητές $\widehat{\beta}_0$ και $\widehat{\beta}_1$ είναι γραμμικές συναρτήσεις των τυχαίων μεταβλητών ε_t, που κατανέμονται κανονικά και ανεξάρτητα μεταξύ τους, έπεται ότι οι εκτιμητές $\widehat{\beta}_0$ και $\widehat{\beta}_1$ θα ακολουθούν την κανονική κατανομή. Συνοψίζοντας, παίρνουμε τις ακόλουθες σχέσεις:

$$\widehat{\beta}_0 \sim N\left(\beta_0, \sigma^2 \frac{\sum X^2}{n \sum \left(X - \overline{X}\right)^2}\right) \tag{1.15}$$

$$\widehat{\beta}_1 \sim N\left(\beta_1, \frac{\sigma^2}{\sum \left(X - \overline{X}\right)^2}\right) \tag{1.16}$$

Από τις σχέσεις (1.15) και (1.16), είναι φανερό ότι οι στατιστικές $\frac{\widehat{\beta}_0-\beta_0}{\sigma_{\widehat{\beta}_0}}$ και $\frac{\widehat{\beta}_1-\beta_1}{\sigma_{\widehat{\beta}_1}}$ ακολουθούν την τυποποιημένη κανονική κατανομή, όπου $\sigma_{\widehat{\beta}_0}$ και $\sigma_{\widehat{\beta}_1}$ είναι τα τυπικά σφάλματα των $\widehat{\beta}_0$ και $\widehat{\beta}_1$ αντίστοιχα. Επειδή η διακύμανση σ^2 είναι άγνωστη, την αντικαθιστούμε με τον αμερόληπτο εκτιμητή της, S^2, οπότε οι διακυμάνσεις των εκτιμητών δίνονται από τις ακόλουθες σχέσεις:

$$S^2_{\widehat{\beta}_0} = S^2 \frac{\sum X^2}{n \sum \left(X - \overline{X}\right)^2}$$

$$S^2_{\widehat{\beta}_1} = \frac{S^2}{\sum \left(X - \overline{X}\right)^2}$$

Αν όμως αντί για $\sigma_{\widehat{\beta}_0}$ χρησιμοποιήσουμε $S_{\widehat{\beta}_0}$ και αντί για $\sigma_{\widehat{\beta}_1}$ χρησιμοποιήσουμε $S_{\widehat{\beta}_1}$, τότε οι στατιστικές $\frac{\widehat{\beta}_0-\beta_0}{S_{\widehat{\beta}_0}}$ και $\frac{\widehat{\beta}_1-\beta_1}{S_{\widehat{\beta}_1}}$ δεν ακολουθούν την κανονική κατανομή, αλλά την κατανομή t με n-2 βαθμούς ελευθερίας. Δηλαδή:

$$\frac{\widehat{\beta}_0 - \beta_0}{S_{\widehat{\beta}_0}} \sim t_{n-2} \tag{1.17}$$

$$\frac{\widehat{\beta}_1 - \beta_1}{S_{\widehat{\beta}_1}} \sim t_{n-2} \tag{1.18}$$

Αυτή η αλλαγή της κατανομής οφείλεται στο εξής: Εάν Z και ω δύο ανεξάρτητες τυχαίες μεταβλητές όπου $Z \sim N(0,1)$ και $\omega \sim X^2_\nu$, τότε η $t = \frac{Z}{\sqrt{\frac{\omega}{\nu}}}$ ακολουθεί την κατανομή t με ν βαθμούς ελευθερίας. Η στατιστική όμως $\frac{(n-2)s^2}{\sigma^2}$ ακολουθεί τη X^2 κατανομή με n-2 βαθμούς ελευθερίας και είναι ανεξάρτητη από τη στατιστική $\frac{\left(\widehat{\beta}_1-\beta_1\right)}{\sigma/\sqrt{\sum\left(X-\overline{X}\right)^2}}$, η οποία ακολουθεί την τυποποιημένη κανονική κατανομή. Επομένως, η στατιστική:

$$\frac{\frac{\left(\widehat{\beta}_1-\beta_1\right)}{\sigma/\sqrt{\sum\left(X-\overline{X}\right)^2}}}{\sqrt{\frac{(n-2)s^2}{(n-2)\sigma^2}}} = \frac{\frac{\left(\widehat{\beta}_1-\beta_1\right)}{\sigma/\sqrt{\sum\left(X-\overline{X}\right)^2}}}{\frac{s}{\sigma}} = \frac{\widehat{\beta}_1 - \beta_1}{s/\sqrt{\sum\left(X-\overline{X}\right)^2}} = \frac{\widehat{\beta}_1 - \beta_1}{s_{\widehat{\beta}_1}}$$

ακολουθεί την t κατανομή με n-2 βαθμούς ελευθείας.

Μπορούμε τώρα χρησιμοποιώντας τις σχέσεις (1.17) και (1.18), να κατασκευάσουμε τα $100\,(1-\alpha)\,\%$ διαστήματα εμπιστοσύνης για τους συντελεστες β_0 και

β_1, ως εξής:

$$\widehat{\beta}_0 - t_{\alpha/2} s_{\widehat{\beta}_0} \leqslant \beta_0 \leqslant \widehat{\beta}_0 + t_{\alpha/2} s_{\widehat{\beta}_0} \tag{1.19}$$

$$\widehat{\beta}_1 - t_{\alpha/2} s_{\widehat{\beta}_1} \leqslant \beta_1 \leqslant \widehat{\beta}_1 + t_{\alpha/2} s_{\widehat{\beta}_1} \tag{1.20}$$

Μέχρι στιγμής, υπολογίσαμε τα διαστήματα εμπιστοσύνης για τους συντελεστές των ελαχίστων τετραγώνων. Ωστόσο για την διαμόρφωση προβλέψεων, θα πρέπει να γνωρίζουμε και το διάστημα εμπιστοσύνης για την προσδοκώμενη τιμή της Y_t.

Όπως ήδη έχουμε αναφέρει, η γραμμή παλινδρομήσεως του δείγματος, $\widehat{Y}_t = \widehat{\beta}_0 + \widehat{\beta}_1 X_t$, είναι μία εκτίμηση της γραμμής παλινδρομήσεως στον πληθυσμό, $E(Y_t) = \beta_0 + \beta_1 X_t$. Με άλλα λόγια, ο $\widehat{Y}_t$ είναι εκτιμητής του (υπό συνθήκη) μέσου της Y_t (δεδομένης της τιμής X_t). Επιπλέον, εφόσον $E\left(\widehat{Y}_t\right) = E\left(\widehat{\beta}_0 + \widehat{\beta}_1 X_t\right) = \beta_0 + \beta_1 X_t$, έπεται ότι ο $\widehat{Y}_t$ είναι αμερόληπτος εκτιμητής του $E(Y_t)$. Επιπλέον, η διακύμανση του $\widehat{Y}_t$ ορίζεται ώς:

$$V\left(\widehat{Y}_t\right) = \sigma^2_{\widehat{Y}_t} = \sigma^2 \left(\frac{1}{n} + \frac{\left(X_t - \overline{X}\right)^2}{\sum \left(X - \overline{X}\right)^2}\right)$$

Επειδή τώρα $\widehat{Y}_t = \widehat{\beta}_0 + \widehat{\beta}_1 X_t$ είναι γραμμική συνάρτηση των τυχαίων μεταβλητών ε_t, που έχουμε υποθέσει ότι κατανέμονται κανονικά και ανεξάρτητα μεταξύ τους, έπεται πως και $\widehat{Y}_t$ ακολουθεί την κανονική κατανομή με μέσο $\beta_0 + \beta_1 X_t$ και διακύμανση την $\sigma^2_{\widehat{Y}_t}$. Επομένως,

$$\frac{\widehat{Y}_t - (\beta_0 + \beta_1 X_t)}{\sigma_{\widehat{Y}_t}} \sim N(0, 1) \tag{1.21}$$

Αν στην (1.21) αντικαταστήσουμε την άγνωστη διακύμανση σ^2 με την αμερόληπτη εκτίμησή της, s^2, τότε:

$$\frac{\widehat{Y}_t - (\beta_0 + \beta_1 X_t)}{s_{\widehat{Y}_t}} \sim t_{n-2} \tag{1.22}$$

όπου

$$s_{\widehat{Y}_t} = s \sqrt{\frac{1}{n} + \frac{\left(X_t - \overline{X}\right)^2}{\sum \left(X - \overline{X}\right)^2}}$$

Συνεπώς, χρησιμοποιώντας την (1.22) μπορούμε να κατασκευάσουμε το $100(1-\alpha)\,\%$ διάστημα εμπιστοσύνης για την προσδοκώμενη τιμή της Y ως εξής:

$$\widehat{Y}_t - t_{\alpha/2} s_{\widehat{Y}_t} \leqslant E(Y_t) \leqslant \widehat{Y}_t + t_{\alpha/2} s_{\widehat{Y}_t}$$

1.2.5 ΔΙΑΜΟΡΦΩΣΗ ΠΡΟΒΛΕΨΕΩΝ

Η διαμόρφωση προβλέψεων βάσει του απλού γραμμικού υποδείγματος είναι η ακόλουθη: Έστω X_F μια γνωστή τιμή της ανεξάρτητης μεταβλητής για μία μελλοντική περίοδο. Με βάση την X_F ενδιαφερόμαστε να κάνουμε μια πρόβλεψη για την αντίστοιχη τιμή Y_F. Εάν υποθέσουμε ότι η ίδια γραμμική σχέση που συνδέει τις μεταβλητές Y και X για την περίοδο του δείγματος, ισχύει και για την περίοδο της προβλέψεως, τότε:

$$Y_F = \beta_0 + \beta_1 X_F + \varepsilon_F$$

Αν επομένως γνωρίζαμε τη γραμμή παλινδρομήσεως στον πληθυσμό και τον τυχαίο όρο, θα μπορούσαμε να βρούμε την Y_F. Εφόσον όμως η Y_F είναι μία τυχαία μεταβλητή, ακόμη και αν γνωρίζαμε την παλινδόμηση στον πληθυσμό, η πρόβλεψη μας θα ήταν πρόβλεψη του μέσου της Y_F, δηλαδή:

$$E(Y_F) = \beta_0 + \beta_1 X_F$$

Επομένως, όταν χρησιμοποιούμε τη γραμμή παλινδρομήσεως του δείγματος, η

$$\widehat{Y}_F = \widehat{\beta}_0 + \widehat{\beta}_1 X_F \tag{1.23}$$

είναι ο ”καλύτερος” εκτιμητής που έχουμε της $E(Y_F)$, δηλαδή του μέσου της Y_F. Η τιμή $\widehat{Y}_F$ στην σχέση (1.23) απεικονίζει την προβλεπόμενη τιμή της εξαρτημένης μεταβλητής που προκύπτει από το χρησιμοποιούμενο εκτιμηθέν υπόδειγμα και δηλώνει την πλέον πιθανή τιμή της εξαρτημένης μεταβλητής που αναμένεται να εμφανιστεί.
Ωστόσο, όταν χρησιμοποιούμε την $\widehat{Y}_F$ ως εκτιμητή της Y_F, κάνουμε το λεγόμενο σφάλμα προβλέψεως που το παριστάνουμε με F. Δηλαδή:

$$F = Y_F - \widehat{Y}_F$$

Το σφάλμα προβλέψεως μπορούμε να το εκφράσουμε και ως εξής:

$$F = Y_F - \widehat{Y}_F = (Y_F - E(Y_F)) - \left(\widehat{Y}_F - E(Y_F)\right) = -\left(\widehat{\beta}_0 - \beta_0\right) + \left(\widehat{\beta}_1 - \beta_1\right) X_F + \varepsilon_F$$

Είναι φανερό λοιπόν, ότι το σφάλμα προβλέψεως είναι συνάρτηση των εκτιμητών $\widehat{\beta}_0$ και $\widehat{\beta}_1$, καθώς και του διαταρακτικού όρου ε_F. Οι εκτιμητές όμως $\widehat{\beta}_0$ και

$\widehat{\beta}_1$, είναι γραμμικές συναρτήσεις των τυχαίων μεταβλητών $\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_n$. Κατά συνέπεια, το σφάλμα πρόβλεψης είναι γραμμικός συνδιασμός των τυχαίων μεταβλητών $\varepsilon_1, \varepsilon_2, \ldots,$
ε_n, οι οποίες είναι ανεξάρτητες μεταξύ τους και ακολουθούν την κανονική κατανομή. Επομένως και το F ακολουθεί την κανονική κατανομή με μέσο μηδέν και διακύμανση σ_F^2, οπότε έχουμε τη σχέση:

$$\frac{F-0}{\sigma_F} = \frac{Y_F - \widehat{Y}_F}{\sigma_F} \sim N\left(0, 1\right) \tag{1.24}$$

όπου η διακύμανση του σφάλματος πρόβλεψης δίνεται από την ακόλουθη σχέση:

$$\sigma_{\widehat{Y}_F}^2 = \sigma^2 \left(1 + \frac{1}{n} + \frac{\left(X_F - \overline{X}\right)^2}{\sum \left(X - \overline{X}\right)^2}\right)$$

Αν στην (1.24) αντικαταστήσουμε την άγνωστη διακύμανση σ^2 με την αμερόληπτη εκτίμηση της, s^2, τότε:

$$\frac{Y_F - \widehat{Y}_F}{s_F} \sim t_{n-2} \tag{1.25}$$

όπου

$$s_F = s\sqrt{1 + \frac{1}{n} + \frac{\left(X_F - \overline{X}\right)^2}{\sum \left(X - \overline{X}\right)^2}}$$

Μπορούμε τώρα να χρησιμοποιήσουμε την κατανομή t για να κατασκευάσουμε το $100\left(1-\alpha\right)\%$ διάστημα εμπιστοσύνης για την Y_F, που δίνεται από τη σχέση:

$$\widehat{Y}_F - t_{\alpha/2} s_F \leqslant Y_F \leqslant \widehat{Y}_F + t_{\alpha/2} s_F$$

Το παραπάνω διάστημα εμπιστοσύνης είναι διάστημα εμπιστοσύνης για μία συγκεκριμένη ή ατομική τιμή της Y, δεδομένης της τιμής X_F. Το αντίστοιχο διάστημα εμπιστοσύνης για την αντίστοιχη μέση τιμή της Y, δηλαδή $E\left(Y_F\right)$, είναι:

$$\widehat{Y}_F - t_{\alpha/2} s_{\widehat{Y}_F} \leqslant E\left(Y_F\right) \leqslant \widehat{Y}_F + t_{\alpha/2} s_{\widehat{Y}_F}$$

όπου

$$s_{\widehat{Y}_F} = s\sqrt{\frac{1}{n} + \frac{\left(X_F - \overline{X}\right)^2}{\sum \left(X - \overline{X}\right)^2}}$$

1.3 ΠΟΛΛΑΠΛΗ ΓΡΑΜΜΙΚΗ ΠΑΛΙΝΔΡΟΜΗΣΗ

Το υπόδειγμα της απλής γραμμικής παλινδρόμησης,, που αναπτύξαμε στην προηγούμενη ενότητα, αναφέρεται σε σχέσεις που περιλαμβάνουν μία μόνο ερμηνευτική μεταβλητή. Ωστόσο, η συμπεριφορά των περισσοτέρων οικονομικών μεταβλητών είναι συνάρτηση όχι μίας αλλά πολλών μεταβλητών. Έστω ότι $Y = f(X_1, X_2, \ldots, X_k)$, δηλαδή η Y είναι συνάρτηση των k ερμηνευτικών μεταβλητών $X_1, X_2, \ldots, X_k$. Αν υποθέσουμε ότι η συναρτησιακή σχέση $Y = f(X_1, X_2, \ldots, X_k)$ είναι γραμμική, για ένα δείγμα από n παρατηρήσεις, μπορούμε να γράψουμε:

$$Y_t = \beta_0 + \beta_1 X_{t1} + \beta_2 X_{t2} + \ldots + \beta_k X_{tk} + \varepsilon_t \tag{1.26}$$

όπου: X_{t1} είναι η t παρατήρηση της ερμηνευτικής μεταβλητής X_1, X_{t2} είναι η t παρατήρηση της ερμηνευτικής μεταβλητής X_2 κ.ο.κ. Ο πρώτος δηλαδή δείκτης αναφέρεται στην παρατήρηση και ο δεύτερος στην ερμηνευτική μεταβλητή.

Η σχέση (1.26) αποτελεί το υπόδειγμα της γραμμικής πολυμεταβλητής παλινδρομήσεως, που είναι επέκταση της απλής παλινδρόμησης. Για $k = 1$, η παραπάνω σχέση γίνεται το υπόδειγμα της απλής γραμμικής παλινδρόμησης.

Οι βασικές υποθέσεις που συνιστούν το κλασικό γραμμικό υπόδειγμα παλινδρόμησης στη γενική του μορφή, δηλαδή με k ερμηνευτικές μεταβλητές είναι σχεδόν οι ίδιες με τις υποθέσεις του διμεταβλητού γραμμικού υποδείγματος. Οι υποθέσεις αυτές, που πρέπει να ισχύουν για όλες τις παρατηρήσεις είναι οι ακόλουθες:

(i) $Y_t = \beta_0 + \beta_1 X_{t1} + \beta_2 X_{t2} + \ldots + \beta_k X_{tk} + \varepsilon_t$
Η υπόθεση αυτή αναφέρεται στη γραμμική σχέση που συνδέει τις μεταβλητές Y και $X_1, X_2, \ldots, X_k$. Δηλαδή, κάθε τιμή της εξαρτημένης μεταβλητής είναι γραμμική συνάρτηση των τιμών των ερμηνευτικών μεταβλητών $X_1, X_2, \ldots, X_k$ και του διαταρακτικού όρου ε_t.

(ii) $E(\varepsilon_t) = 0$
Η μεταβλητή ε_t είναι τυχαία μεταβλητή που παίρνει τιμές θετικές και αρνητικές, αλλά κατά μέσο όρο η τιμή της είναι μηδέν.

(iii) $Var(\varepsilon_t) = E\varepsilon_t^2 = \sigma^2$
Η διακύμανση της τυχαίας μεταβλητής ε_t είναι σταθερή για όλες τις τιμές των μεταβλητών $X_1, X_2, \ldots, X_k$. Δηλαδή η διασπορά των τιμών της ε_t από τον μέσο δεν αλλάζει όταν μεταβάλλονται οι τιμές των $X_1, X_2, \ldots, X_k$, αλλά παραμένει η ίδια. Η υπόθεση αυτή είναι γνωστή ως υπόθεση της ομοσκεδαστικότητας και η παραβίαση της προκαλεί το πρόβλημα της ετεροσκεδαστικότητας.

(iv) $\text{Cov}\,(\varepsilon_t, \varepsilon_s) = 0$
Η υπόθεση αυτή σημαίνει ότι οι διαταρακτικοί όροι είναι ασυσχέτιστοι μεταξύ τους και επομένως, η συνδιακύμανση του διαταρακτικού όρου της παρατήρησης t με το διαταρακτικό όρο οποιασδήποτε άλλης παρατήρησης s, είναι μηδέν. Δηλαδή, $\text{Cov}\,(\varepsilon_t, \varepsilon_s) = E\,(\varepsilon_t - E\varepsilon_t)\,(\varepsilon_t - E\varepsilon_s) = E\varepsilon_t\varepsilon_s = 0$. Η παραβίαση της υπόθεσης αυτής δημιουργεί το πρόβλημα της αυτοσυσχέτισης.

(v) Οι ερμηνευτικές μεταβλητές δεν είναι στοχαστικές. Οι τιμές τους παραμένουν σταθερές και δεν είναι όλες ίσες μεταξύ τους.

(vi) $\varepsilon_t \sim N\,(0, \sigma^2)$
Δηλαδή, οι όροι σφάλματος ή διαταρακτικοί όροι ακολουθούν την κανονική κατανομή. Και όπως ήδη αναφέραμε, με μέσο όρο 0 και διασπορά ίση με σ^2.

(vii) Δεν υπάρχουν ακριβείς γραμμικές σχέσεις ανάμεσα στις ερμηνευτικές μεταβλητές. Η υπόθεση αυτή, αποκλείει την ύπαρξη τέλειας πολλυσυγγραμικότητας μεταξύ των ερμηνευτικών μεταβλητών. Πράγμα που σημαίνει πως καμία από τις k ερμηνευτικές μεταβλητές δεν μπορεί να εκφρασθεί ως γραμμικός συνδιασμός των υπολοίπων.

(viii) Ο αριθμός των παρατηρήσεων του δείγματος είναι μεγαλύτερος από τον αριθμό των συντελεστών του υποδείγματος που θέλουμε να εκτιμήσουμε. Η υπόθεση αυτή εξασφαλίζει τους απαραίτητους βαθμούς ελευθερίας και για την εκτίμηση αλλά και για τον έλεγχο του υποδείγματος. Ο αριθμός των παρατηρήσεων πρέπει να είναι τουλάχιστον ίσος με τους συντελεστές του υποδείγματος, για να είναι δυνατή η εκτίμηση του. Πρέπει ωστόσο να είναι και μεγαλύτερος, για να είναι δυνατός ο έλεγχος υποθέσεων με τις διάφορες στατιστικές ελέγχου που η κατανομή τους εξαρτάται από τους βαθμούς ελευθερίας, όπως η κατανομή t ή η κατανομή F.

Η σχέση (1.26) μπορεί να γραφεί και για λόγους συμμετρίας, και ως εξής:

$$Y_t = \beta_0 X_{t0} + \beta_1 X_{t1} + \beta_2 X_{t2} + \ldots + \beta_k X_{tk} + \varepsilon_t \tag{1.27}$$

όπου $X_{t0} = 1$ για όλες τις τιμές $t = 1, 2, \ldots, n$. Δηλαδή, ο σταθερός όρος β_0 μπορεί να θεωρηθεί ως συντελεστής της μεταβλητής X_0, η οποία είναι ίση με τη μονάδα. Η σχέση (1.27) ισχύει για κάθε παρατήρηση και επομένως, για ένα δείγμα από n παρατηρήσεις έχουμε το ακόλουθο σύστημα:

$$\begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix} = \begin{bmatrix} X_{10} & X_{11} & \cdots & X_{1k} \\ X_{20} & X_{21} & \cdots & X_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ X_{n0} & X_{n1} & \cdots & X_{nk} \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_k \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}$$

ή

$$Y = X\beta + \varepsilon$$

όπου:

- το Y είναι ένα διάνυσμα στήλης, διαστάσεων n x 1, των τιμών της εξαρτημένης μεταβλητής Y.
- το X είναι ένας πίνακας διαστάσεων n x (k+1) των τιμών των ερμηνευτικών μεταβλητών με τα στοιχεία της πρώτης στήλης ίσα με τη μονάδα.
- το β είναι ένα διάνυσμα στήλης διαστάσεων (k+1) x 1 των συντελεστών παλινδρόμησης.
- το ε είναι ένα διάνυσμα στήλης διαστάσεων n x 1 των τιμών του διαταρακτικού όρου

1.3.1 ΜΕΘΟΔΟΣ ΕΛΑΧΙΣΤΩΝ ΤΕΤΡΑΓΩΝΩΝ

Όπως και στην απλή παλινδρόμηση, με τη μέθοδο των ελαχίστων τετραγώνων ελαχιστοποιούμε το άθροισμα των τετραγώνων των καταλοίπων, δηλαδή ελαχιστοποιούμε τη συνάρτηση:

$$\Phi = \sum \widehat{\varepsilon}_t^2 = \sum \left(Y_t - \widehat{Y}_t\right)^2$$

όπου:

$$\widehat{Y}_t = \widehat{\beta}_0 + \widehat{\beta}_1 X_{t1} + \widehat{\beta}_2 X_{t2} + \ldots + \widehat{\beta}_k X_{tk}$$

είναι η παλινδρόμηση στο δείγμα. Τώρα, για να ελαχιστοποιήσουμε την συνάρτηση Φ, υπολογίζουμε τις εξής k+1 μερικές παραγώγους:

$$\begin{aligned} \frac{\partial\Phi}{\partial\widehat{\beta}_0} &= -2\sum\left(Y_t - \widehat{\beta}_0 - \widehat{\beta}_1 X_{t1} - \widehat{\beta}_2 X_{t2} - \ldots - \widehat{\beta}_k X_{tk}\right) \\ \frac{\partial\Phi}{\partial\widehat{\beta}_1} &= -2\sum\left(Y_t - \widehat{\beta}_0 - \widehat{\beta}_1 X_{t1} - \widehat{\beta}_2 X_{t2} - \ldots - \widehat{\beta}_k X_{tk}\right)(X_{t1}) \\ &\vdots \end{aligned}$$

$\frac{\partial \Phi}{\partial \widehat{\beta}_k} = -2 \sum \left(Y_t - \widehat{\beta}_0 - \widehat{\beta}_1 X_{t1} - \widehat{\beta}_2 X_{t2} - \ldots - \widehat{\beta}_k X_{tk} \right) (X_{tk})$

Και εξισώνοντας τις παραπάνω μερικές παραγώγους με το μηδέν, παίρνουμε τι ακόλουθες k+1 εξισώσεις, οι οποίες είναι γνωστές ως ”κανονικές” εξισώσεις.

$$
\begin{aligned}
&\sum Y = \widehat{\beta}_0 n + \widehat{\beta}_1 \sum X_1 + \widehat{\beta}_2 \sum X_2 + \ldots + \widehat{\beta}_k \sum X_k \\
&\sum Y X_1 = \widehat{\beta}_0 \sum X_1 + \widehat{\beta}_1 \sum X_1^2 + \widehat{\beta}_2 \sum X_2 X_1 + \ldots + \widehat{\beta}_k \sum X_k X_1 \\
&\vdots \\
&\sum Y X_k = \widehat{\beta}_0 \sum X_k + \widehat{\beta}_1 \sum X_1 X_k + \widehat{\beta}_2 \sum X_2 X_k + \ldots + \widehat{\beta}_k \sum X_k^2
\end{aligned}
$$

Υπό τη μορφή πινάκων, οι κανονικές εξισώσεις μπορούν να γραφτούν και με την ακόλουθη μορφή:

$$\left(X^T X\right) \widehat{\beta} = X^T Y$$

όπου:
X^T ο ανάστροφος πίνακας του X και

$$X^T Y = \begin{bmatrix} \sum Y \\ \sum Y X_1 \\ \vdots \\ \sum Y X_k \end{bmatrix}, \ \widehat{\beta} = \begin{bmatrix} \widehat{\beta}_0 \\ \widehat{\beta}_1 \\ \vdots \\ \widehat{\beta}_k \end{bmatrix}$$

$$X^T X = \begin{bmatrix} n & \sum X_1 & \cdots & \sum X_k \\ \sum X_1 & \sum X_1^2 & \cdots & \sum X_1 X_k \\ \vdots & \vdots & \ddots & \vdots \\ \sum X_k & \sum X_1 X_k & \cdots & \sum X_k^2 \end{bmatrix}$$

Η συνάρτηση Φ, υπό τη μορφή πινάκων παίρνει και την εξής μορφή:

$$\Phi = \widehat{\varepsilon}^T \widehat{\varepsilon} = \left(Y - X\widehat{\beta}\right)^T \left(Y - X\widehat{\beta}\right) \tag{1.28}$$

όπου $\widehat{\varepsilon}^T$ είναι ανάστροφος του $\widehat{\varepsilon}$. Αναπτύσσωντας, λοιπόν, την (1.28), προκύπτει:

$$\Phi = Y^T Y - \widehat{\beta}^T X^T Y - Y^T X \widehat{\beta} + \widehat{\beta}^T X^T X \widehat{\beta}$$

αλλά $\widehat{\beta}^T X^T Y = Y^T X \widehat{\beta}$. Επομένως, η παραπάνω σχέση γράφεται ως:

$$\Phi = Y^T Y - 2\widehat{\beta}^T X^T Y + \widehat{\beta}^T X^T X \widehat{\beta}$$

Οι εκτιμητές τώρα των ελαχίστων τετραγώνων που θα προκύψουν από την λύση της $\frac{\partial\Phi}{\partial\widehat{\beta}} = 0$, δίνονται από τη σχέση:

$$\widehat{\beta} = \left(X^T X\right)^{-1}\left(X^T Y\right) \tag{1.29}$$

Από τη σχέση (1.29) βλέπουμε ότι για να έχει λύση το σύστημα, θα πρέπει να υπάρχει ο αντίστροφος $\left(X^T X\right)^{-1}$ του συμμετρικού πίνακα $\left(X^T X\right)$. Αυτό σημαίνει πως, εφόσον ο $\left(X^T X\right)$ είναι διαστάσεων (k+1)x(k+1), ο βαθμός του θα πρέπει να είναι (k+1), που με τη σειρά του έπεται ότι ο βαθμός του X είναι (k+1).

Οι εκτιμητές των συντελεστών $\beta_0, \beta_1, \ldots, \beta_k$, που προέκυψαν από τη μέθοδο των ελαχίστων τετραγώνων, τόσο στο απλό γραμμικό υπόδειγμα που έχουμε αναφέρει, όσο και στο πολλαπλό γραμμικό υπόδειγμα, είναι άριστοι, γραμμικοί και αμερόληπτοι, οι δε διακυμάνσεις τους δίνονται από τη σχέση:

$$V\left(\widehat{\beta}\right) = \sigma^2\left(X^T X\right)^{-1} \tag{1.30}$$

Ο πίνακας $V\left(\widehat{\beta}\right)$ είναι διαστάσεων (k+1)x(k+1) και τα διαγώνια στοιχεία δίνουν τις διακυμάνσεις των εκτιμητών $\widehat{\beta}_0, \widehat{\beta}_1, \ldots, \widehat{\beta}_k$, ενώ τα υπόλοιπα στοιχεία τις συνδιακυμάνσεις. Πιο αναλυτικά,

$$V\left(\widehat{\beta}\right) = \begin{bmatrix} V\left(\widehat{\beta}_0\right) & Cov\left(\widehat{\beta}_0, \widehat{\beta}_1\right) & \cdots & Cov\left(\widehat{\beta}_0, \widehat{\beta}_k\right) \\ Cov\left(\widehat{\beta}_1, \widehat{\beta}_0\right) & V\left(\widehat{\beta}_1\right) & \cdots & Cov\left(\widehat{\beta}_1, \widehat{\beta}_k\right) \\ \vdots & \vdots & \ddots & \vdots \\ Cov\left(\widehat{\beta}_k, \widehat{\beta}_0\right) & Cov\left(\widehat{\beta}_k, \widehat{\beta}_1\right) & \cdots & V\left(\widehat{\beta}_k\right) \end{bmatrix}$$

ή

$$V\left(\widehat{\beta}\right) = \sigma^2 \begin{bmatrix} n & \sum X_1 & \cdots & \sum X_k \\ \sum X_1 & \sum X_1^2 & \cdots & \sum X_1 X_k \\ \vdots & \vdots & \ddots & \vdots \\ \sum X_k & \sum X_1 X_k & \cdots & \sum X_k^2 \end{bmatrix}^{-1}$$

1.3.2 ΔΙΑΣΤΗΜΑΤΑ ΕΜΠΙΣΤΟΣΥΝΗΣ

Για το στατιστικό έλεγχο του υποδείγματος, όπως και στην απλή παλινδρόμηση, θα υποθέσουμε ότι ο διαταρακτικός όρος ε_t είναι μία τυχαία μεταβλητή που ακολουθεί την κανονική κατανομή. Δηλαδή, $\varepsilon_t \sim N\left(0, \sigma^2\right)$. Στη πολλαπλή γραμμική παλινδρόμηση έχουμε τα εξής:

- Οι εκτιμητές $\widehat{\beta}_0, \widehat{\beta}_1, \dots, \widehat{\beta}_k$ ακολουθούν την κανονική κατανομή, εφόσον είναι γραμμικές συναρτήσεις των τυχαίων μεταβλητών $\varepsilon_1, \varepsilon_2, \dots, \varepsilon_n$, που κατανέμονται κανονικά. Επομένως, αφού $E\left(\widehat{\beta}_j\right) = \beta_j$, η στατιστική $\frac{\widehat{\beta}_j - \beta_j}{\sigma\sqrt{c_{jj}}}$ ακολουθεί την κανονική κατανομή με μέσο το μηδέν και διακύμανση τη μονάδα, δηλαδή,

$$\frac{\widehat{\beta}_j - \beta_j}{\sigma\sqrt{c_{jj}}} \sim N(0,1) \tag{1.31}$$

  όπου c_{jj} το j διαγώνιο στοιχείο του πίνακα $\left(X^T X\right)^{-1}$.

- Η στατιστική $(n-k-1)\frac{s^2}{\sigma^2}$ ακολουθεί την κατανομή X^2 με $(n-k-1)$ βαθμούς ελευθερίας, δηλαδή,

$$(n-k-1)\frac{s^2}{\sigma^2} \sim X^2_{n-k-1} \tag{1.32}$$

  και είναι ανεξάρτητη από την προηγούμενη στατιστική.

Επομένως, απ' τον ορισμό της κατανομής t έχουμε ότι:

$$\frac{\widehat{\beta}_j - \beta_j}{S\sqrt{c_{jj}}} \sim t_{n-k-1} \tag{1.33}$$

Μπορούμε τώρα χρησιμοποιώντας τη σχέση (1.33), να κατασκευάσουμε το $100(1-\alpha)\%$ διάστημα εμπιστοσύνης για το συντελεστή β_j ως εξής:

$$\widehat{\beta}_j - t_{\alpha/2} S\sqrt{c_{jj}} \leqslant \beta_j \leqslant \widehat{\beta}_j + t_{\alpha/2} S\sqrt{c_{jj}} \tag{1.34}$$

1.3.3 ΔΙΑΜΟΡΦΩΣΗ ΠΡΟΒΛΕΨΕΩΝ

Η λογική της διενέργειας προβλέψεων βάσει του πολλαπλού γραμμικού υποδείγματος είναι η ακόλουθη: Έστω $X_{F1}, \dots, X_{Fk}$ γνωστές τιμές των ανεξάρτητων μεταβλητών έξω από το δείγμα, βάσει του οποίου εκτιμήσαμε την παλινδρόμηση. Με βάση τα $X_{F1}, \dots, X_{Fk}$ ενδιαφερόμαστε να κάνουμε μία πρόβλεψη για την αντίστοιχη τιμή Y_F, της εξαρτημένης μεταβλητής. Συνεπώς, η πραγματική τιμή θα είναι: $Y_F = \beta_0 + \beta_1 X_{F1} + \dots + \beta_k X_{Fk} + \varepsilon_F$ ενώ βάσει της εκτιμημένης παλινδρόμησης η πρόβλεψη θα είναι:

$$\widehat{Y}_F = \widehat{\beta}_0 + \widehat{\beta}_1 X_{F1} + \dots + \widehat{\beta}_k X_{Fk}$$

Η τιμή $\widehat{Y}_F$ απεικονίζει την προβλεπόμενη τιμή της εξαρτημένης μεταβλητής που προκύπτει από το χρησιμοποιούμενο εκτιμηθέν υπόδειγμα και δηλώνει την πλέον πιθανή τιμή της εξαρτημένης μεταβλητής που αναμένεται να εμφανιστεί. Η τιμή αυτή χρησιμοποιείται για να προσδιορίσουμε το $100\,(1{-}\alpha)\,\%$ διάστημα εμπιστοσύνης της πραγματικής τιμής Y_F της εξαρτημένης μεταβλητής, το οποίο υπολογίζεται ως εξής :

$$\widehat{Y}_F - t_{\alpha/2} s_F \leqslant Y_F \leqslant \widehat{Y}_F + t_{\alpha/2} s_F$$

Το αντίστοιχο διάστημα εμπιστοσύνης για την $E\,(Y_F)$ δίνεται από τη σχέση:

$$\widehat{Y}_F - t_{\alpha/2} s_{\widehat{Y}_F} \leqslant E\,(Y_F) \leqslant \widehat{Y}_F + t_{\alpha/2} s_{\widehat{Y}_F}$$

όπου t η κατανομή student και s_F το τυπικό σφάλμα (απόκλιση) της Y_F.

1.4 ΠΡΟΒΛΗΜΑΤΑ ΣΤΗΝ ΑΝΑΛΥΣΗ ΠΑΛΙΝΔΡΟΜΗΣΗΣ

Τα κυριότερα προβλήματα που δύναται να εμφανιστούν κατά την ανάλυση παλινδρόμησης είναι τα εξής:

(i) **Η πολυσυγγραμμικότητα**
Η πολυσυγγραμμικότητα έγκειται στην ύπαρξη γραμμικών σχέσεων ανάμεσα στις ερμηνευτικές μεταβλητές.Στο πολλαπλό γραμμικό υπόδειγμα οι ανεξάρτητες μεταβλητές δεν πρέπει να είναι μεταξύ τους γραμμικά εξαρτημένες. Η παραβίαση της υπόθεσης αυτής συνιστά το πρόβλημα της πολυσυγγραμμικότητας. Δηλαδή οι ανεξάρτητες μεταβλητές συσχετίζονται μεταξύ τους και συμβάλλουν στην δημιουργία πλασματικών πληροφοριών. Το φαινόμενο της πολυσυγγραμμικότητας είναι αρκετά συνηθισμένο όταν χρησιμοποιούνται χρονολογικές σειρές, καθώς οι μεταβλητές που χρησιμοποιούνται, συνήθως οικονομικές, έχουν την τάση να συμμεταβάλλονται διαχρονικά. Επίσης, είναι σχεδόν βέβαιο ότι θα υπάρχει πολυσυγγραμμικότητα σ' ένα υπόδειγμα παλινδρομήσεως, που ανάμεσα στις ερμηνευτικές μεταβλητές περιλαμβάνονται και δύο ή περισσότερες μεταβλητές που αναφέρονται σε συνολικά μακροοικονομικά μεγέθη, όπως οι διάφορες κατηγορίες εισοδήματος και η κατανάλωση. Η πολυσυγγραμμικότητα δεν είναι χαρακτηριστικό μόνο των χρονολογικών σειρών, αλλά συχνά εμφανίζεται και σε διαστρωματικά στοιχεία. Είναι πολύ πιθανό για παράδειγμα να υπάρχει πολυσυγγραμμικότητα σ' ένα δείγμα βιομηχανιών με στοιχεία για τις εισροές των συντελεστών (εργασία, κεφάλαιο). Θα πρέπει λοιπόν να είμαστε ιδιαίτερα προσεκτικοί και να λαμβάνουμε σοβαρά το πρόβλημα της πολυσυγγραμμικότητας, προκειμένου να το αναγνωρίσουμε και να το αντιμετωπίσουμε.

(ii) **Η ετεροσκεδαστικότητα**
Ένα άλλο σημαντικό πρόβλημα που μπορεί να αντιμετωπίσουμε κατά την ανάλυση παλινδρόμησης είναι η ετεροσκεδαστικότητα. Στο κλασσικό γραμμικό υπόδειγμα έχουμε κάνει την υπόθεση ότι η διασπορά των διαταρακτικών όρων ε_t είναι σταθερή για κάθε t, δηλαδή $Var(\varepsilon_t) = \sigma^2$. Όταν παραβιάζεται αυτή η υπόθεση έχουμε το πρόβλημα της ετεροσκεδαστικότητας. Γενικά, το φαινόμενο της ετεροσκεδαστικότητας είναι πιο συνηθισμένο κυρίως σε υποδείγματα που για την εκτίμησή τους χρησιμοποιούνται διαστρωματικά στοιχεία.

(iii) **Η αυτοσυσχέτιση**
Στο κλασσικό γραμμικό υπόδειγμα υποθέσαμε επίσης ότι η συνδιακύμανση των διαταρακτικών όρων είναι μηδέν. Δηλαδή, $Cov(\varepsilon_t, \varepsilon_s) = 0$, για $t \neq s$. Η υπόθεση αυτή σημαίνει πως οι διάφορες τιμές του διαταρακτικού όρου δεν συσχετίζονται. Με άλλα λόγια, ο διαταρακτικός όρος της περιόδου t δε συσχετίζεται με το διαταρακτικό όρο μίας οποιασδήποτε άλλης περιόδου s. Αν αυτή η υπόθεση δεν ικανοποιείται, τότε έχουμε το φαινόμενο της αυτοσυσχέτισης. Η αυτοσυσχέτιση είναι συνηθισμένο φαινόμενο όταν χρησιμοποιούνται στοιχεία χρονολογικών σειρών, χωρίς αυτό να σημαίνει ότι δεν υπάρχει αυτοσυσχέτιση όταν χρησιμοποιούνται διαστρωματικά στοιχεία.

(iv) **Η μη – κανονικότητα των τιμών του τυχαίου σφάλματος**
Τέλος, ένα άλλο πρόβλημα που μπορεί να αντιμετωπίσουμε κατά την ανάλυση παλινδρόμησης είναι η μη – κανονικότητα. Μια από τις πλέον βασικές υποθέσεις τη ανάλυσης παλινδρόμησης είναι η κανονικότητα, δηλαδή οι όροι σφάλματος, ανεξάρτητοι μεταξύ τους, ακολουθούν την κανονική κατανομή, δηλαδή $\varepsilon_t \sim N(0, \sigma^2)$. Η υπόθεση αυτή παίζει καθοριστικό ρόλο στην ανάλυση μας καθώς οποιαδήποτε μορφή στατιστικής αναφοράς στις παραμέτρους του υποδείγματος εξαρτάται από την ισχύ της. Συνεπώς, εάν παραβιάζεται η υπόθεση αυτή, τα στατιστικά αποτελέσματα δεν έχουν απολύτως καμία έννοια, εφόσον οι δειγματικές κατανομές των εκτιμητών των συντελεστών του υποδείγματος δεν ακολουθούν την κανονική κατανομή.

Η απουσία των ανωτέρων μειώνει την αξιοπιστία, την ακρίβεια και την εγκυρότητα των αποτελεσμάτων της παλινδρόμησης και κατά συνέπεια την ακρίβεια των προβλέψεων.

Κεφάλαιο 2

ΜΕΘΟΔΟΙ ΑΝΑΛΥΣΗΣ ΧΡΟΝΟΣΕΙΡΩΝ

2.1 Η ΕΝΝΟΙΑ ΤΗΣ ΧΡΟΝΟΣΕΙΡΑΣ ΚΑΙ ΒΑΣΙΚΑ ΠΡΟΒΛΗΜΑΤΑ ΣΤΗΝ ΑΝΑΛΥΣΗ ΧΡΟΝΟΣΕΙΡΩΝ

Η χρονοσειρά μπορεί να ορισθεί ως μια συλλογή διαδοχικών χρονικών παρατηρήσεων της τιμής κάποιου μεγέθους. Πρόκειται ουσιαστικά για μια στοχαστική διαδικασία, μιας και η εξέλιξη των τιμών του μεγέθους επηρεάζεται από τυχαίους παράγοντες, ενώ η τιμή κάθε χρονικής στιγμής συνιστά και μια ξεχωριστή τυχαία μεταβλητή. Με τον όρο χρονοσειρά δηλαδή, εννοούμε συνήθως μια ακολουθία $\{x_t : t = 0, 1, 2, \ldots\}$, όπου κάθε x_t εκφράζει την κατά την χρονική στιγμή t κατάσταση ενός συστήματος το οποίο εξελίσσεται στο χρόνο κατά τυχαίο εν γένει τρόπο (stochastic system).

Παραδείγματα τέτοιων χρονοσειρών είναι:

(i) Οι ημερήσιες, αεροπορικές και οδικές, αφίξεις τουριστών στην χώρα μας x_t με $t = 1, 2, \ldots$

(ii) Ο αριθμός x_t πελατών μέσα σε ένα πολυκατάστημα κατά τη χρονική στιγμή t με $t \in [0, T]$.

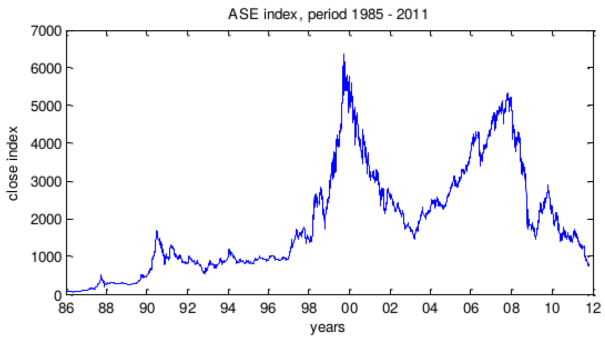
(iii) Ο συνολικός αριθμός τροχαίων ατυχημάτων x_t κατά μήκος μιας οδικής αρτηρίας στο χρονικό διάστημα [0,t] με $t \geq 0$.

(iv) Η ημερήσια κατανάλωση ηλεκτρικού ρεύματος καθώς και η ημερήσια κατανάλωση ύδατος, x_t και y_t αντίστοιχα, σε μια μεγάλη γεωγραφική περιοχή

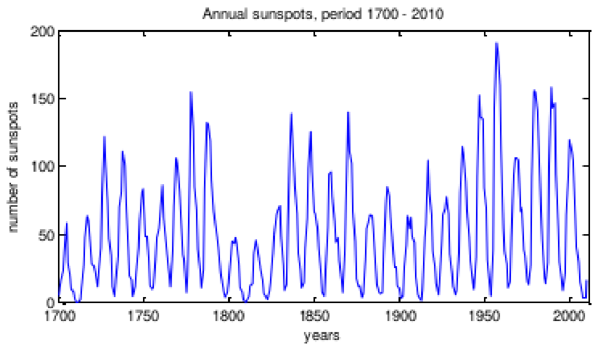
(v) Οι οικονομικές χρονοσειρές, όπως το ετήσιο ακαθάριστο εθνικό προϊόν και ετήσιο ισοζύγιο εξωτερικών συναλλαγών x_t και y_t αντίστοιχα, με $t = 1, 2, \ldots$

(vi) Οι μετεωρολογικές χρονοσειρές, όπως η θερμοκρασία περιβάλλοντος και ατμοσφαιρική πίεση, $x_t y_t$ αντίστοιχα, σε συγκεκριμένη γεωγραφική περιοχή κατά την χρονική στιγμή t.

Επίσης, πολύ γνωστές χρονοσειρές αποτελούν η χρονοσειρά του δείκτη του Χρηματιστηρίου Αξιών Αθηνών (ΧΑΑ), καθώς και η χρονοσειρά των ετήσιων ηλιακών κηλίδων, όπως φαίνονται στα ακόλουθα σχήματα.



Σχήμα 2.1: Η χρονοσειρά του δείκτη του Χρηματιστηρίου Αξιών Αθηνών (ΧΑΑ) την περίοδο 1/1/1986-31/10/2011.



Σχήμα 2.2: Η χρονοσειρά των ετήσιων ηλιακών κηλίδων για την περίοδο 1700-2010.

Για την ανάλυση της χρονοσειράς εντοπίζουμε τα μοτίβα των διαθέσιμων δεδομένων (δηλ. της χρονοσειράς) προκειμένου να καταλάβουμε τον τρόπο με τον οποίο συμπεριφέρονται. Ακριβώς σε αυτή τη συμπεριφορά βασίζεται η πρόβλεψη. Αυτά τα μοτίβα μπορεί να είναι:

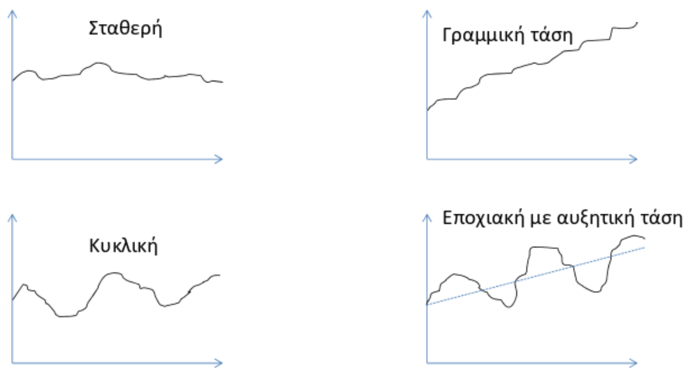
(i) **Το επίπεδο** (level). Ένα τέτοιο μοτίβο υπάρχει όταν τα δεδομένα κυμαίνονται γύρω από μία μέση τιμή.

(ii) **Η τάση** (trend). Γενικά θα μπορούσε να ορισθεί ως η μακροπρόθεσμη μεταβολή του μέσου επιπέδου των τιμών της χρονοσειράς. Έτσι, μπορεί η τάση των τιμών να είναι αυξητική, πτωτική ή σταθερή σ' ένα συγκεκριμένο χρονικό διάστημα.

(iii) **Η εποχικότητα** (seasonality): Είναι η επαναλαμβανόμενη κίνηση των δεδομένων μέσα σε σχετικά μικρό χρονικό διάστημα (συνήθως μέσα σε ένα χρόνο). Τέτοιο μοτίβο μπορεί να παρουσιάζουν οι μεταβλητές που επηρεάζονται από εποχικούς παράγοντες: π.χ. οι πωλήσεις των παγωτών αυξάνονται κατά τους ζεστούς μήνες και μειώνονται κατά τους κρύους.

(iv) **Η κυκλικότητα** (cycle). Είναι (όπως και η εποχικότητα) η επαναλαμβανόμενη αυξομείωση των δεδομένων, με μεγαλύτερο όμως χρονικό ορίζοντα (5-10 χρόνια), και είναι συνήθως συνδεδεμένη με τις διακυμάνσεις στο επίπεδο της συνολικής οικονομίας (γνωστές ως περίοδοι οικονομικής ύφεσης και οικονομικής ανάπτυξης).

(v) **Η τυχαιότητα** (randomness). Είναι η κίνηση των δεδομένων που δεν παρουσιάζει καμία κανονικότητα.



Σχήμα 2.3: Μοτίβα χρονοσειρών

Ο σκοπός της ανάλυσης χρονοσειρών μπορεί να διαφέρει (π.χ. κατανόηση του συστήματος, εκτίμηση διακριτικών χαρακτηριστικών του συστήματος, πρόβλεψη), τα προβλήματα της ανάλυσης όμως είναι λίγο πολύ τα ίδια σε πραγματικές εφαρμογές και έχουν να κάνουν με τη διαθέσιμη πληροφορία. Τα κυριότερα προβλήματα σημειώνονται παρακάτω:

(i) **Μονο-μεταβλητή χρονοσειρά**
Σε πολλά πραγματικά προβλήματα έχουμε μετρήσεις ενός μόνο μεγέθους από το υπό μελέτη σύστημα, όπως για παράδειγμα στο πείραμα πλαστικής παραμόρφωσης, όπου το μέγεθος που μετράμε είναι η ολική τάση.

(ii) **Μια μόνο χρονοσειρά**
Σε χρονοσειρές από πειράματα, έχουμε τη δυνατότητα να επαναλάβουμε το πείραμα και να πάρουμε μετρήσεις κάτω από τις ίδιες συνθήκες, δηλαδή να παρατηρήσουμε περισσότερες από μια πραγματοποιήσεις του ίδιου συστήματος. Ένα τέτοιο παράδειγμα είναι η χρονοσειρά ολικής τάσης από το πείραμα πλαστικής παραμόρφωσης. Σε χρονοσειρές όμως από φυσικές διαδικασίες, δεν υπάρχει αυτή η δυνατότητα καθώς δε μπορούμε να έχουμε επαναλήψεις των μετρήσεων κάτω από τις ίδιες συνθήκες.

(iii) **Περιορισμένο μήκος χρονοσειράς**
Η αναζήτηση μεγάλων δειγμάτων είναι γνωστό πρόβλημα στη στατιστική. Συχνά οι πραγματικές συνθήκες δεν το επιτρέπουν. Δε μπορεί για παράδειγμα να ζητούμε μεγαλύτερη καταγραφή σε επιληπτική κρίση, αφού η διάρκεια της είναι δεδομένη (και ευτυχώς σχετικά σύντομη), όπως δε μπορούμε να έχουμε μετρήσεις του ημερήσιου δείκτη κλεισίματος ΧΑΑ πριν την έναρξη του χρηματιστηρίου ή μετά τη σημερινή ημέρα.

(iv) **Ύπαρξη θορύβου**
Στα πραγματικά δεδομένα πάντα υπάρχει θόρυβος. Αυτός μπορεί να είναι δυναμικός ή θόρυβος συστήματος, έχει να κάνει δηλαδή με την εξέλιξη του συστήματος και περιλαμβάνει όλες εκείνες τις εξωτερικές επιδράσεις στο υπό μελέτη σύστημα που δε μπορούμε να εξηγήσουμε. Για παράδειγμα ο δείκτης ΧΑΑ επηρεάζεται από άλλους χρηματο-οικονομικούς (και όχι μόνο) παράγοντες που μπορούμε ενδεχομένως να συμπεριλάβουμε στη μελέτη μας (πολύ-μεταβλητές χρονοσειρές), αλλά πάντα θα υπάρχουν και κάποιοι άλλοι παράγοντες και τυχαία γεγονότα που επηρεάζουν την εξέλιξη του δείκτη ΧΑΑ και δε μπορούμε να συμπεριλάβουμε, τα οποία αποτελούν το δυναμικό θόρυβο.

(v) **Έλλειψη στασιμότητας**
Αυτό το πρόβλημα σχετίζεται με την ύπαρξη εξωτερικών επιδράσεων ως

προς το σύστημα που μελετάμε, που ενδεχομένως προσθέτουν χαρακτηριστικά ξένα προς το σύστημα. Πολλές φορές το αποτέλεσμα είναι η χρονοσειρά να παρουσιάζει μη-στασιμότητα, δηλαδή αργές μεταβολές (τάσεις) ή/και περιοδικότητα, που δεν σχετίζονται με το μηχανισμό που θέλουμε να μελετήσουμε. Για παράδειγμα, αν θέλουμε να διερευνήσουμε το μηχανισμό που καθορίζει τις μεταβολές του ημερήσιου δείκτη από μέρα σε μέρα, δε μας ενδιαφέρει η στάθμη του δείκτη, αν δηλαδή αναφερόμαστε στην περίοδο της λεγόμενης «φούσκας» του 2000 ή της ύφεσης του 2011. Θα θέλαμε λοιπόν πρώτα να απαλείψουμε τις αργές μεταβολές του δείκτη που έγιναν σε χρονικό ορίζοντα μηνών ή ετών.

2.2 ΚΡΙΤΗΡΙΑ ΑΞΙΟΛΟΓΗΣΗΣ ΤΩΝ ΜΕΘΟΔΩΝ ΠΡΟΒΛΕΨΗΣ

Η ανάλυση χρονοσειρών (time series analysis) ασχολείται αποκλειστικά με τη διερεύνηση της διαχρονικής συμπεριφοράς των τιμών μιας μεταβλητής, οι παρατηρήσεις της οποίας προέρχονται από χρονοσειρά. Η πρόβλεψη των μελλοντικών τιμών της μεταβλητής σύμφωνα με την ανάλυση χρονοσειρών μπορεί να προέλθει απο διάφορες κατηγορίες μεθόδων προβλέψης, όπως η μέθοδος Εξομάλυνσης και η Διάσπαση των χρονοσειρών, με τις οποίες θα ασχοληθούμε στην παρούσα εργασία.

Για την επιλογή της κατάλληλης μεθόδου χρησιμοποιούνται τα κριτήρια αξιολόγησης των μεθόδων προβλέψεων. Τα κριτήρια αυτά βασίζονται στις τιμές των αποκλίσεων των προβλεπόμενων τιμών από τις αντίστοιχες πραγματικές τιμές της χρονοσειράς. Πιο συγκεκριμένα, για μία μεταβλητή Y, η απόκλιση της προβλεπόμενης τιμής της $\widehat{Y}_t$ από την αντίστοιχη πραγματική τιμή της Y_t για την περίοδο t, όπου $t = 1, 2, 3, \ldots, n,$ ονομάζεται σφάλμα της πρόβλεψης (forecast error), συμβολίζεται με e_t και ορίζεται ως:

$$e_t = Y_t - \widehat{Y}_t \tag{2.1}$$

Η σχέση (2.1) εκφράζει για κάθε περίοδο t τη διαφορά μεταξύ της πραγματικής τιμής Y_t και της αντίστοιχης προβλεπόμενης τιμής $\widehat{Y}_t$ που προήλθε από τη μέθοδο πρόβλεψης που χρησιμοποιήθηκε.

Επομένως, για να προσδιορίσουμε την αξιοπιστία μιας συγκεκριμένης μεθόδου πρόβλεψης, θα πρέπει να μελετήσουμε τη διαχρονική συμπεριφορά των τιμών των σφαλμάτων της πρόβλεψης. Αυτό γίνεται με την εφαρμογή διάφορων κριτηρίων, σύμφωνα με τα οποία αξιολογούμε τη χρησιμοποιούμενη μέθοδο πρόβλεψης. Κάθε ένα από τα κριτήρια αυτά ορίζεται από μία συγκεκριμένη συναρτησιακή σχέση των σφαλμάτων της πρόβλεψης και μπορεί να χρησιμοποιηθεί όχι μόνο για την αξιολόγηση μιας μεθόδου πρόβλεψης αλλά και για την επιλογή της “ καλύτερης ” μεταξύ δύο ή περισσοτέρων εναλλακτικών μεθόδων προβλέψεων. Τα κριτήρια αυτά είναι:

- **Μέση απόλυτη απόκλιση MAD** (Mean Absolute Deviation)
  Η μέση απόλυτη απόκλιση ορίζεται ως το άθροισμα των απόλυτων τιμών του σφάλματος της πρόβλεψης διαιρούμενο με τον αριθμό των περιόδων n, στις οποίες έγιναν προβλέψεις, δηλαδή:

  $$\mathsf{MAD} = \frac{1}{n}\sum_{t=1}^{n} |Y_t - \widehat{Y}_t| = \frac{1}{n}\sum_{t=1}^{n} |e_t|$$

  Το MAD εκφράζει τη μέση τιμή των απολύτων αποκλίσεων των προβλεπόμενων τιμών της χρονοσειράς από τις αντίστοιχες πραγματικές και έχει τα ακόλουθα χαρακτηριστικά. Πρώτον, η μονάδα μέτρησης του είναι η ίδια με εκείνη των τιμών της χρονοσειράς και έτσι είναι εύκολη η ερμηνεία του. Δεύτερον, στον υπολογισμό του λαμβάνονται υπ’ όψιν μόνο οι απόλυτες τιμές των σφαλμάτων και όχι οι πραγματικές τιμές τους. Αυτό σημαίνει ότι το MAD είναι ανεξάρτητο από θετικές ή αρνητικές τιμές του σφάλματος, δηλαδή είναι ανεξάρτητο από το αν οι τιμές των προβλέψεων είναι μικρότερες ή μεγαλύτερες των πραγματικών τιμών.Και τέλος, το MAD βασίζεται στην υπόθεση ότι η αξιοπιστία του σφάλματος ή το κόστος που δημιουργείται από το σφάλμα της πρόβλεψης, σχετίζεται γραμμικά με το μέγεθος του σφάλματος.

- **Μέσο σφάλμα τετραγώνου MSE** (Mean Squared Error)
  Το μέσο σφάλμα τετραγώνου ορίζεται ως το άθροισμα των τετραγώνων των σφαλμάτων διαιρούμενο με τον αριθμό των χρονικών περιόδων n, στις οποίες έγιναν οι προβλέψεις, δηλαδή:

  $$\mathsf{MSE} = \frac{1}{n}\sum_{t=1}^{n} \left(Y_t - \widehat{Y}_t\right)^2 = \frac{1}{n}\sum_{t=1}^{n} e_t^2$$

  Το MSE είναι η μέση τιμή των τετραγώνων των αποκλίσεων των προβλεπόμενων τιμών της χρονοσειράς από τις αντίστοιχες πραγματικές. Η μονάδα μέτρησης του MSE είναι εκφρασμένη στη μονάδα μέτρησης των τιμών των παρατηρήσεων υψωμένη στο τετράγωνο.
  Η ύπαρξη προβλέψεων που απέχουν πολύ από τις αντίστοιχες πραγματικές τιμές γίνεται πολύ περισσότερο αισθητή με το κριτήριο MSE από ότι με το κριτήριο MAD, επειδή οι τιμές των σφαλμάτων της πρόβλεψης υψώνονται

στο τετράγωνο. Συνεπώς το κριτήριο MSE είναι στατιστικά περισσότερο αξιόπιστο από το κριτήριο MAD και χρησιμοποιείται συχνότερα για την επιλογή της 'κατάλληλης' μεθόδου πρόβλεψης.

- **Ρίζα μέσου σφάλματος τετραγώνου RMSE** (Root Mean Squared Error)
  Η τετραγωνική ρίζα μέσου σφάλματος τετραγώνου είναι η θετική τιμή της τετραγωνικής του ρίζας, δηλαδή είναι:

$$\mathrm{RMSE} = \sqrt{\mathrm{MSE}} = \sqrt{\frac{1}{n}\sum_{t=1}^{n} e_t^2}$$

  Το RMSE εκφράζεται στην ίδια μονάδα μέτρησης με εκείνη των τιμών της χρονοσειράς.

- **Μέσο απόλυτο ποσοστιαίο σφάλμα MAPE** (Mean Absolute Percentage Error)
  Το μέσο απόλυτο ποσοστιαίο σφάλμα εξετάζει τη συμπεριφορά της απόλυτης τιμής του σφάλματος της πρόβλεψης σε σχέση με την πραγματική τιμή της χρονοσειράς. Το MAPE ορίζεται ως το άθροισμα των απόλυτων τιμών των σφαλμάτων της πρόβλεψης προς τις αντίστοιχες πραγματικές τιμές της χρονοσειράς διαιρούμενο με τον αριθμό των χρονικών περιόδων n, στις οποίες έγιναν προβλέψεις, δηλαδή:

$$\mathrm{MAPE} = \frac{1}{n}\sum_{t=1}^{n} \frac{|Y_t - \widehat{Y}_t|}{Y_t} = \frac{1}{n}\sum_{t=1}^{n} \frac{|e_t|}{Y_t}$$

  Το κριτήριο αυτό είναι απαλλαγμένο από μονάδες μέτρησης και το χρησιμοποιούμε για να συγκρίνουμε την ακρίβεια μιας ή περισσοτέρων μεθόδων προβλέψεων και για περισσότερες από μια χρονοσειρές.

- **Μέσο ποσοστιαίο σφάλμα MPE** (Mean Percentage Error)
  Το μέσο ποσοστιαίο σφάλμα το χρησιμοποιούμε όταν ενδιαφερόμαστε να προσδιορίσουμε αν η μέθοδος πρόβλεψης είναι μεροληπτική, δηλαδή αν οι προβλεπόμενες τιμές είναι συστηματικά μεγαλύτερες ή μικρότερες από τις αντίστοιχες πραγματικές.

$$\mathrm{MPE} = \frac{1}{n}\sum_{t=1}^{n} \frac{Y_t - \widehat{Y}_t}{Y_t} = \frac{1}{n}\sum_{t=1}^{n} \frac{e_t}{Y_t}$$

  Χωρίς αμφιβολία, όσο πιο κοντά στο μηδέν είναι η τιμή του MPE, τόσο πιο αμερόληπτη και καλή είναι η μέθοδος πρόβλεψης που χρησιμοποιήθηκε.

Αντιθέτως, μεγάλες απόλυτες τιμές του MPE φανερώνουν μεγάλη μεροληψία της μεθόδου.

2.3 ΜΕΘΟΔΟΙ ΕΞΟΜΑΛΥΝΣΗΣ

Οι μέθοδοι εξομάλυνσης (smoothing methods) είναι τεχνικές με τις οποίες προσδιορίζονται οι μελλοντικές τιμές μιας μεταβλητής με βάση τον τρόπο εφαρμογής τους. Οι τεχνικές αυτές ονομάζονται μέθοδοι εξομάλυνσης, διότι η δημιουργία των προβλέψεων προέρχεται από την εξομάλυνση της διαχρονικής εξέλιξης των τιμών της μεταβλητής, ώστε να αναγνωριστεί καλύτερα ο τρόπος συμπεριφοράς της. Ορισμένες από τις μεθόδους εξομάλυνσης μπορούν να εφαρμοστούν και σε περιπτώσεις μικρού αριθμού παρατηρήσεων της μεταβλητής. Οι μέθοδοι εξομάλυνσης που θα περιγράψουμε είναι:

- Η μέθοδος του απλού κινητού μέσου
- Η μέθοδος της απλής εκθετικής εξομάλυνσης
- Η μέθοδος του διπλού κινητού μέσου
- Η μέθοδος Brown
- Η μέθοδος Holt
- Η μέθοδος Winters

Εάν μία χρονοσειρά είναι στάσιμη η κατάλληλη μέθοδος πρόβλεψης μελλοντικών τιμών είναι η μέθοδος των κινητών μέσων όρων. Σε μερικές χρονοσειρές όμως οι πρόσφατες παρατηρήσεις μπορεί να περιέχουν περισσότερες πληροφορίες από τις παλαιότερες και αυτό είναι πολύ σημαντικό για τις μελλοντικές προβλέψεις. Σε αυτήν την περίπτωση χρησιμοποιούμε την απλή εκθετική εξομάλυνση. Εάν η χρονοσειρά εμφανίζει κάποιο πρότυπο τάσης τότε χρησιμοποιούμε την μέθοδο διπλού κινητού μέσου, την μέθοδο Brown ή την μέθοδο Holt ενώ εάν η χρονοσειρά εμφανίζει εποχικότητα τότε χρησιμοποιούμε την μέθοδο Winters.

2.3.1 ΑΠΛΟΣ ΚΙΝΗΤΟΣ ΜΕΣΟΣ

Η μέθοδος του απλού κινητού μέσου m-περιόδων (simple moving average) είναι μία πολύ απλή μέθοδος προβλέψεων που χρησιμοποιεί ως πρόβλεψη την τιμή του αριθμητικού μέσου όρου των m πλέον πρόσφατων παρατηρήσεων της χρονοσειράς. Αυτό συμβαίνει διότι οι πλέον πρόσφατες παρατηρήσεις της χρονοσειράς θεωρούνται περισσότερο αντιπροσωπευτικές για τη δημιουργία προβλέψεων από ότι οι πιο απομακρυσμένες στο παρελθόν. Ο μέσος όρος αυτός ονομάζεται κινητός, επειδή η τιμή του δεν είναι σταθερή, αλλά αναπροσαρμόζεται κάθε φορά που μια νέα παρατήρηση της χρονοσειράς γίνεται διαθέσιμη.

Οι προβλέψεις μιας χρονοσειράς Y_t , για $t = 1, 2, \ldots, n$, δημιουργούνται με τη μέθοδο του απλού κινητού μέσου ως εξής:

$$\widehat{Y}_{t+1} = \frac{1}{m}\sum_{j=1}^{m} Y_{t-j+1} = \frac{1}{m}\left(Y_t + Y_{t-1} + \ldots + Y_{t-m+1}\right)$$

όπου $\widehat{Y}_{t+1}$ είναι η πρόβλεψη για την περίοδο $(t + 1)$
και m είναι ο αριθμός των περιόδων που χρησιμοποιούνται για τον υπολογισμό της τιμής του μέσου όρου. Επίσης, για m=1 η πρόβλεψη της επόμενης περιόδου είναι ίση με την πραγματική τιμή της προηγούμενης περιόδου, δηλαδή ισχύει η σχέση: $\widehat{Y}_{t+1} = Y_t$

Απαραίτητη προϋπόθεση για την εφαρμογή του ανωτέρω τύπου είναι ο προσδιορισμός του m. Η τιμή του m δύναται να είναι γνωστή, οπότε χωρίς καμία δυσκολία μπορούμε να δημιουργήσουμε την πρόβλεψη της περιόδου (t+1), δύναται όμως να είναι και άγνωστη. Στην περίπτωση που η τιμή του m δεν είναι προκαθορισμένη (η περίπτωση αυτή αποτελεί το πλέον σύνηθες φαινόμενο) τότε ο ερευνητής καλείται με βάση την κρίση του, την εμπειρία του και τα δεδομένα των παρατηρήσεων της χρονοσειράς να προσδιορίσει την τιμή του m. Συνήθως, για να προσδιοριστεί η τιμή του m για μια συγκεκριμένη χρονοσειρά υπολογίζεται ο απλός κινητός μέσος για διάφορες τιμές του m και επιλέγεται εκείνη η τιμή που ελαχιστοποιεί την τιμή ενός κριτηρίου, από εκείνα που υπάρχουν, για την αξιολόγηση των μεθόδων προβλέψεων (συνήθως του MSE ή του RMSE).

Με τη μέθοδο του απλού κινητού μέσου μπορούμε να εξομαλύνουμε τις τιμές της χρονοσειράς. Αναλυτικότερα, εφαρμόζοντας τον παραπάνω μαθηματικό τύπο στις παρατηρήσεις της χρονοσειράς, εξομαλύνουμε τις τελευταίες (n – m) παρατηρήσεις της χρονοσειράς. Οι τιμές που προκύπτουν χρησιμοποιούνται ως προβλέψεις των αντίστοιχων περιόδων και εν συνεχεία, βάσει αυτών, προσδιορίζουμε τις τιμές των σφαλμάτων της πρόβλεψης.

2.3.2 ΑΠΛΗ ΕΚΘΕΤΙΚΗ ΕΞΟΜΑΛΥΝΣΗ

Ένα μειονέκτημα της μεθόδου του απλού κινητού μέσου m-περιόδων είναι ότι για τον υπολογισμό των προβλέψεων δίνει ίση βαρύτητα σε κάθε παρατήρηση, ανεξάρτητα από το πόσο κοντά ή μακριά βρίσκεται σε σχέση με την προβλεπό-

μενη περίοδο. Το μειονέκτημα αυτό μπορεί να εξαλειφθεί με τη μέθοδο της απλής εκθετικής εξομάλυνσης (simple exponential smoothing), σύμφωνα με την οποία οι προβλέψεις δημιουργούνται με βάση κάποιο σταθμικό μέσο όρο, έτσι ώστε να δίνεται διαφορετική βαρύτητα σε κάθε παρατήρηση. Πιο συγκεκριμένα, με τη μέθοδο αυτή δίνεται πολύ μεγαλύτερη βαρύτητα στις πιο πρόσφατες παρατηρήσεις, από αυτή που δίνεται στις πιο απομακρύσμενες.
Για να κατανοήσουμε το μηχανισμό λειτουργίας της μεθόδου ας θεωρήσουμε ότι οι προβλέψεις της χρονοσειράς δημιουργούνται ως εξής:

$$\widehat{Y}_{t+1} = aY_t + a\,(1{-}a)\,Y_{t-1} + a\,(1{-}a)^2\,Y_{t-2} + \ldots \tag{2.2}$$

που η παράμετρος α ονομάζεται σταθερά εξομάλυνσης (smoothing constant) και λαμβάνει τιμές μεταξύ 0 και 1 δηλαδή: $0 \leqslant a \leqslant 1$
Έτσι, σύμφωνα με την παραπάνω σχέση η πρόβλεψη $\widehat{Y}_{t+1}$ προκύπτει ως ένας σταθμικός μέσος όρος των παρατηρήσεων της χρονοσειράς, αφού το άθροισμα των συντελεστών της σχέσης (2.2) είναι ίσο με τη μονάδα. Όσο πιο μεγάλη είναι η τιμή της παραμέτρου α, τόσο μεγαλύτερη βαρύτητα δίνεται στις πιο πρόσφατες παρατηρήσεις και πολύ μικρή εως μηδαμινή βαρύτητα στις πιο απομακρυσμένες.
Η παραπάνω σχέση μπορεί να γραφεί και με τη μορφή:

$$\widehat{Y}_{t+1} = aY_t + (1{-}a)\,\widehat{Y}_t \tag{2.3}$$

Η σχέση αυτή είναι η μαθηματική έκφραση της μεθόδου της απλής εκθετικής εξομάλυνσης και ορίζεται για $t = 2, 3, \ldots, n$ με αρχική συνθήκη $\widehat{Y}_2 = Y_1$.

Είναι προφανές ότι η πρόβλεψη $\widehat{Y}_{t+1}$ της περιόδου $(t+1)$ είναι ένας σταθμικός μέσος όρος της πραγματικής τιμής Y_t και της προβλεπόμενης τιμής $\widehat{Y}_t$ της περιόδου t με συντελεστές βαύτητας a και $(1-a)$ αντίστοιχα. Πρακτικά, αυτό σημαίνει ότι εάν η τιμή της σταθεράς εξομάλυνσης είναι $a = 0,3$, τότε η πρόβλεψη της επόμενης περιόδου προσδιορίζεται κατά 30% απ' την πραγματική τιμή και κατά 70% απ' την προβλεπόμενη τιμή της τρέχουσας περιόδου. Να σημειωθεί πως όταν η τιμή της σταθεράς εξομάλυνσης λαμβάνει τις ακραίες τιμές ένα και μηδέν τότε:

- για $a = 1$ η πρόβλεψη της επόμενης περιόδου $(t+1)$ είναι η πραγματική τιμή της τρέχουσας περιόδου t
- για $a = 0$ η πρόβλεψη της επόμενης περιόδου $(t+1)$ είναι ίση με την πρόβλεψη της τρέχουσας περιόδου t.

Παρατηρούμε ότι όσο πιο μικρές είναι οι τιμές της σταθεράς α τόσο περισσότερο εξομαλύνονται οι παρατηρήσεις της χρονοσειράς και όσο πιο μεγάλες είναι οι τιμές της σταθεράς α τόσο πιο γρήγορα αντιδρά η εν λόγω μέθοδος στις πραγματικές μεταβολές των παρατηρήσεων της χρονοσειράς. Συνεπώς τίθεται ένα ερώτημα : Μεγαλύτερη εξομάλυνση των παρατηρήσεων της χρονοσειράς και άρα μικρή τιμή για το a ή μεγαλύτερη ανταπόκριση στις πραγματικές μεταβολές των παρατηρήσεων και άρα μεγάλη τιμή για το a ; Η απάντηση στο ερώτημα αυτό δεν είναι καθόλου εύκολη. Θα λέγαμε ότι σπουδαίο ρόλο διαδραματίζει η κρίση του ερευνητή και η τυχόν προηγούμενη εμπειρία που έχει για τη συγκεκριμένη χρονολογική σειρά. Βέβαια, το πιο σωστό είναι η «άριστη» τιμή για το a να προσδιορίζεται από τα δεδομένα της χρονοσειράς. Συγκεκριμένα από τις τιμές του a, $0 \leqslant a \leqslant 1$, επιλέγουμε εκείνη που ελαχιστοποιεί την τιμή του κριτηρίου MSE ή κάποιου άλλου κριτηρίου.

Η μαθηματική σχέση που εκφράζει τη μέθοδο της απλής εκθετικής εξομάλυνσης μπορεί να λάβει και την ακόλουθη μορφή :

$$\widehat{Y}_{t+1} = \widehat{Y}_t + a\left(Y_t - \widehat{Y}_t\right) = \widehat{Y}_t + a e_t \tag{2.4}$$

Η σχέση (2.4) μας δείχνει ότι η πρόβλεψη της περιόδου (t+1) είναι ίση με την πρόβλεψη της περιόδου t συν το σφάλμα της πρόβλεψης e_t , πολλαπλασιασμένο με την τιμή της σταθεράς εξομάλυνσης a. Κατά συνέπεια όσο μεγαλύτερη είναι η τιμή του a, τόσο πιο μεγάλη βαρύτητα δίνεται στο σφάλμα της πρόβλεψης.

2.3.3 ΔΙΠΛΟΣ ΚΙΝΗΤΟΣ ΜΕΣΟΣ

Όταν οι παρατηρήσεις της χρονοσειράς που εξετάζουμε παρουσιάζουν μια ανοδική ή πτωτική εξελικτική πορεία που εκφράζεται από κάποια γραμμική τάση, τότε μπορούμε να χρησιμοποιήσουμε τη μέθοδο του διπλού κινητού μέσου (double moving average) για την πρόβλεψη των τιμών της χρονοσειράς. Για τη διαμόρφωση των προβλέψεων με τη μέθοδο αυτή υπολογίζεται ένας δεύτερος κινητός μέσος από τον απλό κινητό μέσο, ενώ στη συνέχεια λαμβάνεται υπ΄όψιν και η γραμμική τάση των παρατηρήσεων της χρονοσειράς . Για το λόγο αυτό η μέθοδος ονομάζεται πολύ συχνά και μέθοδος του γραμμικού κινητού μέσου (linear moving average).

Η εφαρμογή της μεθόδου του διπλού κινητού μέσου βασίζεται στην ακόλουθη διαδικασία:

(i) Αρχικά, υπολογίζουμε τον απλό κινητό μέσο m-περιόδων, M_t , ως εξής:

$$M_{t+1} = \frac{1}{m}\sum_{j=1}^{m} Y_{t-j+1}$$

(ii) Στη συνέχεια, υπολογίζουμε τον διπλό κινητό μέσο m-περιόδων, M'_t,:

$$M'_{t+1} = \frac{1}{m}\sum_{j=1}^{m} M_{t-j+1}$$

(iii) Ύστερα, βρίσκουμε τη διαφορά a_t:

$$a_t = 2M_t - M'_t$$

(iv) Βρίσκουμε τον παράγοντα προσαρμογής για την τάση, b_t :

$$b_t = \frac{2}{m-1}\left(M_t - M'_t\right)$$

(v) Και τέλος, υπολογίζουμε την πρόβλεψη $\widehat{Y}_{t+h}$ για την h μελλοντική περίοδο ως εξής:

$$\widehat{Y}_{t+h} = a_t + hb_t$$

όπου h είναι ένας ακέραιος θετικός αριθμός.

Η μέθοδος αυτή, σε αντίθεση με τη μέθοδο του απλού κινητού μέσου όρου που παρουσιάσαμε, μπορεί για $h > 1$ να χρησιμοποιηθεί για τη διαμόρφωση προβλέψεων για περισσότερες από μία μελλοντικές περιόδους, ενώ για $h = 1$ δίνει την πρόβλεψη για την επόμενη περίοδο. Βέβαια, η χρήση της προϋποθέτει την ύπαρξη μεγαλύτερου αριθμού παρατηρήσεων, ιδιαίτερα μάλιστα όταν η τιμή του m είναι σχετικά μεγάλη. Όπως και στη μέθοδο του απλού κινητού μέσου, όταν η τιμή του m δεν είναι γνωστή, επιλέγουμε εκείνη την τιμή που ελαχιστοποιεί την τιμή του κριτηρίου MSE ή κάποιου άλλου κριτήριου στα δεδομένα της χρονοσειράς, εφαρμόζοντας τη μέθοδο για διάφορες τιμές του m.

2.3.4 ΔΙΠΛΗ ΕΚΘΕΤΙΚΗ ΕΞΟΜΑΛΥΝΣΗ: ΜΕΘΟΔΟΣ BROWN

Η μέθοδος της διπλής εκθετικής εξομάλυνσης (double exponential smoothing), η οποία ονομάζεται και μέθοδος Brown, είναι μια άλλη μέθοδος προβλέψεων που χρησιμοποιείται σε χρονοσειρές, οι παρατηρήσεις των οποίων παρουσιάζουν τάση. Η βασική φιλοσοφία της μεθόδου αυτής είναι παραπλήσια με εκείνη της μεθόδου του διπλού κινητού μέσου, δηλαδή η εξομάλυνση των παρατηρήσεων της χρονοσειράς γίνεται δύο φορές, ενώ στη διαμόρφωση των προβλέψεων λαμβάνεται υπ'όψιν και η τάση.

Η εφαρμογή της μεθόδου της διπλής εκθετικής εξομάλυνσης στηρίζεται στην ακόλουθη διαδικασία:

(i) Εξομαλύνουμε τις αρχικές παρατηρήσεις της χρονοσειράς με τη μέθοδο της απλής εκθετικής εξομάλυνσης ως εξής:

$$A_t = aY_t + (1 - a)\, A_{t-1}$$

όπου α είναι η σταθερά εξομάλυνσης, για $0 \leqslant a \leqslant 1$, A_t οι εξομαλυνθείσες τιμές της χρονοσειράς που προκύπτουν από την πρώτη εξομάλυνση, για $t = 2, 3, \ldots, n$, ενώ για $t = 1$ ορίζεται ως αρχική συνθήκη $A_1 = Y_1$.

(ii) Εξομαλύνουμε τις εξομαλυνθείσες τιμές A_t της χρονοσειράς με τη μέθοδο της απλής εκθετικής εξομάλυνσης ως εξής:

$$A_t^{'} = aA_t + (1 - a)\, A_{t-1}^{'}$$

όπου $A_t^{'}$ είναι οι εξομαλυνθείσες τιμές της χρονοσειράς που προκύπτουν από τη δεύτερη εξομάλυνση, για $t = 2, 3, \ldots, n$ ενώ για $t = 1$, $A_1^{'} = A_1$.

(iii) Υπολογίζουμε τη διαφορά a_t ως:

$$a_t = 2A_t – A_t^{'}$$

(iv) Υπολογίζουμε τον παράγοντα προσαρμογής για την τάση, b_t, ως:

$$b_t = \frac{a}{1 - a}\left(A_t - A_t^{'}\right)$$

(v) Υπολογίζουμε την πρόβλεψη $\widehat{Y}_{t+h}$ για την h μελλοντική περίοδο ως εξής:

$$\widehat{Y}_{t+h} = a_t + hb_t$$

όπου h είναι ένας ακέραιος θετικός αριθμός.

Η μέθοδος αυτή μπορεί να εφαρμοστεί για τη διαμόρφωση προβλέψεων για περισσότερες από μία μελλοντικές περιόδους σε αντίθεση με τη μέθοδο της απλής εκθετικής εξομάλυνσης, η οποία παρέχει προβλέψεις μόνο για την επόμενη χρονική περίοδο. Επίσης, αν η τιμή της σταθεράς εξομάλυνσης α δεν είναι γνωστή, κάτι που συμβαίνει όταν εφαρμόζουμε τη μέθοδο για πρώτη φορά στα δεδομένα μιας χρονοσειράς, επιλέγουμε κατά τα γνωστά εκείνη την τιμή του α που ελαχιστοποιεί την τιμή του κριτηρίου MSE ή κάποιου άλλου κριτηρίου. Σημειώνουμε ότι ο αριθμός των παρατηρήσεων που απαιτούνται για την εφαρμογή της μεθόδου αυτής είναι αρκετά μικρότερος από τον αντίστοιχο αριθμό της μεθόδου του διπλού κινητού μέσου.

2.3.5 ΕΚΘΕΤΙΚΗ ΕΞΟΜΑΛΥΝΣΗ ΜΕ ΠΡΟΣΑΡΜΟΓΗ ΣΤΗΝ ΤΑΣΗ: ΜΕΘΟΔΟΣ HOLT

Η εκθετική εξομάλυνση με προσαρμογή στην τάση (exponential smoothing adjusted for trend), γνωστή και ως μέθοδος Holt, χρησιμοποιείται κι αυτή όταν υπάρχει τάση στις παρατηρήσεις της χρονοσειράς. Η μέθοδος Holt έχει δύο παραμέτρους εξομάλυνσης, την παράμετρο α για την εξομάλυνση των τιμών της χρονοσειράς και την παράμετρο β για την εξομάλυνση της τάσης, σε αντίθεση με τη μέθοδο της διπλής εξομάλυνσης του Brown που έχει μόνο μια.

Η εφαρμογή της μεθόδου Holt βασίζεται στην ακόλουθη διαδικασία:

(i) Εξομαλύνουμε τις τιμές της χρονοσειράς ως εξής:

$$A_t = \alpha Y_t + (1-\alpha)\left(A_{t-1} + T_{t-1}\right)$$

όπου α είναι η σταθερά για την εξομάλυνση των τιμών της χρονοσειράς, για $0 \leqslant \alpha \leqslant 1$, A_t οι εξομαλυνθείσες τιμές της χρονοσειράς, για $t = 2, 3, \dots, n$, ενώ για $t = 1$ ορίζεται ως αρχική συνθήκη $A_1 = Y_1$.

(ii) Εξομαλύνουμε την τάση ως εξής:

$$T_t = \beta\left(A_t - A_{t-1}\right) + (1-\beta)\, T_{t-1}$$

όπου β, για $0 \leqslant \beta \leqslant 1$, είναι η σταθερά για την εξομάλυνση της τάσης, T_t οι εξομαλυνθείσες τιμές της τάσης, για $t = 2, 3, \dots, n$ ενώ για $t = 1$ ορίζεται ως αρχική συνθήκη $T_1 = 0$.

(iii) Υπολογίζουμε την πρόβλεψη $\widehat{Y}_{t+h}$ για την h μελλοντική περίοδο ως εξής:

$$\widehat{Y}_{t+h} = A_t + hT_t$$

όπου h είναι ένας ακέραιος θετικός αριθμός.

Οι βέλτιστες τιμές των παραμέτρων α και β για μια συγκεκριμένη χρονοσειρά προκύπτουν από την ελαχιστοποίηση της τιμής του κριτηρίου MSE ή κάποιου άλλου κριτηρίου, εφαρμόζοντας τη μέθοδο για όλους τους δυνατούς συνδυασμούς των τιμών α και β. Με τη μέθοδο Holt οι τιμές της τάσης εξομαλύνονται απευθείας ενώ με τη μέθοδο Brown η τάση προσδιορίζεται από τον παράγοντα b_t αφού προηγουμένως εξομαλυνθούν δύο φορές οι τιμές της χρονοσειράς. Αυτό σημαίνει ότι με τη μέθοδο Holt γίνεται καλύτερη εκτίμηση των τιμών της τάσης από ότι με τη μέθοδο Brown που είναι πολύ ευαίσθητη στις τυχαίες διακυμάνσεις της χρονοσειράς. Για το λόγο αυτό, η μέθοδος Holt εφαρμόζεται περισσότερο συχνά στην πράξη, αφού έχει αποδειχθεί ότι παρέχει συνήθως καλύτερα αποτελέσματα από τη μέθοδο Brown.

2.3.6 ΕΚΘΕΤΙΚΗ ΕΞΟΜΑΛΥΝΣΗ ΜΕ ΠΡΟΣΑΡΜΟΓΗ ΣΤΗΝ ΤΑΣΗ ΚΑΙ ΣΤΗΝ ΕΠΟΧΙΚΟΤΗΤΑ: ΜΕΘΟΔΟΣ WINTERS

Σε πολλές χρονοσειρές, οι παρατηρήσεις των οποίων αναφέρονται σε χρονικές περιόδους μικρότερες του έτους, όπως για παράδειγμα μήνες, τρίμηνα κ.α., είναι δυνατόν να παρατηρούνται εποχικές διακυμάνσεις, οι οποίες επαναλαμβάνονται κάθε έτος με την ίδια ή περίπου ίδια μορφή. Η εποχικότητα στις παρατηρήσεις των χρονοσειρών είναι ένα φαινόμενο που εμφανίζεται συχνά κατά τη διερεύνηση των οικονομικών φαινομένων και για να εξεταστεί θα πρέπει να χρησιμοποιούνται μέθοδοι εξομάλυνσης, οι οποίες να τη λαμβάνουν άμεσα υπ'όψιν. Οι μέθοδοι αυτές συντελούν συνήθως στο να μειώνουν το σφάλμα της πρόβλεψης, παρέχοντας έτσι καλύτερες προβλέψεις.

Μια τέτοια μέθοδος είναι εκείνη της εκθετικής εξομάλυνσης με προσαρμογή στην τάση και στην εποχικότητα (exponential smoothing adjusted for trend and seasonality), γνωστή ως μέθοδος Winters, που αποτελεί επέκταση της μεθόδου Holt. Η μέθοδος Winters έχει τρεις παραμέτρους, τις α, β, και γ, οι οποίες χρησιμοποιούνται για την εξομάλυνση των τιμών της χρονοσειράς, της τάσης και της εποχικότητας αντίστοιχα. Αυτό σημαίνει ότι χρησιμοποιεί μια επιπλέον σχέση από τη μέθοδο Holt για την εξομάλυνση της εποχικότητας.

Η εφαρμογή της μεθόδου Winters στηρίζεται στην ακόλουθη διαδικασία:

(i) Η εξομάλυνση των τιμών της χρονοσειράς γίνεται με την ακόλουθη σχέση:

$$A_t = \alpha \frac{Y_t}{S_{t-L}} + (1-\alpha)(A_{t-1} + T_{t-1})$$

όπου α είναι η σταθερά για την εξομάλυνση των τιμών της χρονοσειράς, για $0 \leqslant \alpha \leqslant 1$, A_t οι εξομαλυνθείσες τιμές της χρονοσειράς, ενώ S_t είναι ο εποχικός συντελεστής της περιόδου t και L η περιοδικότητα της εποχικότητας, δηλαδή L=12 για μηνιαία δεδομένα, L=4 για τριμηνιαία δεδομένα κ.ο.κ.

(ii) Η εξομάλυνση της τάσης γίνεται όπως και στη μέθοδο Holt, δηλαδή ως εξής:

$$T_t = \beta (A_t - A_{t-1}) + (1-\beta) T_{t-1}$$

όπου β, για $0 \leqslant \beta \leqslant 1$, είναι η σταθερά για την εξομάλυνση της τάσης ενώ T_t οι εξομαλυνθείσες τιμές της τάσης.

(iii) Η εξομάλυνση της εποχικότητας γίνεται ως ακολούθως:

$$S_t = \gamma \frac{Y_t}{A_t} + (1 - \gamma) S_{t-L}$$

όπου γ, για $0 \leqslant \gamma \leqslant 1$, είναι η σταθερά για την εξομάλυνση της εποχικότητας.

(iv) Η πρόβλεψη $\widehat{Y}_{t+h}$ για τις h μελλοντικές περιόδους του πρώτου έτους προσδιορίζεται ως:

$$\widehat{Y}_{t+h} = (A_t + hT_t) S_{t+h-L}$$

όπου $h = 1, 2, \ldots, L$ και για τις h μελλοντικές περιόδους του δεύτερου έτους από τη σχέση:

$$\widehat{Y}_{t+h} = (A_t + hT_t) S_{t+h-2L}$$

για $h = L + 1, L + 2, \ldots, 2L$ κ.ο.κ.

Οι αρχικές συνθήκες των σχέσεων υπολογίζονται με τον ακόλουθο τρόπο:

- Για $t = 1, 2, ..., L - 1$ δεν προσδιορίζονται οι τιμές A_t , ενώ για $t = L$ το A_L ορίζεται ως:

$$A_L = \frac{Y_1 + Y_2 + \ldots + Y_L}{L}$$

- Για $t = 1, 2, \ldots, L-1$ δεν προσδιορίζονται οι τιμές T_t , ενώ για $t = L$, τίθεται $T_L = 0$.
- Για $t = 1, 2, \ldots, L$ οι τιμές των εποχικών συντελεστών S_t υπολογίζονται ως εξής:

$$S_t = \frac{Y_t}{A_L}$$

Έτσι, όταν προσδιοριστούν οι παραπάνω αρχικές τιμές για το πρώτο έτος, η μέθοδος Winters εφαρμόζεται κανονικά σύμφωνα με τη διαδικασία που περιγράψαμε και μπορεί να προβλέψει τις μελλοντικές τιμές της χρονοσειράς για περισσότερες από μια περιόδους. Οι άριστες τιμές των παραμέτρων α,β και γ προκύπτουν από την ελαχιστοποίηση του κριτήριου MSE ή κάποιου άλλου κριτηρίου, εφαρμόζοντας τη μέθοδο αυτή για όλους τους δυνατούς συνδυασμούς των τιμών των παραμέτρων στα δεδομένα της χρονοσειράς.

2.4 ΜΕΘΟΔΟΣ ΔΙΑΣΠΑΣΗΣ ΧΡΟΝΟΣΕΙΡΩΝ

Οι χαρακτηριστικές κινήσεις μίας χρονοσειράς μπορούν να διακριθούν σε τέσσερα κύρια είδη, τα οποία συχνά ονομάζονται συνιστώσες (components) της χρονοσειράς. Οι κινήσεις αυτές είναι οι μακροπρόθεσμες ή κύριες κινήσεις, οι κυκλικές κινήσεις ή μεταβολές, οι εποχικές κινήσεις ή μεταβολές και οι ακανόνιστες ή τυχαίες κινήσεις. Η διάσπαση χρονοσειρών (time series decomposition) στηρίζεται στην υπόθεση ότι οι τιμές μιας χρονοσειράς σχηματίζονται από τις παραπάνω συνιστώσες που τη συνθέτουν. Για τη δημιουργία των προβλέψεων με τη μέθοδο αυτή, η χρονοσειρά διασπάται στις ανωτέρω τέσσερις συνιστώσες και προσδιορίζεται η επιρροή που έχει καθένα από αυτά στη διαμόρφωση των τιμών της μεταβλητής.

Οι μακροπρόθεσμες ή κύριες κινήσεις ή τάση αναφέρονται στη γενική κατεύθυνση που φαίνεται ότι ακολουθεί η γραφική παράσταση μίας χρονοσειράς κατά μία μεγάλη διάρκεια χρόνου. Σε πολλές περιπτώσεις οι τιμές των παρατηρήσεων ορισμένων χρονοσειρών τείνουν να αυξάνονται ή να μειώνονται με αρκετά σταθερό ρυθμό για μεγάλα χρονικά διαστήματα. Η συμπεριφορά αυτή εκφράζεται από την τάση που φανερώνει τη μακροχρόνια εξέλιξη της χρονοσειράς, η οποία μπορεί να είναι ανοδική ή καθοδική. Η τάση οφείλεται συνήθως σε πληθυσμιακές αλλαγές, σε τεχνολογικές αλλαγές, σε οικονομικούς παράγοντες, όπως π.χ. στον πληθωρισμό, στην αύξηση της παραγωγικότητας κ.α.

Οι κυκλικές κινήσεις ή κυκλικότητα αναφέρονται με μακροπρόθεσμες ταλαντώσεις γύρω από τη γραμμή ή καμπύλη τάσης. Η κυκλικότητα εμφανίζεται ακανόνιστα με κυματοειδή μορφή και διαρκεί για χρονικό διάστημα πολύ μεγαλύτερο του έτους. Η συμπεριφορά αυτή των τιμών των χρονοσειρών αποδίδεται κυρίως στους οικονομικούς κύκλους, οι οποίοι οφείλονται σε μεταβαλλόμενες οικονομικές, τεχνολογικές και άλλες συνθήκες. Επειδή όμως οι οικονομικοί κύκλοι δεν εμφανίζονται με την ίδια περιοδικότητα ή και την ίδια μορφή, για το λόγο αυτό το στοιχείο της κυκλικότητας, σε αντίθεση με την τάση και την εποχικότητα, δεν θεωρείται ότι συμβάλλει άμεσα στη δημιουργία προβλέψεων. Ωστόσο, η κυκλικότητα μπορεί να χρησιμοποιηθεί για να προσδιοριστεί η μέχρι τώρα εξέλιξη των τιμών της χρονοσειράς.

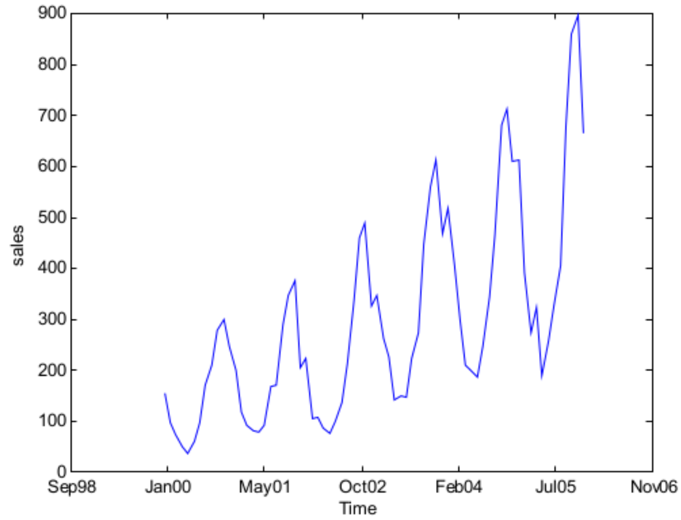
Οι εποχικές κινήσεις ή εποχικότητα αναφέρονται στην ταυτόσημη ή σχεδόν ταυτόσημη εξέλιξη που έχει μία χρονοσειρά κατά τη διάρκεια κάποιων συγκεκριμένων μηνών ή τριμήνων διαδοχικών ετών. Η εποχικότητα οφείλεται σε επαναλαμβανόμενα γεγονότα. Τα δεδομένα ορισμένων χρονοσειρών αναφέρονται σε χρονικές περιόδους μικρότερες του έτους, όπως μήνες ή τρίμηνα, με αποτέλεσμα να παρατηρούνται εποχικές διακυμάνσεις, οι οποίες εμφανίζονται κατά τη διάρκεια του έτους και επαναλαμβάνονται με την ίδια ή περίπου την ίδια μορφή από έτος σε έτος. Για παράδειγμα η μηνιαία κατανάλωση παγωτού είναι μεγαλύτερη κατά την καλοκαιρινή περίοδο και μικρότερη κατά την χειμερινή περίοδο. Γενικά, το φαινόμενο της εποχικότητας οφείλεται κυρίως σε μεταβολές του καιρού, σε πολιτικές της διοίκησης αναφορικά με περιόδους εκπτώσεων, καθώς και σε άλλους παράγοντες όπως θρησκευτικούς, κοινωνικούς κ.α. Οι εποχικές διακυμάνσεις, επειδή παρουσιάζονται με συστηματικό τρόπο συνήθως, μπορούν εύκολα να αναλυθούν και να προσδιοριστούν και κατά συνέπεια να χρησιμοποιηθούν για την πρόβλεψη των μελλοντικών τιμών της χρονοσειράς, κάτι που συμβαίνει άλλωστε και με την τάση.

Οι ακανόνιστες ή τυχαίες κινήσεις αναφέρονται στις σποραδικές, ακανόνιστες (irregular) κινήσεις μιας χρονοσειράς λόγω τυχαίων παραγόντων και γεγονότων. Οι τυχαίες κινήσεις επηρεάζουν τις τιμές των χρονοσειρών κατά ένα τυχαίο και μη συστηματικό τρόπο, ο οποίος δεν μπορεί να προσδιοριστεί. Η συνιστώσα αυτή λοιπόν δεν είναι δυνατόν να χρησιμοποιηθεί στη διαμόρφωση των μελλοντικών τιμών των χρονοσειρών. Οι τυχαίες κινήσεις οφείλονται σε όλους εκείνους τους τυχαίους και απρόσμενους παράγοντες που επηρεάζουν τις τιμές των χρονοσειρών και οι οποίοι δεν προσδιορίζονται από την τάση, την εποχικότητα και την κυκλικότητα. Οι παράγοντες αυτοί μπορεί να είναι πόλεμοι, σεισμοί, απρόσμενες καιρικές μεταβολές,απεργίες, διαδόσεις για συγκεκριμένο προϊον, αιφνίδιες μεταβολές στις προτιμήσεις των καταναλωτών, απρόσμενες αλλαγές στη νομοθεσία κ.α.

Σε μία συγκεκριμένη χρονοσειρά είναι δυνατόν να μην συνυπάρχουν και οι τέσσερις συνιστώσες αλλά μόνο κάποιες από αυτές. Η ανάλυση χρονοσειρών συνίσταται στην περιγραφή (εν γένει με μαθηματικό τρόπο) των συνιστωσών κινήσεων που υπάρχουν. Η γραφική παράσταση μιας χρονοσειράς απεικονίζεται στο παρακάτω διάγραμμα. Τα δεδομένα προέρχονται από τις μηνιαίες πωλήσεις μιας Χ εταιρείας για το χρονικό διάστημα 2000-2005. Οι πωλήσεις είναι σε χιλιάδες ευρώ. Στην παρακάτω χρονοσειρά συνυπάρχουν και οι τέσσερις συνιστώσες.

Για την ανάλυση των χρονοσειρών χρησιμοποιούμε τους ακόλουθους συμβολισμούς:

Y_t= Πραγματική τιμή της χρονοσειρά
T_t= Τάση
S_t= Εποχικότητα
C_t= Κυκλικότητα
I_t= Τυχαίες κινήσεις



Σχήμα 2.4: Γράφημα χρονοσειράς.

όπου $t = 1, 2, 3, \ldots, n$.

Η εξέταση των στοιχείων αυτών γίνεται σύμφωνα με κάποιο μαθηματικό υπόδειγμα που φανερώνει τον τρόπο με τον οποίο οι παρατηρήσεις της χρονοσειράς προσδιορίζονται από τις συνιστώσες της χρονοσειράς. Τα χρησιμοποιούμενα υποδείγματα είναι το προσθετικό μοντέλο (additive model) και το πολλαπλασιαστικό μοντέλο (multiplicative model).

Στο προσθετικό μοντέλο οι πραγματικές τιμές της χρονοσειράς για κάθε περίοδο θεωρούνται ως το άθροισμα των τεσσάρων συνιστωσών και δημιουργούνται με τον ακόλουθο τρόπο:

$$Y_t = T_t + S_t + C_t + I_t$$

Από την παραπάνω σχέση είναι φανερό ότι όλες οι συνιστώσες είναι εκφρασμένες στην ίδια μονάδα μέτρησης με εκείνη των παρατηρήσεων της χρονοσειράς.

Αντίθετα στο πολλαπλασιαστικό μοντέλο οι πραγματικές τιμές της χρονοσειράς προσδιορίζονται από το γινόμενο των τεσσάρων συνιστωσών, δηλαδή ως ακολούθως:

$$Y_t = T_t \cdot S_t \cdot C_t \cdot I_t$$

Στο μοντέλο αυτό μόνο η τάση είναι εκφρασμένη στην ίδια μονάδα μέτρησης με εκείνη της χρονοσειράς Y_t ενώ τα στοιχεία C_t , S_t και $_t$ είναι δείκτες ανεξάρτητοι από μονάδες μέτρησης.

Από τα δύο παραπάνω μοντέλα το προσθετικό μοντέλο χρησιμοποιείται λιγότερο συχνά στην πράξη, επειδή είναι δύσκολο στην ανάλυση του για υπολογιστικούς κυρίως λόγους. Επίσης βασίζεται στην υπόθεση ότι οι συνιστώσες της χρονοσειράς είναι ανεξάρτητες μεταξύ τους, που σημαίνει για παράδειγμα, ότι η τάση δεν επηρεάζει την εποχικότητα στον υπολογισμό των τιμών της χρονοσειράς. Η παραδοχή αυτή μπορεί να είναι σωστή κυρίως για φυσικά φαινόμενα, αλλά σπάνια ισχύει σε επιχειρησιακές και οικονομικές εφαρμογές, στις οποίες συνήθως η τάση επηρεάζει μεταξύ των άλλων και τις εποχικές διακυμάνσεις. Στη συνέχεια της διπλωματικής θα χρησιμοποιήσουμε το πολλαπλασιαστικό μοντέλο, δεδομένου ότι για τους παραπάνω πρακτικούς και θεωρητικούς λόγους το μοντέλο αυτό πλεονεκτεί του προσθετικού για την ανάλυση των οικονομικών χρονοσειρών.

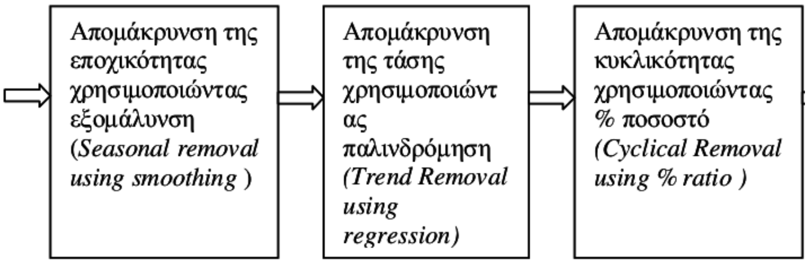
Η διάσπαση χρονοσειρών σε αντίθεση με τις μεθόδους εξομάλυνσης, οι οποίες εφαρμόζονται κυρίως για τη διαμόρφωση βραχυχρόνιων προβλέψεων καθώς και για χρονοσειρές με σχετικά μικρό αριθμό παρατηρήσεων, προϋποθέτει μεγαλύτερο αριθμό παρατηρήσεων και μπορεί να παράγει ακόμα και μακροπρόθεσμες προβλέψεις. Η διάσπαση χρονοσειρών είναι περισσότερο χρονοβόρα σε σχέση με τις προηγούμενες μεθόδους εξομάλυνσης, ωστόσο μας παρέχει τη δυνατότητα να μελετήσουμε πιο διεξοδικά τον τρόπο δημιουργίας των παρατηρήσεων μιας χρονοσειράς.

Η ανάλυση των χρονοσειρών με τη μέθοδο αυτή στηρίζεται στη διάσπαση των παρατηρήσεων τους σε τέσσερα συνθετικά στοιχεία, δηλαδή στην *τάση*, στην *εποχικότητα*, στην *κυκλικότητα* και στη *μη-κανονικότητα*. Σκόπος της διάσπασης των χρονοσειρών είναι η απομόνωση των τεσσάρων παραπάνω συνθετικών στοιχείων, ώστε να προσδιορίσουμε το βαθμό που επηρεάζει κάθε ένα στοιχείο ξεχωριστά τον τρόπο δημιουργίας των παρατηρήσεων των χρονοσειρών. Όταν αυτό επιτευχθεί μπορούμε να χρησιμοποιήσουμε τα αποτελέσματα της ανάλυσης για τη διαμόρφωση προβλέψεων.

Η διαδικασία της διάσπασης (decomposistion) της χρονοσειράς απεικονίζεται σχηματικά με το ακόλουθο διάγραμμα ροής για το πολλαπλασιαστικό μοντέλο.

2.4.1 ΑΝΑΛΥΣΗ ΕΠΟΧΙΚΟΤΗΤΑΣ

Στην ενότητα αυτή θα παρουσιάσουμε τον τρόπο με τον οποίο μπορούμε να απαλλάξουμε τις παρατηρήσεις τις χρονοσειράς από την επίδραση της εποχικότητας. Συγκεκριμένα για τον εντοπισμό και τη μέτρηση της εποχικότητας προβαί-



Σχήμα 2.5: Η διαδικασία της διάσπασης χρονοσειρών.

νουμε στον υπολογισμό εποχικών δεικτών. Εν συνεχεία, διαιρώντας τις πραγματικές τιμές της χρονοσειράς με τους αντίστοιχους εποχικούς δείκτες τις απαλλάσσουμε από το στοιχείο της εποχικότητας, ώστε να μπορέσουμε να διαμορφώσουμε πιο αξιόπιστες προβλέψεις. Ακολουθεί ανάλυση της ανωτέρω διαδικασίας.

Για να μπορέσουμε να υπολογίσουμε τους εποχικούς δείκτες θα πρέπει πρώτα να υπολογίσουμε τον κατάλληλο απλό κινητό μέσο. Ο κινητός μέσος θα πρέπει να περιλαμβάνει τον ίδιο αριθμό περιόδων με αυτόν που υπάρχει στην εποχικότητα που θέλουμε να αναγνωρίσουμε. Με άλλα λόγια, αν τα δεδομένα της χρονοσειράς μας είναι εκφρασμένα σε τριμηνιαία βάση και υποψιαζόμαστε ότι εμφανίζουν κάποια εποχικότητα, τότε θα πρέπει να υπολογίσουμε τον απλό κινητό μέσο τεσσάρων περιόδων. Στην περίπτωση που ο αριθμός των περιόδων του έτους είναι άρτιος, για παράδειγμα 4 για τριμηνιαίες παρατηρήσεις, 6 για διμηνιαίες παρατηρήσεις και 12 για μηνιαίες παρατηρήσεις, τότε θα πρέπει να υπολογίσουμε τον κεντρικό κινητό μέσο. Αυτό συμβαίνει γιατί υπολογίζοντας μόνο τον απλό κινητό μέσο δεν υπάρχει χρονική αντιστοιχία μεταξύ των τιμών του κινητού μέσου και των τιμών της χρονοσειράς. Το πρόβλημα αυτό αντιμετωπίζεται υπολογίζοντας τον κεντρικό κινητό μέσο που δεν είναι τίποτε άλλο παρά ο μέσος δύο διαδοχικών τιμών του απλού κινητού μέσου. Σε περίπτωση που ο αριθμός των περιόδων του έτους είναι περιττός, για παράδειγμα 3 για τετραμηνιαίες παρατηρήσεις, τότε δεν υφίσταται πρόβλημα χρονικής αντιστοιχίας μεταξύ των τιμών του κινητού μέσου και των τιμών της χρονοσειράς και ο υπολογισμός του απλού κινητού μέσου μας αρκεί για τον υπολογισμό των εποχικών δεικτών.

Έχοντας υπολογίσει τον απλό ή τον κεντρικό κινητό μέσο, ανάλογα με τον αριθμό περιόδων του έτους, είμαστε σε θέση να υπολογίσουμε τους εποχικούς δείκτες. Οι εποχικοί δείκτες προκύπτουν από τη διαίρεση των τιμών της χρονοσειράς με τον κινητό μέσο (απλό ή κεντρικό). Έτσι, στην περίπτωση του πολλαπλασιαστικού μοντέλου ο δείκτης εποχικότητας S_t της περιόδου t, για $t = 1, 2, \ldots, n$, καθορίζεται από την ακόλουθη σχέση:

$$S_t = \frac{Y_t}{CA_t} = \frac{T_t \cdot S_t \cdot C_t \cdot I_t}{T_t \cdot C_t \cdot I_t}$$

όπου CA_t είναι η εξομαλυνθείσα τιμή της χρονοσειράς που προέρχεται από τη μέθοδο του κεντρικού κινητού μέσου που χρησιμοποιήθηκε. Έτσι η εποχικότητα προσδιορίζεται από το λόγο των πραγματικών τιμών Y_t της χρονοσειράς προς τις εξομαλυνθείσες τιμές της CA_t , θεωρώντας ότι οι τιμές CA_t εκφράζουν ικανοποιητικά την ταυτόχρονη συμπεριφορά της τάσης, της κυκλικότητας και της μη- κανονικότητας.Επειδή το πολλαπλασιαστικό υπόδειγμα προϋποθέτει ότι το άθροισμα των εποχικών δεικτών είναι ίσο με τον αριθμό των περιόδων, αν κάτι τέτοιο δεν ισχύει τότε θα πρέπει να αναπροσαρμόσουμε τους εποχικούς δείκτες. Το άθροισμα των προσαρμοσμένων εποχικών δεικτών είναι ίσο με τον αριθμό των περιόδων. Η προσαρμογή προκύπτει απ' το γινόμενο των αντίστοιχων μέσων τιμών των δεικτών, με το κλάσμα : αριθμός περιόδων / άθροισμα μέσων εποχικών δεικτών.

Αφού υπολογίσουμε τους προσαρμοσμένους εποχικούς δείκτες, μπορούμε στη συνέχεια να απαλείψουμε την εποχικότητα, διαιρώντας κάθε τιμή Yt της χρονοσειράς με τον προσαρμοσμένο δείκτη SA_i της αντίστοιχης περιόδου, δηλαδή ως εξής:

$$SAY_t = \frac{Y_t}{SA_i}$$

όπου SAY_t είναι οι απαλλαγμένες από εποχικότητα (seasonally adjusted) τιμές της χρονοσειράς της περιόδου t. Οι τιμές αυτές περιέχουν την τάση, την κυκλικότητα και τη μη-κανονικότητα.

2.4.2 ΑΝΑΛΥΣΗ ΜΑΚΡΟΧΡΟΝΙΑΣ ΤΑΣΗΣ

Στην ενότητα αυτή θα χρησιμοποιήσουμε τις απαλλαγμένες από εποχικότητα τιμές της χρονοσειράς για να μελετήσουμε το άλλο συνθετικό της στοιχείο που είναι η τάση. Η τάση φανερώνει τη μακροχρόνια εξέλιξη των τιμών της χρονοσειράς (ανοδική ή πτωτική), η οποία οφείλεται σε δημογραφικούς, τεχνολογικούς, οικονομικούς και άλλους παράγοντες. Για τον προσδιορισμό της θα υποθέσουμε ότι αυτή μπορεί να εκφραστεί ικανοποιητικά από ένα γραμμικό υπόδειγμα στο οποίο ως ανεξάρτητη μεταβλητή θα είναι ο χρόνος. Έστω ότι η τάση δίνεται από το ακόλουθο γραμμικό υπόδειγμα:

$$Y_t = \alpha + \beta \cdot t + \varepsilon_t \tag{2.5}$$

όπου Y_t είναι οι παρατηρήσεις της χρονοσειράς, t η ανεξάρτητη μεταβλητή του χρόνου που λαμβάνει τιμές $1, 2, \ldots, n$ και ε το τυχαίο σφάλμα. Οι εκτιμήσεις των συντελεστών της παραπάνω σχέσης προσδιορίζονται με τη μέθοδο των ελαχίστων τετραγώνων OLS ως εξής:

$$\beta' = \frac{n\sum_{t=1}^{n} t \cdot Y_t - \left(\sum_{t=1}^{n} t\right)\left(\sum_{t=1}^{n} Y_t\right)}{n\sum_{t=1}^{n} t^2 - \left(\sum_{t=1}^{n} t\right)^2}$$

και

$$\alpha' = \frac{1}{n}\sum_{t=1}^{n} Y_t - \beta\frac{1}{n}\sum_{t=1}^{n} t$$

Η τιμή του α είναι η σταθερά της γραμμικής τάσης, δηλαδή η τιμή της τάσης όταν $t = 0$. Αντίθετα, η τιμή του β είναι η κλίση της γραμμικής τάσης και δηλώνει το πόσο θα μεταβληθεί η τιμή της χρονοσειράς, όταν ο χρόνος t μεταβληθεί κατά μια μονάδα. Έτσι όταν η τιμή του β είναι θετική, η μακροχρόνια τάση είναι ανοδική, ενώ όταν η τιμή του είναι αρνητική, η μακροχρόνια τάση είναι πτωτική.

Η τάση είναι το μόνο συνθετικό στοιχείο της χρονοσειράς που μπορούμε να καθορίσουμε, ανεξάρτητα από την ύπαρξη ή όχι εποχικότητας στις τιμές της χρονοσειράς. Όταν δεν υπάρχει εποχικότητα οι συντελεστές της σχέσης (2.5) προσδιορίζονται χρησιμοποιώντας ως εξαρτημένη μεταβλητή τις πραγματικές τιμές Y_t της χρονοσειράς. Αντίθετα, όταν υπάρχει εποχικότητα, η Y_t είναι η SAY_t , δηλαδή οι απαλλαγμένες από εποχικότητα τιμές της χρονοσειράς. Στην περίπτωση αυτή, οι παρατηρήσεις της χρονοσειράς περιέχουν την τάση, την κυκλικότητα και τη μη- κανονικότητα, δηλαδή οι τιμές SAY_t προσδιορίζονται ως:

$$SAY_t = T_t C_t I_t$$

Οι τιμές SAY_t χρησιμοποιούνται στη συνέχεια για την εκτίμηση της γραμμικής τάσης:

$$T_t = \alpha' + \beta' t$$

για όλες τις τιμές του t, δηλαδή για $t = 1, 2, \ldots, n$.

2.4.3 ΑΝΑΛΥΣΗ ΚΥΚΛΙΚΟΤΗΤΑΣ ΚΑΙ ΜΗ ΚΑΝΟΝΙΚΟΤΗΤΑΣ

Για να μελετήσουμε τα στοιχεία της κυκλικότητας και της μη-κανονικότητας της χρονοσειράς πρέπει πρώτα να απομονώσουμε την κυκλικότητα και τη μη- κανονικότητα από τα άλλα δύο συνθετικά στοιχεία της χρονοσειράς. Η απομόνωση αυτή γίνεται με την απαλλαγή της τάσης από τις ήδη απαλλαγμένες από εποχικότητα τιμές της χρονοσειράς. Συνεπώς

$$TAY_t = \frac{SAY_t}{T_t} = C_t I_t$$

όπου TAY_t είναι οι τιμές της χρονοσειράς οι απαλλαγμένες από εποχικότητα και από την τάση (trend adjusted), δηλαδή οι τιμές αυτές περιέχουν μόνο την κυκλικότητα και τη μη-κανονικότητα.

Η κυκλικότητα και η μη-κανονικότητα εκφράζονται ως ποσοστό της τάσης και είναι δείκτες ανεξάρτητοι από τη μονάδα μέτρησης των παρατηρήσεων της χρονοσειράς, όπως οι εποχικοί. Αν ο λόγος SAY_t/T_t είναι ίσος με τη μονάδα για όλες τις παρατηρήσεις της χρονοσειράς, τότε τα στοιχεία της κυκλικότητας και της μη-κανονικότητας δεν εμφανίζονται στις παρατηρήσεις της χρονοσειράς. Διαγραμματικά αυτό σημαίνει ότι όλες οι απαλλαγμένες από εποχικότητα τιμές της χρονοσειράς βρίσκονται στην εκτιμηθείσα γραμμή της τάσης. Αντίθετα, αν ο λόγος SAY_t/T_t δεν ισούται με τη μονάδα, που είναι και η πλέον συνηθισμένη περίπτωση, τότε στις παρατηρήσεις της χρονοσειράς υπάρχει κυκλικότητα και μη-κανονικότητα που προσδιορίζεται από το λόγο αυτό.

Εάν ενδιαφερόμαστε να απομονώσουμε την κυκλικότητα από τη μη- κανονικότητα τότε εφαρμόζουμε τη μέθοδο του σταθμικού κεντρικού κινητού μέσου (weighted centered moving average) στα απαλλαγμένα από εποχικότητα και τάση δεδομένα της χρονοσειράς. Ο σταθμικός κεντρικός κινητός μέσος δίνει μεγαλύτερη βαρύτητα στην κεντρική παρατήρηση και μικρότερη όσο απομακρυνόμαστε χρονικά από αυτή. Οι συντελεστές βαρύτητας ορίζονται από τον ερευνητή, ανάλογα με το είδος των δεδομένων (μηνιαία, διμηνιαία κ.τ.λ.), ενώ το άθροισμα τους πρέπει να είναι ίσο με τη μονάδα.

Για παράδειγμα ένας σταθμικός κεντρικός κινητός μέσος WA_t για τριμηνιαία δεδομένα θα μπορούσε να ήταν:

$$WA_t = \frac{TAY_{t-1} + 2TAY_t + TAY_{t+1}}{4}$$

όπου $t = 1, 2, \dots, n-1$ και Y_t είναι τα απαλλαγμένα από εποχικότητα και τάση δεδομένα της χρονοσειράς. Από την παραπάνω σχέση είναι φανερό ότι το άθροισμα των συντελεστών βαρύτητας ισούται με τη μονάδα και ότι ο συντελεστής του κεντρικού τριμήνου έχει διπλάσια βαρύτητα από τους συντελεστές του προηγούμενου και του επόμενου τριμήνου.

Στην πραγματικότητα, με τη μέθοδο του σταθμικού κεντρικού κινητού μέσου εξομαλύνουμε τα απαλλαγμένα από εποχικότητα και τάση δεδομένα της χρονοσειράς. Οι εξομαλυνθείσες τιμές που προκύπτουν είναι δείκτες, οι τιμές των οποίων είναι μικρότερες, ίσες ή μεγαλύτερες της μονάδας και φανερώνουν την ποσοστιαία μεταβολή της τιμής της χρονοσειράς που οφείλεται στην κυκλικότητα για κάθε χρονική περίοδο. Έτσι αν η τιμή του WA_t είναι ίση με 1,36 , αυτό σημαίνει ότι για τη συγκεκριμένη περίοδο υπάρχει αύξηση κατά 36% στην τιμή της χρονοσειράς, που οφείλεται στην κυκλικότητα. Εάν πάλι η τιμή του WAt είναι 0,88 τότε υπάρχει μείωση κατά 12% στην τιμή της χρονοσειράς που προέρχεται από το συνθετικό στοιχείο της κυκλικότητας.

Τέλος μπορούμε να απομονώσουμε τη μη-κανονικότητα εάν απομακρύνουμε την κυκλικότητα από τα απαλλαγμένα από εποχικότητα και τάση δεδομένα της χρονοσειράς. Αυτό επιτυγχάνεται με τον ακόλουθο τρόπο:

$$CAY_t = \frac{TAY_t}{WA_t} = I_t$$

όπου CAY_t είναι οι τιμές της χρονοσειράς οι απαλλαγμένες και από την κυκλικότητα (cyclical adjusted). Έτσι, οι τιμές αυτές περιέχουν μόνο το στοιχείο της μη κανονικότητας.

2.4.4 ΔΙΑΜΟΡΦΩΣΗ ΠΡΟΒΛΕΨΕΩΝ

Η διαμόρφωση των προβλέψεων εξαρτάται σε μεγάλο βαθμό από τη διαδικασία αναγνώρισης των συνθετικών στοιχείων ή συνιστωσών της χρονοσειράς. Όσο καλύτερη είναι η αναγνώριση των στοιχείων αυτών, τόσο καλύτερη αναμένεται να είναι και η πρόβλεψη των τιμών της χρονοσειράς. Έτσι, κάθε ένα συνθετικό στοιχείο χρησιμοποιείται ξεχωριστά για τον προσδιορισμό των μελλοντικών τιμών της χρονοσειράς. Ειδικότερα, η πρόβλεψη $\widehat{Y}_{t+h}$ της h μελλοντικής περιόδου προσδιορίζεται με βάση το πολλαπλασιαστικό μοντέλο ως εξής:

$$\widehat{Y}_{t+h} = T_{t+h}S_{t+h}C_{t+h}I_{t+h}$$

Η τιμή του I_{t+h} , δηλαδή η συμβολή της μη κανονικότητας για την h μελλοντική περίοδο, δεν μπορεί να καθοριστεί, αφού εξαρτάται από τυχαίους και απρόσμενους παράγοντες και κατά συνέπεια δεν μπορεί να προσδιοριστεί. Για το λόγο αυτό, η μη- κανονικότητα δεν περιλαμβάνεται στη διαμόρφωση των προβλέψεων και έτσι η τιμή του I_{t+h} τίθεται ίση με τη μονάδα, δηλαδή:

$$I_{t+h} = 1$$

Όσον αφορά την κυκλικότητα, αν οι κυκλικές διακυμάνσεις είναι μεγάλες τότε συμπεριλαμβάνονται στην πρόβλεψη ενώ αν οι κυκλικές διακυμάνσεις είναι μικρές, δηλαδή οι τιμές WA_t είναι κοντά στη μονάδα, τότε το C_{t+h} δεν περιλαμβάνεται στη διαμόρφωση των προβλέψεων και θεωρούμε ότι η τιμή του είναι ίση με τη μονάδα, δηλαδή:

$$C_{t+h} = 1$$

Για την τιμή του S_{t+h} , δηλαδή για την τιμή του δείκτη εποχικότητας της h μελλοντικής περιόδου, χρησιμοποιούμε την τιμή του προσαρμοσμένου δείκτη εποχικότητας SA της περιόδου εντός του έτους στην οποία αναφέρεται η h μελλοντική περίοδος, δηλαδή:

$$S_{t+h} = SA_i$$

για $i = 1, 2, \ldots, L$, όπου L είναι η περιοδικότητα της εποχικότητας.

Τέλος, η τιμή του T_{t+h} , δηλαδή η τιμή της τάσης της h μελλοντικής περιόδου, προκύπτει από τη σχέση ως εξής:

$$T_{t+h} = \alpha' + \beta' (t + h)$$

όπου α και β οι συντελεστές της γραμμικής τάσης.

Επομένως οι προβλέψεις των τιμών της χρονοσειράς για την h μελλοντική περίοδο δημιουργούνται από την ακόλουθη σχέση:

$$\widehat{Y}_{t+h} = [\alpha' + \beta' (t + h)] \, SA_i$$

όπου έχουμε θεωρήσει $I_{t+h} = 1$ και $C_{t+h} = 1$, δηλαδή λαμβάνουμε υπ'όψιν μόνο την τάση και την εποχικότητα.

Μέρος ΙΙ

ΕΦΑΡΜΟΓΗ ΤΩΝ ΜΕΘΟΔΩΝ ΓΙΑ ΠΡΟΒΛΕΨΕΙΣ

Κεφάλαιο 3

ΕΦΑΡΜΟΓΕΣ ΤΩΝ ΜΕΘΟΔΩΝ ΕΞΟΜΑΛΥΝΣΗΣ, ΔΙΑΣΠΑΣΗΣ ΧΡΟΝΟΣΕΙΡΩΝ ΚΑΙ ΑΝΑΛΥΣΗΣ ΠΑΛΙΝΔΡΟΜΗΣΗΣ

3.1 ΠΡΟΒΛΕΨΗ ΤΙΜΩΝ ΧΡΟΝΟΣΕΙΡΑΣ ΜΕ ΤΗ ΜΕΘΟΔΟ ΔΙΑΣΠΑΣΗΣ ΧΡΟΝΟΣΕΙΡΩΝ ΚΑΙ ΤΗ ΜΕΘΟΔΟ WINTERS

Παράδειγμα 3.1

Στον πίνακα 3.1 που ακολουθεί παρουσιάζεται η ανάλυση της εποχικότητας, αναφορικά με τις πωλήσεις των αυτοκινήτων στην Ελλάδα. Πιο συγκεκριμένα, δίνονται οι τριμηνιαίες τιμές των πωλήσεων των αυτοκινήτων στην Ελλάδα για τα έτη απο το 2000 έως το 2004. Επίσης, στον ίδιο πίνακα δίνονται ο κινητός μέσος (MA), ο κεντρικός κινητός μέσος (CA_t), οι εποχικοί δείκτες (S_t) και οι τιμές της χρονοσειράς απαλλαγμένες από εποχικότητα (SAY_t). Ακολούθως, στον πίνακα 3.2 παρουσιάζονται οι προσαρμοσμένοι εποχικοί δείκτες.

Έτος	Τριμ	t	Y_t	MA	CA_t	S_t	SAY_t
2000	1ο	1	83.754	-	-	-	76.371,48
	2ο	2	83.121	72.555,50	-	-	72.576,98
	3ο	3	68.976	70.930,25	71.742,88	0,9614	70.538,78
	4ο	4	54.371	70.588,75	70.759,50	0,7684	69.687,79
2001	1ο	5	77.253	70.700,75	70.644,75	1,0935	70.443,51

	2ο	6	81.755	70.053,50	70.377,13	1,1617	71.384,26
	3ο	7	69.424	68.683,25	69.368,38	1,0008	70.996,93
	4ο	8	51.782	67.187,25	67.935,25	0,7622	66.369,44
2002	1ο	9	71.772	67.224,50	67.205,88	1,0679	65.445,64
	2ο	10	75.771	67.122,25	67.173,38	1,1280	66.159,34
	3ο	11	69.573	66.717,50	66.919,88	1,0396	71.149,30
	4ο	12	51.373	65.541,50	66.129,50	0,7769	65.845,22
2003	1ο	13	70.153	63.125,50	64.333,50	1,0905	63.969,34
	2ο	14	71.067	64.323,25	63.724,38	1,1152	62.052,05
	3ο	15	59.909	67.129,25	65.726,25	0,9115	61.266,35
	4ο	16	56.164	70.704,50	68.916,88	0,8150	71.985,89
2004	1ο	17	81.377	72.444,25	71.574,38	1,1370	74.204,00
	2ο	18	85.368	72.422,75	72.433,50	1,1786	74.538,95
	3ο	19	66.868	-	-	-	68.383,02
	4ο	20	56.078	-	-	-	71.875,66

Πίνακας 3.1: Δεδομένα, εποχικοί δείκτες και τιμές απαλλαγμένες από εποχικότητα.

Πίνακας 3.2: Προσαρμοσμένοι Εποχικοί Δείκτες.

Έτος	Τρίμηνο 1ο	2ο	3ο	4ο	Άθροισμα
2000			0,961	0,768	
2001	1,094	1,162	1,001	0,762	
2002	1,068	1,128	1,040	0,777	
2003	1,090	1,115	0,911	0,815	
2004	1,137	1,179			
Άθροισμα	4,389	4,583	3,913	3,122	
Μέσοι	1,097	1,146	0,978	0,781	4,0020
Προσαρμ. Εποχικοί Δείκτες	1,097	1,145	0,978	0,780	4,0000

Ακολουθούν οι πίνακες για τον προσδιορισμό της Μακροχρόνιας Τάσης και των Κυκλικών Διακυμάνσεων για τις πωλήσεις των αυτοκινήτων στην Ελλάδα από το 2000 μέχρι το 2004.

Έτος	Τριμ	t	Y_t	SAY_t	$t * SAY_t$	Τάση
2000	1ο	1	83.754	76.371,48	76.371,48	70.551,67
	2ο	2	83.121	72.576,98	145.153,96	70.415,93
	3ο	3	68.976	70.538,78	211.616,34	70.280,20
	4ο	4	54.371	69.687,79	278.751,16	70.144,46
2001	1ο	5	77.253	70.443,51	352.217,55	70.008,73
	2ο	6	81.755	71.384,26	428.305,56	69.873,00
	3ο	7	69.424	70.996,93	496.978,51	69.737,26
	4ο	8	51.782	66.369,44	530.955,52	69.601,53
2002	1ο	9	71.772	65.445,64	589.010,76	69.465,80
	2ο	10	75.771	66.159,34	661.593,40	69.330,06
	3ο	11	69.573	71.149,30	782.642,30	69.194,33
	4ο	12	51.373	65.845,22	790.142,64	69.058,60
2003	1ο	13	70.153	63.969,34	831.601,42	68.922,86
	2ο	14	71.067	62.052,05	868.728,70	68.787,13
	3ο	15	59.909	61.266,35	918.995,25	68.651,40
	4ο	16	56.164	71.985,89	1.151.774,24	68.515,66
2004	1ο	17	81.377	74.204,00	1.261.468,00	68.379,93
	2ο	18	85.368	74.538,95	1.341.701,10	68.244,20
	3ο	19	66.868	68.383,02	1.299.277,38	68.108,46
	4ο	20	56.078	71.875,66	1.437.513,20	67.972,73

Πίνακας 3.3: Προσδιορισμός Μακοχρόνιας Τάσης για τις πωλήσεις αυτοκινήτων.

Έτος	Τρίμ	t	Y_t	SAY_t	Τάση	TAY_t	WA_t
2000	1ο	1	83.754	76.371,48	70.551,67	1,082	-
	2ο	2	83.121	72.576,98	70.415,93	1,031	1,037
	3ο	3	68.976	70.538,78	70.280,20	1,004	1,008
	4ο	4	54.371	69.687,79	70.144,46	0,993	0,999
2001	1ο	5	77.253	70.443,51	70.008,73	1,006	1,007
	2ο	6	81.755	71.384,26	69.873,00	1,022	1,017
	3ο	7	69.424	70.996,93	69.737,26	1,018	1,003
	4ο	8	51.782	66.369,44	69.601,53	0,954	0,967
2002	1ο	9	71.772	65.445,64	69.465,80	0,942	0,948
	2ο	10	75.771	66.159,34	69.330,06	0,954	0,970
	3ο	11	69.573	71.149,30	69.194,33	1,028	0,991

	4ο	12	51.373	65.845,22	69.058,60	0,953	0,966
2003	1ο	13	70.153	63.969,34	68.922,86	0,928	0,928
	2ο	14	71.067	62.052,05	68.787,13	0,902	0,906
	3ο	15	59.909	61.266,35	68.651,40	0,892	0,934
	4ο	16	56.164	71.985,89	68.515,66	1,051	1,020
2004	1ο	17	81.377	74.204,00	68.379,93	1,085	1,078
	2ο	18	85.368	74.538,95	68.244,20	1,092	1,068
	3ο	19	66.868	68.383,02	68.108,46	1,004	1,039
	4ο	20	56.078	71.875,66	67.972,73	1,057	-

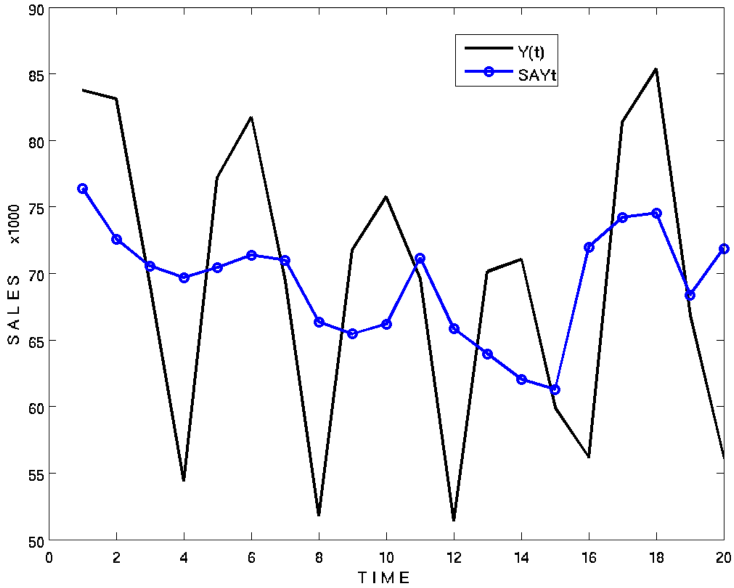
Πίνακας 3.4: Προσδιορισμός Κυκλικών Διακυμάνσεων των πωλήσεων των αυτοκινήτων.

Στον πίνακα που ακολουθεί απεικονίζονται οι προβλέψεις για τα έτη 2005 και 2006.

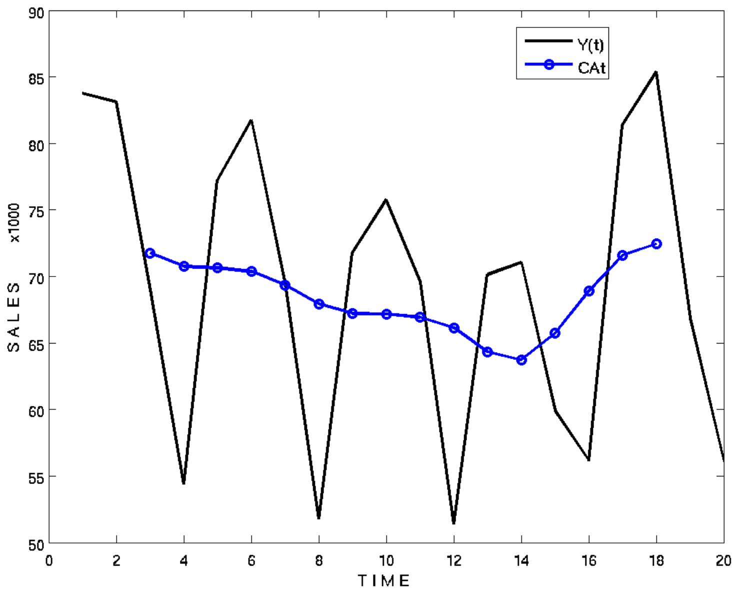
Πίνακας 3.5: Προβλέψεις για τα επόμενα δύο έτη.

Έτος	Τριμ	t	$\widehat{Y}_t$
2005	1ο	21	74.395
	2ο	22	77.537
	3ο	23	66.069
	4ο	24	52.609
2006	1ο	25	73.799
	2ο	26	76.915
	3ο	27	65.538
	4ο	28	52.186

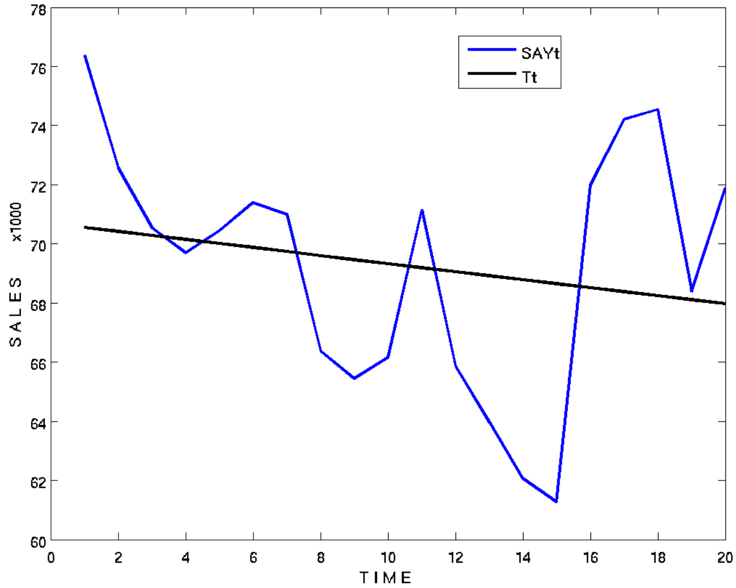
Στα παρακάτω διαγράμματα απεικονίζονται αναλυτικά η ανάλυση της εποχικότητας, της μακροχρόνιας τάσης και των κυκλικών διακυμάνσεων, αναφορικά με τις πωλήσεις των αυτοκινήτων στην Ελλάδα κατά το έτος 2000 έως 2004, καθώς επίσης και η πρόβλεψη μας για τα επόμενα δύο έτη 2005 και 2006.



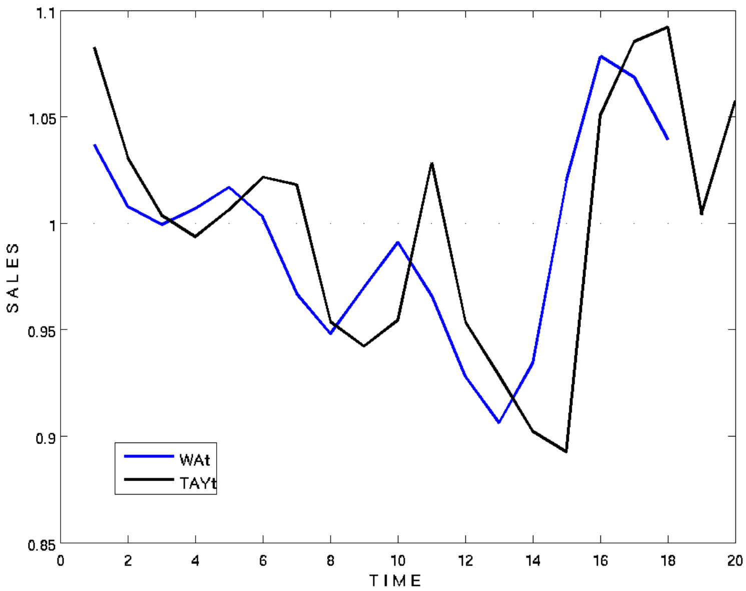
Σχήμα 3.1: Πραγματικές και απαλλαγμένες από εποχικότητα τιμές της χρονοσειράς του παραδείγματος 3.1.



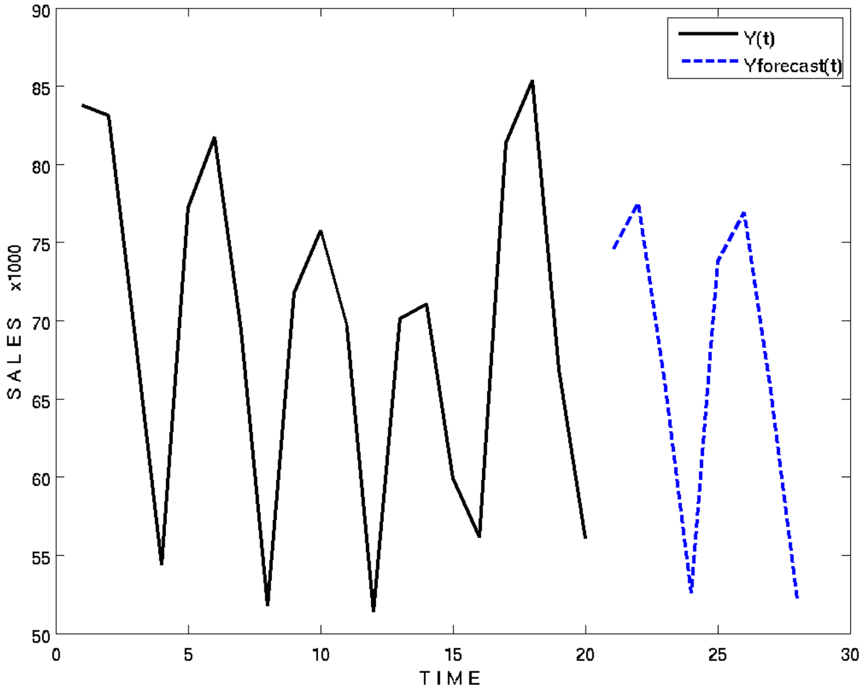
Σχήμα 3.2: Εξομάλυνση των δεδομένων του παραδείγματος 3.1 με κεντρικό κινητό μέσο.



Σχήμα 3.3: Τιμές χρονοσειράς απαλλαγμένες από εποχικότητα και οι αντίστοιχες τιμές της τάσης του παραδείγματος 3.1.



Σχήμα 3.4: Τιμές χρονοσειράς απαλλαγμένες από τάση και από εποχικότητα του παραδείγματος 3.1 και τιμές του σταθμικού κεντρικού κινητού μέσου.

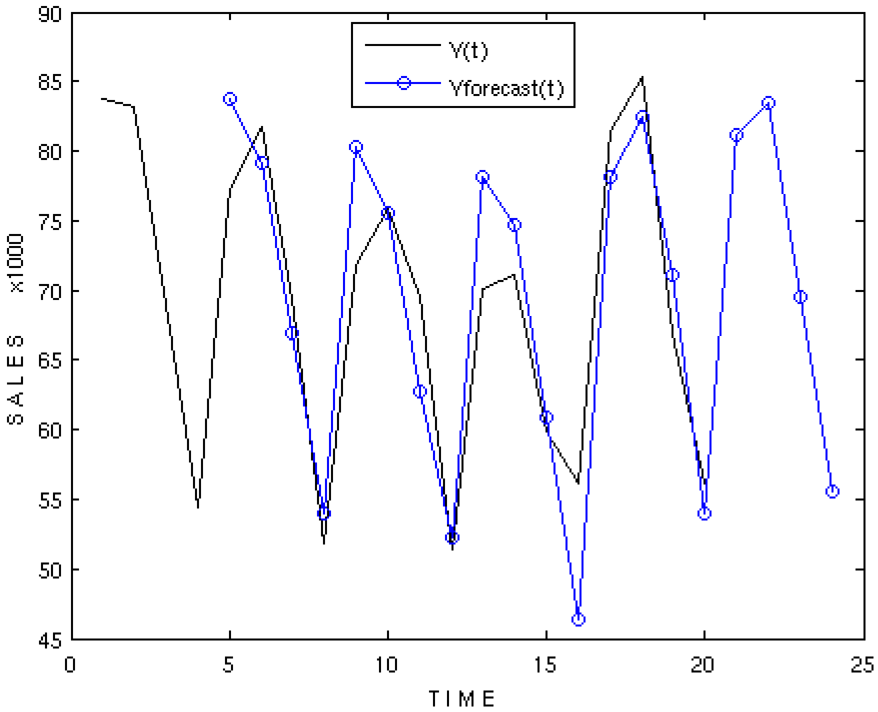


Σχήμα 3.5: Πραγματικές τιμές της χρονοσειράς και οι προβλεπόμενες τιμές για το 2005 και 2006.

Στον πίνακα 3.6 και στο διάγραμμα που ακολουθεί παρατηρούμε τις προβλέψεις για τις πωλήσεις των αυτοκινήτων στην Ελλάδα με τη μέθοδο Winters για $\alpha = 0,61$, $\beta = 0,01$ και $\gamma = 0,32$. Οι τιμές αυτές έχουν επιλεγεί διότι ελαχιστοποιούν την τιμή του κριτηρίου MSE.

Πίνακας 3.6: Προβλέψεις για τις πωλήσεις αυτοκινήτων στην Ελλάδα με τη μέθοδο Wintes για α=0,61, β=0,01 και γ=0,32

Έτος	Τρίμ	t	Y_t	A_t	T_t	S_t	$\widehat{Y}_t$	e_t
2000	1ο	1	83.754			1,15		
	2ο	2	83.121			1,15		
	3ο	3	68.976			0,95		
	4ο	4	54.371	72.555,50	0,00	0,75		
2001	1ο	5	77.253	69.121,08	-34,34	1,14	83.754,00	-6.501,00
	2ο	6	81.755	70.474,95	-20,46	1,15	79.147,11	2.607,89
	3ο	7	69.424	72.023,13	-4,78	0,95	66.978,64	2.445,36
	4ο	8	51.782	70.239,02	-22,57	0,75	53.968,48	-2.186,48
2002	1ο	9	71.772	65.700,57	-67,73	1,13	80.233,57	-8.461,57
	2ο	10	75.771	65.780,78	-66,25	1,15	75.491,97	279,03
	3ο	11	69.573	70.072,20	-22,67	0,97	62.749,69	6.823,31
	4ο	12	51.373	69.354,86	-29,62	0,74	52.222,22	-849,22
2003	1ο	13	70.153	65.020,45	-72,67	1,11	78.106,10	-7.953,10
	2ο	14	71.067	63.002,24	-92,12	1,14	74.738,23	-3.671,23
	3ο	15	59.909	62.327,53	-97,95	0,97	60.832,78	-923,78
	4ο	16	56.164	70.316,76	-17,08	0,76	46.297,77	9.866,23
2004	1ο	17	81.377	72.078,03	0,71	1,12	78.135,81	3.241,19
	2ο	18	85.368	73.649,64	16,42	1,15	82.422,38	2.945,62
	3ο	19	66.868	70.993,35	-10,31	0,96	71.097,92	-4.229,92
	4ο	20	56.078	72.609,38	5,95	0,76	54.047,41	2.030,59
2005	1ο	21					81.115	
	2ο	22					83.404	
	3ο	23					69.558	
	4ο	24					55.556	



Σχήμα 3.6: Γραφική παράσταση πραγματικών και προβλεπόμενων τιμών της χρονοσειράς με τη μέθοδο Winters.

3.2 ΕΦΑΡΜΟΓΗ ΣΤΗΝ ΑΠΛΗ ΓΡΑΜΜΙΚΗ ΠΑΛΙΝΔΡΟΜΗΣΗ

Παράδειγμα 3.2

Στο ακόλουθο παράδειγμα θέλουμε να δούμε τη σχέση που υπάρχει ανάμεσα στις πωλήσεις των αυτοκινήτων στην Ελλάδα και στο ΑΕΠ της χώρας για δεκαπέντε έτη. Στον πίνακα 3.7 απεικονίζονται ο συνολικός αριθμός των πωλήσεων και το ΑΕΠ της χώρας για τα έτη από το 1990 μέχρι το 2004.

Πίνακας 3.7: Οι πωλήσεις των αυτοκινήτων και το ΑΕΠ της Ελλάδας για τα έτη 1990 έως 2004.

Έτος	Περίοδος	Πωλήσεις (Y_t)	ΑΕΠ (X_t)
1990	1	115.480	84.488,3
1991	2	167.737	87.108.9
1992	3	199.094	87.716,2
1993	4	147.789	86.313,5
1994	5	109.544	88.039.6
1995	6	125.023	89.888,3
1996	7	139.821	92.008,2
1997	8	159.867	95.355,1
1998	9	180.145	98.562,6
1999	10	261.711	101.933,1
2000	11	290.222	106.496,7
2001	12	280.214	111.026,5
2002	13	268.489	115.209,5
2003	14	257.293	120.580,7
2004	15	289.691	125.605,0

Στην εφαρμογή αυτή θέλουμε τα εξής:

(i) Να κατασκευαστεί στικτό διάγραμμα (διάγραμμα διασποράς ή διασκόρπισης) για τα δεδομένα του παραδείγματος.

(ii) Να βρεθεί η ευθεία παλινδρόμησης των πωλήσεων των αυτοκινήτων ως προς το ΑΕΠ της χώρας.

(iii) Να βρεθεί σε τι ποσοστό ερμηνεύεται η μεταβλητότητα του αριθμού των πωλήσεων (από το μοντέλο).

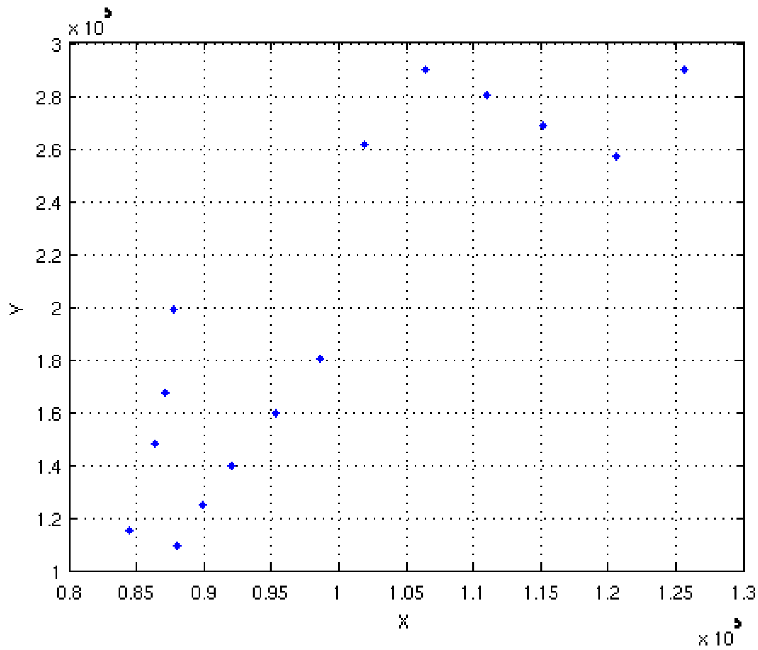
(iv) Να δοθεί μια εκτίμηση της διασποράς των σφαλμάτων.

(v) Να βρεθεί το 95% διάστημα εμπιστοσύνης για τους συντελεστές β_0 και β_1.

(vi) Εάν για τα έτη 2005 και 2006 το ΑΕΠ αναμένεται να λάβει τις τιμές 130373,1 και 134370,2 αντίστοιχα, να προβλέψετε τον αριθμό των πωλήσεων των αυτοκινήτων α) δίνοντας μία σημειακή εκτίμηση και β) χρησιμοποιώντας ένα $100\,(1-\alpha)\,\%$ διάστημα εμπιστοσύνης με επίπεδο σημαντικότητας $\alpha = 0.05$.

Απαντήσεις:

(i) Το διάγραμμα διασποράς για το συγκεκριμένο παράδειγμα, φαίνεται παρακάτω:



Σχήμα 3.7: Διάγραμμα διασποράς των μεταβλητών X,Y όπου: Y ο αριθμός των πωλήσεων και X το ΑΕΠ της χώρας.

(ii) Με την βοήθεια των δεδομένων σχηματίζουμε τον παρακάτω πίνακα:

	Y_t	X_t	X^2	Y^2	$X * Y$
	115480	84488.3	7138272836.89	13335630400	9756708884
	167737	87108.9	7587960459.21	28135701169	14611385559.3
	199094	87716.2	7694131742.44	39638420836	17463769122.8
	147789	86313.5	7450020282.25	21841588521	12756185851.5
	109544	88039.6	7750971168.16	11999887936	9644209942.4
	125023	89888.3	8079906476.89	15630750529	11238104930.9
	139821	92008.2	8465508867.24	19549912041	12864678532.2
	159867	95355.1	9092595096.01	25557457689	15244133771.7
	180145	98562.6	9714586118.76	32452221025	17755559577
	261711	101933.1	10390356875.61	68492647521	26677013534.1
	290222	106496.7	11341547110.89	84228809284	30907685267.4
	280214	111026.7	12326928112.89	78519885796	31111235713.8
	268489	115209.5	13273228890.25	72086343121	30932483445.5
	257293	120580.7	14539705212.49	66199687849	31024570045.1
	289691	125605	15776616025	83920875481	36386638055
SUM	2992120	1490332.4	150622335274.98	661589819198	308374362232.7

Η ευθεία παλινδρόμησης των πωλήσεων των αυτοκινήτων ως προς το ΑΕΠ της χώρας, υπολογίζεται απ' τις εξής σχέσεις με (n=15):

$$\widehat{\beta}_1 = \frac{n\sum_{t=1}^{n} XY - \sum_{t=1}^{n} X \sum_{t=1}^{n} Y}{n\sum_{t=1}^{n} X^2 - \left(\sum_{t=1}^{n} X\right)^2} = \frac{15 * 308374362232.7 - 1490332.4 * 2992120}{15 * 150622335274.98 - 1490332.4^2}$$
$$= 4.35$$

και

$$\widehat{\beta}_0 = \frac{\sum_{t=1}^{n} Y}{n} - \widehat{\beta}_1 \frac{\sum_{t=1}^{n} X}{n} = \frac{2992120}{15} - 4.35\frac{1490332.4}{15} = -232719.28$$

Άρα, η εκτιμηθείσα μορφή του απλού γραμμικού υποδείγματος, που προκύπτει από την ανάλυση των δεδομένων μας ή αλλιώς η ευθεία παλινδρόμησης είναι η εξής:

$$\widehat{Y_t} = -232719.28 + 4.35 * X_t$$

(iii) Ο συντελεστής προσδιορισμού υπολογίζεται ως εξής:

$$R^2 = \widehat{\beta}_1 \frac{\sum_{t=1}^{n} XY - \frac{\sum_{t=1}^{n} X \sum_{t=1}^{n} Y}{n}}{\sum_{t=1}^{n} Y^2 - \frac{\left(\sum_{t=1}^{n} Y\right)^2}{n}} = 4.35 \frac{308374362232.7 - \frac{1490332.4 * 2992120}{15}}{661589819198^2 - \frac{2992120^2}{15}}$$

$= 0.75$

Ο συντελεστής προσδιορισμού παρατηρούμε ότι λαμβάνει υψηλή τιμή, $R^2 = 0.75$, γεγονός που σημαίνει ότι το 75% της μεταβλητότητας των τιμών της εξαρτημένης μεταβλητής (δηλ. του αριθμού των πωλήσεων) ερμηνεύεται απ' τη μεταβλητότητα των τιμών της ανεξάτητης μεταβλητής(δηλ. των τιμών του ΑΕΠ).

(iv) Ο εκτιμητής της διασποράς των σφαλμάτων είναι ίσος με:

$$S^2 = \frac{\sum_{t=1}^{n} Y^2 - \frac{\left(\sum_{t=1}^{n} Y\right)^2}{n} - \widehat{\beta}_1 \left[\sum_{t=1}^{n} XY - \frac{\sum_{t=1}^{n} X \sum_{t=1}^{n} Y}{n}\right]}{n-2}$$
$$= 1268689034.95$$

Συνεπώς, το τυπικό σφάλμα της εκτίμησης είναι:

$$s = \sqrt{S^2} = 35618.66$$

(v) Το 95% διάστημα εμπιστοσύνης για την παράμετρο β_0 είναι ίσο με:

$$\widehat{\beta}_0 - t_{\alpha/2,n-2} s_{\widehat{\beta}_0} \leqslant \beta_0 \leqslant \widehat{\beta}_0 + t_{\alpha/2,n-2} s_{\widehat{\beta}_0}$$

όπου $s_{\widehat{\beta}_0}$ το τυπικό σφάλμα της εκτιμήτριας $\widehat{\beta}_0$ του συντελεστή β_0. Αλλά

$$s_{\widehat{\beta}_0} = s\sqrt{\frac{\sum_{t=1}^{n} X^2}{n\sum_{t=1}^{n}\left(X - \overline{X}\right)^2}} = 35618.66\sqrt{\frac{150622335274.98}{15 * 2549624442.33}}$$
$$= 70686.84$$

γιατί

$$\sum_{t=1}^{n}\left(X - \overline{X}\right)^2 = \sum_{t=1}^{n} X^2 - \frac{\left(\sum_{t=1}^{n} X\right)^2}{n} = 150622335274.98 - \frac{1490332.4^2}{15}$$
$$= 2549624442.33$$

και $t_{\alpha/2,n-2} = t_{0.05/2,15-2} = t_{0.025,13} = 2.16$, οπότε το 95% διάστημα εμπιστοσύνης για την παράμετρο β_0 είναι ίσο με:

$$\widehat{\beta}_0 - t_{\alpha/2,n-2}s_{\widehat{\beta}_0} \leqslant \beta_0 \leqslant \widehat{\beta}_0 + t_{\alpha/2,n-2}s_{\widehat{\beta}_0}$$
$$\Rightarrow -232719.28 - 2.16 * 70686.84 \leqslant \beta_0 \leqslant -232719.28 + 2.16 * 70686.84$$

Άρα,

$$-385402.85 \leqslant \beta_0 \leqslant -80035.72$$

Το 95% διάστημα εμπιστοσύνης για την παράμετρο β_1 είναι ίσο με:

$$\widehat{\beta}_1 - t_{\alpha/2,n-2}s_{\widehat{\beta}_1} \leqslant \beta_1 \leqslant \widehat{\beta}_1 + t_{\alpha/2,n-2}s_{\widehat{\beta}_1}$$

όπου $s_{\widehat{\beta}_1}$ το τυπικό σφάλμα της εκτιμήτριας $\widehat{\beta}_1$ του συντελεστή β_1. Αλλά

$$s_{\widehat{\beta}_1} = \frac{s}{\sqrt{\sum_{t=1}^{n}\left(X - \overline{X}\right)^2}} = \frac{35618.66}{\sqrt{2549624442.33}} = 0.71$$

οπότε το 95% διάστημα εμπιστοσύνης για την παράμετρο β_1 είναι ίσο με:

$$\widehat{\beta}_1 - t_{\alpha/2,n-2}s_{\widehat{\beta}_1} \leqslant \beta_1 \leqslant \widehat{\beta}_1 + t_{\alpha/2,n-2}s_{\widehat{\beta}_1}$$
$$\Rightarrow 4.35 - 2.16 * 0.71 \leqslant \beta_1 \leqslant 4.35 + 2.16 * 0.71$$

Άρα,

$$2.83 \leqslant \beta_1 \leqslant 5.87$$

(vi) Για τα έτη 2005 και 2006 το ΑΕΠ αναμένεται να λάβει τις τιμές 130373.1 και 134370.2 αντιστοίχως. Δηλαδή, $X_{2005} = 130373.1$ και $X_{2006} = 134370.2$ Συνεπώς, χρησιμοποιώντας τη γραμμή παλινδρομήσεως του δείγματος, δηλ. $\widehat{Y_F} = \widehat{\beta}_0 + \widehat{\beta}_1 X_F$, με X_F η γνωστή τιμή του ΑΕΠ για το αντίστοιχο έτος, αναμένεται ο αριθμός των πωλήσεων των αυτοκινήτων (σημειακή εκτίμηση) για τα έτη 2005 και 2006 να είναι:

$$\widehat{Y}_{2005} = \widehat{\beta}_0 + \widehat{\beta}_1 X_{2005} = -232719.28 + 4.35 * 130373.1 = 334400.50$$
$$\widehat{Y}_{2006} = \widehat{\beta}_0 + \widehat{\beta}_1 X_{2006} = -232719.28 + 4.35 * 134370.2 = 351787.78$$

Το 95% διάστημα εμπιστοσύνης για τον αριθμό των πωλήσεων των αυτοκινήτων είναι ίσο με:

$$\widehat{Y}_F - t_{\alpha/2,n-2}s_F \leqslant Y_F \leqslant \widehat{Y}_F + t_{\alpha/2,n-2}s_F$$

όπου

$$s_F = s\sqrt{1 + \frac{1}{n} + \frac{\left(X_F - \overline{X}\right)^2}{\sum\left(X - \overline{X}\right)^2}}$$

Συνεπώς,

Για το 2005:

$$s_{2005} = 35618.66\sqrt{1 + \frac{1}{15} + \frac{(130373.1 - 99355)^2}{2549624442.33}} = 42802.10$$

$$\widehat{Y}_{2005} - t_{\alpha/2,n-2}s_{2005} \leqslant Y_{2005} \leqslant \widehat{Y}_{2005} + t_{\alpha/2,n-2}s_{2005}$$

$$\Rightarrow 334400.50 - 2.16 * 42802.10 \leqslant Y_{2005} \leqslant 334400.50 + 2.16 * 42802.10$$

Άρα,

$$241947.99 \leqslant Y_{2005} \leqslant 426853.02$$

Δηλαδή είμαστε 95% σίγουροι ότι ο αριθμός των πωλήσεων για το έτος 2005, θα είναι από 241947.99 μέχρι 426853.02

Για το 2006:

$$s_{2006} = 35618.66\sqrt{1 + \frac{1}{15} + \frac{(134370.2 - 99355)^2}{2549624442.33}} = 44309.77$$

$$\widehat{Y}_{2006} - t_{\alpha/2,n-2}s_{2006} \leqslant Y_{2006} \leqslant \widehat{Y}_{2006} + t_{\alpha/2,n-2}s_{2006}$$

$$\Rightarrow 351787.78 - 2.16 * 44309.77 \leqslant Y_{2006} \leqslant 351787.78 + 2.16 * 44309.77$$

Άρα,

$$256078.67 \leqslant Y_{2006} \leqslant 447496.89$$

Δηλαδή είμαστε 95% σίγουροι ότι ο αριθμός των πωλήσεων για το έτος 2006, θα είναι από 256078.67 μέχρι 447496.89

3.3 ΣΥΜΠΕΡΑΣΜΑΤΑ

Απ' την εφαρμογή μας μπορούμε να εξάγουμε τα εξής συμπεράσματα.

Απ'τη σύγκριση της μεθόδου Διάσπασης Χρονοσειρών και της μεθόδου Winters έχουμε τους εξής πίνακες αποτελεσμάτων. Όπου:

- Y_t είναι οι πραγματικές τιμές των πωλήσεων των αυτοκινήτων στην Ελλάδα για το έτος 2005
- $\widehat{Y}_t$ είναι οι προσεγγιστικές τιμές απ' την εφαρμογή των μεθόδων
- $|e_t| = |Y_t - \widehat{Y}_t|$ είναι το απόλυτο σφάλμα
- $\frac{|e_t|}{Y_t}$ είναι το απόλυτο ποσοστιαίο σχετικό σφάλμα

Πίνακας 3.9: Αποτελέσματα της μεθόδου Διάσπασης Χρονοσειρών.

Έτος	Τριμ	Y_t	$\widehat{Y}_t$	$\lvert e_t\rvert$	$\frac{\lvert e_t\rvert}{Y_t}$
2005	1ο	77.838	74.395	3.443	4,42 %
	2ο	75.607	77.537	1.930	2,55 %
	3ο	66.615	66.069	546	0,82 %
	4ο	49.670	52.609	2.939	5,91 %
Σύνολο					13,7 %

Πίνακας 3.10: Αποτελέσματα της μεθόδου Winters.

Έτος	Τριμ	Y_t	$\widehat{Y}_t$	$\lvert e_t\rvert$	$\frac{\lvert e_t\rvert}{Y_t}$
2005	1ο	77.838	81.115	3.277	4,21 %
	2ο	75.607	83.404	7.797	10,31 %
	3ο	66.615	69.558	2.943	4,41 %
	4ο	49.670	55.556	5.886	11,85 %
Σύνολο					30,78 %

Υπολογίζοντας το μέσο απόλυτο ποσοστιαίο σφάλμα (MAPE) για κάθε μέθοδο παίρνουμε τα εξής:

Μέθοδος Διάσπασης Χρονοσειρών:

$$\mathsf{MAPE} = \frac{1}{n}\sum_{t=1}^{n}\frac{|e_t|}{Y_t} = \frac{1}{4}\,13,7\% = 3,42\%$$

Μέθοδος Winters:

$$\mathsf{MAPE} = \frac{1}{n}\sum_{t=1}^{n}\frac{|e_t|}{Y_t} = \frac{1}{4}\,30,78\% = 7,7\%$$

Παρατηρούμε λοιπόν, ότι το μέσο απόλυτο ποσοστιαίο σφάλμα της μεθόδου Διάσπασης Χρονοσειρών είναι μικρότερο από το μέσο απόλυτο ποσοστιαίο σφάλμα της μεθόδου Winters. Συνεπώς, το συμπέρασμα στο οποίο καταλήγουμε είναι ότι η μέθοδος Διάσπασης χρονοσειρών έχει μεγαλύτερη ακρίβεια απ' τη μέθοδο Winters για τη συγκεκριμένη εφαρμογή.

Ωστόσο στην παρούσα εργασία είδαμε και τις προβλέψεις για τις τιμές των πωλήσεων των αυτοκινήτων στην Ελλάδα για τα έτη 2005 και 2006, εφαρμόζοντας τη μέθοδο Διάσπασης Χρονοσειρών. Παρατηρώντας λοιπόν, τις πραγματικές τιμές των πωλήσεων έχουμε τον εξής πίνακα αποτελεσμάτων.

Πίνακας 3.11: Αποτελέσματα της μεθόδου Διάσπασης Χρονοσειρών για τα έτη 2005 και 2006.

Έτος	Τριμ	Y_t	$\widehat{Y}_t$	$\lvert e_t\rvert$	$\frac{\lvert e_t\rvert}{Y_t}$
2005	1ο	77.838	74.395	3.443	4,42 %
	2ο	75.607	77.537	1.930	2,55 %
	3ο	66.615	66.069	546	0,82 %
	4ο	49.670	52.609	2.939	5,91 %
2006	1ο	76.059	73.799	2.260	2,97 %
	2ο	76.603	76.915	312	0,41 %
	3ο	64.269	65.538	1.269	1,97 %
	4ο	50.742	52.186	1.444	2,84 %
Σύνολο					21,89 %

Και υπολογίζοντας το μέσο απόλυτο ποσοστιαίο σφάλμα MAPE της μεθόδου παίρνουμε:

$$\mathsf{MAPE} = \frac{1}{n}\sum_{t=1}^{n}\frac{|e_t|}{Y_t} = \frac{1}{8}\,21,89\% = 2,73\%$$

Συνεπώς, παρατηρούμε ότι η μέθοδος αυτή μας έδωσε αρκετά καλά και αξιόπιστα αποτελέσματα για τις τιμές των πωλήσεων των αυτοκινήτων για τα έτη 2005 και 2006.

3.4 ΠΑΡΑΡΤΗΜΑ

Ο κώδικας σε matlab

```
bb=load('sales.txt')
time=bb(:,1);
yt=bb(:,2);
plot(time,yt, 'LineWidth', 2)
xlabel('Time')
ylabel('Sales x1000')
n=length(time);
% moving average
m=4;
movave=[];
ma=0;
for i=1:n-m+1
for j=i:i+m-1
ma=ma+ yt(j);
end
ma=ma/m;
movave=[movave;ma];
ma=0;
end
disp('Oi kinhtoi mesoi oroi einai:')
movave
% centered moving average
l=length(movave);
cma=0;
cenmovave=[];
i=0;
for i=1:l-1
cma=(movave(i)+movave(i+1))/2;
cenmovave=[cenmovave;cma];
end
disp('Oi kentrikoi kinhtoi mesoi einai:')
cenmovave
si=[];
k=0;
a=0;
for k=1+m/2:n-m/2
a= yt(k)./cenmovave(k-m/2);
```

```
si=[si;a];
end
disp('Oi epoxikoi deiktes einai oi ekshs:')
si
sa=[];
i=0;
k=0;
j=0;
sum1=0;
for j=1:m
for i=0:(m-1)
for k=j+4*i
sum1=sum1+si(k);
end
end
sum1=sum1/(m+1);
sa=[sa;sum1];
sum1=0;
end
sum2=0;
disp('To athroisma ston pinaka me tous prosarmosmenous deiktes einai to ekshs:')
sum2=sum(sa)
i=0;
l=length(sa)
for i=1:l
sa(i)=sa(i)*l/sum2;
end
sa
i=0;
k=0;
j=0;
sayt=[];
sa=[sa(m-1);sa(m);sa(m-3);sa(m-2)];
disp('edw kanei anadiataksh twn timwn sa')
sa
for j=1:m
for i=0:m
for k=j+4*i
sayt(k)=yt(k)/sa(j);
end
end
```

```
end
sayt=sayt.' ;
disp('Oi prosarmosmenoi epoxikoi deiktes einai oi ekshs:')
sa
disp('Oi apallagmenes apo epoxikothta times einai oi ekshs:')
sayt
% plots
i=0;
t=[];
for i=1:n-m
t(i)=time(i+m/2);
end
figure
plot(time,yt,'k-',t,cenmovave,'b-o','LineWidth', 2)
legend('Y(t)','CAt')
xlabel('T I M E ');
ylabel('S A L E S x1000');
figure
plot(time,yt,'k-',time,sayt,'b-o', 'LineWidth', 2)
legend('Y(t)','SAYt')
xlabel('T I M E ');
ylabel('S A L E S x1000');
%======================================
%Least Squares
a=0;
b=0;
tt=[];
wt=[];
x2=time.*time;
xy=time.*sayt;
sx=sum(time);
sy=sum(sayt);
sx2=sum(x2);
sxy=sum(xy);
b=(n*sxy-sx*sy)/(n*sx2-sx*sx);
a=(sy/n)-b*(sx/n);
yls=a+b*time;
Tt=yls;
figure
plot(time,sayt,'b-',time,yls,'k-', 'LineWidth', 2)
legend('SAYt','Tt')
```

```
xlabel(’T I M E ’);
ylabel(’S A L E S x1000’);
tayt=sayt./Tt;
figure
plot(time,tayt,’b-’,time,1,’k–’, ’LineWidth’, 2)
legend(’TAYt’)
xlabel(’T I M E ’);
ylabel(’S A L E S’);
i=0;
w=0;
wat=[];
for i=2:n-1
w=0.25*tayt(i-1)+0.5*tayt(i)+0.25*tayt(i+1);
wat=[wat;w];
end
for i=1:n-2
tt(i)=time(i);
end
figure
plot(tt,wat,’b-’,time,tayt,’k-’,time,1,’k–’, ’LineWidth’, 2)
legend(’WAt’,’TAYt’)
xlabel(’T I M E ’);
ylabel(’S A L E S’);
% prediction
ytpred=[];
tpred=[21;22;23;24;25;26;27;28];
f=length(tpred);
i=0;
j=0;k=0;
for j=1:m
for i=0:1
for k=j+4*i
ytpred(k)=(a+b*tpred(k))*sa(j)
end
end
end
figure
plot(time,yt,’k-’,tpred,ytpred,’b–’, ’LineWidth’, 2)
legend(’Y(t)’,’Yforecast(t)’)
xlabel(’T I M E ’);
ylabel(’S A L E S x1000’);
```

Βιβλιογραφία

[1] Αγιακλόγλου Χ.Ν. και Οικονόμου Γ.Σ., *Μέθοδοι προβλέψεων και ανάλυσης αποφάσεων*, Μπένου, Αθήνα, 2004.

[2] Δάρας Ι.Τ. και Σύψας Θ.Π., *Στοχαστικές Ανελίξεις*, Ζήτη, Θεσσαλονίκη, 2003.

[3] Δάρας Ι.Τ. και Σύψας Θ.Π., *Πιθανότητες και Στατιστική*, Ζήτη, Θεσσαλονίκη, 2010.

[4] Κιντής Α.Α., *Στατιστικές και Οικονομετρικές Μέθοδοι*, Gutenberg, Αθήνα, 1999.

[5] Κουγιουμτζής Δ. και Τσιμπίρης Α. , Measures of Analysis of Time Series, *Journal of Statistical Software*, 33(5), 2010.

[6] Λαζαρίδης Α., *Οικονομετρία ΙΙ*, Ζυγός, Θεσσαλονίκη, 2000.

[7] Μαργιά Γ., *Ανάλυση και Πρόβλεψη Χρονοσειρών*, Αριστοτέλειο Πανεπιστήμιο Θεσσαλονίκης, 2009.

[8] Παναγιώτου Σ.Γ., *Προβλέψεις πωλήσεων των Ι.Χ. αυτοκινήτων σε δεκαπέντε χώρες – μέλη της Ευρωπαϊκής Ένωσης*, Πανεπιστήμιο Πειραιώς, 2005.

[9] Πέκος Γ.Δ., *Ασκήσεις Στατιστικής*, Ζυγός, Έκδοση Ε', Θεσσαλονίκη, 1999.

[10] Χρήστου Γ.Κ., *Εισαγωγή στην Οικονομετρία*, Gutenberg, Εκδοση Β', Αθήνα, 2005.

[11] Ψωινός Δ.Π., *Ποσοτική Ανάλυση*, Ζήτη, Θεσσαλονίκη, 1993.

[12] Brockwell P.J. and Davis R.A., *Introduction to time series and forecasting*, Springer, 2002.

[13] Jarret J., *Μέθοδοι Προβλέψεων για Οικονομικές - Επιχειρηματικές Αποφάσεις*, Gutenberg, Αθήνα, 1993.

[14] Kaiser O. , Igdalov D. and Horenko I., Statistical Regression Analysis of Threshold Excesses with Systematically Missing Covariates, *Multiscale Modeling & Simulation*, 13(2), 594–613, 2015.

[15] Montgomery D.C. , Jennings C.L. and Kulahci M., *Introduction to Time Series Analysis and Forecasting*, John Wiley & Sons, New Jersey, 2008.

[16] Spiegel M.R. and Stephens L.J., *Στατιστική*, Τζιόλα, Θεσσαλονίκη, 2000.