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 Ι

ΑΝΑΛΥΣΗ ΧΡΟΝΟΣΕΙΡΩΝ

Κεφάλαιο 1

ΒΑΣΙΚΕΣ ΕΝΝΟΙΕΣ ΧΡΟΝΟΣΕΙΡΩΝ

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 $X_t : t = 0, 1, 2, \dots$, 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 $t = 1, 2, \dots$

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

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

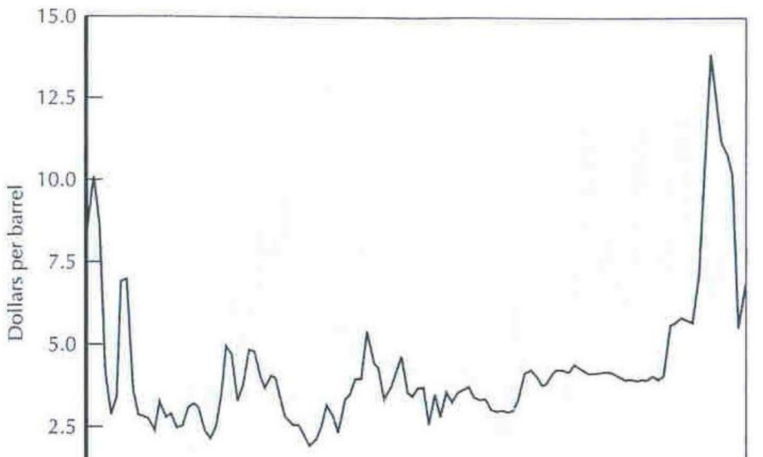
(iv) Η ημερήσια κατανάλωση ηλεκτρικού ρεύματος καθώς και η ημερήσια κατανάλωση ύδατος, X_t και Y_t αντίστοιχα, σε μια μεγάλη γεωγραφική περιοχή της χώρας με $t = 1, 2, \dots$

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

(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 ΣΤΑΤΙΣΤΙΚΑ ΜΕΓΕΘΗ ΧΡΟΝΟΣΕΙΡΩΝ

Ως γνωστόν, τα βασικότερα χαρακτηριστικά μιας τυχαίας μεταβλητής X είναι η μέση τιμή $\mu = E[X]$, η διασπορά $\sigma^2 = V[X]$ και όταν έχουμε να κάνουμε με ζεύγη τυχαίων μεταβλητών, η μικτή ροπή 2ας τάξης, δηλαδή η συνδιακύμανση $\sigma_{xy} = Cov(X, Y)$. Κατ' επέκταση, τα βασικότερα χαρακτηριστικά μιας χρονοσειράς $\{X_t : t \in T\}$ είναι:

- Η συνάρτηση *μέσης τιμής*
- Η συνάρτηση *διασποράς*
- Η συνάρτηση *αυτοσυνδιακύμανσης*

(i) **Η μέση τιμή** ή αναμενόμενη τιμή:

$$\mu(t) = E[X_t], \quad t \in T$$

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 X_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.

(ii) **Η διασπορά**:

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

(iii) **Η αυτοσυνδιακύμανση**:

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

Είναι προφανές ότι ισχύει: $\sigma^2(t) = \gamma(t,0) = \mathrm{Cov}(X_t, X_t), \quad t,h \in T$
Είναι γνωστό επίσης ότι όταν οι διασπορές δύο τυχαίων μεταβλητών X και Y , είναι πεπερασμένες, τότε τόσο οι μέσες τιμές αυτών όσο και η συνδιακύμανση αυτών είναι πεπερασμένες ποσότητες. Τούτο διότι αφενός πρέπει να έχουμε $\mathrm{E}\left[|X|^2\right] < \infty$, $\mathrm{E}\left[|X|^2\right] < \infty$, αφού διαφορετικά δεν ορίζονται οι διασπορές, και αφετέρου, με εφαρμογή της ανισότητας Cauchy-Schwarz

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

προκύπτει:

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

και

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

1.3 ΣΤΑΣΙΜΟΤΗΤΑ ΚΑΙ ΑΥΤΟΣΥΣΧΕΤΗΣΗ

Είναι φανερό ότι έχοντας παρατηρήσει μια χρονοσειρά $\{X_t : t \in T\}$ από κάποια χρονική στιγμή $t = 0$, έστω μέχρι και την παρούσα χρονική στιγμή $t = s$, αν δηλαδή γνωρίζουμε την τροχιά αυτής $\{X_t : 0 \leqslant t \leqslant s\}$, και θέλουμε να προβλέψουμε μελλοντικές τιμές αυτής X_{s+h}, με $h > 0$, θα πρέπει να βασιστούμε στις μέχρι τώρα γνωστές τιμές της και στην εξάρτηση που ενδέχεται να υπάρχει μεταξύ X_{s+h} και των τιμών $\{X_t : 0 \leqslant t \leqslant s\}$ της χρονοσειράς στο παρελθόν. Τούτο βέβαια με την προϋπόθεση ότι όλα τα πιθανοθεωρητικά χαρακτηριστικά μιας χρονοσειράς, ή τουλάχιστον τα βασικότερα εξ αυτών, παραμένουν αναλλοίωτα στον χρόνο.

Όταν όλα τα πιθανοθεωρητικά χαρακτηριστικά μιας χρονοσειράς παραμένουν αναλλοίωτα στο χρόνο τότε μιλάμε για *αυστηρή στασιμότητα*. Συγκεκριμένα έχουμε:

(i) **Αυστηρή Στασιμότητα**

Η χρονοσειρά $\{X_t : t \in T\}$ ονομάζεται *αυστηρά στάσιμη* όταν $\forall n \in \mathbb{N},\ t_i \in T$ με $i = 1, \ldots, n$ και $h \in T$ ισχύει η παρακάτω σχέση ισοδυναμίας

$$(X_{t_1}, \ldots, X_{t_n}) \sim (X_{t_1+h}, \ldots, X_{t_n+h})$$

Εδώ το σύμβολο “ ~ ” διαβάζεται “κατανέμεται όπως”, ή “ισοκατανέμεται με”.

Συνεπώς οι κατανομές πεπερασμένης διάστασης αυστηρώς στάσιμων χρονοσειρών παραμένουν αναλλοίωτες σε χρονικές μεταθέσεις.

(ii) **Ασθενής Στασιμότητα**

Η χρονοσειρά $\{X_t : t \in T\}$ ονομάζεται *ασθενώς στάσιμη*, ή υπό ευρεία έννοια, στάσιμη όταν οι ροπές πρώτης και δεύτερης τάξης είναι σταθερές στο χρόνο, δηλαδή:

- Η μέση τιμή είναι σταθερή: $\mu = \mathsf{E}\left[X_t\right], \quad t \in T$
- $\gamma(h) = \mathsf{Cov}(X_t, X_{t+h}), \quad t, h \in T$

Η συνάρτηση *αυτοσυσχέτισης* μίας χρονοσειράς ορίζεται από τη σχέση:

$$\rho\left(h\right) = \mathsf{Corr}\left(X_0, X_h\right) = \frac{\mathsf{Cov}\left(X_0, X_h\right)}{\sqrt{\mathsf{V}\left[X_0\right]\mathsf{V}\left[X_h\right]}} = \frac{\gamma\left(h\right)}{\gamma\left(0\right)}, \quad h \in T$$

Προφανώς η συνάρτηση αυτοσυσχέτησης έχει όλες τις ιδιότητες της συνάρτησης αυτοσυνδιακύμανσης με την επιπρόσθετη ιδιότητα $|\rho\left(h\right)| \leqslant 1$, ως συνέπεια της ανισότητας Cauchy-Schwarz. Συνεπώς, όλες οι δυνατές τιμές της αυτοσυσχέτισης βρίσκονται εντός του διαστήματος $[-1, 1]$.

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:

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

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.

2.2 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 μέθοδο 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 $t = 2, 3, \dots, n$, 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 $t = 2, 3, \dots, n$ 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:

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

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 $t = 2, 3, \ldots, n$, 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 $t = 2, 3, \ldots, n$ 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

2.3 ΜΕΘΟΔΟΣ ΔΙΑΣΠΑΣΗΣ ΧΡΟΝΟΣΕΙΡΩΝ

2.3.1 ΑΝΑΛΥΣΗ ΕΠΟΧΙΚΟΤΗΤΑΣ

2.3.2 ΑΝΑΛΥΣΗ ΜΑΚΡΟΧΡΟΝΙΑΣ ΤΑΣΗΣ

2.3.3 ΑΝΑΛΥΣΗ ΚΥΚΛΙΚΟΤΗΤΑΣ ΚΑΙ ΜΗ ΚΑΝΟΝΙΚΟΤΗΤΑΣ

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.