

**ΑΕΙ ΠΕΙΡΑΙΑ Τ.Τ.**  
**ΣΧΟΛΗ ΤΕΧΝΟΛΟΓΙΚΩΝ ΕΦΑΡΜΟΓΩΝ**  
**ΤΜΗΜΑ ΜΗΧΑΝΙΚΩΝ ΗΛΕΚΤΡΟΝΙΚΩΝ**  
**ΥΠΟΛΟΓΙΣΤΙΚΩΝ ΣΥΣΤΗΜΑΤΩΝ Τ.Ε.**

**ΠΤΥΧΙΑΚΗ ΕΡΓΑΣΙΑ**

**Ανάλυση ιχνών μακράς μνήμης σε διαταραχές ηλεκτρομαγνητικού πεδίου περιβάλλοντος συχνοτήτων MHz**

**Ασπασία Γ. Γαρδικιώτη**

**Εισηγητής: Δημήτριος Νικολόπουλος, Αν. Καθηγητής**

**ΑΘΗΝΑ**

**ΙΟΥΝΙΟΣ 2016**



**ΠΤΥΧΙΑΚΗ ΕΡΓΑΣΙΑ**

**Ανάλυση ιχνών μακράς μνήμης σε διαταραχές ηλεκτρομαγνητικού πεδίου περιβάλλοντος συχνοτήτων MHz**

**Ασπασία Γ. Γαρδικιώτη**

**A.M. 35961**

**Εισηγητής:**

**Δημήτριος Νικολόπουλος, Αν. Καθηγητής**

**Εξεταστική Επιτροπή:**

**Παναγιώτης Γιαννακόπουλος, Καθηγητής**

**Κωνσταντίνος Κουκουλέτσος, Αν. Καθηγητής**

**Ημερομηνία εξέτασης 14/6/2015**



## ΔΗΛΩΣΗ ΣΥΓΓΡΑΦΕΑ ΠΤΥΧΙΑΚΗΣ ΕΡΓΑΣΙΑΣ

Η κάτωθι υπογεγραμμένη .....Ασπασία Γαρδικιώτη...του .....Γεωργίου....., με αριθμό μητρώου .....35961..... φοιτήτρια του Τμήματος Μηχανικών Η/Υ Συστημάτων Τ.Ε. του Α.Ε.Ι. Πειραιά Τ.Τ. πριν αναλάβω την εκπόνηση της Πτυχιακής Εργασίας μου, δηλώνω ότι ενημερώθηκα για τα παρακάτω:

«Η Πτυχιακή Εργασία (Π.Ε.) αποτελεί προϊόν πνευματικής ιδιοκτησίας τόσο του συγγραφέα, όσο και του Ιδρύματος και θα πρέπει να έχει μοναδικό χαρακτήρα και πρωτότυπο περιεχόμενο.

Απαγορεύεται αυστηρά οποιοδήποτε κομμάτι κειμένου της να εμφανίζεται αυτούσιο ή μεταφρασμένο από κάποια άλλη δημοσιευμένη πηγή. Κάθε τέτοια πράξη αποτελεί προϊόν λογοκλοπής και εγείρει θέμα Ηθικής Τάξης για τα πνευματικά δικαιώματα του άλλου συγγραφέα. Αποκλειστικός υπεύθυνος είναι ο συγγραφέας της Π.Ε., ο οποίος φέρει και την ευθύνη των συνεπειών, ποινικών και άλλων, αυτής της πράξης.

Πέραν των όποιων ποινικών ευθυνών του συγγραφέα σε περίπτωση που το Ίδρυμα του έχει απονείμει Πτυχίο, αυτό ανακαλείται με απόφαση της Συνέλευσης του Τμήματος. Η Συνέλευση του Τμήματος με νέα απόφασης της, μετά από αίτηση του ενδιαφερόμενου, του αναθέτει εκ νέου την εκπόνηση της Π.Ε. με άλλο θέμα και διαφορετικό επιβλέποντα καθηγητή. Η εκπόνηση της εν λόγω Π.Ε. πρέπει να ολοκληρωθεί εντός τουλάχιστον ενός ημερολογιακού βμήνου από την ημερομηνία ανάθεσης της. Κατά τα λοιπά εφαρμόζονται τα προβλεπόμενα στο άρθρο 18, παρ. 5 του ισχύοντος Εσωτερικού Κανονισμού.»



## ΠΕΡΙΛΗΨΗ

Η παρούσα εργασία ασχολείται με τη θεωρία στοχαστικών διαδικασιών και στατιστικών μεθόδων για την πρόβλεψη φρακταλικών χρονοσειρών μακράς μνήμης ληφθείσες από τις γεωφυσικές μελέτες. Η κίνηση Brown είναι ένα στοχαστικό μοντέλο που σε συνδυασμό με τη θεωρία πιθανοτήτων βοηθάει να μελετήσουμε τη στατιστική κίνηση ενός σώματος μέσα σε ένα χώρο πιθανοτήτων. Η κλασματική κίνηση Brown και ο κλασματικός γκαουσιανός θόρυβος αποτελούν βασικά εργαλεία για τη μοντελοποίηση του φρακταλικού ηλεκτρομαγνητικού σήματος, ενώ η εισαγωγή του εκθέτη Hurst σε αυτά τα μοντέλα μας πληροφορεί για το επίπεδο παραμονής (persistence) του σήματος (μακρά μνήμη). Το μοντέλο που ακολουθείται για τη μελέτη των χρονοσειρών είναι το γραμμικό. Οι στατιστικές μέθοδοι που εφαρμόστηκαν για τον υπολογισμό του εκθέτη Hurst για το πειραματικό μέρος είναι οι μέθοδοι R/S, R/L και Variogram.

## ABSTRACT

This thesis addresses the theory of stochastic processes and statistical methods for predicting fractal long memory time series obtained from the geophysical studies. The Brown motion is a stochastic model that in combination with the probability theory helps us to study the statistical movement of a particle in a probability space. The fractional Brownian motion and fractional Gaussian noise are basic tools for modeling of fractal electromagnetic signals, while the introduction of Hurst exponent in these models tells us about the stay level (persistence) of the signal (long memory). The model followed in the study of time series is linear. The statistical methods applied to calculate the Hurst exponent for the experimental part are R / S, R / L and Variogram methods.

ΕΠΙΣΤΗΜΟΝΙΚΗ ΠΕΡΙΟΧΗ: Θεωρία του Χάους, Φράκταλς

ΛΕΞΕΙΣ ΚΛΕΙΔΙΑ: εκθέτης Hurst, κλασματική κίνηση Brown, κλασματικός γκαουσιανός θόρυβος, φράκταλς, μακρά μνήμη

KEYWORDS: Hurst exponent, fractional Brownian motion, fractional Gaussian noise, fractals, long memory

## ΠΕΡΙΕΧΟΜΕΝΑ

<b>1. ΚΙΝΗΣΗ BROWN</b> .....	13
1.1 Πυκνότητες πιθανότητας και η εξίσωση διάχυσης.....	17
1.2 Λευκός Θόρυβος.....	17
1.3 Η κίνηση Brown ως μία γκαουσιανή διαδικασία.....	18
<b>2. ΚΛΑΣΜΑΤΙΚΗ ΚΙΝΗΣΗ BROWN</b> .....	21
2.1 Ορισμός της κλασματικής κίνησης Brown.....	22
2.2 Ιδιότητες αυτο-ομοιότητας.....	23
2.3 Εμπειρικά αποτελέσματα του Hurst για τη συνάρτηση $M(T, \omega)$ .....	24
2.4 Φρακταλικοί θόρυβοι Gauss και προσεγγίσεις.....	25
2.5 Το φάσμα.....	26
2.6 Αυτο-ομοιότητα και κλασματικός γκαουσιανός θόρυβος.....	27
2.7 fBm, fGn και οι συναρτήσεις συσχέτισης.....	28
2.8 Φάσμα Ισχύος fBm και fGn.....	30
2.9 fBm και fGn συνεχούς χρόνου.....	31
2.10 Η φρακταλική γεωμετρία της fBm και fGn.....	33
2.11 Εξάρτηση μεγάλου εύρους.....	35
<b>3. ΧΡΟΝΟΣΕΙΡΕΣ</b> .....	37
3.1 Βασικοί ορισμοί.....	38
3.2 Στασιμότητα και αυτοδιακύμανση.....	39
3.3 Μοντέλα χρονοσειρών γραμμικής διαδικασίας.....	40
3.4 Αυτομειωμένη διαδικασία.....	41
3.5 Διαδικασία μεταβαλλόμενου μέσου όρου.....	44
3.6 Διαδικασία αυτομειωμένου μέσου όρου.....	45
3.7 Η διαδικασία ARIMA.....	46
3.8 Η διαδικασία SARIMA.....	47
3.9 Διάγραμμα συσχέτισης.....	48
<b>4. ΜΑΚΡΑ ΜΝΗΜΗ</b> .....	49
4.1 Προσεγγίσεις ορισμού.....	49
4.2 Αυτο-ομοιότητα και εξάρτηση μεγάλου εύρους.....	52
4.3 Κλασματικός λευκός θόρυβος.....	54
4.4 Σχέσεις με τους ορισμούς της διαδικασίας μακράς μνήμης.....	56
4.5 Η διαδικασία ARFIMA.....	58
4.6 Εκτιμητής Μέγιστης Πιθανότητας.....	59
<b>5. ΦΡΑΚΤΑΛΙΚΗ ΔΙΑΣΤΑΣΗ ΚΑΙ ΕΚΘΕΤΗΣ HURST</b> .....	63
5.1 Συσχέτιση φρακταλικής διάστασης και εκθέτη Hurst.....	63
5.2 Μέθοδος R/S.....	65
5.3 Μέθοδος Variogram.....	66
5.4 Μέθοδος R/L.....	68
<b>6. ΠΕΙΡΑΜΑΤΙΚΟ ΜΕΡΟΣ</b> .....	69
6.1 Κώδικας σε OCTAVE GNU.....	69
6.2 Η εφαρμογή Benoit.....	75
6.2.1 Εύρεση εκθέτη Hurst και συναρτήσεις μετασχηματισμού $\gamma$ χνών.....	76
6.2.2 Υπολογισμός Φρακταλικής Διάστασης και Εκθέτη Hurst.....	76
6.3 Διαγράμματα – Εκτιμήσεις.....	85



<b>ΠΑΡΑΡΤΗΜΑ Α</b> .....	89
<b>7. ΧΡΟΝΟΣΕΙΡΕΣ ΑΝΩΤΕΡΩΝ ΔΙΑΣΤΑΣΕΩΝ</b> .....	89
7.1 Συναρτησιακές χρονοσειρές .....	89
7.2 Χωρικό μοντέλο Hilbert για συναρτησιακά δεδομένα .....	90
7.3 Ο χώρος $L^2$ .....	90
7.4 Βασικές Εμπειρικές Συναρτησιακές Συνιστώσες .....	91
7.5 Συναρτησιακό αυτομειωμένο μοντέλο .....	92
7.6 Υπολογισμός του αυτομειωμένου τελεστή $\Psi$ .....	93
<b>ΒΙΒΛΙΟΓΡΑΦΙΑ</b> .....	98

## ΚΑΤΑΛΟΓΟΣ ΠΙΝΑΚΩΝ

Πίνακας 5.3.1 Σχέση μεταξύ φρακταλικής διάστασης και εκθέτη Hurst .....	68
Πίνακας 6.1.1 Αρχεία .dat.....	69
Πίνακας 6.2.1 Ανάλυση ηλεκτρομαγνητικού σήματος .....	79

## ΚΑΤΑΛΟΓΟΣ ΔΙΑΓΡΑΜΜΑΤΩΝ

<b>Διάγραμμα 6.1.1</b> <u>Ηλεκτρομαγνητικό σήμα μίας ημέρας</u> .....	73
<b>Διάγραμμα 6.1.2</b> <u>Διαχωρισμός σήματος σε 4 τμήματα</u> .....	74
<b>Διάγραμμα 6.3.1</b> <u>Εκθέτης Hurst (κανάλι 9)</u> .....	86
<b>Διάγραμμα 6.3.2</b> <u>Φρακταλική Διάσταση (κανάλι 9)</u> .....	86
<b>Διάγραμμα 6.3.3</b> <u>Εκθέτης Hurst (κανάλι 10)</u> .....	87
<b>Διάγραμμα 6.3.4</b> <u>Φρακταλική Διάσταση (κανάλι 10)</u> .....	87

## ΣΥΝΤΟΜΟΓΡΑΦΙΕΣ

**ARIMA** Auto Regressive Intergrated Moving Average

**AR** Auto Regressive

**MA** Moving Average

**ARMA** Auto Regressive Moving Average

**SARIMA** Seasonal Auto Regressive Intergrated Moving Average

**ARFIMA** Auto Regressive Fractional Intergrated Moving Average

**FAR** Functional Auto Regressive

**EFPC** Empirical Functional Principal Components

**MLE** Maximum Likelihood Estimator

**R/S** Rescale Range

**R/L** Roughness Length

**LRD** Long Range Dependence

**SAW** Self Avoid Random Walk

**ACF** Auto Correlation Function

**PACF** Partial Auto Correlation Function

**IACF** Inverse Auto Correlation Function

**fBm** fractional Brownian motion

**fGn** fractional Gaussian noise

**sssi** self similar stationary increments

**i.i.d** independent identical distributed

## ΚΕΦΑΛΑΙΟ 1

### ΚΙΝΗΣΗ BROWN

Η θεωρία πιθανοτήτων είναι περισσότερο αφιερωμένη στην περιγραφή της μακροσκοπικής εικόνας η οποία προκύπτει από τυχαία συστήματα προσδιοριζόμενα από πλήθος μικροσκοπικών τυχαίων επιδράσεων. Η κίνηση Brown είναι η μακροσκοπική εικόνα που προκύπτει από ένα σωματίδιο κινούμενο τυχαία σε χώρο  $d$  διαστάσεων χωρίς να πραγματοποιεί πολύ μεγάλα άλματα. Σε μικροσκοπικό επίπεδο, σε οποιαδήποτε κλίμακα του χρόνου, το σωματίδιο πραγματοποιεί μια τυχαία μετατόπιση, προκαλούμενη για παράδειγμα είτε από την πρόσκρουση άλλων σωματιδίων είτε από κάποια εξωτερική δύναμη, ώστε, εάν η θέση του σε χρόνο μηδέν είναι  $S_0$ , η θέση του σε χρόνο  $n$  θα είναι  $S_n = S_0 + \sum_{i=1}^n X_i$  όπου οι μετατοπίσεις  $X_1, X_2, X_3$  θεωρούμε ότι είναι ανεξάρτητες, όμοια κατανεμημένες τυχαίες μεταβλητές με τιμές στο  $\mathbb{R}^d$ . Η διαδικασία  $\{S_n : \geq 0\}$  είναι μία τυχαία διαδρομή (random walk), οι μετατοπίσεις αναπαριστούν τις μικροσκοπικές εισόδους – μετατοπίσεις. Όταν μιλάμε για μακροσκοπική εικόνα, τα ερωτήματα που εννοούμε είναι:

- Τείνει το  $S_n$  στο άπειρο; Πόσο συχνά επιστρέφει η  $S_n$  στην αρχική περιοχή της;
- Ποιά είναι η ταχύτητα εξέλιξης της  $\max\{|S_1| \dots |S_n|\}$  καθώς το  $n \rightarrow \infty$ ;
- Ποιός είναι ο ασυμπτωτικός αριθμός των στροφών της  $\{S_n : \geq 0\}$  γύρω από την πηγή;

Είναι προφανές ότι μόνο μερικά από τα χαρακτηριστικά των μικροσκοπικών μετατοπίσεων συμβάλλουν στην μακροσκοπική εικόνα. Εάν υπάρχουν, μόνο ο μέσος (mean) και η συμμεταβλητότητα (covariance) των μετατοπίσεων σχηματίζουν την εικόνα. Με άλλα λόγια, όλες οι τυχαίες διαδρομές (random walks) των οποίων οι μετατοπί -

σεις έχουν τον ίδιο μέσο όρο (mean) και τον ίδιο πίνακα απόκλισης (covariance matrix) προκαλούν την ίδια μακροσκοπική διαδικασία, και ακόμα η υπόθεση ότι όλες οι μετατοπίσεις πρέπει να είναι ανεξάρτητες και όμοια κατανεμημένες δύναται περιστασιακά να μην ισχύει. Αυτή η συνέπεια καλείται universality και η μακροσκοπική

διαδικασία συχνά καλείται universal object. Είναι μια συνηθισμένη προσέγγιση στη θεωρία πιθανοτήτων η μελέτη διαφόρων φαινομένων μέσω των συσχετιζόμενων καθολικών αντικειμένων (universal objects).

Εάν τα άλματα μιας τυχαίας διαδρομής (random walk) είναι επαρκώς προσδιορίσιμα ώστε να θεωρούνται αμελητέα στη μακροσκοπική απεικόνισή τους, ειδικά εάν έχουν πεπερασμένο μέσο (mean) και διακύμανση (variance), οποιαδήποτε συνεχούς χρόνου στοχαστική διαδικασία  $\{B(t): t \geq 0\}$ , η οποία περιγράφει τα μακροσκοπικά χαρακτηριστικά αυτής της τυχαίας διαδρομής πρέπει να έχει τις ακόλουθες ιδιότητες:

- (1) για όλους τους χρόνους  $0 \leq t_1 \leq t_2 \dots \leq t_n$  οι τυχαίες μεταβλητές  $B(t_n) - B(t_{n-1}), B(t_{n-1}) - B(t_{n-2}), \dots, B(t_2) - B(t_1)$  είναι ανεξάρτητες και λέμε ότι η διαδικασία έχει ανεξάρτητες αυξήσεις,
- (2) η κατανομή των αυξήσεων  $B(t+h) - B(t)$  δεν εξαρτάται από το  $t$  και λέμε ότι η διαδικασία έχει σταθερές αυξήσεις (stationary increments),
- (3) η διαδικασία  $\{B(t) : t \geq 0\}$  έχει συνεχή τροχιά.

Προκύπτει από το Κεντρικό Θεώρημα Ορίων (Central Limit Theorem) ότι αυτά τα χαρακτηριστικά υποδηλώνουν ότι υπάρχει ένα διάνυσμα  $\mu \in \mathbb{R}^d$  και ένας πίνακας  $\Sigma \in \mathbb{R}^{d \times d}$  τέτοια ώστε

- (4) για κάθε  $t \geq 0$  και  $h \geq 0$  η αύξηση  $B(t+h) - B(t)$  είναι με πολλές μεταβλητές κανονικά κατανομημένη με μέσο (mean) και πίνακα γινομένου (covariance matrix)  $h\Sigma\Sigma^T$ .

Κάθε διαδικασία με τα χαρακτηριστικά (1)-(3) προσδιορίζεται από τρεις παραμέτρους,

- την αρχική κατανομή, πχ η συνθήκη  $B(0)$ ,
- το διάνυσμα πορείας (drift vector)  $\mu$ ,
- τον πίνακα διασποράς (diffusion matrix)  $\Sigma$ .

Η διαδικασία  $\{B(t) : t \geq 0\}$  καλείται κίνηση Brown με πορεία (drift)  $\mu$  και πίνακα διασποράς (diffusion matrix)  $\Sigma$ . Εάν το διάνυσμα πορείας είναι μηδέν, και ο πίνακας

διασποράς είναι μοναδιαίος λέμε ότι η διαδικασία είναι μια κίνηση Brown. Εάν  $B(0) = 0$ , η κίνηση αρχίζει από την πηγή και χρησιμοποιούμε τον όρο τυπική κίνηση Brown (standard Brownian motion).

Υποθέτουμε ότι έχουμε μια τυπική κίνηση Brown  $\{B(t) : t \geq 0\}$ . Αν  $X$  είναι μια τυχαία μεταβλητή με τιμές στο  $\mathbb{R}^d$ ,  $\mu$  ένα διάνυσμα στο  $\mathbb{R}^d$  και  $\Sigma$  ένας  $d \times d$  πίνακας, τότε η διαδικασία δίνεται από τον τύπο  $B(t) = B(0) + \mu t + \Sigma B(t)$ , για  $t \geq 0$  με ιδιότητες από (1)-(4), με αρχική κατανομή  $X$ , διάνυσμα πορείας  $\mu$  και πίνακα διασποράς (diffusion matrix)  $\Sigma$ . Γι αυτό το λόγο η μακροσκοπική απεικόνιση από μία τυχαία διαδρομή με πεπερασμένη τυπική απόκλιση (variance) δύναται πλήρως να περιγραφεί από την τυπική κίνηση Brown.

Η κίνηση Brown είναι στενά συνδεδεμένη με την κανονική κατανομή (normal distribution). Μια τυχαία μεταβλητή  $X$  είναι κανονικά κατανεμημένη με μέσο  $\mu$  και διακύμανση  $\sigma^2$  εάν

$$\mathbb{P}\{X > x\} = \frac{1}{\sqrt{2\pi\sigma^2}} \int_x^\infty e^{-\frac{(u-\mu)^2}{2\sigma^2}} du \quad \text{για κάθε } x \in \mathbb{R}. \quad (1.1)$$

Μία στοχαστική διαδικασία  $\{B(t) : t \geq 0\}$  πραγματικών τιμών καλείται γραμμική (linear) κίνηση Brown με σημείο εκκίνησης στο  $x \in \mathbb{R}$  εάν ισχύουν τα παρακάτω:

- $B(0) = x$
- η διαδικασία έχει ανεξάρτητες αυξήσεις, πχ για όλα τα  $0 \leq t_1 \leq t_2 \dots \leq t_n$  οι αυξήσεις  $B(t_n) - B(t_{n-1}), B(t_{n-1}) - B(t_{n-2}), \dots, B(t_2) - B(t_1)$  είναι ανεξάρτητες τυχαίες μεταβλητές
- για όλα τα  $t \geq 0$  και  $h > 0$  οι αυξήσεις  $B(t+h) - B(t)$  είναι κανονικά κατανεμημένες με διακύμανση  $\sigma^2$ .
- σχεδόν πάντα, η συνάρτηση  $t \mapsto B(t)$  είναι συνεχής

Η διαδικασία  $\{B(t) : t \geq 0\}$  είναι μια τυπική κίνηση Brown αν  $x = 0$  [14]. Ωστόσο, η ύπαρξη της κίνησης Brown δεν είναι ένα συνηθισμένο γεγονός. Το βασικό ζήτημα είναι να δείχνουν ότι οι γκαουσιανές κατανομές πιθανότητας, οι οποίες συνεπάγονται ότι  $B(t + \Delta t) - B(t)$  είναι τυπικά της τάξης  $\sqrt{\Delta t}$ , χαρακτηρίζονται από τη συνέχεια των δειγματοληπτικών διαδρομών.

Η γκαουσιανή υπόθεση πρέπει, πράγματι, να ικανοποιείται από οποιαδήποτε διαδικασία με ανεξάρτητες αυξήσεις και συνεχείς δειγματοληπτικές διαδρομές. Αυτό είναι μια συνέπεια του Κεντρικού Θεωρήματος Ορίων (Central Limit Theorem), επειδή κάθε αύξηση  $B_t - B_s = \sum_{i=0}^n (B_{t_{i+1}} - B_{t_i})$  με  $s = t_0 < t_1 < \dots < t_n = t$  είναι ένα άθροισμα αυθαίρετα πολλών ανεξάρτητων τυχαίων μεταβλητών με μηδενικό μέσο. Η συνέχεια των δειγματοληπτικών διαδρομών είναι επαρκής να διασφαλίσει ότι ικανοποιείται η υπόθεση του Κεντρικού Θεωρήματος Ορίων. Άλλωστε, από τη στιγμή που οι μέσοι και οι διακυμάνσεις των ανεξάρτητων γκαουσιανών μεταβλητών είναι προσθετικοί, υπάρχουν γραμμικές συναρτήσεις για τη διαφορά στο χρόνο. Μετά από κανονικοποίηση υποθέτουμε ότι ο μέσος των  $B_t - B_s$  είναι μηδέν και η διακύμανση είναι  $t - s$ , σε μια τυπική κίνηση Brown. Η γκαουσιανή κατανομή είναι απείρως διαιρετή, από τη στιγμή που η γκαουσιανή τυχαία μεταβλητή με μέσο  $\mu$  και διακύμανση  $\sigma^2$  είναι ένα άθροισμα από  $n$  ανεξάρτητες, ταυτόσημα κατανεμμένες τυχαίες μεταβλητές με μέσο  $\mu/n$  και διακύμανση  $\sigma^2/n$ .

Από τη στιγμή που η κίνηση Brown είναι ένα άθροισμα από αυθαίρετα πολλών ανεξάρτητων αυξήσεων μέσα σε οποιοδήποτε διάστημα χρόνου, έχει μια τυχαία *δομή φράκταλ* μέσα στην οποία οποιοδήποτε τμήμα της κίνησης, μετά από επαναδιαβάθμιση (rescaling), έχει την ίδια κατανομή με την αρχική κίνηση. Συγκεκριμένα, αν  $c > 0$  είναι μια σταθερά, τότε

$$B_t = \frac{1}{c^{1/2}} B_{ct} \quad (1.2)$$

έχει την ίδια κατανομή με την  $B_t$ , οπότε είναι επίσης μία κίνηση Brown. Άλλωστε, μπορούμε να μεταφέρουμε μια κίνηση Brown  $B_t$  από οποιοδήποτε χρόνο  $s$  ξανά στην πηγή για να πάρουμε μια κίνηση Brown  $\widetilde{B}_t = B_{t+s} - B_s$  και μετά επαναδιαβαθμίζουμε (rescale) τη μεταφερόμενη διαδικασία [22].



## 1.1 Πυκνότητες πιθανότητας και η εξίσωση διάχυσης

Σε αυτή την ενότητα θα μελετήσουμε την περιγραφή της κίνησης Brown σε όρους πυκνοτήτων πιθανότητας πεπερασμένης διάστασης. Η κίνηση Brown είναι μια χρονικά ομογενής διαδικασία Markov, με πυκνότητα μεταβολής

$$p(x, t|y) = \frac{1}{\sqrt{2\pi t}} e^{-(x-y)^2/2t} \quad \text{για } t > 0. \quad (1.1.1)$$

Ως συνάρτηση των  $(x, t)$  η πυκνότητα αλλαγής δίνεται από την εξίσωση διάχυσης ή θερμότητας

$$\frac{\partial p}{\partial t} = \frac{1}{2} \frac{\partial^2 p}{\partial x^2} \quad (1.1.2)$$

και την αρχική συνθήκη  $p(x, 0 | y) = \delta(x - y)$ . Εάν η κίνηση Brown  $B_t$  δεν αρχίζει ακριβώς στο μηδέν και η αρχική τιμή  $B_0$  είναι μια συνεχής τυχαία μεταβλητή, ανεξάρτητη από την υπόλοιπη κίνηση, με πυκνότητα  $p_0(x)$  τότε η πυκνότητα της  $B_t$  για  $t > 0$  δίνεται από την

$$p(x, t) = \frac{1}{\sqrt{2\pi t}} \int e^{-(x-y)^2/2t} p_0(y) dy \quad (1.1.3)$$

Αυτή είναι η έκφραση της συνάρτησης Green για τη λύση της εξίσωσης διάχυσης με αρχικά δεδομένα  $p(x, 0) = p_0(x)$  [22].

## 1.2 Λευκός Θόρυβος

Παρόλο που οι διαδρομές Brown δεν είναι σημειακά διαφορίσιμες, μπορούμε να ερμηνεύσουμε τη χρονική παράγωγο σε μια λογική κατανομή για να λάβουμε μια γενικευμένη στοχαστική διαδικασία που αποκαλείται λευκός θόρυβος. Δηλώνεται ως  $\xi(t, \omega) = \dot{B}(t, \omega)$ . Επίσης συμβολίζεται και ως  $\xi dt = dB$ . Ο όρος λευκός θόρυβος προκύπτει από τη φασματική θεωρία σταθερών τυχαίων διαδικασιών, σύμφωνα με τις οποίες ο λευκός θόρυβος έχει ένα 'επίπεδο' (flat) δυναμοφάσμα το οποίο είναι

ομοιόμορφα κατανομημένο σε όλες τις συχνότητες (όπως το λευκό φως). Αυτό μπορεί να παρατηρηθεί από την παράσταση Fourier της κίνησης Brown.

Το 1923 ο Wiener απέδειξε ότι εάν  $A_0, A_1, \dots, A_n, \dots$  είναι ανεξάρτητες, όμοια κατανομημένες γκαουσιανές μεταβλητές με μέσο μηδέν και διακύμανση ένα, τότε οι σειρές Fourier

$$B(t) = \frac{1}{\sqrt{\pi}} \left( A_0 t + 2 \sum_{n=1}^{\infty} A_n \frac{\sin nt}{n} \right) \quad (1.2.1)$$

έχει σχεδόν σίγουρα μία υπό-ακολουθία μερικών αθροισμάτων τα οποία συγκλίνουν ομοιόμορφα σε μία συνεχή συνάρτηση. Και, επιπλέον, η προκύπτουσα διαδικασία  $B$  είναι μία κίνηση Brown στο  $[0, \pi]$ .

Από τη στιγμή που η κίνηση Brown έχει γκαουσιανές ανεξάρτητες αυξήσεις με μέσο μηδέν, η χρονική παράγωγός της είναι μία γκαουσιανή στοχαστική διαδικασία με μέσο μηδέν της οποίας οι τιμές σε διαφορετικούς χρόνους είναι ανεξάρτητες.

Σε εφαρμογές, η υπόθεση του λευκού θορύβου είναι χρήσιμη για τη μοντελοποίηση φαινομένων στα οποία η χρονική συσχέτιση του θορύβου είναι πολύ σύντομη από οποιαδήποτε άλλη χρονική κλίμακα. Για παράδειγμα, στην περίπτωση της κίνησης Brown, η χρονική συσχέτιση του θορύβου λόγω της επίδρασης των μορίων στο σωματίδιο Brown είναι της τάξης της χρονικής σύγκουσης των ρευστών μορίων μεταξύ τους [22].

### 1.3 Η κίνηση Brown ως μία γκαουσιανή διαδικασία

Παρατηρούμε ότι το διάνυσμα  $(B(t_1), \dots, B(t_n))$  έχει μία πολυμεταβλητή κανονική κατανομή επειδή το γεγονός  $(B(t_1) = x_1, \dots, B(t_n) = x_n)$  δύναται να ξαναγραφεί σε όρους γεγονότων ανεξάρτητων αυξήσεων  $B(t_1) = x_1, B(t_2) - B(t_1) = x_2 - x_1, \dots, B(t_n) - B(t_{n-1}) = x_n - x_{n-1}$  αποδίδοντας την ενσωματωμένη πυκνότητα των  $(B(t_1), \dots, B(t_n))$  ως  $f(x_1, \dots, x_n) = f_{t_1}(x_1) f_{t_2 - t_1}(x_2 - x_1) \dots f_{t_n - t_{n-1}}(x_n)$  όπου

$$f_t(x) = \frac{1}{\sqrt{2\pi t}} e^{-\frac{x^2}{2t}}$$

είναι η πυκνότητα για την κατανομή  $N(0, t)$ .

Η πεπερασμένη διαστατική κατανομή της κίνησης Brown είναι συνεπώς κανονικά πολυμεταβλητή, ή γκαουσιανή, και η κίνηση Brown είναι ένα παράδειγμα γκαουσιανής διαδικασίας δηλαδή, με συνεχής δειγματοληπτικές διαδρομές μέσα στις οποίες οι πεπερασμένες διαστατικές κατανομές είναι κανονικά πολυμεταβλητές, που σημαίνει ότι για οποιαδήποτε ορισμένη επιλογή  $n$  χρονικών σημείων  $0 \leq t_1 < t_2 < \dots < t_n$ , η ενσωματωμένη κατανομή του διανύσματος  $(X(t_1), \dots, X(t_n))$ , είναι κανονικά πολυμεταβλητή [21].



## ΚΕΦΑΛΑΙΟ 2

### ΚΛΑΣΜΑΤΙΚΗ ΚΙΝΗΣΗ BROWN (FRACTIONAL BROWNIAN MOTION)

Η θεωρία της κλασματικής κίνησης Brown (fBm) είναι μια μαθηματική γενίκευση της κλασικής θεωρίας της τυχαίας διαδρομής (random walk) και της κίνησης Brown. Ο όρος κλασματικός (fractional) σχετίζεται με την κλασματική ολοκλήρωση και διαφορίση. Σε αντίθεση με την κλασική κίνηση Brown η οποία έχει ανεξάρτητες αυξήσεις, η κλασματική κίνηση Brown έχει μεγάλη εμβέλεια (long-range) και ισχυρή χωρική και προσωρινή συσχέτιση (correlation) ως ιδιότητες του ορισμού της. Το βασικό χαρακτηριστικό της fBm είναι η «διάρκεια αλληλεξάρτησης» μεταξύ των αυξήσεων της που δύναται να θεωρηθούν άπειρες.

Η κλασματική κίνηση Brown προσφέρει χρήσιμα μοντέλα για ένα πλήθος φυσικών χρονοσειρών. Είναι γνωστό ότι οι χρονοσειρές στην οικονομία «τυπικά» παρουσιάζουν κύκλους όλων των τάξεων μεγέθους, οι αργοί κύκλοι έχουν περιόδους διάρκειας ισοδύναμη με το συνολικό μέγεθος του δείγματος. Το φασματικό δείγμα τέτοιων χρονοσειρών δεν δείχνει τραχύτητα 'καθαρή περίοδος', αλλά μια φασματική πυκνότητα με μια αιχμηρή κορυφή κοντά στις συχνότητες που πλαισιώνουν το αντίστροφο μέγεθος του δείγματος. Ένα άλλο παράδειγμα μπορεί να ληφθεί από τη μελέτη των διακυμάνσεων στα στερεά. Πολλές από αυτές τις διακυμάνσεις καλούνται «1: f θόρυβοι» επειδή η φασματική πυκνότητα των δειγμάτων παίρνει τη μορφή  $\lambda^{1-2H}$ , όπου  $\lambda$  η συχνότητα  $\frac{1}{2} < H < 1$  και το  $H$  συχνотικά κοντά στο 1. Ωστόσο, από τη στιγμή που οι τιμές του  $H$  μακριά από το 1 δύναται να παρατηρηθούν, ο όρος «1: f θόρυβος» είναι ανακριβής και δύσχρηστος. Με μία επιφυλακτικότητα όσον αφορά στη διαθεσιμότητα διάφορων εναλλακτικών εκφράσεων, ο όρος «1: f θόρυβος» μπορεί να προσδιοριστεί και ως κλασματικός θόρυβος. Μια άλλη κατηγορία φαινομένων με υπερβολική αλληλεξάρτηση συναντώνται στην υδρολογία, όπου ο Hurst βρήκε ότι το εύρος της ροής του συσσωρευμένου ύδατος διαφέρει αναλογικά κατά ένα παράγοντα  $t^H$  με  $\frac{1}{2} < H < 1$ .

## 2.1 Ορισμός της Κλασματικής Κίνησης Brown

Ο όρος κλασματική κίνηση Brown χρησιμοποιείται για να δηλώσει μια οικογένεια από γκαουσιανές τυχαίες συναρτήσεις οι οποίες ορίζονται ως εξής: έστω  $B(t)$  μια συνηθισμένη κίνηση Brown, και  $H$  μια παράμετρος που ικανοποιεί τη συνθήκη  $0 < H < 1$ . Τότε η fBm με εκθέτη  $H$  είναι ένας μεταβαλλόμενος μέσος όρος (moving average) της  $dB(t)$  μέσα στην οποία παρελθοντικές αυξήσεις της  $B(t)$  έχουν ισοσταθμιστεί από τον πυρήνα  $(t-s)^{H-1/2}$ . Με  $t$  ορίζεται ο χρόνος ( $-\infty < t < \infty$ ) και ως  $\omega$  (ανήκει μέσα στο δειγματοληπτικό χώρο  $\Omega$ ) ορίζεται το σύνολο όλων των τιμών της τυχαίας συνάρτησης. Η συνήθης κίνηση Brown  $B(t, \omega)$  είναι μια πραγματική τυχαία συνάρτηση με ανεξάρτητες γκαουσιανές αυξήσεις, ώστε η  $B(t_2, \omega) - B(t_1, \omega)$  έχει μέσο μηδέν και διακύμανση  $|t_2 - t_1|$ , και ισχύει εάν τα διαστήματα  $(t_1, t_2)$  και  $(t_3, t_4)$  δεν επικαλύπτονται, επίσης η  $B(t_2, \omega) - B(t_1, \omega)$  είναι ανεξάρτητη της  $B(t_4, \omega) - B(t_3, \omega)$ . Το γεγονός ότι η τυπική απόκλιση της αύξησης  $B(t+T, \omega) - B(t, \omega)$  με  $T > 0$  είναι ίση με  $T^{1/2}$  συχνά αναφέρεται ως νόμος  $T^{1/2}$ .

Έστω  $0 < H < 1$ , και έστω  $b_0$  ένας τυχαία πραγματικός αριθμός. Η ακόλουθη τυχαία συνάρτηση  $B_H(t, \omega)$  καλείται περιορισμένη (reduced) κλασματική κίνηση Brown με παράμετρο  $H$  και σημείο εκκίνησης στο  $b_0$  σε χρόνο  $t = 0$ . Για  $t > 0$ , η  $B_H(t, \omega)$  ορίζεται από τις ιδιότητες

$$B_H(0, \omega) = b_0$$

$$B_H(t, \omega) - B_H(0, \omega) = \frac{1}{\Gamma(H + 1/2)} \left\{ \int_{-\infty}^0 [(t-s)^{H-1/2} - (-s)^{H-1/2}] dB(s, \omega) + \int_0^t (t-s)^{H-1/2} dB(s, \omega) \right\}.$$

Να σημειωθεί ότι αν  $b_0 = 0$ , τότε  $B_{1/2}(t, \omega) = B(t, \omega)$ . Για διαφορετικές τιμές του  $H$ , η  $B_H(t, \omega)$  καλείται κλασματική παράγωγος ή κλασματικό ολοκλήρωμα της  $B(t, \omega)$  (Weyl 1917).

Η εισαγωγή της συνάρτησης Γάμμα  $\Gamma(H + \frac{1}{2})$  στον παρανομαστή διαβεβαιώνει ότι, όταν η ποσότητα  $H - \frac{1}{2}$  είναι ακέραιος, το κλασματικό ολοκλήρωμα γίνεται ένα σύνηθες επαναλαμβανόμενο ολοκλήρωμα.

## 2.2 Ιδιότητες αυτο-ομοιότητας (self-similarity)

Ο συμβολισμός  $\{X(t, \omega)\} \triangleq \{Y(t, \omega)\}$  δηλώνει ότι δύο τυχαίες μεταβλητές  $X(t, \omega)$  και  $Y(t, \omega)$  έχουν τις ίδιες πεπερασμένες συναρτήσεις κατανομής.

Οι αυξήσεις μιας τυχαίας μεταβλητής  $X(t, \omega)$  στο  $-\infty < t < \infty$  θεωρείται ότι είναι αυτο-όμοιες (self-similar) (s-s) με παράμετρο με  $H \geq 0$  αν για κάθε  $h > 0$  και για  $t_0$ , ισχύει

$$\{X(t_0 + \tau, \omega) - X(t_0, \omega)\} \triangleq \{h^{-H} [X(t_0 + h\tau, \omega) - X(t_0, \omega)]\} \quad (2.2.1)$$

Για κάθε  $t_0$  αυτός ο ορισμός δηλώνει ότι, όταν το  $t$  είναι περιορισμένο στο  $t \geq t_0$  η διαφορά  $X(t, \omega) - X(t_0, \omega)$  είναι μία «ημισταθερή στοχαστική διαδικασία». Η ημισταθερότητα είναι ασθενέστερη από τις αυξήσεις της αυτο-ομοιότητας.

Αν η  $X(t, \omega)$  είναι ημισταθερή με παράμετρο και έχει σταθερές αυξήσεις, τότε η  $X(t, \omega)$  είναι περιορισμένη για  $t \geq t_0$  από μια διαδικασία με αυξήσεις s-s με παράμετρο.

**Θεώρημα 2.2.1:** οι αυξήσεις της fGm,  $B_H(\tau, \omega)$  είναι σταθερές και αυτο-όμοιες με παράμετρο.

**Πόρισμα 2.2.1:** ο νόμος  $T^H$  για την τυπική απόκλιση της  $B_H$  δύναται να δηλωθεί ως ακολούθως:

$$E[B_H(t+T, \omega) - B_H(t, \omega)]^2 = T^{2H} V_H$$

$$\text{όπου} \quad V_H = [\Gamma(H + \frac{1}{2})]^{-2} \left\{ \int_{-\infty}^0 [(1-s)^{H-\frac{1}{2}} - (-s)^{H-\frac{1}{2}}]^2 ds + \frac{1}{2H} \right\}$$

**Ορισμός 2.2.1:** έστω  $X(t, \omega)$  μια πραγματική (real-valued) τυχαία συνάρτηση. Το ακολουθιακό εύρος (sequential range) ορίζεται ως

$$M(t, T, \omega) = \sup_{t \leq s \leq t+T} [X(s, \omega) - X(t, \omega) - \inf_{t \leq s \leq t+T} [X(s, \omega) - X(t, \omega)]]$$

Εάν η  $X(t, \omega)$  έχει συνεχή δειγματοληπτική διαδρομή και  $t$  και  $T$  είναι πεπερασμένοι, μπορούμε να αντικαταστήσουμε το  $\sup$  με το  $\max$  και το  $\inf$  με το  $\min$ . Ο νόμος  $T^H$  για το ακολουθιακό εύρος μιας διαδικασίας με αυξήσεις s-s ορίζεται ως εξής:

Αν η  $X(t, \omega)$  έχει αυξήσεις s-s με παράμετρο  $H$  τότε ισχύει

$$M(T, \omega) \triangleq T^H M(1, \omega).$$

Αν για παράδειγμα,  $X(t, \omega) = B(t, \omega)$  έτσι ώστε  $H = \frac{1}{2}$ , τότε  $T^{-1/2} M(t, T, \omega)$  έχει μία κατανομή ανεξάρτητη των  $t$  και  $T$ .

### 2.3 Εμπειρικά αποτελέσματα του Hurst για τη συνάρτηση $M(T, \omega)$

Ο διαχωρισμός της fBm προέκυψε από κάποια εμπειρικά αποτελέσματα του Hurst για τον παράγοντα  $M$ . Ο Hurst μελέτησε την καταγραφή της ροής του νερού στο Νήλο και σε άλλους ποταμούς, την τιμή του σιταριού και άλλων φυσικών σειρών όπως το ύψος της βροχόπτωσης, θερμοκρασίες, πίεση, το πάχος δακτυλίων των δέντρων και τον αριθμό των ηλιακών κηλίδων.

Το εμπειρικό συμπέρασμά του ήταν, σε πρώτη προσέγγιση, ότι το εύρος είναι αναλογικό στο  $T^H$ , όπου  $\frac{1}{2} < H < 1$  αυτό αποτέλεσε έκπληξη για τους στατιστικολόγους επειδή μοντέλα όπως αυτό

$$X(t, \omega) = \int_0^t Y(s, \omega) ds \quad (2.3.1)$$

όπου η  $Y(s, \omega)$  είναι σταθερή με αθροιστική συνάρτηση συνδιακύμανσης (covariance), έχει ακολουθιακό εύρος ασυμπτωτικά αναλογικό με την  $\sqrt{T}$ . Γι' αυτό το λόγο, οι ανακαλύψεις του Hurst έχουν οδηγήσει μερικούς σχολιαστές στο συμπέρασμα ότι η ροή των ποταμών δεν μπορεί να αναπαρασταθεί από στατικές στοχαστικές διαδικασίες.



## 2.4 Φρακταλικοί θόρυβοι Gauss και προσεγγίσεις

Η fBm δεν έχει παράγωγο. Αυτή η δυσκολία επίσης συναντάται και στην περίπτωση της συνηθισμένης κίνησης Brown. Πολλές μέθοδοι, όχι πάντα ακριβής, έχουν χρησιμοποιηθεί για να δώσουν νόημα στην έννοια της παραγώγου της κίνησης Brown. Τα αποτελέσματα που ελήφθησαν καλούνται ‘λευκοί θόρυβοι Gauss’. Ανάλογες προσεγγίσεις μπορούν να εφαρμοστούν για την κλασματική κίνηση Brown και οδηγούν σε αυτό που αποκαλούμε ‘κλασματικοί θόρυβοι Gauss’.

Η πιο στοιχειώδης μέθοδος παράκαμψης της απουσίας παραγώγου της fBm είναι ο μηδενισμός της  $B_H$  και η εισαγωγή της τυχαίας συνάρτησης

$$B_H(t, \omega; \delta) = \delta^{-1} \int_t^{t+\delta} B_H(s, \omega) ds = \int_{-\infty}^{\infty} B_H(s, \omega) \phi_1(t-s) ds, \text{ για } \delta > 0$$

$$\text{όπου } \phi_1(t) = \begin{cases} \delta^{-1} & \text{αν } 0 \leq t \leq \delta, \\ 0, & \text{αλλιού} \end{cases}$$

Η συνάρτηση  $B_H(t, \omega; \delta)$  έχει μια σταθερή παράγωγο

$$B_H'(t, \omega; \delta) = \delta^{-1} [B_H(t+\delta, \omega) - B_H(t, \omega)] = - \int_{-\infty}^{\infty} B_H(s, \omega) d\phi_1(t-s)$$

η οποία είναι σχεδόν πάντα συνεχής, και πάντα μη διαφορίσιμη.

Για ένα  $\delta$  αρκετά μικρό, η  $B_H(t, \omega)$  και η  $B_H(t, \omega; \delta)$  είναι δυσδιάκριτες για πολλούς ‘πρακτικούς λόγους’, για παράδειγμα εξαιρώντας τις επιδράσεις των υψηλών συχνοτήτων, η μη διαφορισιμότητα της  $B_H(t, \omega)$  είναι αναγκαία. Οπότε, εάν χρησιμοποιήσουμε έναν απείρως διαφορίσιμο πυρήνα  $\phi$  ο οποίος μηδενίζει εξωτερικά κάποια διαστήματα και ολοκληρώνει στη μονάδα, τότε η  $k$ -οστή παράγωγος της

$$\int_{-\infty}^{\infty} B_H(s, \omega) \phi(t-s) ds$$

$$\text{είναι } (-1)^k \int_{-\infty}^{\infty} B_H(s, \omega) \phi^{(k)}(t-s) ds$$

η οποία είναι συνεχής και σταθερή για όλους τους θετικούς ακεραίους  $k$ . Ακολουθώντας αυτήν τη προσέγγιση, μπορούμε να ερμηνεύσουμε την  $B_{H'}$  ως μη τυχαία συνάρτηση αλλά γενικευμένη τυχαία συνάρτηση.

Ο κλασματικός γκαουσιανός θόρυβος ορίζεται ως σειράς διαδοχικών αυξήσεων της κλασματικής κίνησης Brown. Με άλλα λόγια ο fGn είναι η διαφόριση της fBm και αντίστροφα, η ολοκλήρωση του fGn δίνει μια fBm. Επομένως, κάθε fBm σχετίζεται με έναν συγκεκριμένο fGn, και οι δυο χαρακτηρίζονται από τον ίδιο εκθέτη  $H$ .

## 2.5 Το φάσμα

Μία πολύ ενδιαφέρουσα απεικόνιση των συχνοτήτων των αυξήσεων της κλασματικής κίνησης Brown που προτάθηκε από τον Hunt είναι η:

$$B_H(t_2, \omega) - B_H(t_1, \omega) = V_H^* \int_0^\infty (e^{2\pi i \lambda t_2} - e^{2\pi i \lambda t_1}) \lambda^{-H-1/2} dB(\lambda, \omega),$$

όπου  $V_H^*$  είναι σταθερά. Αυτή δηλώνει ότι η  $B_H(t, \omega)$  έχει φασματική συχνότητα αναλογική με  $\lambda^{-2H-1}$ . Οι φασματικές συχνότητες μη σταθερών τυχαίων μεταβλητών είναι δύσκολο να ερμηνευτούν. Στη διαφόριση της  $B_H$  υπάρχει ο ισχυρισμός ότι η παράγωγος της  $B_{H'}$  έχει φασματική συχνότητα αναλογική με  $\lambda^{1-2H}$ . Αν  $\frac{1}{2} < H < 1$ , αυτή η τυπική συχνότητα φαίνεται να τείνει στο άπειρο για  $\lambda = 0$ . Η αναλογικότητα της φασματικής συχνότητας στο  $\lambda^{1-2H}$  υποδηλώνει επίσης ότι υπάρχει άπειρη ενέργεια σε υψηλές συχνότητες.

Η φασματική συχνότητα της  $B_{H'}(t, \omega; \delta)$  είναι

$$\begin{aligned} G_{H'}(\lambda; \delta) &= 4 \int_0^\infty C_H(s; \delta) \cos(2\pi \lambda s) ds \\ &= 2V_H \delta^{-2} \int_0^\infty [(s + \delta)^{2H} - 2s^{2H} + |s - \delta|^{2H}] \cos(2\pi \lambda s) ds \end{aligned}$$

Ένα είδος ιδιότητας αυτο-ομοιότητας της  $B_{H'}$  εκφράζεται από το γεγονός ότι μπορούμε να ορίσουμε μία συνάρτηση  $G^*$  γράφοντας  $G_{H'}(\lambda; \delta) = 2V_H \delta^{2H-1} G_{H^*}(\delta \lambda)$ .

Για μικρές τιμές του  $\delta \lambda$  παίρνουμε

$G_{H'}(\delta\lambda) \sim K_H(2\pi\delta\lambda)^{1-2H}$  όπου  $K_H = \frac{\pi H(2H-1)}{\Gamma(2-2H)} [\cos\pi(H-1)]^{-1} > 0$  Και η  $G_{H'}(\lambda; \delta)$  συμπεριφέρεται σαν  $2K_H V_H(2\pi\lambda)^{1-2H}$ . Για ορισμένο  $\lambda > 0$ , το  $\lim_{\delta \rightarrow 0} G_{H'}(\lambda, \delta)$  είναι θετικό και πεπερασμένο και ίσο με την κανονική φασματική συχνότητα της  $B_{H'}$ . Με άλλα λόγια, αλλαγές στην τιμή της  $\delta$  εμπεριέχει λεπτομέρειες της οποίας η ενέργεια είναι κυρίως σε υψηλές συχνότητες [12].

## 2.6 Αυτο-ομοιότητα και κλασματικός γκαουσιανός θόρυβος

Ο κλασματικός γκαουσιανός θόρυβος (fractional Gaussian noise) fGn και η fBm εισήχθησαν από τον Mandelbrot και τον van Ness για την μοντελοποίηση στοχαστικών φρακταλικών διαδικασιών. Ο φρακταλικός γκαουσιανός θόρυβος είναι μια διαδικασία αύξησης διακριτών βημάτων της κλασματικής κίνησης Brown που ορίζεται από την  $X_j = B_H(j) - B_H(j-1)$  για  $j=1,2,\dots$  [27]. Η στασιμότητα των αυξήσεων της fBm δηλώνει ότι είναι στάσιμη γκαουσιανή διαδικασία. Χρησιμοποιώντας τη σχέση  $(a^2 + b^2 - (a-b)^2)/2$  και την αυξητική διακύμανση της fBm προκύπτει η σχέση  $Cov(X_{j+n}, X_j) = \frac{\sigma^2}{2} [(n+1)^{2H} + |n-1|^{2H} - 2n^{2H}]$  για  $j \geq 1, n \geq 0$ . Οπότε προκύπτει,  $\rho_n := Corr(X_{j+n}, X_j) \sim H(2H-1)n^{-2(1-H)}$  καθώς  $n \rightarrow \infty$ . Και ειδικά όταν  $\rho_n \rightarrow 0$ , καθώς το  $n \rightarrow \infty$ . Επιπλέον, από την αυτο-ομοιότητα της fBm, για κάθε  $n$  ισχύει

$$Var(X_1 + \dots + X_n) = Var B_H(n) = \sigma^2 n^{2H} \quad [22] \quad (2.6.1)$$

Ένας fGn με αυτο-ομοιότητα είναι μία σειρά από πανομοιότυπες γκαουσιανές τυχαίες μεταβλητές  $X_1, X_2, X_3, \dots$ , οι οποίες έχουν την ακόλουθη ιδιότητα:

$$A_N[X] \triangleq \frac{X_1 + X_2 + X_3 + \dots + X_N}{N^H} \sim X \quad (2.6.2)$$

όπου ‘ $\sim$ ’ δηλώνει ισότητα από την άποψη κατανομής πιθανοτήτων. Ένας εναλλακτικός αλλά ισοδύναμος ορισμός μπορεί να δοθεί με όρους σχετικής διασποράς:  $RD_N = N^{H-1}RD_1$

όπου

$$RD_N = \frac{\sqrt{\text{Var}[X_1 + X_2 + X_3 + \dots + X_N]}}{NE[X_1]} \quad (2.6.3)$$

και  $\text{Var}[\cdot]$  δηλώνουν αναμενόμενη τιμή και διακύμανση μιας τυχαίας μεταβλητής. Είναι άμεσο επακόλουθο να πούμε ότι όταν  $H = \frac{1}{2}$  τα  $X$  είναι απαραίτητως ανεξάρτητα. Γενικά όταν ισχύει  $0 < H < 1$  τα  $X$  είναι συσχετιζόμενα.

## 2.7 fBm, fGn και οι συναρτήσεις συσχέτισης

Μία κλασματική κίνηση Brown μπορεί να ορισθεί ως μερικό άθροισμα της fGn:

$$B_k^H = X_1 + X_2 + X_3 + \dots + X_k \quad (2.7.1)$$

Οπότε έχουμε:

$$E[(B_h^H - B_k^H)^2] = E[(h - k)^{2H} X_1^2] = (h - k)^{2H} \sigma^2 \quad (2.7.2)$$

οπού έχουμε υποθέσει, χωρίς απώλειες γενικά, ότι  $E[X_1] = 0$  με  $\sigma^2 = \text{Var}[X_1]$

Με ανάπτυξη της εξίσωσης (2.7.2) παίρνουμε

$$\begin{aligned} E[(B_h^H)^2] - 2E[B_h^H B_k^H] + E[(B_k^H)^2] &= h^{2H} \sigma^2 - 2E[B_h^H B_k^H] + k^{2H} \sigma^2 \\ &= (h - k)^{2H} \sigma^2 \end{aligned}$$

Επομένως η συσχέτιση για την fBm είναι:

$$E[B_h^H B_k^H] = \frac{\sigma^2}{2} [h^{2H} - (h - k)^{2H} + k^{2H}]. \quad (2.7.3)$$

Η αυτοσυσχέτιση για τον fGn είναι:

$$\begin{aligned}
E[X_n X_m] &= E[(B_n^H - B_{n-1}^H)(B_m^H - B_{m-1}^H)] \\
&= E[B_n^H B_m^H + B_{n-1}^H B_{m-1}^H - B_{n-1}^H B_m^H - B_n^H B_{m-1}^H] \\
&= \frac{\sigma^2}{2} [(n-m-1)^{2H} - 2(n-m)^{2H} + (n-m+1)^{2H}]
\end{aligned}$$

Χρησιμοποιώντας τη σταθερότητα της  $X$  παίρνουμε:

$$E[X_0 X_k] = \frac{\sigma^2}{2} [(k-1)^{2H} - 2k^{2H} + (k+1)^{2H}]. \quad (2.7.4)$$

Αυτό το αποτέλεσμα μπορεί άμεσα να επιβεβαιωθεί χρησιμοποιώντας την αυτοομοιότητα:

$$\frac{1}{N^{2H}} E \left[ \left( \sum_{l=0}^{N-1} X_l \right)^2 \right] = E[X_0^2] = \sigma^2.$$

Μπορούμε επιπλέον να αποδείξουμε ότι:

$$\frac{1}{N^{2H}} E \left[ \sum_{s=kN}^{(k+1)N-1} X_s \sum_{t=0}^{N-1} X_t \right] = E[X_0 X_k]$$

Ισχύει για κάθε ακέραιο  $N \geq 1$ , η συνάρτηση αυτοσυσχέτισης  $\rho(k) = E[X_0 X_k]$  μονοσήμαντα ικανοποιεί:

$$\rho(k) = N^{-2H} \sum_{l=-(N-1)}^{N-1} (N - |l|) \rho(Nk + l). \quad (2.7.5)$$

Για μία γκαουσιανή στοχαστική διαδικασία, όλες οι ενωμένες κατανομές για την  $B^H$  είναι πολυμεταβαλλόμενες γκαουσιανές κατανομές. Για οποιοδήποτε πολυμεταβαλλόμενες γκαουσιανές τυχαίες μεταβλητές  $\xi_1, \xi_2, \dots, \xi_N$  με  $E[\xi_k] = 0$  ( $k = 1, 2, \dots, N$ ),

οι αποκλίσεις με τάξη μεγαλύτερη από 2 δύνανται όλες να εκφραστούν με όρους συσχέτισης 2<sup>ης</sup> τάξης γνωστή ως θεώρημα Isserlis ή θεώρημα Wick στη στατιστική φυσική:

$$E[\xi_1, \xi_2, \dots, \xi_N] = \sum_P \prod_{j=1}^N E[\xi_j, \xi_{P_j}], \quad (2.7.6)$$

όπου το άθροισμα περνάει από όλες τις μεταθέσεις  $P: j \rightarrow P_j$  για όλους τους θετικούς ακεραίους  $j$ . Δηλαδή,

$$E[B_h^H B_i^H B_j^H B_k^H] = E[B_h^H B_i^H] E[B_j^H B_k^H] + E[B_h^H B_j^H] E[B_i^H B_k^H] + E[B_h^H B_k^H] E[B_i^H B_j^H] \quad (2.7.7)$$

## 2.8 Φάσμα Ισχύος της fBm και fGn

Η εξίσωση (2.7.2) οδηγεί σε μία από τις ιδιότητες του fGn: το φάσμα ισχύος με την ιδιομορφία

$$S(f) = \sum_{k=-\infty}^{\infty} \rho(k) e^{-2\pi i f k}, \quad (2.8.1)$$

έχει αντίστροφο μετασχηματισμό

$$\int_{-1/2}^{1/2} S(f) e^{2\pi i f k} df = \rho(k). \quad (2.8.2)$$

Η ισότητα

$$\sum_{k=-\infty}^{\infty} e^{-2\pi i f k} = \sum_{k=-\infty}^{\infty} \delta(f - k) \quad (2.8.3)$$

η οποία ευθύνεται για φαινόμενα θορύβου (alias), για κάθε θετικό  $N$  και συχνότητα  $f \in [-1/2, 1/2]$  παίρνουμε:

$$\begin{aligned}
N^{2H}S(f) &= \sum_{l=-(N-1)}^{N-1} (N - |l|) \sum_{k=\lfloor f-N/2 \rfloor + 1}^{\lfloor f+N/2 \rfloor} \frac{1}{N} S\left(\frac{f-k}{N}\right) e^{2\pi i l((f-k)/N)} \\
&= \sum_{k=\lfloor f-N/2 \rfloor + 1}^{\lfloor f+N/2 \rfloor} \frac{1}{N} S\left(\frac{f-k}{N}\right) \sum_{l=-(N-1)}^{N-1} (N - |l|) e^{2\pi i l(f-k)/N} \\
&= \sum_{k=\lfloor f-N/2 \rfloor + 1}^{\lfloor f+N/2 \rfloor} \frac{1}{N} S\left(\frac{f-k}{N}\right) \frac{1 - \cos(2\pi(f-k))}{1 - \cos(2\pi(f-k)/N)}
\end{aligned}$$

Και συνοπτικά,

$$\sum_{k=\lfloor f-N/2 \rfloor + 1}^{\lfloor f+N/2 \rfloor} \frac{S(f-k/N)}{1 - \cos(2\pi(f-k)/N)} = \frac{N^{2H+1}}{1 - \cos(2\pi f)} S(f). \quad (2.8.1)$$

Από τη δυναμοσειρά (2.8.1), παίρνουμε

$$S(f) = C(1 - \cos(2\pi f)) \sum_{m=-\infty}^{\infty} \frac{1}{|f+m|^{2H+1}}$$

που ισχύει για οποιοδήποτε  $C$ .

## 2.9 fBm και fGn συνεχούς χρόνου

Ένας κλασματικός θόρυβος Gauss συνεχούς χρόνου είναι μία σταθερή στοχαστική διαδικασία με τις ακόλουθες ιδιότητες:

$$\frac{1}{T^H} \int_0^T X(t) dt \sim X(0) \quad (T > 0),$$

Όπου ‘ $\sim$ ’, δηλώνει ισότητα στην πιθανότητα κατανομής,  $E[X(t)] = 0$ , και  $E[X^2] = \sigma^2$ . Παράλληλα με την παραγωγή της (2.7.3), έχουμε

$$E[B_t^H B_\tau^H] = \frac{\sigma^2}{2} [t^{2H} - 2(\tau - t)^{2H} + \tau^{2H}] \quad (2.9.1)$$

$$\text{όπου} \quad B_t^H = \int_0^t X(\xi) d\xi \quad (2.9.2)$$

είναι μία κλασματική κίνηση Brown συνεχούς χρόνου. Από τη (2.9.2) ο κλασματικός θόρυβος Gauss συνεχούς χρόνου δύναται να γραφεί ως:  $X(t) = dB_t^H/dt$ . Οπότε η συμμεταβλητότητα (covariance) της  $X(t)$  μπορεί να εξαχθεί από την (2.9.1):

$$\begin{aligned} E[X(t) X(\tau)] &= \frac{d^2}{dt d\tau} E[B_t^H B_\tau^H] = -\sigma^2 \frac{d^2 |t - \tau|^{2H}}{dt d\tau} \\ &= 2H(2H - 1) \sigma^2 |t - \tau|^{2H-2} - 2H\sigma^2 |t - \tau|^{2H-1} \frac{d^2 |t - \tau|}{dt d\tau} \\ &= 2H(2H - 1) \sigma^2 |t - \tau|^{2H-2} + 2H\sigma^2 |t - \tau|^{2H-1} \delta(t - \tau), \end{aligned}$$

και από τη στιγμή που η  $X(t)$  είναι σταθερή, έχουμε συνάρτηση αυτοσυσχέτισης για τον κλασματικό θόρυβο Gauss συνεχούς χρόνου:

$$\rho(\tau) = E[X(0)X(\tau)] = 2H(2H - 1) \sigma^2 |\tau|^{2H-2} + 2H\sigma^2 |\tau|^{2H-1} \delta(\tau). \quad (2.9.3)$$

Για  $H = 0.5$  (λευκός θόρυβος Wiener), ο πρώτος όρος στην εξίσωση (2.9.3) είναι μηδέν και ο δεύτερος όρος είναι η συνάρτηση Dirac. Για  $H < 0.5$  ο δεύτερος όρος είναι μηδέν. Η ολοκλήρωση της  $\rho(t)$  δίνει:

$$\int_{-\infty}^{\infty} \rho(t) dt = \begin{cases} \infty & 0.5 < H < 1 \\ \sigma^2 & H = 0.5 \\ 0 & 0 < H < 0.5. \end{cases}$$

Για  $H < 0.5$  η  $\rho(t)$  είναι αρνητική για όλα τα  $t \neq 0$ . Η ολοκλήρωση και των δύο όρων στην εξίσωση (2.9.3) δεν συγκλίνει με τον παραδοσιακό τρόπο. Δύναται να πάρουμε μηδέν με την ακύρωση των δύο απείρων  $\infty$ .



Η φασματική συνάρτηση πυκνότητας για τον συνεχούς χρόνου κλασματικό θόρυβο Gauss υπολογίζεται ως εξής:

$$S(f) = \int_{-\infty}^{\infty} \rho(t) e^{-2\pi i f t} dt \quad (2.9.4)$$

Για  $H > 0.5$ , έχουμε:

$$S(f) = 2H(2H - 1) \sigma^2 C f^{1-2H} \quad (2.9.5)$$

όπου η σταθερά ισούται με

$$C = \int_{-\infty}^{\infty} |\xi|^{2H-2} e^{-2\pi i \xi} d\xi.$$

Επομένως, ο συνεχούς χρόνου κλασματικός θόρυβος Gauss έχει μία απλή εκθετική φασματική συνάρτηση πυκνότητας σε όλες τις συχνότητες  $f$ .

Για μη-συσχετιζόμενο (anti-correlated) κλασματικό θόρυβο Gauss με  $H < 0.5$ , παρατηρούμε ότι ο μετασχηματισμός Fourier της  $\rho(t)$  δεν υπάρχει με την παραδοσιακή έννοια.

## 2.10 Η φρακταλική γεωμετρία της fBm και fGn

Υπάρχουν πολλά φρακταλικά γεωμετρικά χαρακτηριστικά που σχετίζονται με μία fBm ή με τον fGn. Για παράδειγμα, η γραφική απεικόνιση της φρακταλικής διάστασης της fBm στις  $(E + d)$  διαστάσεις είναι  $(2 - H)$ . Επίσης, η φρακταλική διάσταση μπορεί να ληφθεί από τη δειγματοληπτική διαδρομή του Ευκλείδειου χώρου διάστασης  $E$ . Ας θεωρήσουμε μία τροχιά  $N$ -βημάτων μιας γκαουσιανής διαδικασίας με μέσο μηδέν και διακύμανση  $\rho_n = \sigma^2 n^{2H}$  ( $n=1,2,\dots,N$ ). Αυτό ορίζει μία fBm με συντελεστή Hurst  $H$  ( $0 < H < 1$ ). Γι' αυτό το λόγο, μέσα στον Ευκλείδειο χώρο διάστασης  $E$  η πιθανότητα της fBm να ξεκινάει από την πηγή και στο  $\nu$ -οστό βήμα της να φτάνει σε μία σφαίρα  $\varepsilon$  σε ένα χώρο γύρω από το  $x (x \in R^E)$  είναι:

$$\frac{\exp[-x^2/(2\rho_n)]}{(2\pi\rho_n)^{E/2}} V_\varepsilon, \quad (2.10.1)$$

όπου  $V_\varepsilon$  είναι ο όγκος της σφαίρας  $\varepsilon$ . Τα σημεία  $N$  έχουν μία ακτίνα περιστροφής  $\sim N^H$ , όπου στατιστικά χαρακτηρίζει το μέσο μέγεθος όλων των δυνατών τροχιών  $N$  βημάτων. Αντιθέτως, για μία δοθείσα σφαίρα ακτίνας  $R$ , το μέσο μήκος της τροχιάς μέσα στη σφαίρα είναι  $L \sim R^{1/H}$ . Επομένως, σύμφωνα με την έννοια του Mandelbrot για ένα σύνολο φράκταλ, η τροχιά έχει μία φρακταλική διάσταση

$$d_f = \frac{d \ln(L)}{d \ln(R)} = 1/H. \quad (2.10.2)$$

Τελικά μπορεί μία fBm να φτάσει σε κάθε σημείο του διαστασιακού χώρου  $E$ ; Για να απαντήσουμε σε αυτήν την ερώτηση, πρέπει να σημειώσουμε ότι εάν το  $\varepsilon$  είναι επαρκώς μικρό, μπορούμε επιπλέον να υποθέσουμε ότι η εξίσωση (2.10.1) είναι η πιθανότητα της κλασματικής κίνησης Brown να φτάσει την σφαίρα  $\varepsilon$  την πρώτη στιγμή από την πιθανότητα να την φτάσει σε κάποιο δεύτερο χρόνο είναι  $\propto V_\varepsilon^2$ . Επομένως έχουμε την πιθανότητα της fBm άσχετα από το  $n$ , να φτάσει στη γειτονιά του  $x$

$$V_\varepsilon \sum_{n=1}^{\infty} \frac{\exp[-x^2/(2\rho_n)]}{(2\pi\rho_n)^{E/2}}, \quad \text{όπου ασυμπτωτικά,} \quad \sim V_\varepsilon \sum_{n=1}^{\infty} \frac{1}{n^{HE}} \quad (2.10.3)$$

Για επαρκώς μικρό  $\varepsilon$ , αυτή η πιθανότητα θα είναι λιγότερο από μονάδα αν  $HE > 1$ . Επομένως, υπάρχει μία πεπερασμένη πιθανότητα όπου η fBm δεν θα φτάσει την γειτονιά του  $x$ . Σημειώνουμε ότι  $H = 1/d_f$  οπότε  $d_f < E$  δηλώνει ότι η φρακταλική διάσταση της τροχιάς είναι μικρότερη από τον Ευκλείδιο χώρο. Αντιθέτως, αν  $d_f \geq E$  τότε η fBm με πιθανότητα 1 θα επισκευτεί οποιαδήποτε μικρή σφαίρα  $\varepsilon$  στο κέντρο του  $x$ .

Αυτό το αποτέλεσμα έχει μία ισχυρή φυσική ερμηνεία: εάν η δειγματοληπτική διαδρομή μιας συσχετιζόμενης τυχαίας διαδρομής (random walk) έχει μία φρακταλική διάσταση  $d_f$ , δύναται να γεμίσει τον Ευκλείδιο χώρο με διάσταση  $E \leq d_f$ , αλλά όχι για  $E > d_f$ . Για την κλασική τυχαία διαδρομή με  $H = 0.5$  το παραπάνω αποτέλεσμα είναι γνωστό ως θεώρημα Polya, το οποίο δηλώνει ότι η

τυχαία διαδρομή θα επισκευτεί οποιοδήποτε χώρο μέσα σε μία ή δύο διαστάσεις, αλλά όχι σε τρεις.

Η μαθηματική έρευνα της κλασματικής κίνησης Brown αναδεικνύει εκ των έσω το πρόβλημα της αυτο-αποφυγής της τυχαίας διαδρομής (self-avoid random walk SAW) της οποίας η τροχιά δεν μπορεί να τέμνει τον εαυτό της στο χώρο. Αυτό παραμένει ένα σημαντικά άλυτο πρόβλημα στη στατιστική φυσική των πολυμερών. Είναι γνωστό ότι η αυτο-αποφυγή της τυχαίας διαδρομής έχει ασυμπτωτική διακύμανση  $\sim n^6/(E+2)$ , όπου  $E$  είναι η διάσταση του Ευκλείδειου χώρου της τυχαίας διαδρομής. Παρόλο που η κλασματική κίνηση Brown δεν είναι πιστό μοντέλο για την SAW αυτά τα δύο μοντέλα συμφωνούν ασυμπτωτικά αν  $H = 3/(E+2)$ . Για  $E=2$  αντιστοιχεί σε  $H = 3/4$ . Αυτά τα αποτελέσματα προσδιορίζουν μία προσεγγιστική μέτρηση και είναι πολύ στενά σχετιζόμενα με τη μαθηματική έννοια του διατμηματικού τοπικού χρόνου (intersectional local time), στον οποίο η ιδιαιτερότητα του  $H = 3/4$  είναι αξιοσημείωτη [27].

## 2.11 Εξάρτηση μεγάλου εύρους (Long range dependence)

Η στατιστική των φράκταλς από τις φυσικές διαδικασίες συχνά παρουσιάζει δύο ιδιότητες: (1) την τάση να έχει μεγάλο εύρος διακυμάνσεων που απαντώνται στα διακριτά άλματα και (2) την τάση για χρονοσειρές να έχουν στατιστικά χαρακτηριστικά και επαναλήψεις ενώ παραμένουν χαοτικές. Από τη στιγμή που η κλασματική κίνηση Brown είναι αυτό-όμοια (self-similar) με σταθερές αυξήσεις (stationary increments) με δεικτη  $H$  ( $H$ -sssi), για οποιοδήποτε  $0 < H < 1$ , με εφαρμογή του κριτηρίου του Kolmogorov (μία στοχαστική διαδικασία  $\{X(t), t \in \mathbb{R}\}$  έχει μία εκδοχή με συνεχείς διαδρομές αν υπάρχει  $p \geq 1, \eta \geq 1$  και μία σταθερά  $c$  τέτοια ώστε  $\forall t_1, t_2$  έχουμε  $E|X(t_2) - X(t_1)|^p \leq c|t_2 - t_1|^\eta$ ) στην fBm  $B_H(t)$  με  $p \geq 1/H$ , παίρνουμε:

$$E|B_H(t_2) - B_H(t_1)|^p = E|B_H(1)|^p |t_2 - t_1|^{pH} \quad (2.11.1)$$

Που σημαίνει ότι οι διαδρομές της fBm είναι συνεχής. Ωστόσο,  $E \left| \frac{B_H(t_2) - B_H(t_1)}{t_2 - t_1} \right|^2 = \sigma^2 |t_2 - t_1|^{2H-2} \rightarrow \infty$ , όταν  $t_1 \rightarrow t_2$  με την έννοια του  $L^2(\Omega)$  σημαίνει ότι η fBm δεν είναι διαφορίσιμη.

Συνεπώς, από την προηγούμενη εξίσωση, συμπεραίνουμε ότι οι διαδρομές της fBm έχουν λιγότερα ζιγκ-ζαγκ καθώς το  $H$  τείνει από το 0 στο 1. Συγκεκριμένα, όταν  $0 < H < 1/2$  η αυτοσυσχέτιση των αυξήσεων της fBm είναι αρνητική, οπότε οι αυξήσεις της fBm έχουν αρνητικό πρόσημο. Γι' αυτό το λόγο, το σωματίδιο τείνει να επιστρέφει. Μια τέτοια fBm καλείται αντι - παραμένουσα (anti-persistent). Αντιθέτως, όταν  $1/2 < H < 1$  η αυτοσυσχέτιση των αυξήσεων της fBm είναι θετική, οπότε το σωματίδιο τείνει να επιμένει στην ίδια κατεύθυνση. Μία τέτοια fBm καλείται παραμένουσα (persistent).

Πράγματι, με τα παραμένον και μη παραμένον χαρακτηριστικά, η fBm έχει άπειρη εξάρτηση από το παρελθόν, οπότε δεν διαχωρίζεται καθαρά σε περιοδικές συνιστώσες και συνιστώσες τυχαίου θορύβου. Γι' αυτό το λόγο προκειμένου να παρεμβληθεί η αναμενόμενη τιμή της συνάρτησης σε ένα σημείο, πρέπει να γνωρίζουμε την τιμή σε χρόνους πολύ πριν ή μετά από τα δύο τελευταία χρονικά σημεία. Αυτό αποτελεί ακόμα μία μεγάλη διαφορά μεταξύ της κλασματικής κίνησης Brown και της συνήθους κίνησης Brown.

Οι ιδιότητες παραμονής (persistence) και μη-παραμονής (anti-persistence) είναι γνωστές ως εξάρτηση μεγάλου εύρους (long-range dependence) [29].

## ΚΕΦΑΛΑΙΟ 3

### ΧΡΟΝΟΣΕΙΡΕΣ

Υπάρχουν διάφοροι τρόποι συλλογής δεδομένων. Αυτό συμβαίνει ιδιαίτερα στην περίπτωση όπου επαναλαμβανόμενες παρατηρήσεις ενός δεδομένου συστήματος πραγματοποιούνται διαδοχικά στο χρόνο. Η ανάλυση χρονοσειρών ασχολείται με βραχυπρόθεσμες και μακροπρόθεσμες αλλαγές και με μετρήσεις συσχέτισης ή εξάρτησης μεταξύ παρελθόντος και μέλλοντος. Τα δεδομένα που συγκεντρώνονται ακολουθιακά στο χρόνο καλούνται χρονοσειρές. Οι χρονοσειρές προκύπτουν με φυσικό τρόπο σε πολλά πεδία, όπως:

- στην οικονομία
- στη μοντελοποίηση του περιβάλλοντος
- στη δημογραφία
- στη μηχανική
- στον έλεγχο ποιότητας

Οι χρόνοι παρατήρησης δεν είναι απαραίτητο να είναι ίσοι. Οι παρατηρήσεις δύναται να παίρνουν τιμές από διακριτές κατανομές. Εάν παρατηρούμε με επανάληψη ένα δεδομένο σύστημα σε τακτά χρονικά διαστήματα είναι πολύ πιθανό οι παρατηρήσεις που θα κάνουμε να είναι συσχετιζόμενες (correlated). Οπότε δεν μπορούμε να θεωρήσουμε ότι τα δεδομένα συνιστούν ένα τυχαίο δείγμα. Η χρονική διαδοχή μέσα στην οποία πραγματοποιούνται οι παρατηρήσεις είναι κρίσιμη. Οι χρονοσειρές μπορούν να θεωρηθούν από δύο πλευρές: στο πεδίο του χρόνου και στο πεδίο της συχνότητας. Η ανάλυση στο πεδίο του χρόνου βασίζεται στις συναρτήσεις αυτοσυσχέτισης και αντίστροφης συσχέτισης (cross-correlation functions) για να περιγράψει και να εξηγήσει την αστάθεια των χρονοσειρών. Στο πεδίο της συχνότητας αναπαρίστανται προσωρινές διακυμάνσεις ως άθροισμα ημιτονοειδών συνιστωσών και οι συναρτήσεις συσχέτισης και αντίστροφης συσχέτισης αντικαθίστανται από τους μετασχηματισμούς Fourier, οι οποίοι είναι γνωστοί ως συναρτήσεις φασματικής και αντίστροφης φασματικής πυκνότητας (γνωστή και ως φασματική ανάλυση).

Η φασματική ανάλυση χρησιμοποιείται για την ανίχνευση κρυμμένων περιόδων. Με την ανάλυση των χρονοσειρών επιδιώκεται:

- περιγραφή – στατιστικός προσανατολισμός και απεικόνιση σε διάγραμμα
- ανάλυση και ερμηνεία – εύρεση του μοντέλου για την περιγραφή του χρόνου εξάρτησης των δεδομένων
- πρόβλεψη, από ένα δείγμα της σειράς, της επόμενης τιμής ή των επόμενων τιμών
- έλεγχος και προσαρμογή διαφόρων παραμέτρων ώστε να διαμορφώσουμε τις σειρές να βρίσκονται κοντά σε κάποιο στόχο
- προσαρμογή - σε ένα γραμμικό μοντέλο τα λάθη δύναται να σχηματίσουν μία χρονοσειρά από συσχετιζόμενες παρατηρήσεις και έτσι μπορούμε να διευθετήσουμε τις διακυμάνσεις που προκύπτουν.

### 3.1 Βασικοί ορισμοί

Υποθέτουμε ότι οι σειρές  $X_t$  τρέχουν κατά μήκος του χρόνου, αυτό σημαίνει ότι  $(X_t)_{t=0, \pm 1, \pm 2, \dots}$ , ωστόσο παρατηρείται μόνο σε χρόνους  $t = 1, \dots, n$ , οπότε έχουμε  $(X_1, \dots, X_n)$ . Η θεωρία των χρονοσειρών βασίζεται στην υπόθεση της ‘στασιμότητας 2<sup>ης</sup> τάξης’. Τα δεδομένα που λαμβάνουμε από τη φύση συνήθως δεν είναι στάσιμα: εκδηλώνουν μία γραμμική τάση (trend) – μεγάλης διάρκειας χρονικές μετατοπίσεις μέσα στο μέσο (mean), ή εκδηλώνονται εποχιακά (seasonal) [19] - κυκλικές διακυμάνσεις αναφορικές με ημερολόγιο. Οι μη στάσιμες χρονοσειρές παρουσιάζουν συστηματικά τάσεις που μπορεί να είναι γραμμικές, δευτεροβάθμιες κοκ. Οι μη στάσιμες χρονοσειρές που μετασχηματίζονται σε στάσιμες με διαφορίση καλούνται κατά μία έννοια ομοιογενείς μη στάσιμες χρονοσερές. Ο όρος  $d$  της τάσης (trend) χρειάζεται για να μετατρέψει μία μη στάσιμη χρονοσειρά σε στάσιμη (stationary). Ένα μοντέλο χρονοσειράς με δύο όρους τάσης ( $d = 2$ ) χρειάζεται να διαφορισθεί δύο φορές για να γίνει στάσιμη. Η πρώτη διαφορίση αφαιρεί τη γραμμική τάση, η δεύτερη τη δευτεροβάθμια τάση κοκ. [23]. Επίσης, στην ανάλυση των χρονοσειρών ο κύκλος εργασίας (business cycle) - κυκλικές διακυμάνσεις - παίζει σημαντικό ρόλο. Εάν η περίοδος έχει οριστεί τότε ο κύκλος δύναται να αναπαρασταθεί από ένα εποχιακό (seasonal) ή περιοδικό μοντέλο. Για την ντετερμινιστική συνάρτηση  $f(\cdot)$  λέμε ότι

είναι περιοδική με περίοδο  $s$  εάν ισχύει  $f(t) = f(t + k \times s)$   $k = 0, \pm 1, \pm 2, \dots$   
 Τυπικό παράδειγμα αποτελούν οι τριγωνομετρικές σειρές πχ  $\cos(\theta) = \cos(\theta + 2k\pi)$ .  
 Μια στοχαστική διαδικασία  $X_t$  καλείται εποχιακή (seasonal) χρονοσειρά με περίοδο  $s$  αν  $X_t$  και  $X_{t+ks}$  έχουν την ίδια κατανομή [26].

Τα Υπόλοιπα (Residuals)  $E_t$  είναι άλλες τυχαίες ή συστηματικές διακυμάνσεις [19].

### 3.2 Στασιμότητα και αυτοδιακύμανση

Η διαδικασία καλείται ασθενής στασιμότητα (weakly stationary) ή 2<sup>ης</sup> τάξης στασιμότητα (second order stationary) αν για όλους τους ακεραίους  $t, \tau$  ισχύει:

$$E(X_t) = \mu \text{ και } Cov(X_{t+\tau}, X_t) = \gamma_t \text{ όπου } \mu \text{ και } \gamma_t \text{ δεν εξαρτάται από το } \tau.$$

Η διαδικασία είναι αυστηρά στάσιμη αν  $(X_{t_1}, \dots, X_{t_k})$  και  $(X_{t_1+\tau}, \dots, X_{t_k+\tau})$  έχουν την ίδια κατανομή για όλα τα σύνολα των χρονικών σημείων  $t_1, \dots, t_k$  και για όλους τους ακεραίους  $\tau$ . Η διαδικασία που είναι αυστηρά στάσιμη, αυτόματα είναι ασθενώς στάσιμη. Το αντίστροφο γενικά δεν ισχύει. Ωστόσο, αν η διαδικασία είναι γκαουσιανή, δηλαδή αν  $(X_{t_1}, \dots, X_{t_k})$  έχει πολλαπλές μεταβλητές κανονικής κατανομής για όλα τα  $t_1, \dots, t_k$ , τότε η ασθενής στασιμότητα συνεπάγεται ισχυρή στασιμότητα. Σημειώνουμε ότι  $var(X_t) = \gamma_0$ , και από στασιμότητα,  $\gamma - t = \gamma_t$ . Η ακολουθία  $(\gamma_t)$  καλείται συνάρτηση αυτοδιακύμανσης (autocovariance function).

Η συνάρτηση αυτοσυσχέτισης (autocorrelation function ACF)  $\rho_t$  δίνεται από τη σχέση

$$\rho_t = corr(X_{t+\tau}, X_t) = \frac{\gamma_t}{\gamma_0}. \quad (3.2.1)$$

Η ACF περιγράφει τις ιδιότητες 2<sup>ης</sup> τάξης των χρονοσειρών.

Υπολογίζουμε την αυτοδιακύμανση  $\gamma_t$  ως  $c_t$  και την αυτοσυσχέτιση  $\rho_t$  ως  $r_t$ , όπου

$$c_t = \frac{1}{n} \sum_{s=\max(1,1-t)}^{\min(n-t,n)} [X_{s+t} - \bar{X}][X_s - \bar{X}] \quad \text{και} \quad r_t = \frac{c_t}{c_0}.$$

Για  $t > 0$  η διακύμανση  $Cov(X_{t+t}, X_t)$  υπολογίζεται από  $n - t$  παρατηρηθέντα ζευγάρια  $(X_{t+1}, X_1), \dots, (X_n, X_{n-t})$ . Εάν υπολογίσουμε τη συνδιακύμανση αυτών των ζευγαριών θα χρησιμοποιούσαμε διαφορετικές εκτιμήσεις του μέσου (mean) και των διακυμάνσεων για κάθε ένα από τα υποσύνολα των σειρών  $(X_{t+1}, \dots, X_n)$  και  $(X_1, \dots, X_{n-t})$ , αν και έχουμε υποθέσει ότι είναι στάσιμες έχουν τον ίδιο μέσο και διακύμανση. Το  $\bar{X}$  δηλώνει την προηγούμενη σχέση.

Ένα διάγραμμα της  $r_t$  ως προς  $t$  καλείται διάγραμμα συσχέτισης (correlogram).

Μία χρονοσειρά  $(X_t)$  θεωρείται ότι βρίσκεται σε καθυστέρηση (lagged) εάν ο άξονας του χρόνου είναι μετατοπισμένος κατά  $\tau$  καθυστερήσεις δίνοντας μία σειρά  $(X_{t-\tau})$ .

Οπότε η  $r_t$  είναι η εκτιμηθείσα αυτοσυσχέτιση σε καθυστέρηση  $t$ , επίσης καλείται συνάρτηση δειγματοληπτικής αυτοσυσχέτισης [19].

### 3.3 Μοντέλα χρονοσειρών γραμμικής διαδικασίας

Η διαδικασία  $(X_t)$  καλείται γραμμική εάν έχει έκφραση της μορφής

$$X_t = \mu + \sum_{r=-\infty}^{\infty} c_r \epsilon_{t-r} \quad (3.3.1)$$

όπου  $\mu$  δηλώνει τον μέσο (mean),  $c_r$  είναι η ακολουθία από προσδιορισμένες σταθερές και  $\epsilon_t$  είναι η ανεξάρτητη τυχαία μεταβλητή με μέσο  $0$  και απλή διακύμανση. Θεωρούμε ότι  $\sum c_r^2 < \infty$  ώστε η διακύμανση  $X_t$  να είναι πεπερασμένη. Αν η  $\epsilon_t$  είναι όμοια κατανομημένη, τότε μία τέτοια διαδικασία είναι αυστηρά στάσιμη. Αν η  $c_r = 0$  για  $r < 0$  χαρακτηρίζεται ως τυχαία (casual), για παράδειγμα η διαδικασία σε χρόνο  $t$  δεν εξαρτάται από το μέλλον καθώς δεν έχουν παρατηρηθεί οι τιμές της  $\epsilon_t$ .

Οι διαδικασίες **AR**, **MA**, **ARMA**, και **ARIMA** είναι όλες ειδικές περιπτώσεις της τυχαίας γραμμικής διαδικασίας και ενσωματώνουν ένα μεγάλο εύρος στάσιμων (stationary) και μη στάσιμων (non stationary) χρονοσειρών οι οποίες έχουν πολλές πρακτικές εφαρμογές. Όλες οι διαδικασίες **MA** είναι στάσιμες. Αντιθέτως, οι χρονοσειρές που προκύπτουν από το μοντέλο **ARIMA** είναι μη στάσιμες. Τα αμιγώς **AR** μοντέλα σε συνδυασμό με τα **ARMA** μοντέλα είναι είτε στάσιμα είτε μη στάσιμα [19].



Οι Box και Jenkins (1976) ανέπτυξαν μία βασική ιδέα για την προσαρμογή των μοντέλων *AR*, *MA*, *ARMA*, και *ARIMA*, η οποία είναι γνωστή ως μοντέλο Box – Jenkins.

Η διαδικασία έχει τρία βασικά βήματα: (1) προσδιορισμός μοντέλου (πχ. επιλογή παραμέτρου  $p$ ,  $d$  και  $q$ ), (2) εκτίμηση μοντέλου (πχ. εκτίμηση των παραμέτρων  $\phi_1, \phi_2, \dots, \phi_p$  και  $\theta_1, \theta_2, \dots, \theta_q$ ), και (3) εξακρίβωση μοντέλου.

Επειδή τα μοντέλα έχουν ιδανικά χαρακτηριστικά αυτοσυσχέτισης και εξαρτώνται από τις τιμές των  $p$ ,  $d$  και  $q$  το διάγραμμα συσχέτισης (correlogram) είναι ένα σημαντικό εργαλείο για την ταυτοποίηση του μοντέλου. Η συνάρτηση αυτοσυσχέτισης γενικά χρησιμοποιείται για τη σύνδεσή της με την συνάρτηση μερικής αυτοσυσχέτισης (partial autocorrelation function PACF), η οποία μετράει την ποσότητα αυτοσυσχέτισης που παραμένει απροσδιόριστη μετά την εφαρμογή της διαδικασίας της αυτομείωσης (autoregressive process) τάξεως  $k = 1, 2$  (πχ.  $PACF(k)$  είναι η ποσότητα αυτοσυσχέτισης που δεν μπορεί να εξηγηθεί από τη διαδικασία της αυτομείωσης τάξεως  $k$ ). Αυτές οι δύο συναρτήσεις παρέχουν συμπληρωματικές πληροφορίες για το εν λόγω μοντέλο: για μία  $AR(p)$  διαδικασία, η ACF ελαττώνεται και η PACF σταματάει να υπολογίζει μετά από μία καθυστέρηση  $p$  (πχ. η PACF είναι μηδέν εάν η τάξη της διαδικασίας αυτομείωσης είναι μεγαλύτερη ή ίση από μία ορισμένη τιμή  $p$ ), για μία διαδικασία  $MA(q)$ , η ACF σταματάει να υπολογίζει μετά από μία καθυστέρηση  $q$  και η PACF ελαττώνεται. Ένα άλλο εργαλείο για την ταυτοποίηση του μοντέλου είναι η συνάρτηση ανάστροφης αυτοσυσχέτισης (inverse autocorrelation function IACF) μίας αυτομειωμένης διαδικασίας μεταβαλλόμενου μέσου όρου, η οποία είναι η ACF «αναστρεφόμενης» διαδικασίας ληφθείσα από την ανταλλαγή των παραμέτρων  $\phi_1, \phi_2, \dots, \phi_p$  και  $\theta_1, \theta_2, \dots, \theta_q$  και άλλων διάφορων μοντέλων [11].

### 3.4 Αυτομειωμένη διαδικασία (Autoregressive process)

Η αυτομειωμένη διαδικασία (η μεταβλητή  $X$  μεώνεται από τον εαυτό της) τάξης  $p$  (δηλώνει το βαθμό καθυστέρησης) δηλώνεται ως  $AR(p)$  και ορίζεται ως:

$$X_t = \sum_{r=1}^p \phi_r X_{t-r} + \epsilon_t \quad (3.4.1)$$

όπου  $\phi_1 \dots \phi_p$  είναι προσδιορισμένες σταθερές και  $\epsilon_t$  είναι μία χρονοσειρά λευκού θορύβου (η ακολουθία  $\epsilon_t$  αποτελείται από ανεξάρτητες ή μη συσχετιζόμενες τυχαίες μεταβλητές με μέσο 0 και διακύμανση  $\sigma^2$  που καλείται λευκός θόρυβος. Είναι μία δεύτερης τάξης σειρά στασιμότητας με  $\gamma_0 = \sigma^2$  και  $\gamma_k = 0, k \neq 0$ ). Η καθυστέρηση είναι 1.

Η διαδικασία **AR(1)** ορίζεται ως

$$X_t = \phi_1 X_{t-1} + \epsilon_t \quad (3.4.2)$$

Για να υπολογίσουμε τη συνάρτηση αυτοδιακύμανσης κάνουμε αντικαταστάσεις της μορφής

$$X_t = \epsilon_t + \phi_1 (\epsilon_{t-1} + \phi_1 (\epsilon_{t-2} + \dots)) = \epsilon_t + \phi_1 \epsilon_{t-1} + \phi_1^2 \epsilon_{t-2} + \dots$$

Το γεγονός ότι η διαδικασία  $X_t$  είναι στάσιμη 2<sup>ης</sup> τάξης προκύπτει από την παρατήρηση της  $\mathbb{E}(x_t) = 0$  και έτσι η συνάρτηση αυτοδιακύμανσης δύναται να υπολογιστεί ως εξής:

$$\gamma_0 = \mathbb{E} \left( \epsilon_t + \phi_1 \epsilon_{t-1} + \phi_1^2 \epsilon_{t-2} + \dots \right)^2 = \left( 1 + \phi_1^2 + \phi_1^4 + \dots \right) \sigma^2 = \frac{\sigma^2}{1 - \phi_1^2}$$

$$\gamma_k = \mathbb{E} \left( \sum_{r=0}^{\infty} \phi_1^r \epsilon_{t-r} \sum_{s=0}^{\infty} \phi_1^s \epsilon_{t+k-s} \right) = \frac{\sigma^2}{1 - \phi_1^2} \quad (3.4.3)$$

Ένας άλλος τρόπος για να πάρουμε αυτά τα αποτελέσματα είναι να πολλαπλασιάσουμε την εξίσωση (3.4.2) με  $X_{t-k}$  και συνυπολογίζοντας την αναμενόμενη τιμή, παίρνουμε:

$$\mathbb{E}(X_t X_{t-k}) = \mathbb{E}(\phi_1 X_{t-1} X_{t-k}) + \mathbb{E}(\epsilon_t X_{t-k}). \quad (3.4.4)$$

Επομένως,  $\gamma_k = \phi_1 \gamma_{k-1}, k = 1, 2, \dots$

Παρομοίως, υψώνοντας στο τετράγωνο την εξίσωση (3.4.2) και συνυπολογίζοντας την αναμενόμενη τιμή, παίρνουμε:

$$\mathbb{E}(X_t^2) = \phi_1 \mathbb{E}(X_{t-1}^2) + 2\phi_1 \mathbb{E}(X_{t-1} \epsilon_t) + \mathbb{E}(\epsilon_t^2) = \phi_1^2 \mathbb{E}(X_{t-1}^2) + 0 + \sigma^2$$

Επομένως  $\gamma_0 = \sigma^2 / (1 - \phi_1^2)$ . Γενικά, η διαδικασία  $AR(p)$  ορίζεται ως

$$X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + \dots + \phi_p X_{t-p} + \epsilon_t. \quad (3.4.5)$$

Τώρα, η συνάρτηση αυτοσυσχέτισης μπορεί να υπολογιστεί πολλαπλασιάζοντας την εξίσωση (3.4.5) με  $X_{t-k}$  συνυπολογίζοντας την αναμενόμενη τιμή και διαιρώντας με  $\gamma_0$  και με αυτόν τον τρόπο εισάγεται η εξίσωση Yule – Walker

$$\rho_k = \phi_1 \rho_{k-1} + \phi_2 \rho_{k-2} + \dots + \phi_p \rho_{k-p}, k = 1, 2, \dots$$

Υπάρχουν γραμμικές σχέσεις επανάληψης, με γενική λύση της μορφής

$$\rho_k = C_1 \omega_1^{|k|} + \dots + C_p \omega_p^{|k|}, \quad (3.4.6)$$

όπου  $\omega_1, \dots, \omega_p$  είναι οι ρίζες της  $\omega^p - \phi_1 \omega^{p-1} - \phi_2 \omega^{p-2} - \dots - \phi_p = 0$

και  $C_1, \dots, C_p$  προσδιορίζεται από την  $\rho_0 = 1$  και τις εξισώσεις για  $k = 1, \dots, p-1$ .

Είναι φυσικά απαραίτητο να χρειάζεται η  $\gamma_k = 0$  καθώς το  $k \rightarrow \infty$  στην οποία περίπτωση οι ρίζες πρέπει να εκτείνονται μέσα στον μοναδιαίο κύκλο, που σημαίνει ότι  $|\omega_i| < 1$ . Γι' αυτό το λόγο, υπάρχει περιορισμός στις τιμές των  $\phi_1, \dots, \phi_p$  που δύναται να επιλεγουν [24].

### 3.5 Διαδικασία μεταβαλλόμενου μέσου όρου (Moving average process)

Η διαδικασία του μεταβαλλόμενου μέσου όρου τάξης  $q$  δηλώνεται ως  $MA(q)$  (η μέγιστη καθυστέρηση  $q$  είναι η τάξη του μεταβαλλόμενου μέσου όρου) καθώς το  $X$  εκφράζεται ως ισοσταθμισμένος μέσος όρος παρελθοντικών τιμών της σειράς λευκού θορύβου και ορίζεται ως

$$X_t = \sum_{s=0}^q \theta_s \epsilon_{t-s} \quad (3.5.1)$$

όπου  $\theta_1, \dots, \theta_q$  είναι προσδιορισμένες σταθερές,  $\theta_0=1$  και  $\epsilon_t$  είναι η ακολουθία των ανεξάρτητων (ή μη συσχετιζόμενων) τυχαίων μεταβλητών με μέσο 0 και διακύμανση  $\sigma^2$ . Εδώ η καθυστέρηση του μεταβαλλόμενου μέσου όρου είναι 1. Μπορούμε να θεωρήσουμε τη σειρά λευκού θορύβου ως μια παρεμβολή ή μία αιφνίδια ταραχή: νέα μη συσχετιζόμενη πληροφορία εμφανίζεται στοχαστικά σε κάθε χρονικό βήμα, το οποίο συνδυαζόμενο με άλλες παρεμβολές ή ταραχές δίνουν την παρατηρούμενη σειρά  $X$ .

Είναι προφανές από τον ορισμό ότι αυτή η διαδικασία είναι  $2^{\text{ης}}$  τάξης στάσιμη και έτσι

$$\gamma_k = \begin{cases} 0, & |k| > q \\ \sigma^2 \sum_{s=0}^{q-|k|} \theta_s \theta_{s+k}, & |k| \leq q \end{cases} \quad (3.5.2)$$

Δύο διαδικασίες μεταβαλλόμενου μέσου όρου δύναται να έχουν την ίδια συνάρτηση αυτοσυσχέτισης. Για παράδειγμα,

$$X_t = \epsilon_t + \theta_{\epsilon_{t-1}} \quad \text{και} \quad X_t = \epsilon_t + (1/\theta)_{\epsilon_{t-1}}$$

και οι δύο έχουν  $\rho_1 = \theta/(1 + \theta^2)$ ,  $\rho_k = 0, |k| > 1$ . Επομένως η πρώτη δίνει

$$\epsilon_t = X_t - \theta_{\epsilon_{t-1}} = X_t - \theta(X_{t-1} - \theta_{\epsilon_{t-2}}) = X_t - \theta X_{t-1} + \theta^2 X_{t-2} - \dots \quad (3.5.3)$$

Αυτό ισχύει μόνο για  $|\theta| < 1$ , γνωστή ως αντιστρέψιμη διαδικασία (invertible process). Επιπλέον, δυο αντιστρέψιμες διαδικασίες δεν μπορούν να έχουν την ίδια συνάρτηση αυτοσυσχέτισης [24].

### 3.6 Διαδικασία αυτομειωμένου μέσου όρου (Autoregressive moving average process)

Η αυτομειωμένη διαδικασία και η διαδικασία μεταβαλλόμενου μέσου όρου δύναται να συνδυασθούν για την παραγωγή ενός τρίτου μοντέλου γνωστού ως διαδικασία αυτομειωμένου μέσου όρου. Η διαδικασία αυτομειωμένου μέσου όρου (autoregressive moving average)  $ARMA(p, q)$  ορίζεται ως

$$X_t - \sum_{r=1}^p \phi_r X_{t-r} = \sum_{s=0}^q \theta_s \epsilon_{t-s} \quad (3.6.1)$$

όπου  $\theta_0 = 1$  και  $\epsilon_t$  είναι λευκός θόρυβος.

Γενικά ένας πιο λεπτός ορισμός της διαδικασίας  $ARMA$  περιλαμβάνει μία μη μηδενική τιμή  $\mu$  για το μέσο και μπορεί να εξαχθεί αντικαθιστώντας την  $X_t$  με  $X_t - \mu$  και την  $X_{t-i}$  με  $X_{t-i} - \mu$ .

Από τον ορισμό της διαδικασίας  $MA(q)$  βλέπουμε ότι για οποιοδήποτε  $\theta_1, \dots, \theta_q$  είναι 2<sup>ης</sup> τάξης στάσιμη. Ωστόσο, οι μέθοδοι  $AR(p)$  και  $ARMA(p, q)$  δεν ορίζουν απαραίτητα στάσιμη χρονοσειρά 2<sup>ης</sup> τάξης (εξαρτάται από τον συνδυασμό της αυτομειωμένης παραμέτρου  $\phi_r$ ). Για παράδειγμα, είδαμε ότι για την  $AR(1)$  είναι απαραίτητος ο περιορισμός  $|\omega_1| < 1$ . Αυτή είναι η συνθήκη στασιμότητας για τη διαδικασία  $AR(1)$ . Όλες οι διαδικασίες  $AR$  απαιτούν αυτού του τύπου τη συνθήκη.

Οριζόμενη για οποιοδήποτε μιγαδικό αριθμό  $z$  το πολυώνυμο αυτομείωσης είναι

$$\phi_\alpha(z) = 1 - \alpha_1 z - \dots - \alpha_p z^p. \quad (3.6.2)$$

Οπότε η συνθήκη στασιμότητας είναι:

όλα τα μηδενικά της συνάρτησης  $\phi_\alpha(z)$  βρίσκονται εκτός μοναδιαίου κύκλου στο μιγαδικό επίπεδο.

Αυτή ακριβώς είναι η συνθήκη που απαιτείται στο  $\alpha_1, \dots, \alpha_p$  να διασφαλίσει ότι η διαδικασία είναι καλά ορισμένη και στάσιμη [19].

### 3.7 Η διαδικασία ARIMA (Auto-Regressive Integrated Moving Average)

Η διαδικασία **ARIMA** (auto-regressive integrated moving average) ορίζεται από τις μεταβλητές  $(p, d, q)$ . Η αναγνώριση ενός μοντέλου χρονοσειράς είναι η διαδικασία εύρεσης ενός ακεραίου, συνήθως πολύ μικρού  $p \in \{0, 1, 2\}$ , για κάθε μία μεταβλητή  $p$ ,  $d$  και  $q$  που μοντελοποιεί το δείγμα των δεδομένων. Όταν η τιμή είναι 0, το στοιχείο δεν χρειάζεται στο μοντέλο. Η μεσαία μεταβλητή  $d$  (δηλώνει την τάξη διαφορίσης) ερευνάται πριν τις  $p$  και  $q$ . Στόχος είναι να ελεγχεί εάν η διαδικασία είναι στάσιμη (stationary) πριν ελεγχούν οι τιμές των καθυστερήσεων  $p$  και  $q$  [23].

Εάν η αρχική διαδικασία  $Y_t$  δεν είναι στάσιμη υπολογίζουμε τη διαφορά 1<sup>ης</sup> τάξης από τη σχέση

$X_t = \nabla Y_t = Y_t - Y_{t-1}$  ενώ τη διαφορά 2<sup>ης</sup> τάξης από την

$$X_t = \nabla^2 Y_t = \nabla(\nabla Y)_t = Y_t - 2Y_{t-1} + Y_{t-2} \text{ κοκ} \quad (3.7.1)$$

Ο γενικός τύπος διαφορίσης είναι  $(1 - B)^d X_t = Y_t$  για  $t = 1, 2, \dots$

όπου  $\{Y_t\}$  μία συνήθης **ARMA**( $p, q$ ) διαδικασία [4].

Όταν βρεθεί ότι η διαδικασία διαφορίσης είναι στάσιμη τότε ψάχνουμε για το μοντέλο **ARMA** αυτής.

Η διαδικασία  $Y_t$  θεωρείται ότι είναι αυτομειωμένη ολοκληρώσιμη μεταβαλλόμενου μέσου όρου διαδικασία, **ARIMA**( $p, d, q$ ), εάν το  $X_t = \nabla^d Y_t$  είναι μία διαδικασία **ARMA**( $p, q$ ) [23].

### 3.8 Η διαδικασία SARIMA (Seasonal Auto-Regressive Integrated Moving Average)

Το μοντέλο *ARIMA* εφαρμόζεται σε μη-εποχιακά μη-στάσιμα (no-seasonal no-stationary) δεδομένα. Οι Box και Jenkins γενίκευσαν αυτό το μοντέλο ώστε να υπολογίζει και εποχιακά φαινόμενα. Το μοντέλο που πρότειναν είναι γνωστό ως *SARIMA*( $p, d, q$ )  $\times$  ( $P, D, Q$ )<sup>s</sup><sup>1</sup> (seasonal auto-regressive integrated moving average). Σε αυτό το μοντέλο υπολογίζουμε την τάξη της εποχιακής διαφορίσης για να αφαιρεθεί η μη-στασιμότητα (no-stationarity) από τη χρονοσειρά. Η εποχιακή διαφορά 1<sup>ης</sup> τάξης είναι η διαφορά μεταξύ της παρατήρησης και της αντίστοιχης παρατήρησης προηγούμενου έτους και υπολογίζεται από τη σχέση  $z_t = y_t - y_{t-s}$ . Για μηνιαίες χρονοσειρές  $s = 12$ , και για τετραμελής χρονοσειρές  $s = 4$  [1]. Η διαδικασία SARIMA είναι πλήρως ανάλογη με την διαδικασία *ARIMA*.

$$X_{n+h} = Y_{n+h} + \sum_{j=1}^{d+Ds} \alpha_j X_{n+h-j} \quad (3.8.1)$$

Υπό την προϋπόθεση ότι οι πρώτες  $d + Ds$  παρατηρήσεις της  $X_{-d-Ds+1}, \dots, X_0$  είναι μη συσχετιζόμενες (uncorrelated) με την  $\{Y_t, t \geq 1\}$ , μπορούμε να προσδιορίσουμε τις καλύτερες γραμμικές προβλέψεις  $P_n X_{n+h}$  της  $X_{n+h}$  βασιζόμενοι στη σειρά  $\{1, X_{-d-Ds+1}, \dots, X_n\}$  και εφαρμόζοντας το πολυώνυμο  $P_n$  σε κάθε μέρος προκύπτει ο γενικός τύπος [1]

$$P_n X_{n+h} = P_n X_{n+h} + \sum_{j=1}^{d+Ds} \alpha_j P_n X_{n+h-j}. \quad (3.8.2)$$

---

<sup>1</sup> Τα εποχιακά μοντέλα δύναται να είναι είτε προσθετικά (additive) είτε πολλαπλασιαστικά (multiplicative). Τα προσθετικά εποχιακά μοντέλα έχουν τη μορφή (0,2,0) με διαφορά καθυστέρησης (lag) 1 και 4. Ο συμβολισμός για το πολλαπλασιαστικό εποχιακό μοντέλο *SARIMA*( $p, d, q$ )  $\times$  ( $P, D, Q$ )<sup>s</sup>, όπου  $s$  είναι ο εποχιακός κύκλος.

Οπότε ο συμβολισμός για το εποχιακό μοντέλο με τοπική τάση (trend) καθυστέρησης 1 και εποχιακή τάση 4 είναι για το εποχιακό μοντέλο με τοπική τάση (trend) καθυστέρησης 1 και εποχιακή τάση 4 είναι (0,1,0)x(0,1,0)<sup>4</sup> [1]

### 3.9 Διάγραμμα συσχέτισης (correlogram)

Το διάγραμμα συσχέτισης, ή πλιό απλά συνάρτηση αυτοσυσχέτισης (autocorrelation function) λαμβάνεται αντικαθιστώντας την  $Cov(y_t, y_{t+k})$  και την  $Var(y_t)$  στην αμγιώς συνάρτηση αυτοσυσχέτισης

$$ACF(k) = \frac{Cov(y_t, y_{t+k})}{Var(y_t)} \quad (3.9.1)$$

με τα αντίστοιχα δείγματα συνδιακύμανσης και διακύμανσης:

$$\widehat{ACF(k)} = \frac{\sum_{t=1}^{n-k} (y_t - \bar{y})(y_{t+k} - \bar{y})}{\sum_{t=1}^n (y_t - \bar{y})^2} = r_k \quad (3.9.2)$$

και σχεδιάζοντας τον συντελεστή αυτοσυσχέτισης  $r_k$  ως προς  $k$  (χρονικές καθυστερήσεις).

Ένα διάγραμμα συσχέτισης περιέχει πολλές πληροφορίες. Η συνάρτηση αυτοσυσχέτισης για καθαρά τυχαίες (λευκός θόρυβος) σειρές (πχ. σειρές ανεξάρτητων ταυτόσημα κατανεμημένων παρατηρήσεων) αναμένεται να είναι προσεγγιστικά 0 για όλες τις μη μηδενικές καθυστερήσεις. Εάν μία χρονοσειρά έχει κάποια τάση (trend) τότε η ACF ελαττώνεται αργά (πχ. γραμμικά) με αυξανόμενες καθυστερήσεις.

Εάν η χρονοσειρά περιέχει εποχιακές ή κυκλικές συνιστώσες, το διάγραμμα συσχέτισης επίσης θα εμφανίζει διακυμάνσεις. Το διάγραμμα συσχέτισης για εποχιακές χρονοσειρές με μηνιαία διαστήματα, μπορεί για παράδειγμα να έχει μεγάλες αρνητικές τιμές σε καθυστερήσεις των 6,18 κτλ (επειδή οι μετρήσεις που πραγματοποιούνται το καλοκαίρι και το χειμώνα είναι αρνητικά σχετιζόμενες) και μεγάλες θετικές τιμές σε καθυστερήσεις των 12,24 κτλ (επειδή οι μετρήσεις που πραγματοποιούνται στην ίδια εποχή είναι θετικά σχετιζόμενες) [11].



## ΚΕΦΑΛΑΙΟ 4

### ΜΑΚΡΑ ΜΝΗΜΗ

#### 4.1 Προσεγγίσεις Ορισμού

Η έννοια της εξάρτησης μεγάλου εύρους συνδέεται άμεσα με τη μνήμη σε μία στοχαστική διαδικασία [20]. Η έννοια της μακράς μνήμης αναφέρεται σε μη περιοδικές διαδικασίες [6]. Η μακρά μνήμη εκφράζει μία τάση για μία σταδιακή εξασθένιση του πλάτους των συσχετίσεων (correlations) των χρονοσειρών ως μία συνάρτηση του μεγέθους καθυστέρησης (lag), και ακόμα να διατηρούν στασιμότητα. Ουσιαστικά, ορίζεται ως μια έκφραση συμπεριφοράς τάσης (trend), δηλαδή, συνέχειας (persistence) (θετική μακρά μνήμη). Κατά μία γενική έννοια, δύναται να εκφράζει είτε παραμονής (persistence) ή έλλειψη παραμονής (anti-persistence), η οποία είναι μία μεταβαλλόμενη συμπεριφορά εντονότερη από εκείνη της τυχαίας διαδικασίας (αρνητική μακρά μνήμη). Η τραχύτητα (roughness) είναι ένα χαρακτηριστικό των χρονοσειρών που περιγράφει την τάση για μία τυχαία συμπεριφορά με συχνές και ετερογενείς αλλαγές.

Ενώ η μακρά μνήμη υπολογιζόμενη από τον συντελεστή  $H$  του Hurst, είναι ένα γενικό χαρακτηριστικό των σειρών, η τραχύτητα, υπολογιζόμενη από την φρακταλική διάσταση  $D$ , είναι ένα τοπικό χαρακτηριστικό των σειρών. Για μία διαδικασία αυτοομοιότητας, δοθείσα από την  $X(ct) = c^H X(t)$  επιβεβαιώνεται ότι  $D + H = 2$ . Σε αυτές τις διαδικασίες, τα χαρακτηριστικά των σειρών της γενικής μακράς μνήμης είναι μία έκφραση των τοπικών χαρακτηριστικών της τραχύτητας. Μία γενίκευση δίνεται από τις πολυφρακταλικές (multifractal) διαδικασίες, από τη σχέση

$$E[X(t) = c(q)t^{\tau(q)+1}] \quad (4.1.1)$$

όπου  $\tau(q)$  είναι μία κοίλη συνάρτηση. Σε μία μονοφρακταλική (monofractal) διαδικασία, έχουμε

$$\tau(q) + 1 - Hq \quad (4.1.2)$$

οπότε έχουμε μία γραμμική κλίμακα.

Η διαδικασία της μακράς μνήμης αρχικά προτάθηκε και ορίστηκε, αντίστοιχα, από τους McLeod, Hipel και Hall, ως μία στάσιμη διαδικασία για την οποία οι απόλυτες τιμές των αυτοσυσχετίσεων (autocorrelations) δεν είναι αθροίσιμες στην περίπτωση των διακριτών τιμών

$$\sum_{r_k=-\infty}^{+\infty} |\rho(r_k)| = +\infty \quad (4.1.3)$$

όπου  $\rho(r_k)$  είναι η καθυστέρηση (lag),  $r_k$ , η αυτοσυσχέτιση.

Στο κείμενο, υποθέτουμε ότι η παράμετρος  $H$  επαληθεύει την ανισότητα  $\frac{1}{2} < H < 1$ .

Ο περιορισμός  $H < 1$  εγγυάται ότι η διαδικασία είναι στάσιμη (stationary).

Επιπλέον, η ανισότητα  $H > \frac{1}{2}$  εγγυάται ότι η διαδικασία είναι παραμένουσα (persistent) (για  $H = \frac{1}{2}$  έχουμε μικρή μνήμη, ενώ για  $0 < H < \frac{1}{2}$  έχουμε μακρά μνήμη, αν και αρνητική).

Ένας διαφορετικός ορισμός της διαδικασίας μακράς μνήμης επιτυγχάνεται από το χαρακτηρισμό ασυμπτωτικών αυτοσυσχετίσεων (asymptotic autocorrelations) :

$$\rho(r_k) \sim r_k^{2h-2} l_1(r_k) \quad \text{καθώς το } r_k \rightarrow +\infty \quad (4.1.4)$$

όπου  $l_1$  είναι μία συνάρτηση αργής διακύμανσης (δηλαδή μία μετρήσιμη συνάρτηση (measurable function) η οποία είναι θετική σε μία γειτονιά του  $\infty$  και για την οποία

$\forall c > 0 \frac{f(cx)}{f(x)} \rightarrow 1$  καθώς το  $x \rightarrow +\infty$ ), και  $H$  είναι ο συντελεστής Hurst.

Παρατηρούμε ότι η συνάρτηση αυτοσυσχέτισης έχει αργή εξασθένιση, ως μία συνάρτηση εκθετικής δύναμης. Στην περίπτωση των διακριτών τιμών αυτό υποδηλώνει τη μη αθροισμότητα των αυτοσυσχετίσεων.

Ένας τρίτος εναλλακτικός ορισμός δύναται να εκφραστεί στο πεδίο των συχνοτήτων:

$$f(\lambda) \sim \lambda^{1-2h} l_2\left(\frac{1}{|\lambda|}\right) \quad \text{καθώς το } \lambda \rightarrow 0 \quad (4.1.5)$$

όπου το  $l_2$  είναι συνάρτηση αργής διακύμανσης, το  $\lambda$  δηλώνει μία συχνότητα και  $f$  είναι μία φασματική πυκνότητα. Ωστόσο, η φασματική πυκνότητα τείνει στο άπειρο καθώς η συχνότητα πλησιάζει στο μηδέν.

**Σημείωση 1:** (i) Δεν υπάρχει καμία ισότητα μεταξύ αυτού και του προηγούμενου ορισμού, αλλά συνεπάγεται από αυτόν εάν η  $l_1$  είναι σχεδόν μονότονη, δηλαδή

$$\exists \alpha > 0 \int_0^x t^\alpha |dl_1(t)| = O(x^\alpha l_1(x)), \quad x \rightarrow +\infty \quad (4.1.6)$$

Ένας τελευταίος ορισμός για μία διαδικασία μακράς μνήμης διακριτού χρόνου  $\{y_t\}$  που προκύπτει από την ανάλυση του Wold είναι :

$$y_t = u + \sum_{k=0}^{+\infty} \Psi_k \varepsilon_{t-k} \quad (4.1.7) \quad \text{όπου} \quad \Psi_0 = 1, \sum_{k=1}^{+\infty} (\Psi_k)^2 < +\infty$$

και  $\varepsilon_t$  είναι ένας λευκός θόρυβος, δηλώνοντας ότι  $\Psi_j \sim j^{H-\frac{3}{2}} l_3(j)$  (4.1.8) όπου η  $l_3$  είναι μία συνάρτηση αργής διακύμανσης.

(ii) Η συνθήκη (4.1.8) υποδηλώνει τη συνθήκη (4.1.4).

(iii) Αν  $H > \frac{1}{2}$  έχουμε διαδικασία μικρής μνήμης, αν  $H \geq 1$  η διαδικασία είναι ασταθής (non-stationary). Επομένως, η περίπτωση της μακράς μνήμης είναι μία ενδιαμέση περίπτωση μεταξύ μικρής μνήμης και αστάθειας. Για  $H \leq \frac{1}{2}$  έχουμε συμπεριφορά ασυνέχειας (anti-persistence).

(iv) Η μακρά μνήμη χαρακτηρίζεται από μία συμπεριφορά ασυμπτωτικής εκθετικής δύναμης, όπου ο συντελεστής Hurst παίζει σημαντικό ρόλο.

Σε άλλους ορισμούς της μακράς μνήμης, θεωρούμε ασυμπτωτικές σχέσεις βασισμένες σε συναρτήσεις δυνάμεων, είτε στο πεδίο του χρόνου είτε των συχνοτήτων. Εφαρμόζοντας ένα λογαριθμικό μετασχηματισμό στις μεταβλητές, λαμβάνουμε γραμμικές εμπειρικές σχέσεις μεταξύ τους, οι οποίες δύναται να

εναρμονιστούν από μικρές τετραγωνικές προσεγγίσεις. Αυτός ο μηχανισμός χρησιμοποιείται για να δημιουργηθούν διάφορες μέθοδοι εκτίμησης του συντελεστή Hurst  $H$ .

Έστω, μία διαδικασία μακράς μνήμης της μορφής  $y_t = u + \varepsilon_t(1-L)^{-H+\frac{1}{2}}$ , έχουμε

$f_y(\lambda) = [4\sin^2(\frac{\lambda}{2})]^{-H+\frac{1}{2}}f_\varepsilon(\lambda)$ , όπου  $\varepsilon_t$  ένας λευκός θόρυβος,  $u = E(y_t)$ , το  $\lambda$  δηλώνει συχνότητα και το  $L$  είναι ο τελεστής καθυστέρησης (lag). Λογαριθμίζοντας την  $\ln f_y(\lambda) = \ln f_\varepsilon(\lambda) - (H - \frac{1}{2}) \ln [4\sin^2(\frac{\lambda}{2})]$ , έχουμε τη γραμμική εμπειρική σχέση των μεταβλητών  $\ln \hat{f}_y(\lambda_k) = \beta - (H - \frac{1}{2}) \ln [4\sin^2(\frac{\lambda_k}{2})] + \varepsilon_k$ ,  $k = 1, 2, \dots, \eta_f(T)$  (7) όπου  $\beta = \ln f_\varepsilon(\lambda)$  και  $\eta_f(T)$  είναι ο αριθμός αρκετών συχνοτήτων.

Η στατιστική του επανακλιμακούμενου εύρους της τυπικής απόκλισης ( rescale range over standard deviation R/S), αρχικά προτάθηκε από τον Hurst, δίνεται από την

$$Q_T = \frac{1}{S_T} \left[ \max_{1 \leq k \leq T} \sum_{j=1}^k (y_t - \bar{y}) - \min_{1 \leq k \leq T} \sum_{j=1}^k (y_j - \bar{y}) \right] \quad (4.1.9)$$

όπου  $\bar{y}$  και  $S_T$  είναι, αντιστοίχως, το δείγμα του μέσου (mean) και της τυπικής απόκλισης (standard deviation). Αυτός είναι ένας αναδυόμενος τρόπος υπολογισμού της μακράς μνήμης, και για i.i.d. γκαουσιανές τυχαίες μεταβλητές έχουμε

$$\frac{Q_T}{\sqrt{T}} \Rightarrow V = V_s - V_{s-1} \quad (4.1.10)$$

όπου  $V_t$  είναι μία γέφυρα Brown και το  $\Rightarrow$  δηλώνει ασθενής σύγκλιση [17].

## 4.2 Αυτο-ομοιότητα και εξάρτηση μεγάλου εύρους

Μία στάσιμη (stationary) διαδικασία  $X$  έχει εξάρτηση μεγάλου εύρους εάν οι αυτοσυσχετίσεις (autocorrelations) της τείνουν στο μηδέν τόσο αργά που το άθροισμά τους δεν συγκλίνει, δηλαδή,  $\sum_{k=1}^{\infty} |\rho(k)| = +\infty$ . Διαισθητικά, η εξάρτηση

μεταξύ ευρύτατα διαχωρίσιμων τιμών από τη διαδικασία εξάρτησης μεγάλου εύρους δεν τείνει στο μηδέν ακόμα και πέρα από απείρως μεγάλων χρονικών μετατοπίσεων. Μία στοχαστική διαδικασία  $X$  καλείται αυτο-όμοια αν

$$X(at) = a^H X(t) \quad (4.2.1)$$

όπου η παράμετρος αυτο-ομοιότητας  $H$ , είναι ο εκθέτης Hurst. Η αυτο-ομοιότητα περιγράφει το φαινόμενο όπου συγκεκριμένες ιδιότητες για τη διαδικασία διατηρούνται ανεξάρτητα της κλίμακας του χώρου ή του χρόνου.

Η αυτο-ομοιότητα δεύτερης τάξης περιγράφει την ιδιότητα όπου η δομή της συσχέτισης (correlation) (ACF) μιας χρονοσειράς διατηρείται ανεξάρτητα του συνυπολογισμού του χρόνου. Πιο απλά, η ACF μίας αυτο-όμοιας χρονοσειράς δεύτερης τάξης είναι ίδια είτε στις μεγάλες είτε στις μικρές χρονικές κλίμακες. Μία στασιμη διαδικασία  $X_t$  είναι δεύτερου βαθμού αυτο-όμοια αν

$$\rho(k) = \frac{1}{2} [(k+1)^{2H} - 2k^{2H} + (k-1)^{2H}], \quad 0,5 < H < 1, \quad (4.2.2)$$

και ακριβώς ασυμπτωτικά αυτο-όμοια αν

$$\lim_{k \rightarrow \infty} \rho(k) = \frac{1}{2} [(k+1)^{2H} - 2k^{2H} + (k-1)^{2H}], \quad 0,5 < H < 1, \quad (4.2.3)$$

Η αυτο-όμοια διαδικασία δεύτερου βαθμού χαρακτηρίζεται από μία συνάρτηση αυτοσυσχέτισης με υπερβολική (hyperbolically) εξασθένιση, και σε μεγάλο βαθμό χρησιμοποιείται στη μοντελοποίηση διαδικασιών εξάρτησης μεγάλου εύρους.

Αντιθέτως, η εξάρτηση μικρού εύρους (short range dependence) χαρακτηρίζεται από συσχετίσεις σύντομης εξασθένισης (πχ. Η διαδικασία **ARMA**). Από αυτούς τους ορισμούς συνεπάγεται ότι μία χρονοσειρά χαρακτηρίζεται από εξάρτηση μεγάλου εύρους, αν  $0,5 < H < 1$ . Καθώς το  $H \rightarrow 1$  η εξάρτηση είναι ισχυρή. Πιο εκτεταμένα χρησιμοποιούνται οι διαδικασίες αυτο-ομοιότητας, στην προσομοίωση μεγάλης εξάρτησης του κλασματικού γκαουσιανού θορύβου (fGn) και των κλασματικών

διαδικασιών *ARIMA*. Η διαδικασίες *ARFIMA* περιγράφουν χρονοσειρές εξάρτησης μεγάλου εύρους όταν  $0 < d < 0.5$ , όπου  $H = d + 0.5$  [9].

Συνήθως, ένα μοντέλο  $X_t$  μακράς μνήμης δύναται να προσδιοριστεί από μία παράμετρο μνήμης  $d \in (0, 1/2)$  που καλείται βαθμός της μνήμης του μοντέλου, ο οποίος ρυθμίζει το σχήμα του φάσματος της συχνότητας κοντά στο μηδέν και την υπερβολική (hyperbolic) εξασθένηση της συνάρτησης αυτοσυσχέτισης. Πιο συγκεκριμένα, η φασματική πυκνότητα  $f_x(\lambda)$  του μοντέλου μακράς μνήμης προσεγγίζεται σε μία περιοχή μηδενικών συχνοτήτων από την  $f_x(\lambda) \sim c\lambda^{-2d}$  καθώς το  $\lambda \rightarrow 0_+$ ,  $0 < c < \infty$  συνεπώς,  $f_x(\lambda) \rightarrow \infty$  καθώς το  $\lambda \rightarrow 0_+$ . Με επιπρόσθετες υποθέσεις ακρίβειας για την  $f_x(\lambda)$ , η συνάρτηση αυτοσυσχέτισης  $\rho(k)$ , του μοντέλου μακράς μνήμης έχει την ακόλουθη ασυμπτωτική συμπεριφορά:  $\rho(k) \sim ck^{2d-1}$  καθώς το  $k \rightarrow \infty$ . Ως αποτέλεσμα σύγκλισης, για  $0 < d < 0.5$ , το άθροισμα  $\sum_{-\infty}^{+\infty} |\rho(k)| = \infty$ . Αυτή η ιδιότητα της έλλειψης αθροιστικότητας των αυτοσυσχετίσεων θεωρείται μακρά μνήμη και ικανοποιείται από τα μοντέλα *ARFIMA*.

Οι ιδιότητες του μοντέλου  $X_t$  εξαρτώνται από την τιμή της παραμέτρου  $d$  [3].

### 4.3 Κλασματικός Λευκός Θόρυβος (Fractional White Noise)

Η μέθοδος *ARFIMA(0, d, 0)*, ή διαδικασία λευκού κλασματικού θορύβου είναι μία διακριτού χρόνου στοχαστική διαδικασία  $Y_t$ , η οποία δηλώνεται ως

$$(1 - L)^d Y_t = \varepsilon_t \quad (4.3.1)$$

όπου  $\varepsilon_t$  είναι ένας λευκός θόρυβος μηδενικού μέσου και  $d$  πιθανών ένας μη ακέραιος. Το ακόλουθο θεώρημα δίνει κάποιες βασικές ιδιότητες για τη διαδικασία, θεωρώντας ότι η συμμεταβλητότητα (covariance) ισούται με  $\sigma_\varepsilon^2 = 1$ .

**Θεώρημα 4.3.1:** Έστω  $Y_t$  μία διαδικασία *ARFIMA(0, d, 0)*.

(α) Όταν  $d < \frac{1}{2}$  η  $Y_t$  είναι μία στάσιμη διαδικασία και έχει έκφραση άπειρου μεταβαλλόμενου μέσου όρου (moving average)

$$Y_t = \varphi(L)\varepsilon_t = \sum_{k=0}^{\infty} \varphi_k \varepsilon_{t-k}, \quad (4.3.2)$$

όπου 
$$\varphi_k = \frac{d(1+d) \dots (k-1+d)}{k!} = \frac{(k+d-1)!}{k!(d-1)!} = \frac{\Gamma(k+d)}{\Gamma(k+1)\Gamma(d)}$$

Όπου  $\Gamma(\cdot)$ , η συνάρτηση Γάμμα. Καθώς το

$$k \rightarrow \infty, \varphi_k \sim k^{d-1}/(d-1)! \equiv \frac{1}{\Gamma(d)} \cdot k^{d-1}.$$

(β) Όταν  $d > -\frac{1}{2}$ , η  $Y_t$  είναι αντιστρέψιμη και έχει έκφραση άπειρης αυτομείωσης (autoregressive)

$$\phi(L)Y_t = \sum_{k=0}^{\infty} \varphi_k Y_{t-k} = \varepsilon_t, \quad (4.3.3)$$

όπου 
$$\varphi_k = \frac{-d(1-d) \dots (k-1-d)}{k!} = \frac{(k-d-1)!}{k!(-d-1)!} = \frac{\Gamma(k-d)}{\Gamma(k+1)\Gamma(-d)}$$

Καθώς το  $k \rightarrow \infty, \varphi_k \sim k^{-d-1}/(-d-1)! \equiv \frac{1}{\Gamma(-d)} \cdot k^{-d-1}$ .

(γ) Όταν  $-\frac{1}{2} < d < \frac{1}{2}$ , η αυτοδιακύμανση (autocovariance) της  $Y_t (\sigma_\varepsilon^2 = 1)$  είναι

$$Y_t (\sigma_\varepsilon^2 = 1) \gamma_k = E(Y_t Y_{t-k}) = \frac{\Gamma(k+d)\Gamma(1-2d)}{\Gamma(k+1-d)\Gamma(1-d)\Gamma(d)} \quad (4.4.4)$$

και οι συναρτήσεις αυτοσυσχέτισης (autocorrelation) είναι

$$r_k = \frac{\gamma_k}{\gamma_0} = \frac{\Gamma(k+d)\Gamma(1-d)}{\Gamma(k-d+1)\Gamma(d)} \quad (4.4.5)$$

Καθώς το  $k \rightarrow \infty$ ,  $r_k \sim \frac{\Gamma(1-d)}{\Gamma(d)} \cdot k^{2d-1}$ .

#### 4.4 Σχέσεις με τους ορισμούς της διαδικασίας μακράς μνήμης

Για  $0 < d < \frac{1}{2}$ , για την κλασματικά ολοκληρωμένη διαδικασία  $I(d)$ , η  $Y_t$  είναι μακρά μνήμη, οι αυτοσυσχετίσεις είναι όλες θετικές  $\left(\frac{\Gamma(1-d)}{\Gamma(d)} \cdot k^{2d-1}\right)$ , έτσι ώστε ο ορισμός (4.1.3) μεταβάλλεται και εξασθενεί με υπερβολικό ρυθμό.

Για  $-\frac{1}{2} < d < 0$ , το άθροισμα των απόλυτων τιμών των συσχετίσεων των διαδικασιών τείνουν σε μία σταθερά, οπότε έχουμε σύντομη μνήμη σύμφωνα με τον ορισμό. Με αυτά τα δεδομένα, η διαδικασία  $ARFIMA(0, d, 0)$  θεωρείται μη παραμένουσα (anti-persistent) ή ότι έχει 'ενδιάμεση μνήμη', και όλες οι αυτοσυσχετίσεις της, εκτός της μηδενικής υστέρησης (lag), είναι αρνητικές και εξασθενούν υπερβολικά στο μηδέν.

Μια άλλη προσέγγιση του ορισμού της μακράς μνήμης με τη διαδικασία  $I(d)$  δύναται να εκφραστεί με τη συμπεριφορά του μερικού αθροίσματος  $S_T$  από τη σχέση

$$\sigma^2 = \lim_{T \rightarrow \infty} E(T^{-1}S_T^2) \quad (4.4.1)$$

(μικρή μνήμη) (αν δεν είναι μηδέν και

$$\frac{T^{-1/2}S_{[Tr]}}{\sigma} \Rightarrow B(r) \quad (4.4.2)$$

για όλα τα  $r \in [0,1]$ , όπου  $B(r)$  η τυπική κίνηση Brown) όταν η  $Y_t$  είναι ένας κλασματικός λευκός θόρυβος. Ο Sowell (1990) έδειξε ότι το

$$\lim_{T \rightarrow \infty} \text{Var}(S_T)T^{-(1+2d)} = \lim_{T \rightarrow \infty} E(S_T^2)T^{-(1+2d)} = \sigma_\varepsilon^2 \frac{\Gamma(1-2d)}{\Gamma(1+2d)\Gamma(1+d)\Gamma(1-d)}.$$

Οπότε,  $\text{Var}(S_T) = O(T^{1+2d})$ , που σημαίνει ότι η διακύμανση των μερικών αθροισμάτων μιας διαδικασίας  $I(d)$ , με  $d = 0$  αυξάνεται γραμμικά, δηλαδή με ρυθμό  $O(T^1)$ . Για μία διαδικασία με ενδιάμεση μνήμη με  $-\frac{1}{2} < d < 0$ , η



διακύμανση των μερικών αθροισμάτων αυξάνεται βραδύτερα και όχι γραμμικά, ενώ για τη διαδικασία της μακράς μνήμης με  $0 < d < \frac{1}{2}$ , ο ρυθμός αύξησης είναι γρηγορότερος. Μία τρίτη προσέγγιση σχετίζεται με την κλασματική κίνηση Brown και τον κλασματικό γκαουσιανό θόρυβο. Το ανάλογο του διακριτού χρόνου της κίνησης Brown είναι η τυχαία διαδρομή (random walk),  $X_t$  η οποία ορίζεται από τη σχέση

$$(1 - L)X_t = \varepsilon_t \quad (4.4.3)$$

όπου  $\varepsilon_t$  είναι μία i.i.d. Η πρώτη διαφώριση της  $X_t$  είναι η διαδικασία λευκού θορύβου διακριτού χρόνου  $\varepsilon_t$ . Αναλογικά με τον παραπάνω ορισμό του λευκού θορύβου συνεχούς χρόνου ορίζουμε ως κλασματικό διαφορίσιμο λευκό θόρυβο με παράμετρο  $H$  την  $(1/2 - H)^{\eta}$  κλασματική διαφώριση ή διακριτού χρόνου λευκό θόρυβο. Ο τελεστής κλασματικής διαφώρισης  $(1 - L)^d$  ορίζεται με φυσικό τρόπο από τη διωνυμική σειρά:

$$\begin{aligned} (1 - L)^d &= \sum_{k=0}^{\infty} \binom{d}{k} (-L)^k = 1 - dL - \frac{1}{2}d(1-d)L^2 - \frac{1}{6}d(1-d)(2-d)L^3 - \dots \\ &= \sum_{k=0}^{\infty} \frac{\Gamma(k-d)z^k}{\Gamma(-d)\Gamma(k+1)}. \end{aligned} \quad (4.4.4)$$

Θέτουμε  $d = H - \frac{1}{2}$  ώστε ο κλασματικός λευκός θόρυβος συνεχούς χρόνου με παράμετρο  $H$  να έχει ως ανάλογο διακριτού χρόνου τη διαδικασία  $X_t = (1 - L)^{-d}\varepsilon_t$  ή  $(1 - L)^d X_t = \varepsilon_t$ , όπου  $\varepsilon_t$  είναι μία διαδικασία λευκού θορύβου. Με τα παραπάνω αποτελέσματα, η κλασματική λευκή διαδικασία  $I(d)$  είναι επίσης μία διαδικασία μακράς μνήμης αν στην εξίσωση (4.4.5) θέσουμε όπου  $d = H - \frac{1}{2}$ .

#### 4.5 Η διαδικασία **ARFIMA** (Auto Regressive Fractional Intergrated Moving Average)

Μιά φυσική επέκταση του μοντέλου του κλασματικού λευκού θορύβου (4.3.1) είναι το κλασματικό μοντέλο **ARMA** ή **ARFIMA**( $p, d, q$ )

$$\phi(L)(1-L)^d Y_t = \theta(L)\varepsilon_t \quad (4.5.1)$$

όπου  $d$  δηλώνει την κλασματική παράμετρο διαφορίσης,

$$\phi(L) = 1 - \phi_1 L - \phi_2 L^2 - \dots - \phi_p L^p, \theta(L) = 1 + \theta_1 L + \theta_2 L^2 + \dots + \theta_q L^q \text{ και } \varepsilon_t$$

λευκός θόρυβος [10]. Η παράμετρος  $d$  περιορίζεται στο διάστημα  $[-0.5:0.5]$ , που σημαίνει ότι το μοντέλο **ARFIMA** λαμβάνει μόνο στάσιμα (stationary) σήματα μεταξύ -1 και 1 (fGn). Για να λάβουμε σήματα της fBm εφαρμόζουμε κανόνες διαφορίσης στα αντίστοιχα fGn σήματα. Οπότε μπορούμε να εκτιμήσουμε τη θεωρητική κλασματική παράμετρο των fBm σειρών προσθέτοντας την τιμή 1 στην παράμετρο  $d$  [25].

Μερικές βασικές ιδιότητες της διαδικασίας **ARFIMA** συνοψίζονται στο θεώρημα 4.5.1.

##### **Θεώρημα 4.5.1:**

Έστω  $Y_t$  μία διαδικασία **ARFIMA**( $p, d, q$ ). Τότε

(α) η  $Y_t$  είναι στάσιμη αν  $d < \frac{1}{2}$  και όλες οι ρίζες της  $\phi(L) = 0$  βρίσκονται εκτός του μοναδιαίου κύκλου.

(β) η  $Y_t$  είναι αντιστρέψιμη αν  $d > -\frac{1}{2}$  και όλες οι ρίζες της  $\theta(L) = 0$  βρίσκονται εκτός του μοναδιαίου κύκλου.

(γ) αν  $-\frac{1}{2} < d < \frac{1}{2}$ , η αυτοδιακύμανση της  $Y_t$ ,  $\gamma_k = E(Y_t Y_{t-k}) \sim B \cdot k^{2d-1}$  καθώς το  $k \rightarrow \infty$ , όπου  $B$  είναι μία συνάρτηση της  $d$ .

Η διαδικασία **ARFIMA**( $p, d, q$ ) χρησιμοποιείται ως μέθοδος μοντελοποίησης επειδή η επίδραση της παραμέτρου  $d$  στην απόσταση των παρατηρήσεων εξασθενεί υπερβολικά καθώς αυξάνεται η καθυστέρηση, ενώ οι επιδράσεις των παραμέτρων  $\phi_i$

και  $\theta_j$  εξασθενούν εκθετικά. Για το λόγο αυτό η  $d$  δύναται να επιλεγεί για την περιγραφή της δομής συσχέτισης υψηλών καθυστερήσεων των χρονοσειρών ενώ οι  $\phi_i$  και  $\theta_j$  επιλέγονται για την περιγραφή δομών συσχέτισης χαμηλών καθυστερήσεων [10].

#### 4.6 Εκτιμητής Μέγιστης Πιθανότητας (Maximum Likelihood Estimator)

Εφόσον έχουμε πιστοποιήσει ένα δοκιμαστικό μοντέλο, το επόμενο βήμα είναι να εκτιμήσουμε τις παραμέτρους του μοντέλου [28].

Η μέθοδος του εκτιμητή μέγιστης πιθανότητας (MLE) παρέχει εκτιμήσεις του  $H$  σε κάθε χρονική κλίμακα. Η παραμονή (persistence) μεγάλου εύρους χαρακτηρίζεται από μία συνάρτηση αυτοσυσχέτισης (autocorrelation) δυναμοεκθέτη (power law). Μία συνάρτηση αυτοσυσχέτισης αυτού του είδους δηλώνει αργή εξασθένιση των αυτοσυσχετίσεων οι οποίες δείχνουν ότι παρατηρήσεις καλά χωρισμένες στο χρόνο είναι γραμμικά σχετιζόμενες (correlated). Αυτή η ιδιότητα καλείται παραμονή μεγάλου εύρους (long range persistence) ή μακρά μνήμη (long range memory). Μία συνάρτηση αυτοσυσχέτισης (autocorrelation) δυναμοεκθέτη (power law)  $C(l)$  δύναται να εκφραστεί με όρους χρονικών καθυστερήσεων (time lag)  $l$ , ως  $C(l) \sim kl^{2(H-1)}$ ,  $0.5 \leq H < 1$  όπου  $k$  είναι μία σταθερά και  $\sim$  (διαβάζεται ‘κλίμακα ως’) αντανακλά το γεγονός ότι η ισότητα σε αυτή τη σχέση είναι θεωρητική. Στην πράξη, δείγματα συναρτήσεων αυτοσυσχέτισης  $C(l)$ , είναι μόνο προσεγγιστικά  $kl^{2(H-1)}$ . Το  $H$  αναφέρεται εδώ ως εκθέτης παραμονής (persistence), και υπολογίζει την παραμονή των αυτοσυσχετίσεων των χρονοσειρών και ποσοτικοποιεί την εξασθένιση.

Σύμφωνα με τη θεωρία της τυχαίας διαδρομής (random walk), για χρονοσειρές με συνάρτηση συσχέτισης δυναμοεκθέτη  $C(l) \sim l^{2(H-1)}$ , οι διακυμάνσεις  $F(l)$  μιάς χρονικής καθυστέρησης μήκους  $l$  αυξάνουν καθώς η σχέση του δυναμοεκθέτη

$$F(l) = cl^H, \quad 0.5 \leq H < 1 \quad (4.6.1)$$

όπου  $c$  μία σταθερά και  $H$  ο εκθέτης σταθερότητας. Η  $F(l)$  μπορεί να υπολογιστεί χρησιμοποιώντας μία απλή μέθοδο γνωστή ως τυπική ανάλυση διακυμάνσεων.

Έστω  $X_t, t = 1, \dots, N$  μία χρονοσειρά από μία στάσιμη (stationary) διαδικασία με τη σκέψη ότι εκφράζει παραμονή μεγάλου εύρους και με τρέχων άθροισμα  $Y_m = \sum_{i=1}^m (X_i - \bar{X})$  όπου  $\bar{X}$  ο δειγματοληπτικός μέσος (mean). Αρχικά, διαιρούμε τη χρονοσειρά σε κουτιά μήκους καθυστέρησης  $l, \{Y_1, \dots, Y_{l+1}\}, \{Y_{l+2}, \dots, Y_{2(l+1)}\}, \dots$  και θεωρούμε τις παρατηρήσεις στα δύο άκρα των μη επικαλυπτόμενων κουτιών,  $Y_t$  και  $Y_{t+l}$ . Η  $F^2(l)$  (4.6.1) μπορεί να υπολογιστεί ως ο μέσος όρος των τετραγωνισμένων διαφορών των παρατηρήσεων και στα δύο άκρα των κουτιών,

$$\hat{F}^2(l) = \frac{1}{n(l)} \left[ (Y_1 - Y_{l+1})^2 + (Y_{l+2} - Y_{2(l+1)})^2 + \dots \right] \quad (4.6.2)$$

όπου  $n(l)$  ο αριθμός των μη επικαλυπτόμενων κουτιών μήκους  $l$  και οι διαφορές  $Y_1 - Y_{l+1}, t = 1, l+2, \dots$  είναι ουσιαστικά οι αποκλίσεις ενός σημείου  $l$  μεταβαλλόμενου μέσου όρου της  $X(t)$  από ένα συνολικό μέσο (mean) της  $X_t: Y_t - Y_{t+l} = l\bar{X} - \sum_{i=1}^l X_{t+i}$ . Ο εκθέτης  $H$  εκτιμάται από την κλίση της καλύτερης προσαρμοζόμενης γραμμής του  $\log(l)$  από τον  $\log(\hat{F}(l))$ .

Έστω  $X(t)$  μία χρονοσειρά από μία διαδικασία θεωρούμενη ότι παρουσιάζει παραμονή (persistence) μεγάλου εύρους και  $Y_t$  το τρέχων άθροισμα. Για μία ορισμένη καθυστέρηση  $l$ , υποθέτουμε ότι οι αυξήσεις  $Y_t - Y_{t+l}, t = 1, l+2, 2l+3, \dots$  είναι ανεξάρτητες και πανομοιότυπα κατανομημένες (i.i.d) με μέσο μηδέν και διακύμανση  $\sigma_l^2$

$$Z_{i,l} = Y_t - Y_{t+l} \sim N(0, \sigma_l^2), \quad i = 1, 2, \dots, n(l) \quad (4.6.3)$$

για καθυστέρηση  $l$ , όπου  $\sigma_l^2$  η διακύμανση, η οποία εκτιμάται από την

$$\hat{\sigma}_l^2 = \frac{1}{n(l)} \sum_{i=1}^{n(l)} Z_{i,l}^2 = \frac{1}{n(l)} \left[ (Y_1 - Y_{l+1})^2 + (Y_{l+2} - Y_{2(l+1)})^2 + \dots \right]$$

η οποία για διαδικασίες με παραμονή (persistence) μεγάλου εύρους (από εξισώσεις 4.6.1 και 4.6.2) είναι

$$\sigma_l^2 = \sigma^2 l^{2H} \quad (4.6.4)$$

Ορίζουμε  $J_l$  ως το τετραγωνισμένο άθροισμα των i.i.d γκαουσιανών τυχαίων μεταβλητών  $n(l)$  (η ακολουθία στην εξίσωση (4.6.3) με αναμενόμενο μέσο 0) διαιρούμενων από τη διακύμανσή τους

$$J_l = \frac{1}{\sigma_l^2} \sum_{i=1}^{n(l)} \frac{Z_{i,l}^2}{\sigma_l^2} = \frac{1}{\sigma_l^2} \sum_{i=1}^{n(l)} Z_{i,l}^2 = \frac{1}{\sigma^2 l^{2H}} \sum_{i=1}^{n(l)} Z_{i,l}^2 \sim \chi_{n(l)-2}^2 \quad (4.6.5)$$

$J_l$  είναι το άθροισμα των τυχαίων μεταβλητών  $n(l)x^2$  με ένα βαθμό ελευθερίας και γι' αυτό το λόγο έχει μία κατανομή  $x^2$  με  $n(l) - 2$  βαθμούς ελευθερίας.

Έχοντας αυτή την πληροφορία μπορούμε να υπολογίσουμε τον MLE του  $H$  για μία ορισμένη καθυστέρηση  $l$ .

Η συνάρτηση πιθανότητας, η οποία είναι ενωμένη με τη συνάρτηση πυκνότητας πιθανότητας του δείγματος  $z_{1,l}, \dots, z_{n(l),l}$  είναι η συνάρτηση πυκνότητας πιθανότητας στο  $x^2$  της μεταβλητής  $J_l$  και δίνεται από την

$$P(z_{1,l}, \dots, z_{n(l),l} | H) = \frac{j_l^{((n(l)-2)/2)-1} e^{-j_l/2}}{e^{(n(l)-2)/2} \Gamma\left(\frac{n(l)-2}{2}\right)} \quad (4.6.6)$$

με  $J_l = 1/c^2 l^{2H} \sum_{i=1}^{n(l)} z_{i,l}^2$  όπως στην εξίσωση (4.6.5). Για ένα δεδομένο  $l$ , μεγιστοποιούμε την πιθανότητα από την (4.6.6) αποδίδοντας την εκτίμηση μέγιστης πιθανότητας MLE του  $H$  στην καθυστέρηση  $l$ .

Στο προηγούμενο αποτέλεσμα θεωρείται ότι η τιμή της σταθεράς  $c$  στην έκφραση  $\sigma_l^2 = c^2 l^{2H}$  είναι γνωστή. Πράγματι, δύναται εύκολα να υπολογιστεί χρησιμοποιώντας τις παρατηρήσεις στο  $l = 1$

$$\hat{\sigma}_1^2 = \frac{1}{n(l)} \sum_{i=1}^{n(l)} (Y_t - Y_{t+1})^2 = \hat{c}^2 1^{2H} = \hat{c}^2 \quad (4.6.7)$$

Ο εκτιμητής MLE, του δείγματος  $\hat{H}$ , για τον εκθέτη  $H$  έχει κάποια σημαντικά πλεονεκτήματα. Αρχικά, παρέχει την εκτίμηση  $\hat{H}(l)$  (δηλ. του  $\hat{H}$  στο  $l$ ) η οποία είναι η πιο υποσχόμενη τιμή στο συγκεκριμένο δείγμα. Δεύτερον, παρέχει μία εκτίμηση για κάθε χρονική καθυστέρηση  $l$ , η οποία προκύπτει από την έκταση στην οποία η παραμονή μεγάλης διάρκειας ( long-range persistence) θα εκτιμηθεί σε μεγάλο βαθμό. Τέλος, το βασικό πλεονέκτημα είναι ότι με τη συνάρτηση πυκνότητας πιθανότητας (4.6.6) δεν είναι μόνο δυνατόν να λάβουμε την τιμή του εκθέτη παραμονής  $\hat{H}(l)$  που αυξάνει αυτή την πιθανότητα, αλλά επίσης και να μετρήσουμε την αβεβαιότητα της εκτίμησης με τη βοήθεια της κατανομής  $\hat{H}(l)$  [8].

## ΚΕΦΑΛΑΙΟ 5

### ΦΡΑΚΤΑΛΙΚΗ ΔΙΑΣΤΑΣΗ ΚΑΙ ΕΚΘΕΤΗΣ HURST

Η φρακταλική διάσταση είναι μία μέτρηση της τραχύτητας (roughness), και σε αντίθεση με τα χαρακτηριστικά της μακράς μνήμης, μετράει την τοπική μνήμη των σειρών. Όταν μοντελοποιούμε τη δυναμική συμπεριφορά μίας μεταβλητής, από τη λύση των ντετερμινιστικών εξισώσεων, το σύνολο όλων των στιγμιαίων καταστάσεων του συστήματος είναι ο χώρος των φάσεων. Το υποσύνολο των φάσεων του χώρου προς τον οποίο το σύστημα συγκλίνει, καλείται ελκυστής (attractor) και μπορεί να είναι φράκταλ. Ένα φράκταλ είναι μία ανώμαλη γεωμετρική μορφή, το σχέδιο της εξέλιξης, όντας ίδια σε διάφορες κλίμακες του χρόνου (αυτο-ομοιότητα). Όταν η αυτο-ομοιότητα της γεωμετρικής μορφής μέσω της κλίμακας δεν είναι τέλεια, καλείται στατιστική (statistical). Η φρακταλική διάσταση δύναται να εξαχθεί ως εκθέτης μίας ποσότητας με κλιμακωτή συμπεριφορά, μετρώντας τον όγκο του αντικειμένου ως προς τη μέτρηση του αντίστοιχου μεγέθους (γραμμική απόσταση) :

όγκος  $\sim$  μέγεθος<sup>διάσταση</sup> από την οποία λαμβάνουμε

$\ln(\text{όγκος}) \sim \text{διάσταση} \times \ln(\text{μέγεθος})$ , οπότε η διάσταση δίνεται από τον τύπο

$$\lim_{\text{μέγεθος} \rightarrow 0} \frac{\ln(\text{όγκος})}{\ln(\text{μέγεθος})} \quad [12]$$

#### 5.1 Συσχέτιση φρακταλικής διάστασης και εκθέτη Hurst

Η εξάρτηση μεγάλου εύρους ή παραμονή (persistence) στις χρονοσειρές συνδέεται με συσχετίσεις δυναμοεκθέτη (power law), και συχνά αναφέρεται ως φαινόμενο Hurst. Αρχικά, η κλασματική διάσταση και ο συντελεστής Hurst είναι ανεξάρτητα μεταξύ τους: η κλασματική διάσταση είναι μία τοπική ιδιότητα, και η μακρά μνήμη

είναι ένα καθολικό (global) χαρακτηριστικό. Ωστόσο, οι δύο έννοιες είναι στενά συνδεδεμένες στο μεγαλύτερο μέρος της επιστημονικής βιβλιογραφίας. Αυτό προκύπτει από την επιτυχία των μοντέλων αυτο-συγγένειας (self-affine) όπως ο φρακταλικός γκαουσιανός θόρυβος και η φρακταλική κίνηση Brown στην εξήγηση και τη μοντελοποίηση του εκάστοτε φαινομένου. Για διαδικασίες αυτο-συγγένειας, οι τοπικές ιδιότητες αντανακλούν τις καθολικές (global), οι οποίες προκύπτουν από την περίφημη σχέση

$$D + H = n + 1 \quad (5.1.1)$$

μεταξύ της φρακταλικής διάστασης  $D$ , και του συντελεστή  $H$ , για μία επιφάνεια αυτο-συγγένειας στον διαστασιακό χώρο  $n$ . Η μακρά μνήμη, ή παραμονή (persistence), σχετίζεται με την περίπτωση  $H \in (1/2, 1)$  και συνδέεται σε επιφάνειες με μικρές φρακταλικές διαστάσεις. Πολύ ανώμαλες επιφάνειες με πολύ υψηλές φρακταλικές διαστάσεις συναντώνται σε διαδικασίες με έλλειψη παραμονής (antipersistence) με  $H \in (0, 1/2)$ .

Οι στάσιμες (stationary) γκαουσιανές διαδικασίες χαρακτηρίζονται από τη συνάρτηση συσχέτισης (correlation),

$$c(k) = E[Z(x)Z(x+k)], \quad k \in \mathbb{R}^n. \quad (5.1.2)$$

Η συμπεριφορά της συνάρτησης συσχέτισης στο  $k = 0$  προσδιορίζει τις τοπικές ιδιότητες της διαδικασίας. Ειδικά, αν

$$1 - c(k) \sim |k|^\alpha \quad \text{καθώς το } k \rightarrow 0. \quad (5.1.3)$$

Για κάποιο  $\alpha \in (0, 2]$  η διαδικασία της τυχαίας μεταβλητής έχει φρακταλική διάσταση

$$D = n + 1 - \frac{\alpha}{2} \quad (5.1.4)$$



με πιθανότητα 1. Παρομοίως, η ασυμπτωτική συμπεριφορά της συνάρτησης συσχέτισης στο άπειρο δηλώνει την παρουσία ή απουσία εξάρτησης μεγάλου εύρους. Αν ο δυναμοεκθέτης (power-law)  $c(k) \sim |k|^{-\beta}$  καθώς το  $|k| \rightarrow \infty$ , ευσταθεί για κάποιο  $\beta \in (0,1)$  η διαδικασία θεωρείται ότι έχει μακρά μνήμη με συντελεστή Hurst

$$H = 1 - \frac{\beta}{2}. \quad (5.1.5)$$

Από τη συνάρτηση συσχέτισης (4.2.2) του κλασματικού γκαουσιανού θορύβου (fGn) και θεωρώντας ότι  $H \neq \frac{1}{2}$ , έχουμε  $1 - c(k) \sim |k|^{2H}$  καθώς το  $k \rightarrow 0$  και

$$\frac{c(k)}{H(2H-1)|k|^{-(2-2H)}} \rightarrow 1 \quad \text{καθώς το } k \rightarrow \infty.$$

Για το λόγο αυτό η γραμμική σχέση (5.1.1) ευσταθεί για  $n = 1$ . Η περίπτωση  $H \in (\frac{1}{2}, 1)$  σχετίζεται με θετικές συσχετίσεις, παραμένουσες (persistent) διαδικασίες και χαμηλές φρακταλικές διαστάσεις. Αυτή είναι η περίπτωση της μακράς μνήμης. Αν  $H \in (0, \frac{1}{2})$  βρίσκουμε αρνητικές συσχετίσεις, έλλειψη παραμένουσων (antipersistent) διαδικασιών, και υψηλές φρακταλικές διαστάσεις [7].

## 5.2 Μέθοδος R/S

Αρχικά οι χρονοσειρές διαιρούνται σε  $d$  συναφής υπό – σειρές μήκους  $n$  όπου  $d \times n = N$ , το συνολικό μήκος των χρονοσειρών. Για κάθε μία από αυτές τις χρονοσειρές  $m$ , όπου  $m = 1, \dots, d$ :

- (1) προσδιορίζουμε το μέσο (mean),  $E_m$ , κάθε υπο-σειράς.
- (2) προσδιορίζουμε την τυπική απόκλιση,  $S_m$ , για κάθε υπο-σειρά.
- (3) κανονικοποιούμε τα δεδομένα ( $Z_{i,m}$ ), αφαιρώντας το μέσο από κάθε δεδομένο:

$$X_{i,m} = Z_{i,m} - E_m, \quad \text{για κάθε } i = 1, \dots, n.$$

(4) χρησιμοποιώντας τα κανονικοποιημένα δεδομένα δημιουργείται μία αθροιστική χρονοσειρά με συνεχόμενο άθροισμα των δεδομένων:

$$Y_{i,m} = \sum_{j=1}^i X_{j,m} \text{ για κάθε } i = 1, \dots, n.$$

(5) χρησιμοποιώντας τις νέες αθροιστικές σειρές βρίσκουμε το εύρος αφαιρώντας την ελάχιστη τιμή από τη μέγιστη τιμή:

$$R_m = \max\{Y_{1,m}, \dots, Y_{n,m}\} - \min\{Y_{1,m}, \dots, Y_{n,m}\}$$

(6) επαναλαμβάνουμε την κλίση (rescale) του εύρους,  $R_m/S_m$  διαιρώντας το εύρος με την τυπική απόκλιση (standard deviation).

(7) υπολογίζουμε το μέσο (mean) του επαναδιαβαθμιζόμενου (rescaled) εύρους για όλες τις υποσειρές μήκους  $n$ :

$$(R/S)_n = \frac{1}{d} \sum_{m=1}^d R_m/S_m$$

(8) το μήκος  $n$  πρέπει να αυξηθεί στην επόμενη υψηλή τιμή, όπου  $d \times n = N$  και  $d$  είναι μία ακέραια τιμή. Τα βήματα από το 1 μέχρι το 7 είναι επαναλαμβανόμενα και πρέπει να επαναληφθούν μέχρι  $n = N/2$ .

(9) τελικά η τιμή του  $H$  λαμβάνεται χρησιμοποιώντας μία σύνθετης εμπειρική σχέση ελάχιστων τετραγώνων με  $\log(n)$  ως ανεξάρτητη μεταβλητή και  $\log(R/S)_n$  ως εξαρτημένη μεταβλητή, η κλίση της προκύπτουσας εξίσωσης είναι η εκτίμηση του εκθέτη Hurst. Η επανάληψη τρέχει πάνω από τιμές του  $n$  μεγαλύτερες από 10, καθώς μικρές τιμές του  $n$  παράγουν ασταθείς εκτιμήσεις, όταν το δειγματοληπτικό μέγεθος είναι μικρό [13].

### 5.3 Μέθοδος Variogram

Η μέθοδος variogram χρησιμοποιείται για την εκτίμηση της κλασματικής διάστασης ( $D$ ). Αυτό βασίζεται στην υπόθεση ότι το υπολογιζόμενο μέγεθος έχει στατιστικές ιδιότητες παρόμοιες με εκείνες της κλασματικής κίνησης Brown (fBm) εκφραζόμενες από ένα δυναμοεκθέτη (power law) variogram στο πεδίο ορισμού του χώρου:

$$\gamma(h) = \frac{1}{2N(h)} \sum_{i=1}^{N(h)} [Z(x_i) - Z(x_i + h)]^2 = ch^a \quad (5.3.1)$$

Όπου,

$\gamma(h)$  = ημι-διακύμανση με καθυστέρηση  $h$ .

$h$  = απόσταση (καθυστέρηση (lag))

$x$  = θέση σε μία διάσταση

$c$  = σταθερά αναλογίας

$a$  = κλίση

Για το μοντέλο της fBm, το  $\alpha$  ανήκει στο διάστημα  $0 < \alpha < 2$  και σχετίζεται με τον εκθέτη Hurst  $H$  και διαδοχικά στο  $D$  από την ακόλουθη σχέση:

$$\alpha = 2H$$

$$D = E_d + 1 - H$$

Όπου,  $E_d$  = Ευκλείδεια διάσταση: ο αριθμός δύο συντεταγμένων που απαιτείται για να προσδιορίσουμε το αντικείμενο.

$H$  = ο συντελεστής Hurst με εύρος  $0 < H < 1$ . Κάτω από προϋποθέσεις, η σχέση μεταξύ  $D$  και  $H$  είναι  $H = 2 - D$

Ο εκθέτης Hurst μετράει τις ιδιότητες διαβάθμισης (scaling) της fBm. Υπό αυτή την έννοια, η υπόθεση ήταν ότι ο εκθέτης Hurst σχετίζεται αρκετά με την κλασματική διάσταση, η οποία επίσης υπολογίζει τις ιδιότητες διαβάθμισης (scaling). Καθώς το  $H$  αυξάνεται, η fBm αναμένεται να είναι περισσότερο παραμένουσα (persistent), και το διάγραμμα της συνάρτησης πίο ομαλό με μία διαδοχική μείωση του  $D$ . Αντιστρόφως, καθώς το  $H$  μειώνεται, η fBm γίνεται περισσότερο αντι-παραμένουσα (anti-persistent) και το διάγραμμα της συνάρτησης, γι αυτό το λόγο, τείνει να γίνει πίο τραχύ, και συνεπώς το  $D$  αυξάνεται. Η σχέση μεταξύ φρακταλικής διάστασης και εκθέτη Hurst υπογραμμίζει τέλεια τη σχέση της φρακταλικής διάστασης ως μέτρηση της ανομοιογένειας ενός αντικειμένου.

Πίνακας 5.3.1 Σχέση μεταξύ φρακταλικής διάστασης και εκθέτη Hurst

H	D	Φύση Διαδικασίας
> 0.5	< 1.5	Παραμένουσα
= 0.5	= 1.5	Τυχαία κίνηση Brown
< 0.5	> 1.5	Αντι-παραμένουσα

[16]

#### 5.4 Μέθοδος R/L

Η μέθοδος R/L βασίζεται στην έννοια της φρακταλικής γεωμετρίας η οποία χρησιμοποιείται για τον ακριβή υπολογισμό του εκθέτη Hurst. Ο υπολογισμός του εκθέτη Hurst μέσω της μεθόδου R/L υλοποιείται με τον υπολογισμό της τυπικής απόκλισης  $S(n)$  των υψηλών τιμών ενός τμήματος μήκους  $n$  ενός αυτο-συγγενικού (self-affine) προφίλ από τον τύπο

$$S(n) = A \cdot n^H \quad (5.4.1)$$

Στην εξίσωση (5.4.1) το  $A$  είναι η σταθερά αναλογίας που περιγράφει το μήκος της διακύμανσης και  $H$  ο εκθέτης Hurst. Η  $S(n)$  υπολογίζεται από τον τύπο

$$S(n) = \frac{1}{\xi_n} \sum_{i=1}^{\xi_n} \sqrt{\frac{1}{m_i - 2} \sum_{j \in n_i} (x_j - \langle x \rangle_{n_i})^2}$$

όπου  $\xi_n$  είναι ο συνολικός αριθμός των τμημάτων πλάτους  $n$  μέσα στο οποίο έχει διαιρεθεί το προφίλ της  $S(n)$ ,  $m_i$  είναι ο αριθμός των σημείων συμπεριλαμβανομένου του  $i$ -οστού τμήματος  $n_i$ ,  $x_j$  είναι το διάστημα του σημείου σύγκλισης από την καλύτερη προσαρμοζόμενη γραμμή και  $\langle x \rangle_{n_i}$  είναι η μέση τιμή του  $x_j$  στο τμήμα  $n_i$ . Απεικονίζοντας τα ζευγάρια  $\{n, S(n)\}$  σε ένα διπλό λογαριθμικό διάγραμμα, ο εκθέτης Hurst υπολογίζεται μέσω ενός ελάχιστα προσαρμοζόμενου τετραγώνου [15].

## ΚΕΦΑΛΑΙΟ 6

### ΠΕΙΡΑΜΑΤΙΚΟ ΜΕΡΟΣ

Στο πειραματικό μέρος περιγράφεται η διαδικασία που ακολούθησα για την ανάλυση των ηλεκτρομαγνητικών σημάτων, ο κώδικας και η εφαρμογή *Benoit* που χρησιμοποίησα ως μέρος της διαδικασίας. Ο εκθέτης Hurst καθώς και η φρακταλική διάσταση των σημάτων υπολογίστηκαν με τις μεθόδους R/L, R/S και Variogram.

#### 6.1 Κώδικας σε octave GNU

Σε αυτή την παράγραφο παρατίθενται ενδεικτικά κάποιες βασικές γραμμές του κώδικα της ανάλυσης των ιχνών μακράς μνήμης με στόχο ο αναγνώστης να λάβει μία γενική εικόνα και να κατανοήσει τα βήματα που ακολουθήθηκαν για την επεξεργασία των σημάτων. Ο πλήρης κώδικας θα κάλυπτε μεγάλη έκταση και ξεφεύγει από τους στόχους της παρούσας εργασίας.

Η ανάλυση αναφέρεται σε ηλεκτρομαγνητικά σήματα που λήφθηκαν το έτος 2009 από την 152<sup>η</sup> ημέρα μέχρι την 182<sup>η</sup> ημέρα.

Το σήμα μίας ημέρας που λαμβάνεται από το όργανο μέτρησης χωρίζεται σε 10 τμήματα-κανάλια για την εύκολη επεξεργασία του. Ένα μικρό μέρος των καναλιών απεικονίζεται στον πίνακα 6.1.1. Το κανάλι που χρησιμοποιώ ως παράδειγμα στην περιγραφή μου είναι το κανάλι των 46 MHz. Τα δεδομένα του πίνακα είναι σε μορφή .dat.

Πίνακας 6.1.1 Αρχεία .dat

Id/data loger	Julian Day	hour	sec	3KHz EW	3KHz NS	10KHz EW	10KHz NS	41MHz	46MHz	Volt
102	0153.	0000.	01.00	28.15	384.3	89.35	171.4	590.0	585.1	11.93
102	0153.	0000.	02.00	30.60	381.9	97.92	176.3	525.1	585.1	11.93
102	0153.	0000.	03.00	28.15	385.6	85.68	143.2	552.0	585.1	11.93

102	0153.	0000.	04.00	22.03	385.6	105.3	176.3	559.4	586.3	11.93
102	0153.	0000.	05.00	19.58	388.0	102.8	166.5	564.3	587.5	11.93
102	0153.	0000.	06.00	19.58	389.2	118.7	190.9	565.5	591.2	11.92
102	0153.	0000.	07.00	18.36	400.2	93.02	178.7	560.6	588.7	11.93
102	0153.	0000.	08.00	18.36	396.6	82.01	165.2	565.5	588.7	11.93
102	0153.	0000.	09.00	18.36	401.5	101.6	178.7	565.5	590.0	11.93
102	0153.	0000.	10.00	17.14	402.7	94.25	166.5	581.4	587.5	11.93
102	0153.	0000.	11.00	28.15	400.2	145.7	198.3	574.1	585.1	11.93
102	0153.	0000.	12.00	39.17	403.9	126.1	181.2	585.1	586.3	11.93
102	0153.	0000.	13.00	26.93	403.9	108.9	148.1	576.5	592.4	11.93
102	0153.	0000.	14.00	22.03	401.5	108.9	145.7	555.7	586.3	11.93
102	0153.	0000.	15.00	22.03	401.5	108.9	145.7	555.7	586.3	11.93
102	0153.	0000.	16.00	17.14	401.5	108.9	179.9	559.4	591.2	11.93
102	0153.	0000.	17.00	17.14	397.8	90.58	178.7	559.4	586.3	11.93
102	0153.	0000.	18.00	18.36	396.6	96.70	173.8	554.5	588.7	11.93
102	0153.	0000.	19.00	17.14	394.1	99.14	186.0	561.8	586.3	11.93
102	0153.	0000.	20.00	17.14	394.1	93.02	187.3	548.4	588.7	11.93
102	0153.	0000.	21.00	36.72	394.1	108.9	187.3	556.9	586.3	11.93
102	0153.	0000.	22.00	36.72	394.1	108.9	187.3	556.9	586.3	11.93
102	0153.	0000.	23.00	24.48	391.7	105.3	168.9	569.2	588.7	11.93
102	0153.	0000.	24.00	20.81	390.5	107.7	173.8	556.9	588.7	11.93
102	0153.	0000.	25.00	18.36	395.4	101.6	168.9	550.8	586.3	11.93
102	0153.	0000.	26.00	18.36	388.0	111.4	156.7	555.7	586.3	11.93
102	0153.	0000.	27.00	17.14	385.6	115.1	184.8	547.1	585.1	11.93
102	0153.	0000.	28.00	17.14	384.3	117.5	187.3	556.9	587.5	11.93

Στόχος της ανάλυσης είναι ο διαχωρισμός του ηλεκτρομαγνητικού σήματος μακράς μνήμης σε επιμέρους τμήματα ώστε να ανιχνεύσουμε μέσα σε αυτά τα τμήματα των χρονοσειρών τη φρακταλική διάσταση και τον εκθέτη Hurst. Η διαδικασία που ακολουθείται για την ανάλυση είναι η εξής:

Αρχικά, δηλώνουμε αν θα επεξεργαστούμε ξεχωριστά αρχεία (επιλογή 0) ή ένα ολόκληρο αρχείο (επιλογή 1). Η επιλογή που μας ενδιαφέρει είναι η επιλογή 1 οπότε και προσδιορίζουμε κάποια βασικά χαρακτηριστικά του σήματος όπως το έτος, την ημέρα ή τις ημέρες και το βήμα (increment) που θέλουμε να διαχωρίσουμε την επεξεργασία της ανάλυσης, ώστε να τα χρησιμοποιήσουμε αργότερα ως πληροφορία στο διάγραμμα. Στη συνέχεια δηλώνουμε το κανάλι και τέλος το όνομα του αρχείου από το οποίο θα ληφθούν τα δεδομένα. Το αρχείο του σήματος (στην περίπτωση που εξετάζουμε, το κανάλι 9) που έχει προστεθεί για επεξεργασία μετατρέπεται σε .TS για να χρησιμοποιηθεί αργότερα στην εφαρμογή *Benoit*.

```
disp('=====');
disp('          Program to perform R/S analysis');
disp(' and plot the Hurst exponent evolution with time');
disp('=====');
disp('Development:');
disp('For Octave by Dr.Dimitrios Nikolopoulos, June 2014');
disp('Sliding version');
disp('=====');
disp('');

%=====
%          read the file
%=====

batchread=0;
disp('          The Signals are expected to have a *.dat extension');

batchread=input('Manipulate one or more EM Signal files (0) or a
certain Signal File (1)...> ');

year = input('Please provide year...> ');
station_code = input('Please provide station code...> ','s');
```

```

disp('Please provide days range');
startday=input('Start day ...> ');
endday=input('End day ...> ');
increment=input('Increment...> ');

fileextension=['.dat'];
disp('');
disp('Loading Signal(s).....');
.
.
.
.

filename = [mainfilename,fileextension];
commandstr = ['load ',datadir,filename];

commandstr= ['Signal=',mainfilename,'];'];
eval(commandstr);

clc;
Channel= input('Please provide channel to process...> ');
AvData=Signal(:,Channel);

mainfilename = input('Please provide the name of *.dat file to be
investigated ...> ', 's');
fileextension=['.dat'];
filename = [mainfilename,fileextension];

%read file
disp('');
disp('Loading signal....');
commandstr = ['load ',datadir,filename];

clear Signal;
disp('Signal is loaded....');
end

```



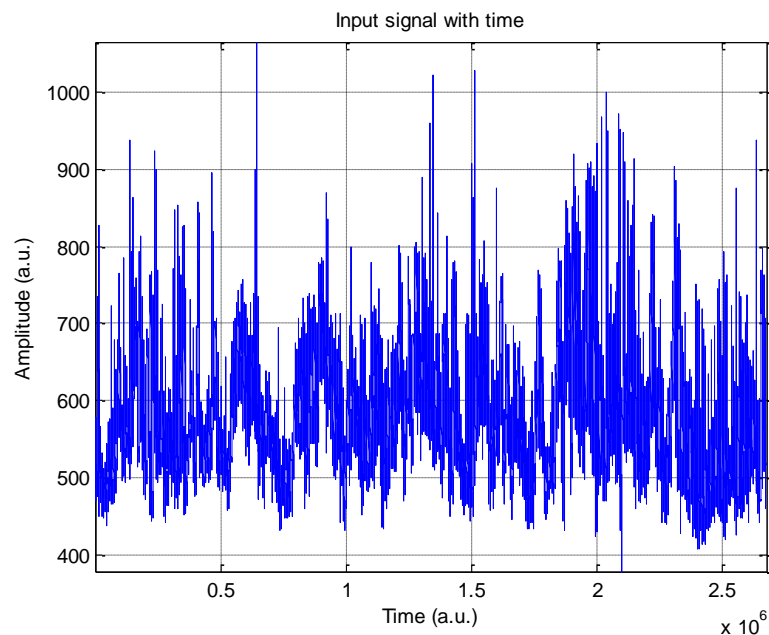
Στο τμήμα του κώδικα που αφορά το διάγραμμα η βασική εντολή αναφέρεται στον προσδιορισμό των επιμέρους τμημάτων που θα διαχωρίσουμε το σήμα. Ο διαχωρισμός έγινε σε τέσσερα τμήματα.

```

%=====
%           PLOT TS
%=====
dummy=input('Please press any key to plot the signal and select
areas','s');
figure(1)

grid;
xlabel('Time (a.u)');
if (batchread==0)
    titlestring=['Year: ', num2str(year), ' Start day:
',num2str(startday), ' End day: ', num2str(endday)];
    ylabelstring=[station_code,num2str(Channel), ' (a.u.)'];
else
    titlestring=filename;
    ylabelstring='(a.u.)';
end

```



Διάγραμμα 6.1.1 Ηλεκτρομαγνητικό σήμα μίας ημέρας

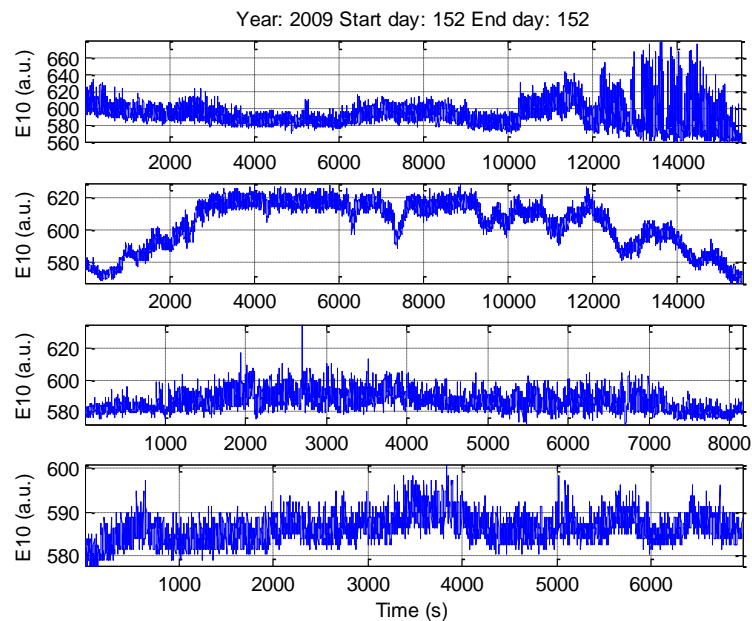
```

%=====
% GET regions of interest from user
%=====
% prompt for the number of areas to be analyzed
ar_num = input( 'Region of Interest (ROI) determination. Please give
number of ROI: ');
num = ar_num*2;
% Get areas from user, by mouse pointing and selection
disp(['You will have to select ', num2str(num),...
      ' points by positioning and left-clicking the mouse'])

for n=1:num
    [x,y] = ginput(1);
    xi(n) = round(x);
    message=['Point ', num2str(n), ' is selected']
    disp(message);

clc;
disp('ROI stored in memory');

```



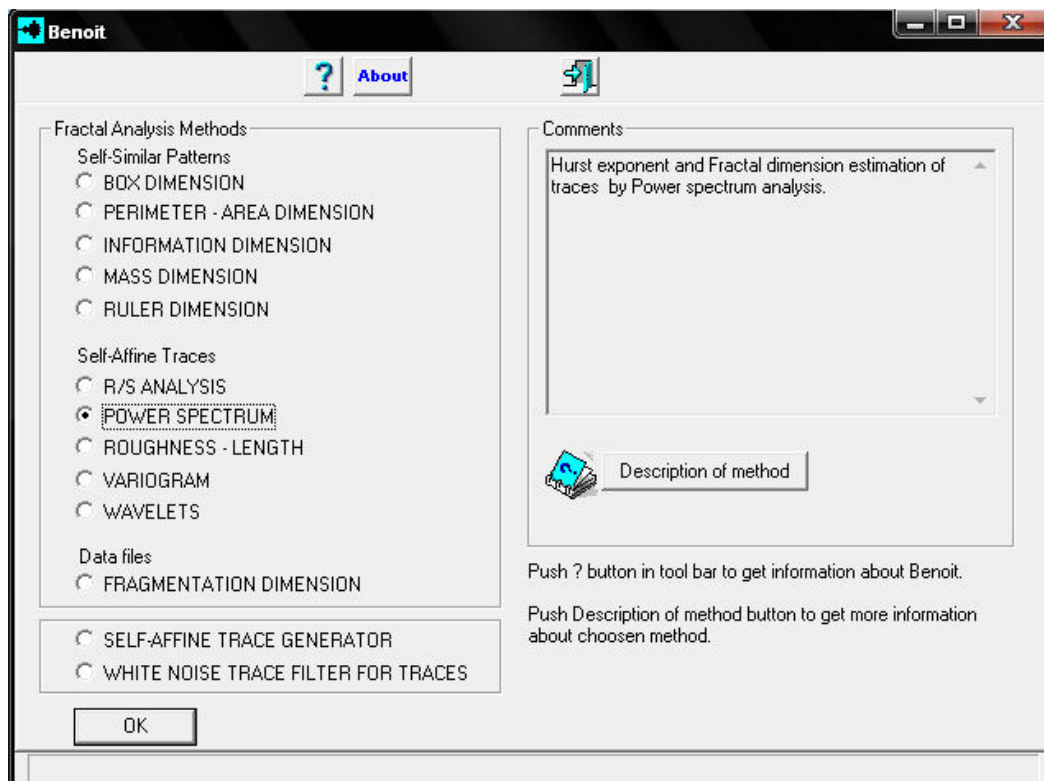
Διάγραμμα 6.1.2. Διαχωρισμός σήματος σε 4 τμήματα

## 6.2 Η εφαρμογή BENOIT

Το Λογισμικό BENOIT™ είναι μία εφαρμογή υπολογιστών η οποία δίνει τη δυνατότητα να μετρήσουμε τη φρακταλική διάσταση ή /και τον εκθέτη Hurst των συνόλων δεδομένων μας χρησιμοποιώντας έντεκα μεθόδους:

- κανόνα, κουτιού, πληροφορίας, περιμέτρου-περιοχής για αυτο-όμοια (self-similar) μοτίβα (2D δεδομένα)
- R/S, variogram, R/L, ανάλυση φάσματος ισχύος (power spectral), και wavelet για αυτο-συγγενικά ίχνη (self-affine) (1D δεδομένα)
- θρυμματισμού (fragmentation) για δεδομένα συχνότητας

Το φιλτράρισμα παρέχεται από το πρόγραμμα για την αφαίρεση του λευκού θορύβου από τα ίχνη με τη μέθοδο είτε Fourier είτε Wavelet [30].



Εικ 6.2.1 Βασικό μενού της εφαρμογής Benoit

### 6.2.1 Εύρεση εκθέτη Hurst και συναρτήσεις μετασχηματισμού ιχνών

Αν τα ίχνη δεδομένων έχουν μία κλίση μεταξύ -1 και 1 σε ένα διάγραμμα log-log τότε το *Benoit* δεν θα μπορέσει να τα διαχωρίσει και θα δώσει ως αποτέλεσμα  $H=0$  και  $D=2$ . Για το λόγο αυτό χρησιμοποιεί δυο απλούς μετασχηματισμούς που μετατοπίζουν την κλίση και συνεπώς το  $H$  και το  $D$ . Ο μετασχηματισμός που χρησιμοποιήθηκε στην ανάλυση είναι ο μετασχηματισμός *Sum*:

$$z_i = \sum_{j=i}^i (y_j - \bar{y})$$

$z_i$  είναι η  $i$ -οστή τιμή του άξονα  $y$  στο τρέχων άθροισμα

$y_j$  είναι η  $j$ -οστή τιμή του άξονα  $y$  στο αρχικό ίχνος

$\bar{y}$  είναι η μέση τιμή του άξονα  $y$  στο αρχικό ίχνος

Η επίδραση αυτού του μετασχηματισμού είναι η μετατόπιση της κλίσης της αρχικής σειράς κατά ένα παράγοντα +2. Έτσι, εάν η κλίση της αρχικής χρονοσειράς ήταν -1, τότε η κλίση της μετασχηματισμένης χρονοσειράς είναι 1 [2].

### 6.2.2 Υπολογισμός Φρακταλικής Διάστασης και Εκθέτη Hurst

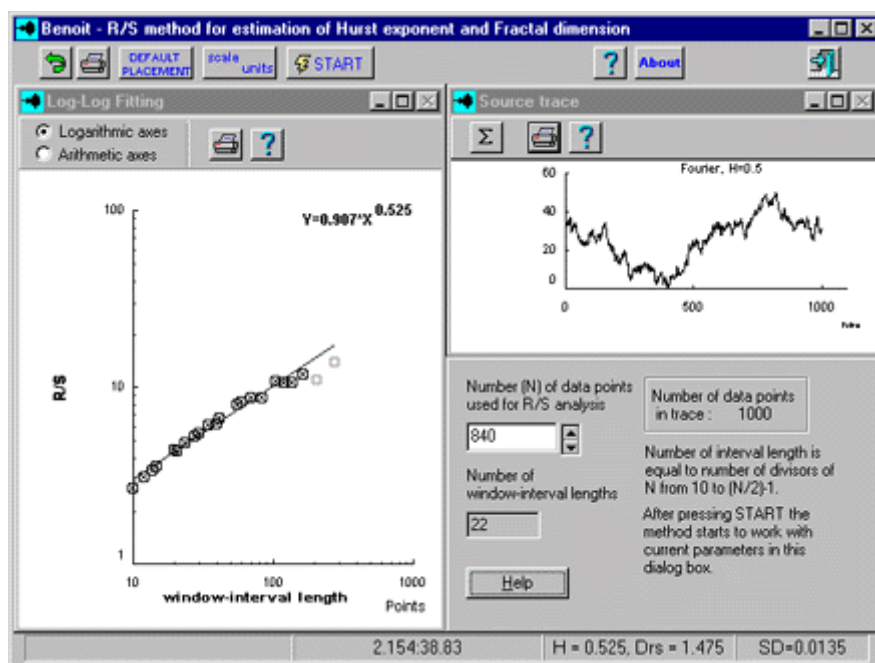
Για τον υπολογισμό του εκθέτη Hurst και της φρακταλικής διάστασης μέσω της R/S μεθόδου ακολουθούμε την παρακάτω διαδικασία:

- επιλέγουμε τη μέθοδο R/S από το βασικό μενού της εφαρμογής (εικ.1.1),
- φορτώνουμε το αρχείο .TS (εφόσον το έχουμε αποθηκεύσει σε κάποιο φάκελο του *Benoit*) και
- επιλέγουμε το μετασχηματισμό που θέλουμε να εφαρμόσουμε στη χρονοσειρά, στην περίπτωση μας είναι ο μετασχηματισμός *Sum*

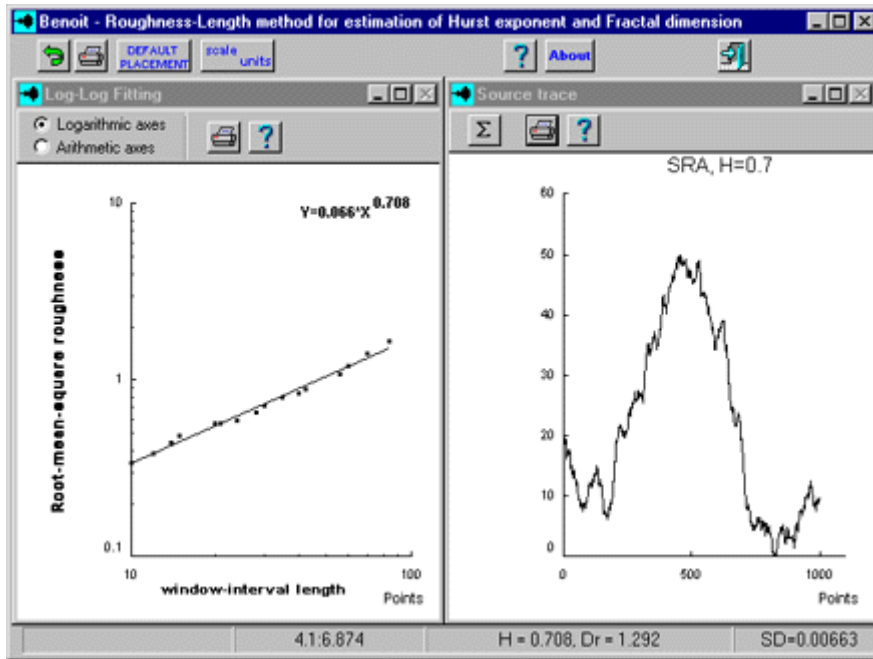
και στη συνέχεια το *Benoit* εμφανίζει το παράθυρο της μεθόδου R/S με το αποτέλεσμα του υπολογισμού της φρακταλικής διάστασης και του εκθέτη Hurst. Τα αποτελέσματα τα αντιγράφουμε σε ένα αρχείο excel.

Η ίδια διαδικασία ακολουθείται και για τις μεθόδους R/L και Variogram.

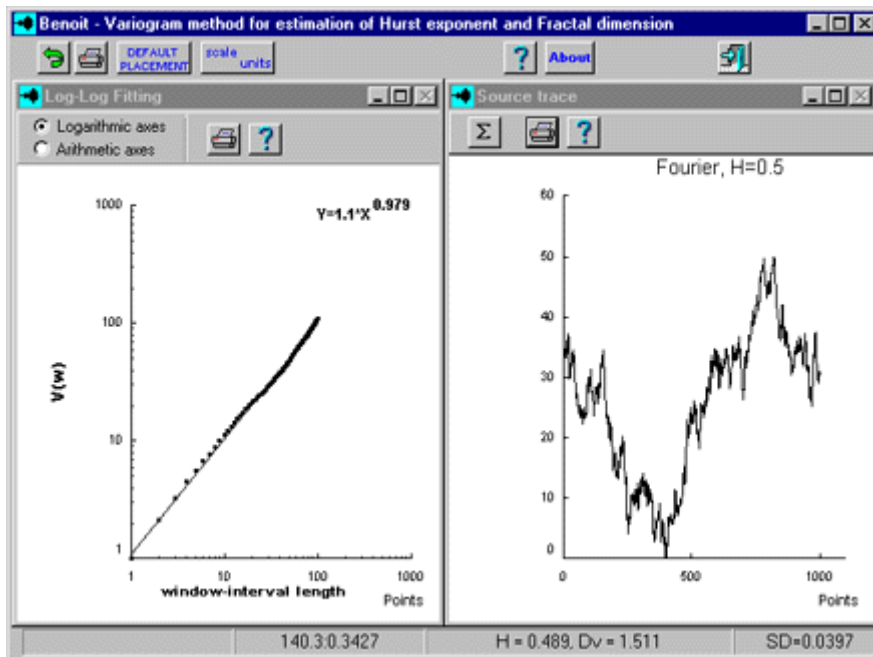
Στις εικόνες (6.2.2.1),(6.2.2.2),(6.2.2.3) βλέπουμε τα παράθυρα για τις αντίστοιχες μεθόδους, ενώ στον πίνακα (6.2.1) που ακολουθεί βλέπουμε τις τιμές της φρακταλικής διάστασης και του εκθέτη Hurst, από ένα αρχείο excel, για το κανάλι 9.



Εικ 6.2.2.1 Το παράθυρο της μεθόδου R/S



Εικ 6.2.2.2 Το παράθυρο της μεθόδου R/L



Εικ 6.2.2.3 Το παράθυρο της μεθόδου Variogram [2]

Πίνακας 6.2.1 Ανάλυση ηλεκτρομαγνητικού σήματος

	R/S		R/L		Variogram	
	Hurst	FD	Hurst	FD	Hurst	FD
<b>152</b>						
signal 1	1,016	0.984	0.969	1,031	0.941	1,059
signal 2	0.970	1,030	0.948	1,052	0.898	1,102
signal 3	0.881	1,119	0.943	1,057	0.892	1,108
signal 4	0.981	1,019	0.980	1,020	0.945	1,055
<b>153</b>						
signal 1	0.945	1,055	0.924	1,076	0.926	1,074
signal 2	0.938	1,062	0.879	1,121	0.988	1,012
signal 3	0.906	1,094	0.851	1,149	0.916	1,084
signal 4	0.987	1,013	0.868	1,132	0.951	1,049
<b>154</b>						
signal 1	0.975	1,025	0.974	1,026	0.981	1,019
signal 2	0.955	1,045	0.887	1,113	0.977	1,023
signal 3	0.970	1,030	0.963	1,037	0.976	1,024
signal 4	0.877	1,123	0.858	1,142	0.882	1,118
<b>155</b>						
signal 1	0.836	1,164	0.838	1,162	0.878	1,122
signal 2	1,005	0.995	0.993	1,007	0.932	1,068
signal 3	0.983	1,017	0.933	1,067	0.948	1,052
signal 4	0.752	1,248	0.704	1,296	0.831	1,169
<b>156</b>						
signal 1	0.889	1,111	0.870	1,130	0.897	1,103
signal 2	0.969	1,031	1,005	0.995	0.926	1,074
signal 3	0.969	1,031	0.893	1,107	0.954	1,046
signal 4	0.914	1,086	0.918	1,082	0.914	1,086

<b>157</b>						
signal 1	1,013	0.987	1,031	0.969	0.968	1,032
signal 2	0.917	1,083	0.838	1,162	0.963	1,037
signal 3	0.942	1,058	0.799	1,201	0.966	1,034
signal 4	0.942	1,058	0.998	1,002	0.938	1,062
<b>158</b>						
signal 1	0.817	1,183	0.662	1,338	0.888	1,039
signal 2	0.988	1,012	0.973	1,027	0.945	1,055
signal 3	0.888	1,112	0.843	1,157	0.896	1,104
signal 4						
<b>159</b>						
signal 1	0.834	1,166	0.821	1,179	0.903	1,097
signal 2	0.877	1,123	0.806	1,194	0.948	1,052
signal 3	0.948	1,052	0.915	1,085	0.870	1,003
signal 4	0.856	1,144	0.782	1,218	0.964	1,036
<b>160</b>						
signal 1	0.853	1,147	0.797	1,203	0.956	1,044
signal 2	0.950	1,050	0.930	1,070	0.941	1,059
signal 3	0.963	1,037	0.914	1,086	0.977	1,023
signal 4	0.992	1,008	0.992	1,008	0.950	1,050
<b>161</b>						
signal 1	1,015	0.985	1,011	0.989	0.952	1,048
signal 2	0.947	1,053	0.972	0.972	0.965	1,035
signal 3	1,021	0.979	1,004	0.996	0.964	1,036
signal 4	1,630	0.370	0.878	1,122	0.901	1,099



<b>162</b>						
signal 1	0.888	1,112	0.837	1,163	0.893	1,107
signal 2	0.918	1,082	0.911	1,089	0.846	1,154
signal 3	0.883	1,117	0.884	1,116	0.916	1,084
signal 4	0.865	1,135	0.784	1,216	0.968	1,032
<b>163</b>						
signal 1	0.876	1,124	0.863	1,137	0.924	1,076
signal 2	0.890	1,110	0.898	1,102	0.883	1,117
signal 3	0.949	1,051	0.930	1,070	0.910	1,090
signal 4	0.883	1,117	0.807	1,193	0.926	1,074
<b>164</b>						
signal 1	0.860	1,140	0.781	1,219	0.948	1,052
signal 2	0.884	1,116	0.836	1,164	0.935	1,065
signal 3	0.950	1,050	0.923	1,077	0.946	1,054
signal 4	0.867	1,133	0.842	1,158	0.968	1,032
<b>165</b>						
signal 1	0.917	1,083	0.858	1,142	0.960	1,040
signal 2	0.979	1,021	0.981	1,019	0.979	1,021
signal 3	0.946	1,054	0.966	1,034	0.919	1,081
signal 4	0.986	1,014	0.996	1,004	0.901	1,099
<b>166</b>						
signal 1	0.906	1,094	0.904	1,096	0.898	1,102
signal 2	1,022	0.978	1,025	0.975	0.949	1,051
signal 3	0.982	1,018	0.993	1,007	0.967	1,033
signal 4	0.975	1,025	0.980	1,020	0.879	1,121

<b>167</b>						
signal 1	0.951	1,049	0.925	1,075	0.958	1,042
signal 2	0.983	1,017	0.992	1,008	0.919	1,081
signal 3	0.866	1,134	0.838	1,162	0.865	1,135
signal 4	0.957	1,043	0.893	1,107	0.956	1,044
<b>168</b>						
signal 1	1,026	0.974	1,037	0.963	0.966	1,034
signal 2	0.904	1,096	0.856	1,144	0.894	1,106
signal 3	1,000	1,000	1,031	0.969	0.892	1,108
signal 4	0.975	1,025	0.960	1,040	0.925	1,075
<b>169</b>						
signal 1	0.851	1,149	0.774	1,226	0.952	1,048
signal 2	0.968	1,032	0.985	1,015	0.915	1,085
signal 3	0.946	1,054	0.962	1,038	0.878	1,122
signal 4	0.935	1,065	0.965	1,035	0.861	1,139
<b>170</b>						
signal 1	0.829	1,171	0.805	1,195	0.905	1,095
signal 2	0.997	1,003	1,030	0.970	0.939	1,061
signal 3	1,050	0.950	1,010	0.990	0.997	1,003
signal 4	0.989	1,011	0.969	1,031	0.944	1,056
<b>171</b>						
signal 1	0.896	1,104	0.873	1,127	0.913	1,087
signal 2	1,003	0.997	1,020	0.980	0.956	1,044
signal 3	0.854	1,146	0.829	1,171	0.853	1,147
signal 4	0.932	1,068	0.910	1,090	0.964	1,036

<b>172</b>						
signal 1	0.869	1,131	0.826	1,174	0.920	1,080
signal 2	0.937	1,063	0.856	1,144	0.977	1,023
signal 3	0.982	1,018	0.969	1,031	0.974	1,026
signal 4	0.978	1,022	0.954	1,046	0.885	1,115
<b>173</b>						
signal 1	1,007	0.993	1,025	0.975	0.923	1,077
signal 2	0.805	1,195	0.775	1,225	0.883	1,117
signal 3	0.972	1,028	0.991	1,009	0.976	1,024
signal 4	0.827	1,173	0.749	1,251	0.964	1,036
<b>174</b>						
signal 1	0.868	1,132	0.762	1,238	0.980	1,020
signal 2	0.732	1,268	0.677	1,323	0.917	1,083
signal 3	0.913	1,087	0.827	1,173	0.991	1,009
signal 4	0.864	1,136	0.851	1,149	0.870	1,130
<b>175</b>						
signal 1	0.812	1,188	0.818	1,182	0.894	1,106
signal 2	0.798	1,202	0.777	1,223	0.824	1,176
signal 3	0.874	1,126	0.808	1,192	0.975	1,025
signal 4	0.761	1,239	0.694	1,306	0.830	1,170
<b>176</b>						
signal 1	0.934	1,066	0.940	1,060	0.911	1,089
signal 2	0.985	1,015	0.995	1,005	0.936	1,064
signal 3	0.915	1,085	0.845	0.845	0.880	1,120
signal 4	0.757	1,243	0.729	1,271	0.892	1,108

<b>177</b>						
signal 1	0.873	1,127	0.794	1,206	0.889	1,111
signal 2	1,037	0.963	1,025	0.975	0.968	1,032
signal 3	0.936	1,064	0.854	1,146	0.931	1,069
signal 4	0.895	1,105	0.904	1,096	0.914	1,086
<b>178</b>						
signal 1	0.995	1,005	1,037	0.963	0.854	1,146
signal 2	0.970	1,030	0.952	1,048	0.939	1,061
signal 3	0.934	1,066	0.962	1,038	0.959	1,041
signal 4	0.920	1,080	0.876	1,124	0.891	1,109
<b>179</b>						
signal 1	1,039	0.961	1,037	0.963	0.863	1,137
signal 2	0.866	1,134	0.818	1,182	0.902	1,098
signal 3	1,028	0.972	1,006	0.994	0.937	1,063
signal 4	0.938	1,062	0.942	1,058	0.886	1,114
<b>180</b>						
signal 1	0.988	1,012	0.987	1,013	0.860	1,140
signal 2	0.928	1,072	0.916	1,084	0.929	1,071
signal 3	0.954	1,046	0.115	1,885	0.877	1,123
signal 4	0.996	1,004	0.946	1,054	0.110	1,890
<b>181</b>						
signal 1	0.906	1,094	0.938	1,062	0.847	1,153
signal 2	0.975	1,025	0.917	1,083	0.961	1,039
signal 3	0.840	1,160	0.772	1,228	0.957	1,043
signal 4	0.856	1,144	0.829	1,171	0.925	1,075

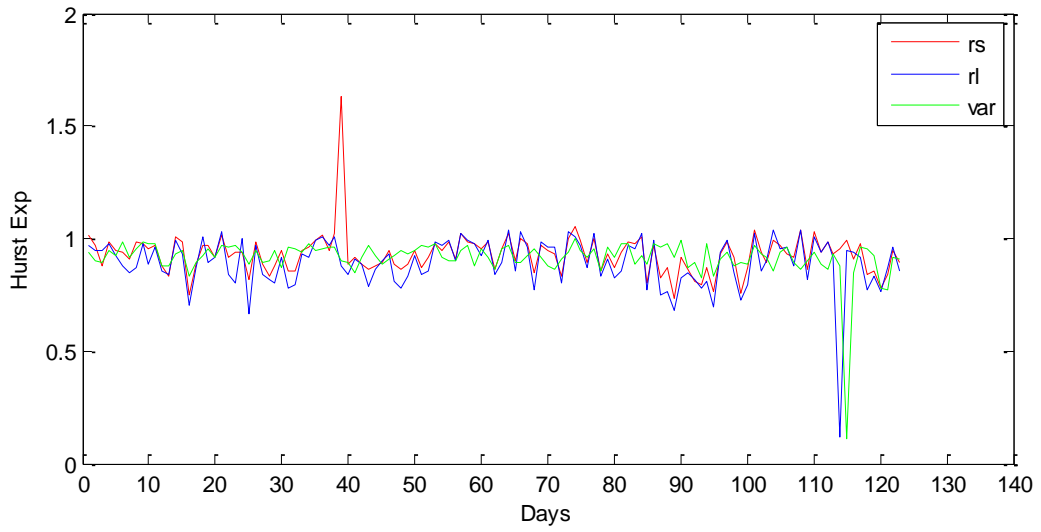
<b>182</b>						
signal 1	0.779	1,221	0.765	1,235	0.778	1,222
signal 2	0.844	1,156	0.873	1,127	0.773	1,227
signal 3	0.944	1,056	0.964	1,036	0.918	1,082
signal 4	0.896	1,104	0.854	1,146	0.906	1,094

Όπως έχει αναφερθεί, το σήμα μίας ημέρας έχει χωριστεί σε τέσσερα μικρότερα τμήματα για τη μελέτη της μακράς μνήμης, και στον πίνακα 6.2.1 βλέπουμε για κάθε ημέρα την τιμή του εκθέτη Hurst καθώς και της φρακταλικής διάστασης (FD) για κάθε ένα τμήμα του σήματος.

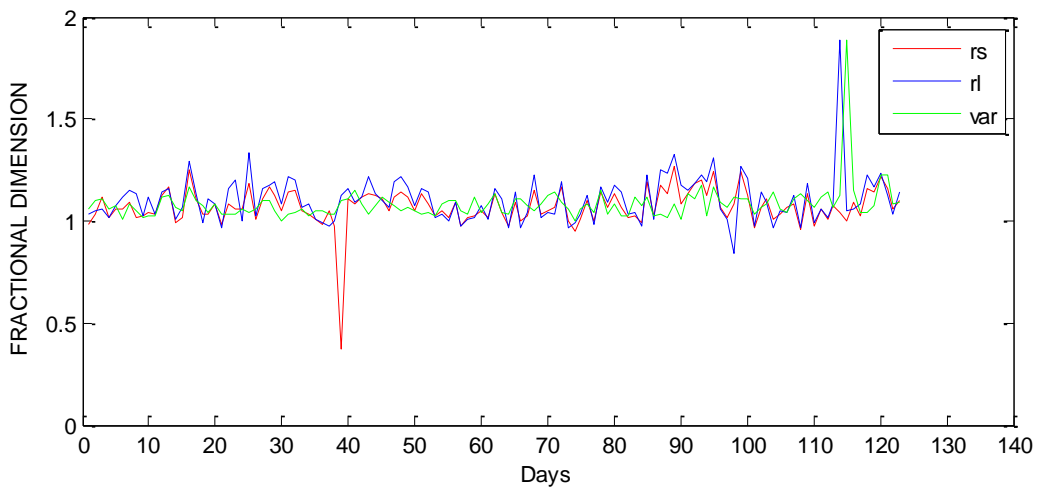
Στη συνέχεια, οι τιμές του πίνακα 6.2.1 μεταφέρονται σε πίνακες στο MATLAB για την εξαγωγή του τελικού διαγράμματος.

### 6.3 Διαγράμματα – Εκτιμήσεις

Τα αποτελέσματα της ανάλυσης ακολουθούν τη θεωρία. Παρατηρούμε ότι καθώς αυξάνεται ο εκθέτης Hurst η κλασματική διάσταση μειώνεται. Οι μέθοδοι R/S, R/L και Variogram ακολουθούν μία κανονικότητα. Το σύνολο τιμών του εκθέτη Hurst  $H$  και της κλασματικής διάστασης  $D$  βρίσκεται μεταξύ του  $(0,2)$ . Συγκεκριμένα για το κανάλι 9 (διάγραμμα 6.3.1) ο εκθέτης Hurst κυμαίνεται μεταξύ 0.5 και 1. Αυτό μας πληροφορεί ότι το σήμα έχει μακρά μνήμη (persistence). Αντιστρόφως, η κλασματική διάσταση του σήματος (διάγραμμα 6.3.2) βρίσκεται μεταξύ 1 και 1.5 και μας πληροφορεί ότι η χρονοσειρά ακολουθεί μία κανονικότητα και είναι προβλέψιμη [31].

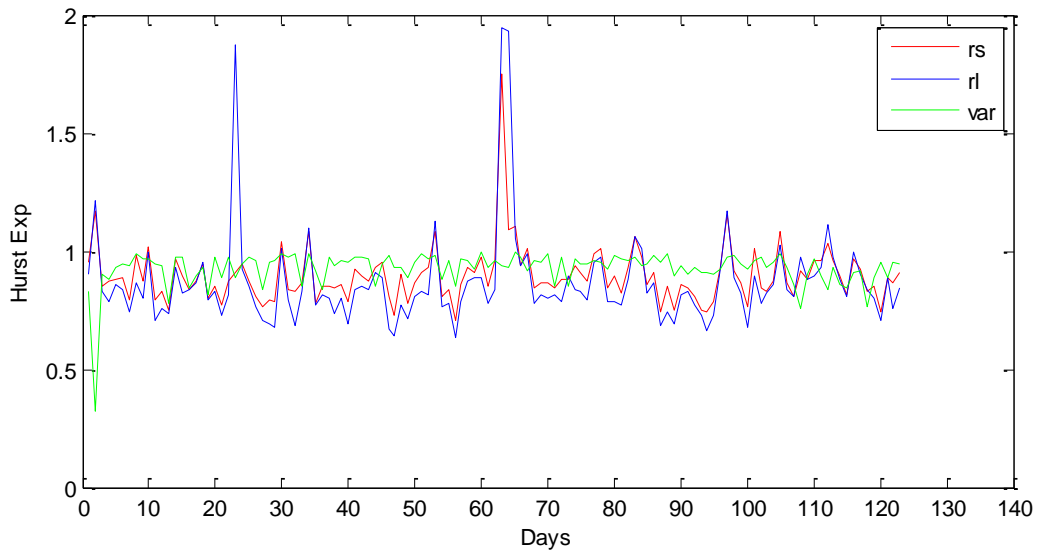


Διάγραμμα 6.3.1 Εκθέτης Hurst (κανάλι 9)

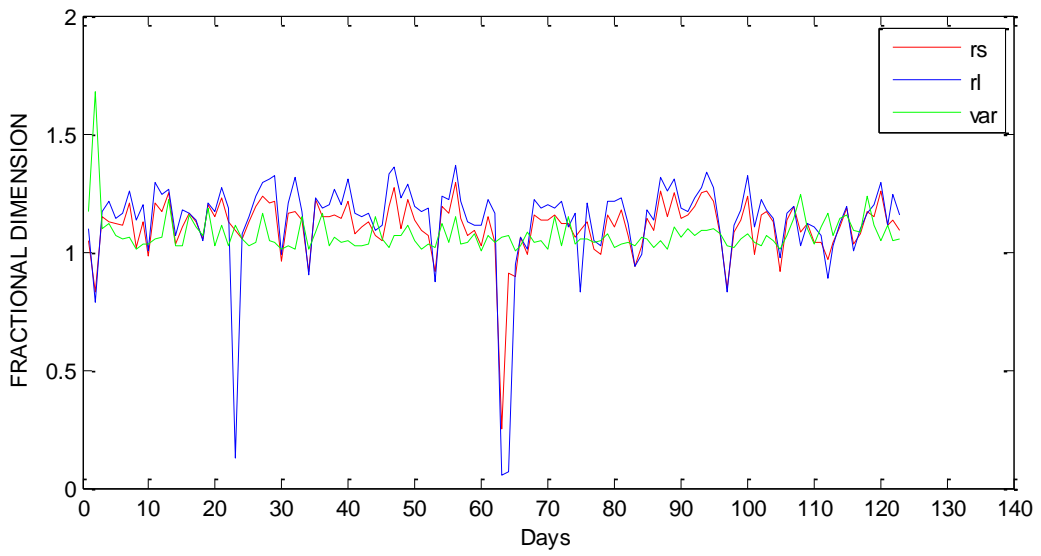


Διάγραμμα 6.3.2 Φρακταλική Διάσταση (κανάλι 9)

Ανάλογα συμπεράσματα προκύπτουν και για το κανάλι 10.



Διάγραμμα 6.3.3 Εκθέτης Hurst (κανάλι 10)



Διάγραμμα 6.3.4 Φρακταλική Διάσταση (κανάλι 10)





## ΠΑΡΑΡΤΗΜΑ Α

### ΚΕΦΑΛΑΙΟ 7

#### ΧΡΟΝΟΣΕΙΡΕΣ ΑΝΩΤΕΡΩΝ ΔΙΑΣΤΑΣΕΩΝ

Η στατιστική ανωτέρων διαστάσεων αναφέρεται σε στατιστικά αποτελέσματα όταν ο αριθμός των άγνωστων παραμέτρων  $p$  είναι πολύ μεγαλύτερης τάξης από το δειγματοληπτικό μέγεθος  $n, p \gg n$ . Αυτό περιλαμβάνει, επισταμένο έλεγχο στις μεταβολές των μεταβλητών και στην κατάταξη των μοντέλων, όπου ο αριθμός των μεταβολών είναι κατά πολύ μεγαλύτερης τάξης από την  $n$ , μη ελεγχόμενο περιβάλλον, όπως συλλογές, ή γραφική μοντελοποίηση με περισσότερες μεταβλητές από τις παρατηρήσεις ή πολλαπλές δοκιμές όπου ο αριθμός των ελέγχων είναι μεγαλύτερος από τον αριθμό του δείγματος [5].

##### 7.1 Συναρτησιακές χρονοσειρές (Functional time series)

Τα συναρτησιακά δεδομένα προκύπτουν συχνά από μετρήσεις που έχουν ληφθεί από το διαχωρισμό μιας σχεδόν συνεχούς χρόνου καταγραφής σε φυσικά διαστήματα, πχ ημέρες. Παραδείγματα αποτελούν τα δεδομένα από τις καθημερινές καμπύλες οικονομικών συναλλαγών και τα καθημερινά δείγματα γεωφυσικών και περιβαντολλογικών δεδομένων. Οι ληφθείσες συναρτήσεις σχηματίζουν χρονοσειρές  $\{X_k, k \in \mathbb{Z}\}$  όπου κάθε  $X_k$  είναι μία τυχαία συνάρτηση  $\{X_k(t), t \in [a, b]\}$ . Κεντρικό θέμα στην ανάλυση τέτοιων δεδομένων είναι η προσωρινή εξάρτηση των παρατηρήσεων, πχ η εξάρτηση μεταξύ γεγονότων που έχουν προσδιοριστεί από την  $\{X_k, k \leq m\}$  και από την  $\{X_k, k \geq m + h\}$ . Η εστίαση της ανάλυσης συναρτησιακών δεδομένων βρίσκεται περισσότερο στις ανεξάρτητες και πανομοιότητα κατανεμημένες (independent and identically distributed i.i.d) συναρτησιακές παρατηρήσεις.

## 7.2 Χωρικό μοντέλο Hilbert για συναρτησιακά δεδομένα

Συχνά θεωρείται ότι οι παρατηρήσεις  $X_k$  είναι στοιχεία ενός διαχωρίσιμου χώρου Hilbert (H) (πχ. ένας χώρος Hilbert με μετρήσιμη βάση  $\{e_k, k \in \mathbb{Z}\}$ ) με εσωτερικό γινόμενο  $\langle \cdot, \cdot \rangle$  το οποίο δημιουργεί μία νόρμα  $\|\cdot\|$ . Ένα σημαντικό παράδειγμα του χώρου Hilbert είναι  $L^2 = L^2([0,1])$ . Συμβατικά τα δεδομένα  $H$  τιμών θα αναφέρονται ως συναρτησιακές παρατηρήσεις για ένα γενικό χώρο Hilbert. Όλες οι τυχαίες συναρτήσεις ορίζονται σε ένα κοινό χώρο πιθανοτήτων  $(\Omega, \mathcal{A}, P)$ . Το  $X$  είναι ολοκληρώσιμο αν  $E\|X\| < \infty$  και τετραγωνικά ολοκληρώσιμο αν  $E\|X\|^2 < \infty$ . Αν  $E\|X\|^p < \infty, p > 0$ , τότε  $X \in L_H^p = L_H^p(\Omega, \mathcal{A}, P)$ . Η σύγκλιση της  $\{X_n\}$  στη  $X$  στο χώρο  $L_H^p$  ισοδυναμεί με  $E\|X_n - X\|^p \rightarrow 0$ , όπου  $\|X_n - X\| \rightarrow 0$  αναφέρεται σχεδόν πάντα ως σύγκλιση.

## 7.3 Ο χώρος $L^2$

Ο χώρος  $L^2$  είναι ένα σύνολο από μετρητικές (measurable) συναρτήσεις  $x$  πραγματικών τιμών στο  $[0,1]$  ικανοποιώντας τη σχέση  $\int_0^1 x^2(t) dt < \infty$ . Είναι ένας διαχωρίσιμος χώρος Hilbert με εσωτερικό γινόμενο

$$\langle x, y \rangle = \int x(t) y(t) dt. \quad (7.3.1)$$

Το σύμβολο του ολοκληρώματος χωρίς τα όρια ολοκλήρωσης εννοείται ότι δηλώνει όλο το διάστημα  $[0,1]$ . Αν  $x, y \in L^2$  τότε η ισότητα  $x = y$  πάντα ισοδυναμεί με  $\int [x(t) - y(t)]^2 dt = 0$ . Μία σημαντική ομάδα τελεστών στο  $L^2$  είναι οι τελεστές ολοκλήρωσης οριζόμενοι από την

$$\Psi(x)(t) = \int \psi(t, s) x(s) ds, \quad x \in L^2 \quad (7.3.2)$$

Με τον πραγματικό πυρήνα  $\psi(\cdot, \cdot)$ . Τέτοιοι τελεστές καλούνται Hilbert – Schmidt αν και μόνο αν  $\iint \psi^2(t, s) dt ds < \infty$ , στην οποία περίπτωση ισχύει

$$\|\Psi\|_S^2 = \iint \psi^2(t, s) dt ds. \quad (7.3.3)$$

Ένας συμμετρικά θετικά ορισμένος Hilbert – Schmidt τελεστής  $\Psi$  είναι ο

$$\Psi(x) = \sum_{j=1}^{\infty} \lambda_j \langle x, v_j \rangle v_j, \quad x \in H, \quad (7.3.4)$$

Αν  $\psi(s, t) = \psi(t, s)$  και  $\iint \psi(t, s)x(t)x(s) dt ds \geq 0$  ο ολοκληρωτικός τελεστής  $\Psi$  είναι συμμετρικά και θετικά ορισμένος και προκύπτει από την εξίσωση (7.3.4) ότι

$$\psi(t, s) = \sum_{j=1}^{\infty} \lambda_j v_j(t)v_j(s) \quad \text{στο} \quad L^2([0,1] \times [0,1]).$$

Αν το  $\Psi$  είναι συνεχές τότε, η παραπάνω ανάπτυξη ισχύει για κάθε  $s, t \in [0,1]$  και η σειρά συγκλίνει ομοιόμορφα.

#### 7.4 Βασικές Εμπειρικές Συναρτησιακές Συνιστώσες (Empirical Functional Principal Components)

Έστω οι συναρτήσεις  $x_1, x_2, \dots, x_N$ . Θεωρούμε ότι βρίσκονται στο κέντρο, δηλαδή,  $\sum_{i=1}^N x_i = 0$ . Για έναν ορισμένο ακέραιο  $p < N$  θεωρούμε τον  $p$  πολύ μικρό από το  $N$ . Η ορθοκανονική βάση  $u_1, u_2, \dots, u_p$  με

$$\tilde{S}^2 = \sum_{i=1}^N \left\| x_i - \sum_{k=1}^p \langle x_i, u_k \rangle u_k \right\|^2 \quad (7.4.1)$$

ελαχιστοποιείται. Οπότε η σχέση  $\sum_{k=1}^p \langle x_i, u_k \rangle u_k$  είναι μία προσέγγιση στη  $x_i$ . Για το

$p$  που έχουμε επιλέξει αυτή η προσέγγιση είναι ομοιόμορφα βέλτιστη, με την έννοια ότι μικραίνει το  $\widehat{S}^2$ . Αυτό σημαίνει ότι αντι να δουλεύουμε με απείρων διαστάσεων καμπύλες  $x_i$ , μπορούμε να εργαστούμε με διανύσματα  $p$  – διαστάσεων

$$x_i = [\langle x_i, u_1 \rangle, \langle x_i, u_2 \rangle, \dots, \langle x_i, u_p \rangle]^T. \quad (7.4.2)$$

Αυτή είναι μία κεντρική ιδέα της συναρτησιακής ανάλυσης δεδομένων, καθώς εκτελεί πρακτικούς υπολογισμούς, πρέπει να τη μειώσουμε από άπειρο σε πεπερασμένο αριθμό. Οι συναρτήσεις  $u_j$  αθροιστικά ονομάζονται βασικές ορθοκανονικές εμπειρικές βελτιώσεις (optimal empirical orthonormal basis) ή φυσικές ορθοκανονικές συνιστώσες (natural orthonormal components), οι λέξεις “εμπειρικές” και “φυσικές” δίνουν έμφαση στον άμεσο υπολογισμό από τα συναρτησιακά δεδομένα. Οι συναρτήσεις  $u_1, u_2, \dots, u_p$  ελαχιστοποιώντας την  $\widehat{S}^2$  είναι ισοδύναμες (μέχρι ένα σημείο) στην κανονικοποίηση των ιδιοσυναρτήσεων  $\widehat{v}_1, \widehat{v}_2, \dots, \widehat{v}_p$ , του δειγματοληπτικού τελεστή συμμεταβλητότητας (covariance)  $C(y) = E[(X - EX, y)(X - EX)]$ ,  $y \in H$ , δηλαδή  $\widehat{C}(u_i) = \widehat{\lambda}_i u_i$ , όπου  $\widehat{\lambda}_1 \geq \widehat{\lambda}_2 \geq \dots \geq \widehat{\lambda}_p$ . Οι ιδιοσυναρτήσεις  $\widehat{v}_i$  καλούνται βασικές εμπειρικές συναρτησιακές συνιστώσες (EFPCs) των δεδομένων  $x_1, x_2, \dots, x_N$ . Γι’ αυτό το λόγο, οι  $\widehat{v}_i$  είναι οι φυσικές ορθοκανονικές συνιστώσες και σχηματίζουν τη βέλτιστη εμπειρική ορθοκανονική βάση.

## 7.5 Συναρτησιακό αυτομειωμένο μοντέλο (Functional autoregressive model FAR)

Η ακολουθία  $\{X_n, -\infty < n < \infty\}$  με συναρτήσεις μηδενικού μέσου στον  $H$  ακολουθεί ένα συναρτησιακό  $AR(1)$  μοντέλο αν  $X_n = \Psi(X_{n-1}) + \varepsilon_n$  (7.5.1) όπου  $\Psi \in \mathcal{L}$  και  $\varepsilon_n, -\infty < n < \infty$  είναι η ακολουθία των ανεξάρτητων και πανομοιότυπα κατανομημένων (i.i.d.) λαθών στον  $H$  ικανοποιώντας τη σχέση  $E\|\varepsilon_n\|^2 < \infty$ . Η θεωρία του υπολογισμού της διαδικασίας (7.5.1) αναπτύχθηκε με την υπόθεση ότι τα λάθη είναι i.i.d.

Οι εξισώσεις της βαθμωτής AR(1) διαδικασίας αποδέχονται τη μοναδική τυχαία λύση  $X_n = \sum_{j=0}^{\infty} \psi^j \varepsilon_{n-j}$  εάν  $|\psi| < 1$ .

**Λήμμα 7.5.1:** Για οποιοδήποτε  $\Psi \in \mathcal{L}$  οι επόμενες δύο συνθήκες (conditions) είναι ισοδύναμες:

C0: Η ύπαρξη ενός ακεραίου  $j_0$  τέτοιος ώστε  $\|\psi\|^{j_0} < 1$ .

C1: Υπάρχει  $a > 0$  και  $0 < \beta < 1$  τέτοιο ώστε για κάθε  $j \geq 0$ ,  $\|\psi^j\| \leq a\beta^j$ .

Η συνθήκη C1 είναι επαρκώς ισχυρή ώστε να διασφαλίσει τη σύγκλιση της σειράς  $\sum_j \Psi^j(\varepsilon_{n-j})$  και την ύπαρξη μιας στάσιμης τυχαίας λύσης στις συναρτησιακές εξισώσεις AR(1). Αν η συνθήκη C0 ευσταθεί, τότε υπάρχει μια μοναδική τυχαία λύση της εξίσωσης (7.5.1). Η λύση δίνεται από την

$$X_n = \sum_{j=0}^{\infty} \Psi^j(\varepsilon_{n-j}). \quad (7.5.2)$$

Η σειρά συγκλίνει σχεδόν σίγουρα και στον  $L_H^2$ .

## 7.6 Υπολογισμός του αυτομειωμένου τελεστή $\Psi$

**Θεώρημα 7.6.1:** υποθέτουμε ότι οι παρατηρήσεις  $x_1, x_2, \dots, x_N$  είναι i.i.d. στον  $H$  και ότι έχει την ίδια κατανομή με την  $X$  η οποία θεωρούμε ότι βρίσκεται στον  $L_H^4$  με  $EX = 0$ . Θεωρούμε ότι  $\lambda_1 > \lambda_2 > \dots > \lambda_d > \lambda_{d+1}$ . (7.6.1)

Τότε, για κάθε  $1 \leq j \leq d$ , έχουμε

$$E \left[ \|\hat{\varepsilon}_j - \hat{v}_j - \hat{v}_j\|^2 \right] = O(N^{-1}), \quad E \left[ |\lambda_j - \tilde{\lambda}_j|^2 \right] = O(N^{-1}). \quad (7.6.2)$$

**Θεώρημα 7.6.2:** έστω ότι ο τελεστής  $\Psi$  στην εξίσωση (7.5.1) ικανοποιεί τη συνθήκη C0 στο λήμα 1 και η λύση  $\{X_n\}$  ικανοποιεί την  $E\|X_0\|^4 < \infty$ . Εάν η (7.6.1) ευσταθεί, τότε για κάθε  $1 \leq j \leq d$ , οι σχέσεις στην εξίσωση (7.6.2) ευσταθούν.

Πρώτα θα επικεντρωθούμε στην σταθερή (univariate) περίπτωση  $X_n = \psi X_{n-1} + \varepsilon_n$  στην οποία όλες οι ποσότητες είναι βαθμοτές. Υποθέτουμε ότι  $|\psi| < 1$  ώστε να υπάρχει μία στάσιμη (stationary) λύση, τέτοια ώστε το  $\varepsilon_n$  να είναι ανεξάρτητο της  $X_{n-1}$ . Τότε, πολλαπλασιάζοντας την εξίσωση AR(1) με  $X_{n-1}$ , λαμβάνουμε  $\gamma_1 = \psi\gamma_0$ , όπου  $\gamma_k = E[X_n X_{n+k}] = Cov(X_n, X_{n+k})$ . Οι αυτοσυμμεταβολές (autocovarianes)  $\gamma_k$  εκτιμώνται από το δείγμα συμμεταβολών  $\hat{\gamma}_k$  οπότε ο κανονικός υπολογισμός του  $\psi$  είναι  $\hat{\psi} = \hat{\gamma}_1 / \hat{\gamma}_0$ .

Αυτός ο υπολογισμός είναι βέλτιστος με κάθε τρόπο, και η διαδικασία που περιγράψαμε πιά πάνω είναι γνωστή ως υπολογισμός Yule –Walker και εφαρμόζεται στη διαδικασία αυτομειωμένου μέσου όρου σε ανώτερες διαστάσεις και πολυμεταβλητών (multivariate) αυτομειωμένων διαδικασιών.

*Εφαρμογή της τεχνικής στο συναρτησιακό μοντέλο που περιγράφεται από την (7.5.1) υπό τη συνθήκη C0 του λήματος 7.5.1*

Έχουμε  $E[\langle X_n, x \rangle X_{n-1}] = E[\langle \Psi(X_{n-1}), x \rangle X_{n-1}]$ ,  $x \in H$ .

Ο τελεστής αυτοσυμμεταβολής καθυστέρησης -1 δίνεται από την  $C_1(x) = E[\langle X_n, x \rangle X_{n+1}]$  και δηλώνοντας τον εκθέτη  $T$  προκύπτει η ισότητα  $C_1^T = C\Psi^T$ . Από αυτήν την ισότητα μπορούμε να εξάγουμε μία εκτίμηση του  $\Psi$  χρησιμοποιώντας ένα δείγμα της σχέσης  $\Psi = C_1 C^{-1}$ . Ωστόσο, ο τελεστής  $C$  δεν έχει μία οριοθετημένη αντιστροφή σε όλο το  $H$ . Αυτό κάνει δύσκολο τον υπολογισμό του οριοθετημένου (bounded) τελεστή  $\Psi$  χρησιμοποιώντας τη σχέση  $\Psi = C_1 C^{-1}$ . Μία πρακτική λύση είναι να χρησιμοποιήσουμε μόνο τα πρώτα  $p$ , τις πιο σημαντικές συνιστώσες  $\hat{v}_j$  των EFPC και να ορίσουμε τη σχέση

$$\tilde{I}\tilde{C}_p(x) = \sum_{j=1}^p \tilde{\lambda}_j^{-1}(x, \hat{v}_j) \hat{v}_j. \quad (7.6.3)$$

Ο τελεστής  $\tilde{I}\tilde{C}_p$  ορίζεται σε όλο το  $L^2$  και οριοθετείται αν  $\hat{\lambda}_j > 0$  για  $j \leq p$ . Επιλέγοντας τα  $p$  βρίσκουμε μία ισορροπία μεταξύ της παραμονής της σχετικής πληροφορίας στο δείγμα και του κινδύνου να δουλεύουμε με τις αντιστροφές (reciprocals) μικρών ιδιοτιμών  $\hat{\lambda}_j$ . Για να εξάγουμε μία υπολογίσιμη εκτίμηση του  $\Psi$  χρησιμοποιούμε μια εμπειρική έκδοση της σχέσης  $C_1 = \Psi C$ . Επειδή η  $C_1$  εκτιμάται από την

$$\tilde{C}_1(x) = \frac{1}{N-1} \sum_{k=1}^{N-1} \langle X_k, x \rangle X_{k+1}, \quad (7.6.4)$$

λαμβάνουμε δεδομένα για οποιοδήποτε  $x \in H$ ,

$$\begin{aligned} \tilde{C}_1 \tilde{I}\tilde{C}_p(x) &= \tilde{C}_1 \left( \sum_{j=1}^p \hat{\lambda}_j^{-1} \langle x, \hat{v}_j \rangle \hat{v}_j \right) \\ &= \frac{1}{N-1} \sum_{k=1}^{N-1} \langle X_k, \sum_{j=1}^p \hat{\lambda}_j^{-1} \langle x, \hat{v}_j \rangle \hat{v}_j \rangle X_{k+1} \\ &= \frac{1}{N-1} \sum_{k=1}^{N-1} \sum_{j=1}^p \hat{\lambda}_j^{-1} \langle x, \hat{v}_j \rangle \langle X_k, \hat{v}_j \rangle X_{k+1}. \end{aligned}$$

Ο εκτιμητής  $\tilde{C}_1 \tilde{I}\tilde{C}_p$  μπορεί να χρησιμοποιηθεί αρχικά, αλλά τυπικά εισάγεται ένα επιπλέον βήμα εξομάλυνσης με την προσέγγιση  $X_{k+1} \approx \sum_{i=1}^p \langle X_{k+1}, \hat{v}_i \rangle \hat{v}_i$ . Αυτός ο τύπος οδηγεί στον εκτιμητή

$$\hat{\Psi}_p(x) = \frac{1}{N-1} \sum_{k=1}^{N-1} \sum_{j=1}^p \sum_{i=1}^p \hat{\lambda}_j^{-1} \langle x, \hat{v}_j \rangle \langle X_k, \hat{v}_j \rangle \langle X_{k+1}, \hat{v}_i \rangle \hat{v}_i. \quad (7.6.5)$$

Για να εξακριβώσουμε τη συνοχή του εκτιμητή, υποθέτουμε ότι  $p = pN$  είναι μία συνάρτηση δειγματοληπτικού μεγέθους  $N$ . Αν  $H = L^2$ , ο εκτιμητής (7.6.5) είναι ένας τελεστής πυρήνας με πυρήνα

$$\hat{\psi}_p(t,s) = \frac{1}{N-1} \sum_{k=1}^{N-p} \sum_{j=1}^p \sum_{i=1}^p \hat{\lambda}_j^{-1} (X_k, \hat{v}_j)(X_{k+1}, \hat{v}_i) \hat{v}_j(s) \hat{v}_i(t). \quad (7.6.6)$$

Αυτή η σχέση επαληθεύεται από την

$$\hat{\Psi}_p(x)(t) = \int \hat{\psi}_p(t,s) x(s) ds. \quad (7.6.7)$$

Πολυάριθμα αριθμητικά πειράματα έχουν διεξαχθεί για να προσδιορίσουν πόσο η εκτιμώμενη επιφάνεια  $\hat{\psi}_p(t,s)$  είναι κοντά στην  $\psi_p(t,s)$  η οποία χρησιμοποιείται για την προσομοίωση μίας FAR(1) διαδικασίας. Για παράδειγμα, για  $N \leq 100$  οι αποκλίσεις είναι πολύ μεγάλες τόσο σε μέγεθος όσο και σε σχήμα [18].



## BIBΛΙΟΓΡΑΦΙΑ

- [1] Adhikari, R., & Agrawal, R. (2013). An introductory study on time series modeling and forecasting. Retrieved from:  
<https://arxiv.org/ftp/arxiv/papers/1302/1302.6613.pdf>
- [2] Benoit menus. (2010). Retrieved from:  
[https://sunandclimate.files.wordpress.com/2010/01/benoit\\_menus.pdf](https://sunandclimate.files.wordpress.com/2010/01/benoit_menus.pdf)
- [3] Boutahar, M., & Khalfaoui, R. (2011). Estimation of the long memory parameter in non stationary models: a simulation study, HAL archives-ouvertes. Retrieved from:  
<https://halshs.archives-ouvertes.fr/halshs-00595057/document>
- [4] Brockwell, P. J., & Davis, R. A. (2002). Introduction to time series and forecasting (2nd ed). New York: Springer
- [5] Bühlmann, P., & Geer S. van de. (2011). Statistics for high dimensional data, methods, theory and applications. New York: Springer
- [6] Douglas, M. von. (2006). What can we learn from climate Data? Methods for fluctuation, time/scale and phase analysis. Institut für Physik. Retrieved from:  
[https://publishup.uni-potsdam.de/files/811/maraun\\_diss.pdf](https://publishup.uni-potsdam.de/files/811/maraun_diss.pdf)
- [7] Gneiting, T., & Schlather, M. (2003). Stochastic Models That Separate Fractal Dimension and Hurst Exponent. Retrieved from:  
<https://www.stat.washington.edu/research/reports/2003/tr425.pdf>
- [8] Guerrero, A., & Smith, A. L. (2005). A maximum likelihood estimator for long range persistence, Physica A 355, 619-632. Retrieved from:  
[http://www.lse.ac.uk/CATS/Publications/Publications%20PDFs/71\\_MaximumLikelihood\\_2005.pdf](http://www.lse.ac.uk/CATS/Publications/Publications%20PDFs/71_MaximumLikelihood_2005.pdf)
- [9] Karagiannis, T., Molle, M. & Faloutsos, M. (2006). Understanding the limitations of estimation methods for long-range dependence, retrieved from:  
<http://research.microsoft.com/pubs/71482/ucr-cs-2006-10245.pdf>
- [10] Lee, C. (2005). Long memory. Retrieved from:  
[http://econ.nsysu.edu.tw/ezfiles/124/1124/img/Chapter25\\_LongMemoryProcess.pdf](http://econ.nsysu.edu.tw/ezfiles/124/1124/img/Chapter25_LongMemoryProcess.pdf)
- [11] Linnell N. A. F. (1995). Analysis of repeated measures and time series: An Introduction with Forestry Examples. Biometrics Information Handbook No.6: Province of British Columbia
- [12] Mandelbrot, B., & J. W. Van Ness. (1968). Fractional brownian motions, fractional noises and applications. SIAM Review: 10, 422-437. Retrieved from:  
[http://users.math.yale.edu/~bbm3/web\\_pdfs/052fractionalBrownianMotions.pdf](http://users.math.yale.edu/~bbm3/web_pdfs/052fractionalBrownianMotions.pdf)

- [13] Millen, S. , & Beard, R. (2003). Estimation of the Hurst Exponent for the Burdekin River using the Hurst-Mandelbrot Rescaled Range Statistic. Retrieved from: <https://espace.library.uq.edu.au/view/UQ:10831/beard.pdf>
- [14] Mörters, P., & Peres, Y. (2010). Brownian motion. Cambridge University Press
- [15] Nikolopoulos, D. Petraki, E. Vogiannis, E. Chaldeos, Y. Yannakopoulos, P. Kottou, ....., Stonham, (2013). Traces of self-organisation and long-range memory in variations of environmental radon in soil: comparative results from monitoring in Lesvos Island and Ileia (Greece), Journal of radioanalytical and nuclear chemistry, Vol. 299, pp 203-219
- [16] Oppong, T. H. (2015). Fractal scaling of the hydraulic and hydrologic properties of an Acrisol. Applied Research Journal, Vol.1, Issue, 5, pp.320-326
- [17] Pascoal, R., & Monteiro, A. M. (2014). Market Efficiency, Roughness and Long Memory in PSI20 Index, Returns: Wavelet and Entropy Analysis, Entropy, 16, 2768-2788; doi:10.3390/e16052768
- [18] Rao, C.R. (2012). Handbook of statistics vol.30.Oxford: Elsevier
- [19] Reinert, G. (2010). Time series. Retrieved from: <http://www.stats.ox.ac.uk/~reinert/time/notesht10short.pdf>
- [20] Samorodnitsky, G. (2006). Long range dependence. doi:10.1561/09000000004
- [21] Sigman, K. (2006). IEOR 4700: Notes of Brownian motion, retrieved from: <http://www.columbia.edu/~ks20/FE-Notes/4700-07-Notes-BM.pdf>
- [22] [https://www.math.ucdavis.edu/~hunter/m280\\_09/ch5.pdf](https://www.math.ucdavis.edu/~hunter/m280_09/ch5.pdf)
- [23] <https://datajobs.com/data-science-repo/Time-Series-Analysis-Guide.pdf>
- [24] <http://www.statslab.cam.ac.uk/~rrw1/timeseries/t.pdf>
- [25] Torre, K., Delignières, D., & Lemoine, L. (2007). Detection of long-range dependence and estimation of fractal exponents through ARFIMA modeling, British Journal of Mathematical and Statistical Psychology, 60, 85-106. Retrieved from: <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.463.3311&rep=rep1&type=pdf>
- [26] Tsay,R. Seasonal Time Series, Trend Analysis & Component Model.(1990). Retrieve from: <http://faculty.chicagobooth.edu/ruey.tsay/teaching/uts/lec4-08.pdf>

[27] Qian, H. (2003). Fractional Brownian motion and fractional Gaussian noise. In processes with long-range correlations: Theory and applications, vol.621, 22-33, retrieved from: <http://faculty.washington.edu/hqian/reprints/fbm.pdf>

[28] Wei, W.S. (2006). Time series analysis: Univariate and multivariate methods (2<sup>nd</sup> ed.). Boston: Pearson

[29] Yang, B. (2008). LRD of fractional Brownian motion and application in data network. Retrieved from: <http://math.arizona.edu/~flaschka/Topmatter/527files/termpapers/bole.pdf>

[30] <http://www.trusoft-international.com/benoit.html#>

[31] Petraki, E., Nikolopoulos, D., Fotopoulos, A., Panagiotaras, D., Nomicos, C., Yiannakopoulos, P., ... Stonham. (2013), Long range memory patterns in variations of environmental radon in soil, The royal society of chemistry, doi:10.1039/c3ay26486f

