ΠΟΛΥΤΕΧΝΕΙΟ ΚΡΗΤΗΣ
ΓΕΝΙΚΟ ΤΜΗΜΑ

ΠΡΟΓΡΑΜΜΑ ΜΕΤΑΠΤΥΧΙΑΚΩΝ ΣΠΟΥΔΩΝ
ΕΦΑΡΜΟΣΜΕΝΕΣ ΕΠΙΣΤΗΜΕΣ ΚΑΙ ΤΕΧΝΟΛΟΓΙΑ

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

ΕΦΑΡΜΟΓΗ ΤΗΣ ΜΕΘΟΔΟΥ ΧΡΟΝΟΣΕΙΡΩΝ ΣΤΗΝ ΠΡΟΒΛΕΨΗ ΠΑΡΑΓΩΓΗΣ ΠΕΤΡΕΛΑΙΟΥ

ΠΑΠΑΔΟΠΟΥΛΟΥ ΓΕΩΡΓΙΑ

Επιβλέπων: Επίκουρος Καθηγητής **Δάρας Τρύφων**

ΧΑΝΙΑ, 2015

Περιεχόμενα

ΕΙΣΑΓΩΓΗ

Εδώ η Γεωργία θα γράψει την εισαγωγή της. Στην πορεία θα συζητήσουμε την δομή της εισαγωγής [3]. Πρέπει η εισαγωγή να έχει παρόμοιο θεματικό σκελετό με την [2, 1] περίληψη και τα συμπεράσματα στο τέλος της εργασίας.

Το αντικείμενο είναι λοιπόν η ανάλυση χρονοσειρών, δηλαδή η χρήση μεθόδων που θα μας επιτρέψουν να διερευνήσουμε το μηχανισμό (στοχαστική διαδικασία ή δυναμικό σύστημα) που παράγει τη χρονοσειρά, να εκτιμήσουμε χαρακτηριστικά του, να αναπτύξουμε μοντέλο για να το περιγράψουμε και να κάνουμε προβλέψεις της εξέλιξης του, δηλαδή τις επόμενες τιμές στη χρονοσειρά.

Γ. Παπαδοπούλου, Χανιά 2015.

Μέρος I

ΑΝΑΛΥΣΗ ΧΡΟΝΟΣΕΙΡΩΝ

Κεφάλαιο 1

ΒΑΣΙΚΕΣ ΕΝΝΟΙΕΣ ΧΡΟΝΟΣΕΙΡΑΣ

1.1 Η ΕΝΝΟΙΑ ΤΗΣ ΧΡΟΝΟΣΕΙΡΑΣ

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 $t \geq 0$.

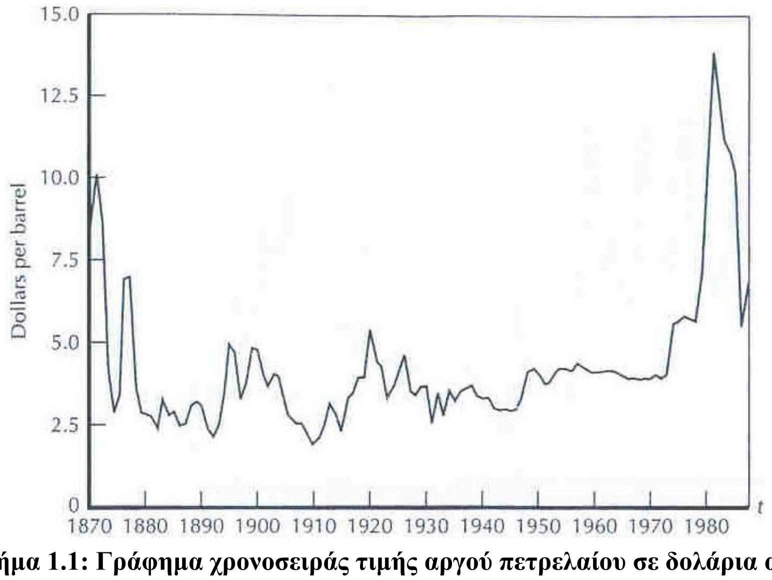
(iv) Η ημερήσια κατανάλωση ηλεκτρικού ρεύματος καθώς και η ημερήσια κατανάλωση ύδατος, x_t και y_t αντίστοιχα, σε μια μεγάλη γεωγραφική περιοχή της χώρας με $t = 1, 2, \ldots$

(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 και y_t αντίστοιχα, σε συγκεκριμένη γεωγραφική περιοχή με γεωγραφικές συντεταγμένες (l, a, h) κατά την χρονική στιγμή t. Εδώ η χρησιμοποιούμενη παράμετρος t είναι περισσότερο σύνθετη και συγκεκριμένα $t = (l, a, h, t)$.

Όπως διαπιστώνει κανείς από τα παραπάνω παραδείγματα, οι χρονοσειρές μπορούν να αφορούν διακριτά μεγέθη x_t σε διακριτό χρόνο t, περίπτωση (i), διακριτά μεγέθη x_t σε συνεχή χρόνο t, περιπτώσεις (ii) και (iii), συνεχή μεγέθη x_t σε διακριτό χρόνο t, περιπτώσεις (iv) και (v) και συνεχή μεγέθη x_t σε συνεχή χρόνο t, περίπτωση (vi). Το πρόβλημα είναι η "πρόβλεψη" μελλοντικών τιμών της χρονοσειράς με βάση τις μέχρι σήμερα τιμές τις ίδιας χρονοσειράς, περιπτώσεις (i)-(iii), είτε ακόμα και σε συνδυασμό με τις μέχρι σήμερα τιμές μιας άλλης χρονοσειράς η οποία εξελίσσεται παράλληλα με την πρώτη και επιδρά πάνω σ΄ αυτή, περιπτώσεις (iv)-(vi), οπότε μιλάμε για πολυμεταβλητές χρονοσειρές. Το σύνολο των δυνατών καταστάσεων ονομάζεται χώρος καταστάσεων και συμβολίζεται με S, ένα (μονοδιάστατο) υποσύνολο του R , ή γενικότερα ένα πολυδιάστατο υποσύνολο του R^d, ενώ το σύνολο τιμών του t ονομάζεται παραμετρικός χώρος, συμβολίζεται με T και μπορεί επίσης να είναι υποσύνολο του R^k, όταν χρειάζεται ένα πολυδιάστατο t για να καθορίσουμε πέραν του χρόνου t και γεωγραφικές π.χ. συντεταγμένες (l, a, h) σε χωρο-χρονοσειρές (spatial time series), βλ. παράδειγμα (vi) παραπάνω. Σημειώνεται οι όροι διακριτά και συνεχή μεγέθη είναι σε αντιστοιχία με τους όρους διακριτές και συνεχείς τυχαίες μεταβλητές.

Ένα παράδειγμα μίας χρονοσειράς απεικονίζεται στο ακόλουθο σχήμα 1.1, όπου παρουσιάζεται το γράφημα της ετήσιας τιμής του αργού πετρελαίου σε αμερικανικά δολάρια ανά βαρέλι. Ο χρονικός ορίζοντας είναι 117 χρόνια, από το έτος 1870 έως το έτος 1987. Το γράφημα ελήφθη από το βιβλίο των R. Pindyck και D. Rubinfeld (1998).



Σχήμα 1.1: Γράφημα χρονοσειράς τιμής αργού πετρελαίου σε δολάρια ανά βαρέλι.

Η συστηματική μελέτη μιας χρονοσειράς ξεκινάει με την επισκόπηση του γραφήματός της στο πεδίο του χρόνου, από το οποίο μπορούν αρχικά να ανιχνευθούν τρία βασικά ποιοτικά χαρακτηριστικά της: Η τάση, η εποχικότητα και οι ακραίες παρατηρήσεις.

- Η **τάση** (trend) γενικά θα μπορούσε να ορισθεί ως η μακροπρόθεσμη μεταβολή του μέσου επιπέδου των τιμών μιας χρονοσειράς. Έτσι, μπορεί η τάση των τιμών να είναι αυξητική, πτωτική ή σταθερή σε ένα συγκεκριμένο χρονικό διάστημα, ενώ μπορεί και να έχει τη μορφή κάποιας συνάρτησης στο εν λόγω διάστημα. Να σημειωθεί ότι η έννοια “μακροπρόθεσμη μεταβολή” εξαρτάται από την εκάστοτε εφαρμογή που εξετάζεται.

- Η **εποχικότητα** (seasonal) μπορεί να ορισθεί σαν μια περιοδική διακύμανση που έχει σταθερό μήκος. Η εν λόγω διακύμανση τις περισσότερες φορές διακρίνεται εύκολα και μπορεί να ερμηνευθεί στα πλαίσια του υπό μελέτη φαινομένου. Φερ' ειπείν, αν κανείς επιθυμούσε να αναλύσει τη χρονοσειρά των τιμών των καυσίμων σε βάθος χρόνων, είναι λογικό να περιμένει να παρατηρήσει μια σχετική άνοδο κατά τους χειμερινούς μήνες κάθε έτους.

- Οι **ακραίες παρατηρήσεις** (outliers) είναι οι απομονωμένες παρατηρήσεις που εμφανίζονται στο γράφημα κάποιας χρονοσειράς ως απότομες αλλαγές στο πρότυπο συμπεριφοράς της. Τα outliers μελετήθηκαν αρχικά κατά κύριο λόγο από τον A. J. Fox (1972), ο οποίος μάλιστα εισήγαγε δύο τύπους. Ο τύπος I αφορά περιπτώσεις όπου η ύπαρξη μιας ακραίας τιμής δεν έχει καμία επίδραση στις ακόλουθες παρατηρήσεις. Αντιθέτως, ο τύπος II αφορά περιπτώσεις όπου υπάρχει επίδραση στη μετέπειτα συμπεριφορά των τιμών της χρονοσειράς, παράγοντας μια σειρά λιγότερο ή περισσότερο ακραίων παρατηρήσεων ή αλλάζοντας εξ' ολοκλήρου τα χαρακτηριστικά της.

1.2 ΣΤΟΧΑΣΤΙΚΕΣ ΑΝΕΛΙΞΕΙΣ

Όπως ήδη αναφέραμε, χρονοσειρά είναι μια ακολουθία τιμών $\{x_t : t = 0, 1, 2, \dots\}$ η οποία εκφράζει την εξέλιξη ενός στοχαστικού συστήματος, ενός συστήματος δηλαδή με τυχαία κατά μάλλον ή ήττον συμπεριφορά, σε αντίθεση με αυτήν ενός προσδιοριστικού (deterministic) συστήματος, η οποία συνήθως περιγράφεται από ένα σύστημα διαφορικών εξισώσεων. Όπως είναι γνωστό από την Θεωρία των Στοχαστικών Ανελίξεων, η εξέλιξη ενός στοχαστικού συστήματος μπορεί να περιγραφεί πιθανοθεωρητικά μέσω μιας στοχαστικής ανέλιξης (σ.α.), μονοδιάστατης ή πολυδιάστατης, δηλαδή μιας ακολουθίας τυχαίων μεταβλητών $\{X_n : n \in N\}$, ή γενικότερα μιας (υπεραριθμήσιμης) οικογένειας τυχαίων μεταβλητών $\{X_t : t \in T\}$, που ορίζεται πάνω σε ένα χώρο πιθανότητας (Ω, F, P) .

Σημειώνεται ότι για δεδομένο $\omega \in \Omega$, η $\{x_t = X_t(\omega) : t \in T\}$ εκφράζει μια συνάρτηση του χρόνου t και αποτελεί την τροχιά ή πραγματοποίηση (sample path, realization) της σ.α. $\{X_t : t \in T\}$. Συνεπώς, ο όρος χρονοσειρά χρησιμοποιείται τόσο για μια στοχαστική ανέλιξη $\{X_t : t \in T\}$, όσο και για μία τροχιά $\{x_t = X_t(\omega) : t \in T\}$.

1.3 ΚΑΤΑΝΟΜΕΣ ΚΑΙ ΡΟΠΕΣ ΣΤΟΧΑΣΤΙΚΗΣ ΔΙΑΔΙΚΑΣΙΑΣ

Η ακολουθία των τυχαίων μεταβλητών Y_t για κάθε χρονική στιγμή t ορίζει τη στοχαστική διαδικασία $\{Y_t\}_{t=-\infty}^{+\infty}$(και αντίστοιχα για τη X_t και $\{X_t\}_{t=-\infty}^{+\infty}$). Θα αναφερόμαστε στη στοχαστική διαδικασία και ως χρονοσειρά εννοώντας την άγνωστη ακολουθία των τυχαίων μεταβλητών και όχι τις παρατηρήσεις.

Η πλήρης περιγραφή μιας στοχαστικής διαδικασίας $\{Y_t\}_{t=-\infty}^{+\infty}$ απαιτεί ότι οι κοινές κατανομές όλων των τάξεων είναι γνωστές για κάθε χρονική στιγμή t. Η κατανομή τάξης ένα αντιστοιχεί στη στατική περιγραφή της στοχαστικής διαδικασίας και είναι η (περιθώρια) κατανομή της $\{Y_t\}_{t=-\infty}^{+\infty}$

$$\forall t \in Z, \quad f_{Y_t}(y) = f_Y(y,t),$$

δηλαδή ορίζεται ως συνάρτηση όχι μόνο της κάθε τιμής y αλλά και του χρόνου t. Κατά τον ίδιο τρόπο η κοινή κατανομή δύο μεταβλητών της $\{Y_t\}_{t=-\infty}^{+\infty}$ (κατανομή τάξης 2) είναι

$$\forall t_1, t_2 \in Z, \quad f_{Y_{t1},Y_{t2}}(y_1, y_2) = f_Y(y_1, y_2, t_1, t_2)$$

.

Δεδομένου ότι μια χρονοσειρά αποτελεί μια στοχαστική διαδικασία, δηλαδή κάθε παρατήρησή της αποτελεί μια τυχαία μεταβλητή, υποθέτουμε ότι μελετούμε την χρονική εξέλιξη των τιμών ενός μεγέθους . Έτσι, την χρονική στιγμή t η τιμή του μεγέθους θα είναι μια τυχαία μεταβλητή Y_t με συνάρτηση πυκνότητας πιθανότητας $f_{Y_t}(y_t)$. Έστω ακόμη ότι έχουμε παρατηρήσει ένα δείγμα τιμών μεγέθους T, δηλαδή έχουμε συλλέξει τις τιμές για $t = 1, 2, \ldots, T$

$$\{Y_1, Y_2, \ldots, Y_T\}$$

Το παραπάνω δείγμα τιμών αναπαριστά μια συγκεκριμένη "έκβαση" της στοχαστικής διαδικασίας που γεννά τα δεδομένα. Αν τώρα η εν λόγω στοχαστική διαδικασία Υ επαναληφθεί Ν ανεξάρτητες φορές, θα έχουμε ένα σύνολο Ν "πραγματοποιήσεων" όλων των παρατηρήσεων – τυχαίων μεταβλητών $_t$ που συνθέτουν τη χρονοσειρά. Συνεπώς, όσον αφορά το παραπάνω δείγμα τιμών μεγέθους Τ, για την κάθε μία παρατήρηση Y_t θα έχουμε το εξής σύνολο

$$\{Y_t^{(1)}, Y_t^{(2)}, \ldots, Y_t^{(N)}\}, \quad \forall t = 1, 2, \ldots, T$$

Η μέση τιμή (ροπή πρώτης τάξης) ή αναμενόμενη τιμή μ_t της t–οστής παρατήρησης (τυχαίας μεταβλητής) Y_t της παραπάνω χρονοσειράς Υ δίδεται από τη σχέση:

$$\mu = E[Y_t] = \int_{-\infty}^{+\infty} y_t f_{Y_t}(y_t)\, dy$$

Στο σημείο αυτό θα πρέπει να αναφέρουμε ότι η μέση τιμή μ_t σχετίζεται άμεσα με την έννοια της *τάσης* της χρονοσειράς, εφόσον εκφράζεται ως συνάρτηση της χρονικής στιγμής t της παρατήρησης Y_t . Συγκεκριμένα, αν μια χρονοσειρά παρουσιάζει αυξητική ή πτωτική τάση αντιστοίχως σε ένα χρονικό διάστημα, αυτό θα αποτυπώνεται και στη μέση τιμή της μ_t ως συνάρτηση του χρόνου. Ακόμη, η μη ύπαρξη τάσης σε ένα χρονικό διάστημα αποτυπώνεται στη σταθερή αναμενόμενη τιμή μ_t στο εν λόγω διάστημα.

Η διασπορά:

$$\sigma^2(t) = V[X_t] = E\left[(X_t - \mu_t)^2\right], \quad t \in T$$

Η αυτοσυνδιακύμανση:

$$\gamma(t,h) = Cov(X_t, X_{t+h}) = E[(X_t - \mu_t)(X_{t+h} - \mu_{t+h})], \quad t, h \in T$$

Είναι προφανές ότι ισχύει: $\sigma^2(t) = \gamma(t, 0) = Cov(X_t, X_t), \quad t, h \in T$
Είναι γνωστό επίσης ότι όταν οι διασπορές δύο τυχαίων μεταβλητών Χ και Υ , είναι πεπερασμένες, τότε τόσο οι μέσες τιμές αυτών όσο και η συνδιακύμανση αυτών είναι

πεπερασμένες ποσότητες. Τούτο διότι αφενός πρέπει να έχουμε
$E\left[|X|^2\right] < \infty, E\left[|X|^2\right] < \infty$, αφού διαφορετικά δεν ορίζονται οι διασπορές, και αφετέρου, με εφαρμογή της ανισότητας Cauchy-Schwarz

$$E\left[|X||Y|\right] \leqslant \sqrt{E\left[X^2\right] E\left[Y^2\right]},$$

προκύπτει:

$$|E[X]| \leqslant E[|X|] \leqslant \sqrt{E[X^2]} < \infty, \quad |E[Y]| \leqslant E[|Y|] \leqslant \sqrt{E[Y^2]} < \infty,$$

και

$$|Cov(X,Y)| \leqslant \sqrt{V[X]\,V[Y]} < \infty.$$

1.4 ΣΤΑΣΙΜΟΤΗΤΑ ΚΑΙ ΑΥΤΟΣΥΣΧΕΤΗΣΗ

Η στατιστική περιγραφή της στοχαστικής διαδικασίας απλουστεύεται αν θεωρήσουμε ότι οι στατιστικές της ιδιότητες παραμένουν σταθερές στο χρόνο και τότε η στοχαστική διαδικασία ορίζεται ως στάσιμη. Αυτή είναι μα υπόθεση που δύσκολα μπορεί να υιοθετηθεί σε πολλά πραγματικά προβλήματα, αλλά μπορεί να χρησιμοποιηθεί ως υπόθεση εργασίας για την εξαγωγή χρήσιμων συμπερασμάτων.

Ειδικότερα ορίζονται δύο μορφές στασιμότητας:

☞ **Αυστηρή Στασιμότητα**

Η στοχαστική διαδικασία $\{Y_t\}_{t=-\infty}^{+\infty}$ είναι αυστηρά στάσιμη όταν οι κατανομές της για κάθε τάξη (ή ισοδύναμα όλες οι ροπές) είναι σταθερές στο χρόνο, δηλαδή όταν ισχύει:

$$\forall t \in Z, \quad f_{Y_t}(y) = f_Y(y,t) = f_Y(y)$$
$$\forall t_1, t_2 \in Z, \quad f_{Y_{t1},Y_{t2}}(y_1,y_2) = f_{Y_t,Y_{t-\tau}}(y_1,y_2)$$
$$\forall t_1, t_2, t_3 \in Z, \quad f_{Y_{t1},Y_{t2},Y_{t3}}(y_1,y_2,y_3) = f_{Y_t,Y_{t-\tau1},Y_{t-\tau2}}(y_1,y_2,y_3)$$

και αντίστοιχα για κατανομές μεγαλύτερης τάξης.

Για ροπές τάξης μεγαλύτερης του ένα, οι κατανομές δίνονται ως συνάρτηση όχι των χρονικών στιγμών, π.χ. t_1, t_2 , αλλά της υστέρησης μεταξύ των χρονικών στιγμών, π.χ. $\tau = t_2 - t_1$, δηλαδή για οποιεσδήποτε δύο χρονικές στιγμές που απέχουν μεταξύ τους τ χρονικά βήματα. Ο έλεγχος της αυστηρής στασιμότητας απαιτεί τη διερεύνηση κοινών κατανομών ή ροπών όλων των τάξεων και δεν αποτελεί μια πρακτικά χρήσιμη ιδιότητα. Για αυτό συχνά χαλαρώνουμε τη συνθήκη στασιμότητας περιορίζοντας την στις δύο πρώτες ροπές.

☞ **Ασθενής Στασιμότητα**

Η στοχαστική διαδικασία ή χρονοσειρά $\{Y_t\}_{t=-\infty}^{+\infty}$ είναι ασθενώς στάσιμη όταν οι ροπές πρώτης και δεύτερης τάξης είναι σταθερές στο χρόνο, δηλαδή:

(i) Η μέση τιμή είναι σταθερή: $\forall t \in Z, \quad E[Y_t] = \mu$

(ii) Η αυτοδιασπορά ορίζεται μόνο ως προς την υστέρηση και όχι τις χρονικές στιγμές:

$$\forall t_1, t_2 \in Z, \quad \gamma(t_1, t_2) = \gamma(t, t-\tau) = \gamma(\tau) \equiv \gamma_\tau$$

.
(Το σύμβολο $\equiv$ δηλώνει ισοδυναμία).
Το ii) προκύπτει από τη συνθήκη ότι η δεύτερη ροπή είναι σταθερή, δηλαδή ισχύει

$$E[Y_{t_1} Y_{t_2}] = E[Y_t, Y_{t-\tau}] = \kappa(t, t-\tau) = \kappa(\tau)$$

Από τις συνθήκες i) και ii) προκύπτει ότι η διασπορά είναι επίσης σταθερή. Πράγματι, για $\tau = 0$,ισχύει $E\left[Y_t^2\right] = \kappa(0)$ και άρα:

$$\sigma_\gamma^2 = \gamma(0) = E\left[Y_t^2\right] - (E[Y_t])^2 = \kappa(0) - \mu^2$$

.

Στην πράξη, η συνθήκη ασθενούς στασιμότητας ερμηνεύεται συχνά ως σταθερή μέση τιμή και διασπορά (απλή ροπή δεύτερης τάξης), που δεν είναι σωστό αφού η συνθήκη αναφέρεται στην κοινή ροπή δεύτερης τάξης (αυτοδιασπορά).

Για τη μελέτη συσχετίσεων σε στάσιμες χρονοσειρές χρησιμοποιείται η αυτοσυσχέτιση, που είναι η κανονικοποίηση της αυτοδιασποράς με την διασπορά.
Θεωρούμε την (ασθενώς) στάσιμη στοχαστική διαδικασία (ή χρονοσειρά) $\{Y_t\}_{t=-\infty}^{+\infty}$

Η **αυτοσυσχέτιση** μίας χρονοσειράς,για υστέρηση τ ορίζεται από τη σχέση:

$$\rho_\tau \equiv \rho(\tau) = \frac{\gamma(\tau)}{\gamma(0)}$$

Η αυτοσυσχέτιση μετράει τη συσχέτιση μεταβλητών της $\{Y_t\}_{t=-\infty}^{+\infty}$ που βρίσκονται σε χρονική υστέρηση τ και είναι ένα χρήσιμο μέτρο της «μνήμης» της στοχαστικής διαδικασίας. Επιπλέον, ισχύει ότι $|\gamma_\tau| \leqslant \gamma_0$ και άρα $|\rho_\tau| \leqslant 1$ για κάθε υστέρηση τ. Επίσης η αυτοδιασπορά και η αυτοσυσχέτιση είναι άρτιες συναρτήσεις της υστέρησης τ, ισχύει δηλαδή $\gamma_\tau = \gamma_{-\tau}$ και $\rho_\tau = \rho_{-\tau}$

Παράδειγμα 1: Έστω $X_t = U\cos(\vartheta t) + V\sin(\vartheta t), \quad \vartheta \in (-\pi, \pi]$ και U, V δύο ασυσχέτιστες τ.μ με μηδενικούς μέσους και μοναδιαίες διασπορές, δηλαδή $\mu_U = \mu_V = 0$, $\sigma_U^2 = \sigma_V^2 = 1$ και $\rho(X, Y) = 0$.

Η χρονοσειρά $\{X_t : t \in R\}$ είναι στάσιμη (υπό ευρεία έννοια). Πράγματι έχουμε:
$E[X_t] = E[U]\cos(\vartheta t) + E[V]\sin(\vartheta t) = 0, \quad \forall t \in R.$

$$\begin{aligned} Cov(X_t, X_{t+h}) &= E[X_t X_{t+h}] - E[X_t] \cdot E[X_{t+h}] = E[X_t X_{t+h}] \\ &= E\{U^2 \cos(\vartheta t)\cos(\vartheta(t+h)) + UV\cos(\vartheta t)\sin(\vartheta(t+h)) \end{aligned}$$

$$
\begin{aligned}
&\quad + UV\sin(\vartheta t)\cos(\vartheta(t+h)) + V^2\sin(\vartheta t)\sin(\vartheta(t+h))\} \\
&= E\left[U^2\right]\cos(\vartheta t)\cos(\vartheta(t+h)) + E[UV]\cos(\vartheta t)\sin(\vartheta(t+h)) \\
&\quad + E[UV]\sin(\vartheta t)\cos(\vartheta(t+h)) + E\left[V^2\right]\sin(\vartheta t)\sin(\vartheta(t+h)) \\
&= \cos(\vartheta t)\cos(\vartheta(t+h)) + \sin(\vartheta t)\sin(\vartheta(t+h))\} \\
&= \cos(\vartheta t - \vartheta(t+h)) = \cos(\vartheta h)
\end{aligned}
$$

το οποίο είναι ανεξάρτητο του t.

Παράδειγμα 2: Έστω $\{X_n : n \in N\}$ τυχαίος περίπατος. Έχουμε δηλαδή

$$X_n = Z_1 + Z_2 + \ldots + Z_n \quad (n = 1, 2, \ldots)$$

με Z_ν $(\nu = 1, 2, \ldots)$ ανεξάρτητες και ισόνομες τ.μ. με μέση τιμή $\mu = 0$ και διασπορά σ^2. Η χρονοσειρά $\{X_n : n \in N\}$ δεν είναι στάσιμη. Τούτο διότι ναι μεν έχει σταθερή μέση τιμή, αφού $E[X_n] = nE[Z] = 0$, όμως η συνδιακύμανση εξαρτάται από το n αφού

$$
\begin{aligned}
Cov(X_n, X_{n+k}) &= Cov\left(X_n, X_n + \sum_{\nu=n+1}^{n+k} Z_\nu\right) \\
&= Cov(X_n, X_n) + Cov\left(\sum_{\nu=1}^{n} Z_\nu, \sum_{\nu=n+1}^{n+k} Z_\nu\right) \\
&= V[X_n] = nV[Z] = n\sigma^2
\end{aligned}
$$

Κεφάλαιο 2

ΜΕΘΟΔΟΙ ΠΡΟΒΛΕΨΗΣ

2.1 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.
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: $e_t = Y_t - \widehat{Y}_t$
Η παραπάνω σχέση 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:

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

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:

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

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:

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

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.

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

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.

2.2 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 της διπλής εκθετικής εξομάλυνσης, την μέθοδο Brown ή την μέθοδο Holt ενώ εάν η χρονοσειρά εμφανίζει εποχικότητα τότε χρησιμοποιούμε την μέθοδο Winters.

2.2.1 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:

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

όπου $\widehat{Y}_{t+1}$ είναι η πρόβλεψη για την περίοδο (t+1) και m 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: $\widehat{Y}_{t+1} = Y_t$

Συνήθως για να προσδιορίσουμε την τιμή του m για τη δημιουργία προβλέψεων σε μια χρονοσειρά, εφαρμόζουμε τη μέθοδο του απλού κινητού μέσου στη χρονοσειρά για διαφορετικές τιμές του m και επιλέγουμε εκείνη την τιμή του m που ελαχιστοποιεί την τιμή του κριτηρίου MSE ή κάποιου άλλου κριτηρίου.

2.2.2 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 προβλεπόμενη περίοδο. Το μειονέκτημα αυτό μπορεί να εξαλειφθεί με τη μέθοδο της απλής εκθετικής εξομάλυνσης (simple exponential smoothing), σύμφωνα με την οποία οιπροβλέψεις 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:

$$\widehat{Y}_{t+1} = aY_t + a\,(1-a)\,Y_{t-1} + a\,(1-a)^2\,Y_{t-2} + \ldots \tag{2.1}$$

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 (2.1) 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:

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

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

Η τιμή της παραμέτρου α καθορίζεται από τον ερευνητή. Ωστόσο, πιο αντικειμενικό είναι η "άριστη" τιμή του α να προσδιορίζεται από τα δεδομένα της χρονοσειράς. Ειδικότερα, εφαρμόζοντας τη μέθοδο της απλής εκθετικής εξομάλυνσης στις παρατηρήσεις της χρονοσειράς, για τιμές του α από το μηδέν μέχρι τη μονάδα επιλέγουμε εκείνη την τιμή του α που ελαχιστοποιεί την τιμή του κριτηρίου MSE ή κάποιου άλλου κριτηρίου.

2.2.3 ΔΙΠΛΟΣ ΚΙΝΗΤΟΣ ΜΕΣΟΣ

Η μέθοδος του διπλού κινητού μέσου (double moving average) μπορεί να χρησιμοποιηθεί για την πρόβλεψη των τιμών μιας χρονοσειράς, οι παρατηρήσεις της οποίας παρουσιάζουν ανοδική ή πτωτική πορεία που εκφράζεται από κάποια γραμμική τάση. 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) 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) Υπολογίζουμε τον διπλό κινητό μέσο m-περιόδων, $M_t^{'}$, ως εξής:

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

(iii) Υπολογίζουμε τη διαφορά a_t ως:

$$a_t = 2M_t - M_t^{'}$$

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

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

(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, σε αντίθεση με τις δύο προηγούμενες μεθόδους προβλέψεων που παρουσιάσαμε, μπορεί για $h > 1$ να χρησιμοποιηθεί για τη διενέργεια 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.

2.2.4 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.

2.2.5 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:

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

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, για $t = 2, 3, \dots, n$, ενώ για $t = 1$ ορίζεται ως αρχική συνθήκη $A_1 = Y_1$.

(ii) Εξομαλύνουμε την τάση ως εξής:

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

όπου β, για $0 \leqslant \beta \leqslant 1$, είναι η σταθερά για την εξομάλυνση της τάσης, 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.

2.2.6 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:

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

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, 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:

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

όπου β, για $0 \leqslant \beta \leqslant 1$, είναι η σταθερά για την εξομάλυνση της τάσης ενώ $_t$ 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:

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

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 τίμες 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.

2.3 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, οικυκλικές 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.

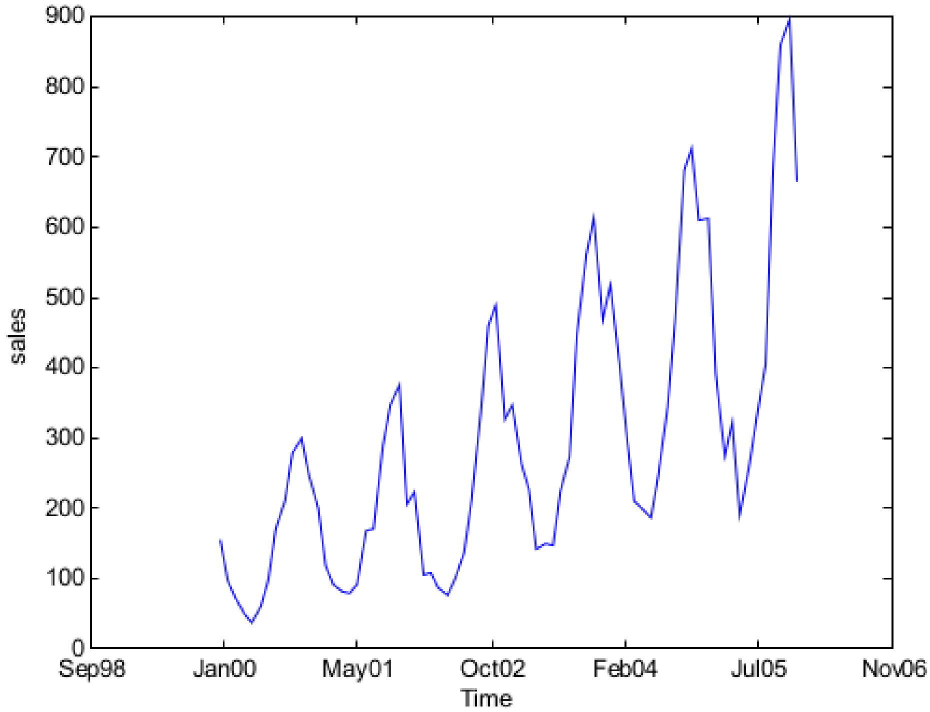
Οι μακροπρόθεσμες ή κύριες κινήσεις ή τάση αναφέρονται στη γενική κατεύθυνση που φαίνεται ότι ακολουθεί η γραφική παράσταση μίας χρονοσειράς κατά μία μεγάλη διάρκεια χρόνου. Σε πολλές περιπτώσεις οι τιμές των παρατηρήσεων ορισμένων χρονοσειρών τείνουν να αυξάνονται ή να μειώνονται με αρκετά σταθερό ρυθμό 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.



Σχήμα 2.1: Γράφημα χρονοσειράς

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

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.

2.3.1 ΑΝΑΛΥΣΗ ΕΠΟΧΙΚΟΤΗΤΑΣ

Η εποχικότητα είναι ένα από τα τέσσερα συνθετικά στοιχεία των χρονοσειρών που πρέπει να μελετηθεί, όταν στις παρατηρήσεις τους εμφανίζεται κάποιο εποχικό πρότυπο. Η εποχικότητα μετριέται με τους δείκτες εποχικότητας (seasonal indices), σκοπός των οποίων είναι η ανίχνευση του τρόπου συμπεριφοράς των παρατηρήσεων της χρονοσειράς που προκαλείται από αυτό το εποχικό φαινόμενο. Ο προσδιορισμός των δεικτών αυτών συμβάλλει στην απαλλαγή των τιμών της χρονοσειράς από το στοιχείο της εποχικότητας, ώστε να δημιουργηθούν πιο αξιόπιστες βραχυπρόθεσμες και μεσοπρόθεσμες προβλέψεις.

Οι δείκτες εποχικότητας προσδιορίζονται με την εφαρμογή της μεθόδου του κεντρικού κινητού μέσου (centered moving average) στις παρατηρήσεις της χρονοσειράς. Με τη μέθοδο αυτή προσπαθούμε να απομονώσουμε την εποχικότητα από τα άλλα τρία συνθετικά στοιχεία της χρονοσειράς, δηλαδή από την τάση, την κυκλικότητα και τη μη-κανονικότητα.

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.

Το πολλαπλασιαστικό μοντέλο προϋποθέτει να είναι το άθροισμα των εποχικών δεικτών ίσο με τον αριθμό των περιόδων εντός του έτους. Εάν αυτό δεν ισχύει θα πρέπει να γίνει κατάλληλη αναπροσαρμογή τους ώστε το άθροισμα των τιμών των δεικτών να ισούται με τον αριθμό των περιόδων. Οι δείκτες που προκύπτουν στην περίπτωση αυτή ονομάζονται προσαρμοσμένοι εποχικοί δείκτες (adjusted seasonal indices).

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 του αντίστοιχου τριμήνου, 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.

2.3.2 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:

$$Y_t = \alpha + \beta \cdot t + \varepsilon_t \tag{2.2}$$

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 (2.2) προσδιορίζονται χρησιμοποιώντας ως εξαρτημένη μεταβλητή τις πραγματικές τιμές Y_t της χρονοσειράς. Αντίθετα, όταν υπάρχει εποχικότητα, η Y_t είναι η SY_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 $t = 1, 2, \ldots, n$.

2.3.3 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 TA_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:

$$WA_t = \frac{Y_{t-1} + 2Y_t + Y_{t+1}}{4}$$

όπου $t = 1, 2, \ldots, n-1$ 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.

2.3.4 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 ΤΗΝ ΠΡΟΒΛΕΨΗ ΠΑΡΑΓΩΓΗΣ ΠΕΤΡΕΛΑΙΟΥ

Κεφάλαιο 3

Όνομα τρίτου κεφαλαίου

3.1 Όνομα πρώτης ενότητας

3.1.1 Όνομα πρώτης υποενότητας

3.2 Όνομα δεύτερης ενότητας

Βιβλιογραφία

[1] C. Li. Chen and A.C. Reynolds. Robust constrained optimization of short- and long-term net present value for closed-loop reservoir management. *SPE Journal*, 17(3):849–864, 2012.

[2] Jan Dirk Jansen. *A Systems Description of Flow Through Porous Media*. SpringerBriefs in Earth Sciences, 2013.

[3] Drosos Kourounis, Louis J. Durlofsky, Jan Dirk Jansen, and Khalid Aziz. Adjoint formulation and constraint handling for gradient-based optimization of compositional reservoir flow. *Computational Geosciences*, pages 1–21, 2014.