

## 4 Μοντέλα χρονοσειρών και πρόβλεψη

Η ανάλυση των χρηματο-οικονομικών δεικτών ως προς την κατανομή και τις συσχετίσεις μας οδηγεί στα παρακάτω συμπεράσματα:

1. Η περιθώρια κατανομή των μεταβολών (αποδόσεων) των δεικτών δεν είναι Γκαουσιανή αλλά έχει παχιές ουρές, που πρακτικά σημαίνει πως μεγάλες μεταβολές των δεικτών εμφανίζονται με μη-μηδενική πιθανότητα.
2. Δεν υπάρχουν σημαντικές συσχετίσεις στις μεταβολές για μικρές υστερήσεις αλλά ενδέχεται οι μεταβολές να έχουν συσχετίσεις μακράς διάρκειας. Το πρώτο πρακτικά σημαίνει πως η βραχυπρόθεσμη πρόβλεψη (με γραμμικά μοντέλα) δε μπορεί να γίνει, τουλάχιστον όχι με σημαντική ακρίβεια (πλήρως αναμενόμενο!). Το δεύτερο αφήνει ανοιχτό το ενδεχόμενο να υπάρχει απόκλιση από τον τυχαίο περίπατο, δηλαδή η αγορά να αποτελεί κάποιο δυναμικό σύστημα ή συσχετισμένη στοχαστική διαδικασία που προκαλεί τις μεταβολές του δείκτη.

Στο τελευταίο μέρος θα μελετήσουμε μοντέλα που έχουν χρησιμοποιηθεί στην ανάλυση και πρόβλεψη χρονοσειρών στα χρηματο-οικονομικά. Η μελέτη θα ξεκινήσει με γραμμικά αυτοπαλινδρομούμενα (AR) μοντέλα και μοντέλα τύπου ARIMA, υποθέτοντας πως η χρονοσειρά (με αναφορά στις μεταβολές του χρηματο-οικονομικού δείκτη) προέρχεται από μια γραμμική στοχαστική διαδικασία με συσχετίσεις βραχείας κλίμακας. Για τα μοντέλα AR και κατ' επέκταση ARIMA γνωρίζουμε καλά τις στατιστικές τους ιδιότητες και τους περιορισμούς τους. Αν και δεν προσφέρονται για τη μοντελοποίηση χρονοσειρών με μη-Γκαουσιανή κατανομή και συσχετίσεις μακράς κλίμακας χρησιμοποιούνται ως απλά μοντέλα αναφοράς. Στη συνέχεια θα μελετήσουμε μη-γραμμικά μοντέλα και ειδικότερα τοπικά μοντέλα πρόβλεψης που βασίζονται στη θεωρία των δυναμικών συστημάτων και χάους. Έρευνες τα τελευταία χρόνια έδειξαν πως αυτά τα μοντέλα μπορούν να πετύχουν καλύτερες προβλέψεις σε κάποιες περιπτώσεις χρηματο-οικονομικών δεικτών. Τέλος θα μελετήσουμε μοντέλα για τη διασπορά (αυξομείωση) της μεταβολής των δεικτών ως γραμμικές συναρτήσεις προηγούμενων τετραγώνων μεταβολών και διασπορών. Τα μοντέλα αυτά δεν προσφέρονται για προβλέψεις των χρηματοοικονομικών δεικτών αλλά για το μέγεθος της μεταβολής (απόδοσης) των δεικτών.

### 4.1 Γραμμικά μοντέλα πρόβλεψης χρονοσειρών

Θεωρούμε τη χρονοσειρά των μεταβολών του (χρηματο-οικονομικού) δείκτη  $\{x_t\}_{t=1}^N$ , είτε αυτή δίνεται από τις πρώτες διαφορές ή από τις αποδόσεις του δείκτη. Η χρονοσειρά  $\{x_t\}_{t=1}^N$  θεωρείται στάσιμη, καθώς τάσεις στη (μη-στάσιμη) χρονοσειρά του δείκτη  $\{y_t\}_{t=0}^N$  έχουν εξαλειφθεί με το μετασχηματισμό των πρώτων διαφορών στις τιμές του δείκτη ή στους λογαρίθμους των τιμών του δείκτη. Σε αυτήν την παράγραφο θα θεωρήσουμε ότι η χρονοσειρά  $\{x_t\}_{t=1}^N$  αποτελεί την πραγματοποίηση μιας στοχαστικής διαδικασίας, δηλαδή κάποιου συστήματος που οδηγείται από θόρυβο (τυχαία βήματα). Ειδικότερα θα θεωρήσουμε ότι η στοχαστική διαδικασία και άρα και η χρονοσειρά που εξετάζουμε είναι γραμμική, δηλαδή δεν περιέχει μη-γραμμικές συσχετίσεις. Θα προσπαθήσουμε να προσαρμόσουμε γραμμικά μοντέλα στη χρονοσειρά για να κάνουμε προβλέψεις. Θα μελετήσουμε επίσης τη διαδικασία

Box-Jenkins για την πρόβλεψη μη-στάσιμων γραμμικών χρονοσειρών, όπως η χρονοσειρά του χρηματο-οικονομικού δείκτη.

#### 4.1.1 Μοντέλα AR, ARMA και ARIMA

Τα μοντέλα παλινδρόμησης (regression models) ορίζουν μια μεταβλητή (εξαρτημένη) ως συνάρτηση κάποιων άλλων ανεξάρτητων μεταβλητών. Στα γραμμικά μοντέλα παλινδρόμησης η συνάρτηση αυτή είναι γραμμική δηλαδή η εξαρτημένη μεταβλητή δίνεται ως γραμμικός συνδυασμός των ανεξάρτητων μεταβλητών. Τα **αυτοπαλινδρομούμενα μοντέλα** (AutoRegressive models, AR) είναι μοντέλα γραμμικής παλινδρόμησης, όπου θεωρούμε ως εξαρτημένη μεταβλητή την τυχαία μεταβλητή της χρονοσειράς σε μια χρονική στιγμή  $t$ ,  $x_t$ , και ως ανεξάρτητες μεταβλητές θεωρούμε την τυχαία μεταβλητή της χρονοσειράς σε προηγούμενους χρόνους, δηλαδή τις  $x_{t-1}, \dots, x_{t-p}$ . Ο αριθμός των υστερήσεων που συμπεριλαμβάνουμε λέγεται η **τάξη** (order) του αυτοπαλινδρομούμενου μοντέλου. Ένα αυτοπαλινδρομούμενο μοντέλο τάξης  $p$  συμβολίζεται AR( $p$ ) και ορίζεται ως

$$x_t = \phi_0 + \phi_1 x_{t-1} + \dots + \phi_p x_{t-p} + z_t, \quad (26)$$

όπου  $\phi_0, \phi_1, \dots, \phi_p$  είναι οι συντελεστές του μοντέλου και  $\{z_t\} \sim \text{iid}$  με μέση τιμή 0 και διασπορά  $\sigma_z^2$ . Άρα το AR μοντέλο είναι γνωστό αν γνωρίζουμε τους συντελεστές και τη διασπορά του λευκού θορύβου. Στην πράξη οι συντελεστές του AR( $p$ ) μοντέλου, καθώς και η διασπορά του λευκού θορύβου (iid) εκτιμώνται από τη χρονοσειρά και οι εκτιμήσεις τους χρησιμοποιούνται για την πρόβλεψη της χρονοσειράς τις επόμενες χρονικές στιγμές.

Σύμφωνα με το μοντέλο AR( $p$ ) η μεταβλητή της χρονοσειράς τη χρονική στιγμή  $t$  κατά ένα μέρος εξηγείται από το γραμμικό συνδυασμό των  $p$  τελευταίων τιμών της χρονοσειράς  $x_{t-1}, x_{t-2}, \dots, x_{t-p}$ . Το υπόλοιπο μέρος που δεν εξηγείται από τις προηγούμενες τιμές της χρονοσειράς είναι καθαρά στοχαστικό και οφείλεται σε εξωγενείς επιδράσεις τη χρονική στιγμή  $t$ , που συνοψίζονται στην τυχαία μεταβλητή  $z_t$ . Στα χρηματο-οικονομικά  $z_t$  αναφέρεται ως το *σοκ* (shock) της χρονοσειράς. Σε κάποιες περιπτώσεις υποθέτουμε πως εξωγενείς παράγοντες σε προηγούμενους χρόνους μπορούν επίσης να επηρεάζουν τη μεταβλητή της χρονοσειράς τη χρονική στιγμή  $t$ . Συμπεριλαμβάνοντας και αυτό το μέρος που λέγεται μέρος **κινούμενου μέσου** (moving average), το γενικό γραμμικό μοντέλο για την πρόβλεψη στάσιμης χρονοσειράς είναι το **αυτοπαλινδρομούμενο μοντέλο κινούμενου μέσου** (AutoRegressive Moving Average, ARMA) που δίνεται ως

$$x_t = \phi_0 + \phi_1 x_{t-1} + \dots + \phi_p x_{t-p} + z_t - \theta_1 z_{t-1} - \dots - \theta_q z_{t-q}. \quad (27)$$

Το αυτοπαλινδρομούμενο μέρος (AR) είναι τάξης  $p$  και το μέρος του κινούμενου μέσου (MA) είναι τάξης  $q$  και το μοντέλο συμβολίζεται ARMA( $p, q$ ).

Όταν η χρονοσειρά των μεταβολών του δείκτη  $\{x_t\}_{t=1}^N$  προσδιορίζεται από ένα μοντέλο ARMA( $p, q$ ) τότε η χρονοσειρά του δείκτη  $\{y_t\}_{t=0}^N$  προσδιορίζεται από το **ολοκληρωμένο αυτοπαλινδρομούμενο μοντέλο κινούμενου μέσου** ή **ολοκληρωμένο μικτό μοντέλο** (autoregressive integrated moving average model, ARIMA). Το μοντέλο συμβολίζεται ARMA( $p, 1, q$ ), όπου ο δείκτης 1 δηλώνει ότι οι πρώτες διαφορές της χρονοσειράς  $\{y_t\}_{t=0}^N$ , δηλαδή η χρονοσειρά  $\{x_t\}_{t=1}^N$ , προσδιορίζεται από το μοντέλο ARMA( $p, q$ ). Κατά αυτήν την έννοια τα ARIMA μοντέλα ορίζονται και για επανάληψη του μετασχηματισμού των πρώτων διαφορών  $d$

φορές μέχρι να πετύχουμε η παραγόμενη χρονοσειρά να είναι στάσιμη και το μοντέλο συμβολίζεται ARIMA( $p,d,q$ ).

Το αυτοπαλινδρομούμενο μοντέλο AR( $p$ ) είναι σε άμεση σχέση με την αυτοσυσχέτιση. Μικρή αυτοσυσχέτιση βραχείας κλίμακας σημαίνει αδυναμία πρόβλεψης με μοντέλο AR και άρα περιμένουμε τα AR μοντέλα να μην έχουν προβλεπτική ικανότητα σε χρονοσειρές μεταβολών χρηματο-οικονομικών δεικτών που δεν έχουν σημαντικές συσχετίσεις.

Ας πάρουμε ως παράδειγμα το μοντέλο AR(1). Αν η αυτοσυσχέτιση για υστέρηση 1 είναι στατιστικά μη-μηδενική, αυτό σημαίνει ότι τα στοιχεία  $x_t$  και  $x_{t-1}$  είναι γραμμικά συσχετισμένα και άρα μπορούμε να θεωρήσουμε πως υπάρχει κάποια εξάρτηση του  $x_t$  από το  $x_{t-1}$ . Το πιο απλό μοντέλο που μπορούμε να σχηματίσουμε τότε είναι το **αυτοπαλινδρομούμενο μοντέλο τάξης 1**, AR(1) (autoregressive model of order 1)

$$x_t = \phi_0 + \phi_1 x_{t-1} + z_t. \quad (28)$$

Η συνθήκη για το συντελεστή  $|\phi_1| < 1$  εξασφαλίζει τη στασιμότητα της χρονοσειράς που περιγράφεται από το μοντέλο AR(1). Αν η χρονοσειρά έχει μέση τιμή 0 (δηλαδή αν υποθέσουμε ότι η στοχαστική διαδικασία που την παράγει έχει μέση τιμή 0,  $\mu = 0$ ) τότε  $\phi_0 = 0$ . Αλλιώς μπορούμε να αφαιρέσουμε από κάθε στοιχείο της χρονοσειράς τη μέση τιμή,  $x'_t = x_t - \mu$ , και το AR(1) γίνεται

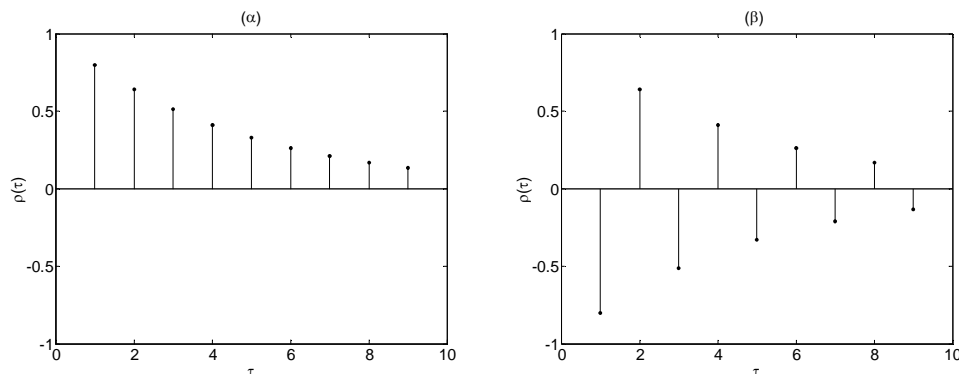
$$x'_t = \phi x'_{t-1} + z_t, \quad (29)$$

από το οποίο προκύπτει ότι  $\phi_1 = \phi$  και  $\phi_0 = \mu - \phi\mu$ . Για  $\phi = 1$  η AR(1) χρονοσειρά είναι τυχαίος περίπατος και φυσικά μη-στάσιμη.

Η αυτοσυσχέτιση χρονοσειράς που περιγράφεται από στάσιμο AR(1) μοντέλο φθίνει εκθετικά στο 0 και μπορεί να δειχθεί πως ισχύει

$$\rho_\tau = \phi^\tau \quad \text{και} \quad \gamma_0 = \sigma_x^2 = \frac{\sigma_z^2}{1 - \phi^2} \quad (30)$$

Στο Σχήμα 31 δίνεται η συνάρτηση αυτοσυσχέτισης για θετικό και αρνητικό συντελεστή  $\phi$ . Παρατηρούμε ότι για θετικό  $\phi$  το  $\rho_\tau$  φθίνει προς το 0 μονότονα ενώ για αρνητικό  $\phi$  εναλλασσόμενα γύρω από το 0.



**Σχήμα 31** Αυτοσυσχέτιση AR(1) μοντέλου με συντελεστή  $\phi = 0.8$  στο (α) και  $\phi = -0.8$  στο (β).

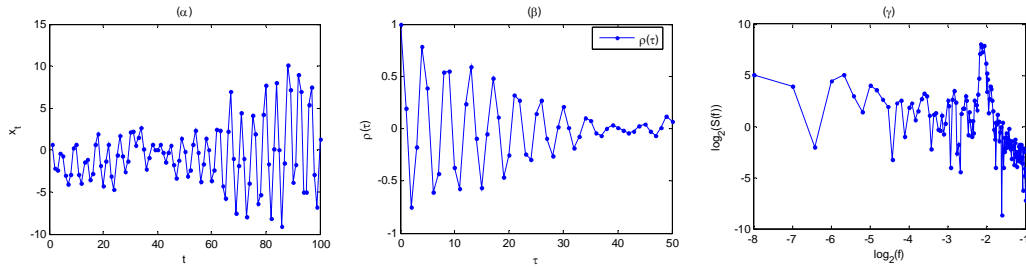
Αυξάνοντας την τάξη του μοντέλου AR μπορούμε να πετύχουμε ποικίλες μορφές της συνάρτησης αυτοσυσχέτισης αλλά πάντα αυτή φθίνει εκθετικά προς το 0.

### Παράδειγμα: Μοντέλο AR(4) και αυτοσυσχέτιση

Θεωρούμε τη χρονοσειρά που παράγεται από το AR(4) μοντέλο

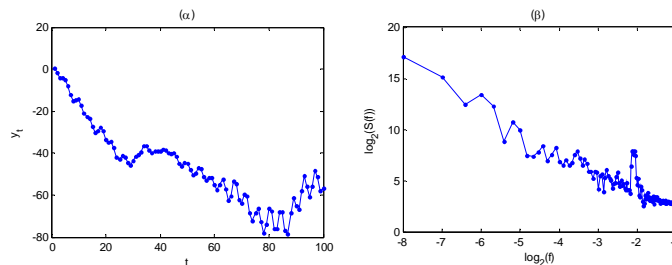
$$x_t = 1.4x_{t-1} - 1.53x_{t-2} + 1.228x_{t-3} - 0.3104x_{t-4} + z_t \quad (31)$$

με  $\sigma_z^2 = 1$ . Η χρονοσειρά παρουσιάζει ταλαντώσεις, όπως δίνεται στο Σχήμα 32α (αυτή η συμπεριφορά οφείλεται σε μιγαδικές ρίζες του χαρακτηριστικού πολυωνύμου που έχει συντελεστές τους συντελεστές του μοντέλου, αλλά δε θα προχωρήσουμε σε αυτό το θέμα). Η αυτοσυσχέτιση παρουσιάζει επίσης ταλαντώσεις με περίοδο 4 ή 5 (Σχήμα 32β). Αυτό φαίνεται και από την κορυφή του φάσματος ισχύος στην αντίστοιχη συχνότητα (για  $\log_2(f) \approx -2$  και άρα  $f \approx 1/4$ ).



Σχήμα 32 Χρονοσειρά από μοντέλο AR(4) στο (α), η αυτοσυσχέτιση της στο (β) και το φάσμα ισχύος της στο (γ).

Μεταβαίνοντας στη χρονοσειρά  $\{y_t\}_{t=1}^N$  του προφίλ (ολοκλήρωσης) της  $\{x_t\}_{t=1}^N$  ( $y_t = \sum_{i=1}^t x_i$  ή  $y_t = y_{t-1} + x_t$ ), αυτή είναι τυχαίος περίπατος αλλά με συσχετισμένα τυχαία βήματα. Όπως παρατηρούμε στο Σχήμα 33 το προφίλ έχει αργές τάσεις αλλά παρουσιάζει επίσης ταλαντώσεις και αυτό το χαρακτηριστικό φαίνεται στο φάσμα ισχύος που έχει φθίνουσα κλίση τύπου  $1/f^2$  αλλά με κορυφή σε συχνότητες γύρω στο  $1/4$ . Σημειώνεται πως το αντίστοιχο μοντέλο για τη  $\{y_t\}_{t=1}^N$  είναι το ARIMA(4,1,0), που σημαίνει απλά πως η χρονοσειρά των πρώτων διαφορών της, δηλαδή η  $\{x_t\}_{t=1}^N$ , προσδιορίζεται από το μοντέλο AR(4) στην (31).



Σχήμα 33 Χρονοσειρά του προφίλ της χρονοσειράς από μοντέλο AR(4) στο (α) και το φάσμα ισχύος της στο (β).

#### 4.1.2 Πρόβλεψη χρονοσειρών

Ένα από τα σημαντικότερα ερωτήματα στα χρηματο-οικονομικά είναι κατά πόσο μπορούν να προβλεφθούν οι χρηματο-οικονομικοί δείκτες. Εδώ δε θα προσπαθήσουμε να δώσουμε απάντηση σε αυτό το ερώτημα αλλά θα παρουσιάσουμε τη διαδικασία της πρόβλεψης γενικά και ειδικότερα με τα γραμμικά μοντέλα για στάσιμες και μη-στάσιμες χρονοσειρές. Το πρόβλημα της πρόβλεψης στις χρονοσειρές τίθεται ως η εκτίμηση της τιμής του δείκτη ενδιαφέροντος για  $T$  χρονικά

βήματα μπροστά από μια δεδομένη χρονική στιγμή  $n$ , όταν γνωρίζουμε την τιμή του δείκτη για όλες τις χρονικές στιγμές μέχρι το χρόνο  $n$ .

Θεωρούμε τη χρονοσειρά  $n$  παρατηρήσεων κάποιου δείκτη  $\{y_0, y_1, \dots, y_n\}$  και έστω ότι θέλουμε να προβλέψουμε την τιμή του δείκτη  $T$  χρονικές στιγμές μπροστά,  $y_{n+T}$ . Μαθηματικά το πρόβλημα της πρόβλεψης  $\hat{y}_{n+T} = y_n(T)$  δίνεται από την εύρεση της δεσμευμένης μέσης τιμής

$$y_n(T) = E[y_{n+T} | y_n, y_{n-1}, \dots], \quad (32)$$

όπου στο συμβολισμό της πρόβλεψης  $T$  βημάτων μπροστά  $y_n(T)$  ( $T$ -step ahead forecast), ο δείκτης  $n$  δηλώνει την *αφετηρία της πρόβλεψης* (forecast origin) και ο δείκτης  $T$  στην παρένθεση δηλώνει τον *ορίζοντα της πρόβλεψης* (forecast horizon). Φυσικά η πρόβλεψη γίνεται πιο πολύπλοκη (και αβέβαιη!) με το άνοιγμα του ορίζοντα πρόβλεψης  $T$ .

#### 4.1.2.1 Πρόβλεψη χρονοσειράς με τάση και περιοδικότητα

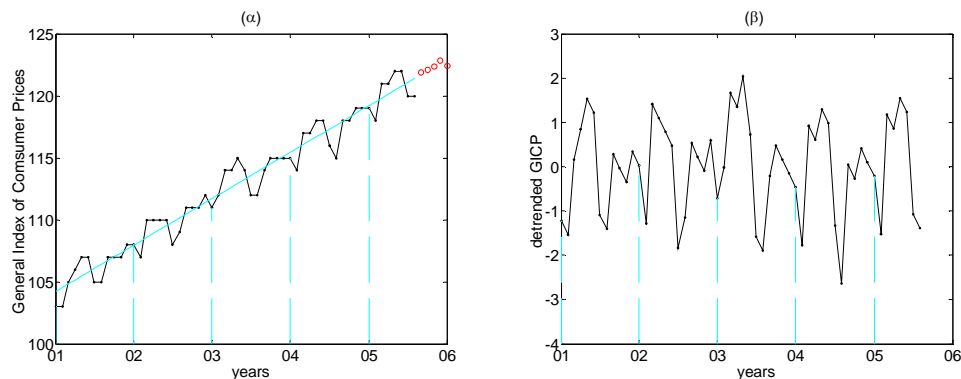
Στη γενική περίπτωση μια χρονοσειρά μπορεί να περιέχει τάσεις, όπως πληθωρισμός ή άνοδος και πτώση χρηματιστηρίου, αλλά και περιοδικότητα ή εποχικότητα σταθερής περιόδου  $d$ , όπως ομαλοί επιχειρησιακοί κύκλοι (regular business cycles). Μια χρονοσειρά μπορεί να αναλυθεί σε τρεις συνιστώσες ως

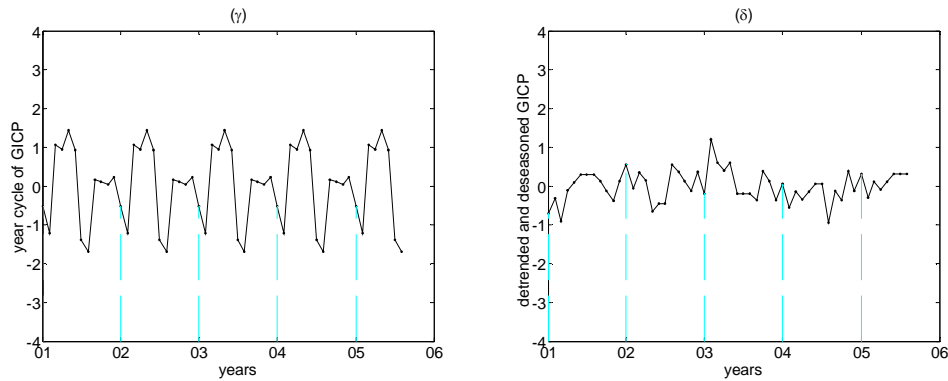
$$y_t = \mu_t + s_t + x_t, \quad (33)$$

όπου  $\mu_t$  είναι η τάση ως συνάρτηση του χρόνου,  $s_t$  είναι η περιοδική ή εποχική συνάρτηση περιόδου  $d$  (οι τιμές της  $s_t$  επαναλαμβάνονται κάθε  $d$  χρονικές στιγμές) και  $x_t$  είναι η χρονοσειρά των υπολοίπων που μπορεί να είναι iid ή κάποια στάσιμη χρονοσειρά με συσχετίσεις. Η απαλοιφή της τάσης και περιοδικότητας μετατρέπει τη μη-στάσιμη χρονοσειρά  $\{y_t\}_{t=1}^N$  στη στάσιμη χρονοσειρά των υπολοίπων  $\{x_t\}_{t=1}^N$  την οποία μπορούμε να μελετήσουμε στη συνέχεια και να εκτιμήσουμε μοντέλα πρόβλεψης.

##### Παράδειγμα: Απαλοιφή τάσης και εποχικότητας στη χρονοσειρά GICP

Τη διαδικασία απαλοιφής τάσης και περιοδικότητας θα τη δούμε μέσα από ένα πραγματικό παράδειγμα. Στο Σχήμα 34α, δίνεται ο γενικός δείκτης τιμών καταναλωτή (general index for consumer price, GICP) σε μηνιαίες τιμές από τον Ιανουάριο 2001 ως τον Αύγουστο 2005 (οι ανοιχτοί κύκλοι στο σχήμα είναι προβλέψεις που θα δούμε παρακάτω). Η χρονοσειρά έχει μήκος  $n = 56$ .





**Σχήμα 34 (α) Μηνιαίες τιμές γενικού δείκτη τιμών καταναλωτή (GICP) την περίοδο Ιανουάριος 2001 – Αύγουστος 2005 (η κάθετη διακεκομμένη γραμμή δηλώνει την αρχή του έτους). Στο διάγραμμα φαίνεται η προσαρμογή γραμμικού μοντέλου τάσης. (β) Η χρονοσειρά που προκύπτει από την αφαίρεση της γραμμικής τάσης στη χρονοσειρά GICP. (γ) Ο εκτιμημένος ετήσιος κύκλος για την GICP. (δ) Η χρονοσειρά που προκύπτει από την αφαίρεση του εκτιμημένου ετήσιου κύκλου από τη χρονοσειρά στο (β).**

Φαίνεται καθαρά, πως ο GICP παρουσιάζει σταθερή πληθωριστική τάση για όλη την περίοδο παρατήρησης αλλά και ασθενή ετήσια περιοδικότητα. Η πληθωριστική τάση μπορεί να εξηγηθεί ικανοποιητικά από γραμμική συνάρτηση  $\mu_t$  του χρόνου  $t$  (μήνα). Η προσαρμογή απλού γραμμικού μοντέλου παλινδρόμησης του  $\mu_t$  ως προς το  $t$ , θεωρώντας τις παρατηρήσεις  $\{y_t\}_{t=1}^{56}$  ως τιμές του  $\mu_t$ , έδωσε  $\mu_t = 103.9 + 0.31t$  και φαίνεται με γκριζα γραμμή στο Σχήμα 34α. Αφαιρώντας αυτή τη γραμμική τάση από τη χρονοσειρά GICP βρίσκουμε τη χρονοσειρά *απαλλαγμένη από τάση* (detrended)  $y'_t = y_t - \mu_t$  που δίνεται στο Σχήμα 34β. Αυτή είναι η κατάλληλη χρονοσειρά για ανάλυση αν θέλουμε να μελετήσουμε την μεταβολή του GICP απαλλαγμένη από τον πληθωρισμό. Εδώ φαίνεται καλύτερα η ετήσια περιοδικότητα του GICP αλλά δεν είναι προφανές ποια είναι η αναλυτική μορφή της περιοδικής συνάρτησης του χρόνου του ετήσιου κύκλου. Μπορούμε όμως να εκτιμήσουμε τον *ετήσιο κύκλο* (annual cycle) του GICP από τις μέσες τιμές του κάθε μήνα για τα έτη 2001 – 2005 (ως το 2004 για τους μήνες Σεπτέμβριο – Δεκέμβριο). Υπολογίζουμε για τον Ιανουάριο το μέσο όρο από τις τιμές Ιανουαρίου για τα 5 χρόνια και την τιμή αυτή την αφαιρούμε από τις 5 παρατηρήσεις του Ιανουαρίου (για τα έτη 2001 – 2005). Το ίδιο κάνουμε και για τους άλλους μήνες. Η περιοδική χρονοσειρά  $\{s_t\}_{t=1}^n$  (του επαναλαμβανόμενου ετήσιου κύκλου) δίνεται στο Σχήμα 34γ όπου η περίοδος είναι  $d = 12$ .

Αν μας ενδιαφέρει να εξετάσουμε αν μπορούμε να αντλήσουμε κάποια επιπλέον πληροφορία από τη χρονοσειρά GICP πέρα από τη γραμμική τάση, που απαλείψαμε, και τον ετήσιο κύκλο που εκτιμήσαμε, τότε από τη χρονοσειρά χωρίς τάση  $\{y'_t\}_{t=1}^n$  αφαιρούμε την περιοδική χρονοσειρά  $\{s_t\}_{t=1}^n$ . Η χρονοσειρά των υπολοίπων  $\{x_t\}_{t=1}^n$  δίνεται ως (δες επίσης τη σχέση (33))

$$x_t = y'_t - s_t = y_t - \mu_t - s_t,$$

και παρουσιάζεται γραφικά στο Σχήμα 34δ. Παρατηρούμε ότι αυτή η χρονοσειρά είναι *απαλλαγμένη από τάση* και *περιοδικότητα* και δε φαίνεται να έχει κάποια κανονικότητα ή δομή.

Αν η χρονοσειρά  $\{x_t\}_{t=1}^n$  είναι εντελώς τυχαία, δηλαδή δεν έχει να μας δώσει καμιά πληροφορία, τότε η ανάλυση σταματάει εδώ και περιοριζόμαστε να περιγράψουμε τη χρονική μεταβολή του GICP ως μια πληθωριστική γραμμική τάση σε συνδυασμό με έναν ετήσιο κύκλο. Έτσι αν θέλουμε να προβλέψουμε το GICP για το μήνα Σεπτέμβριο θα *επεκτείνουμε* (extrapolate) την συνάρτηση τάσης για το μήνα Σεπτέμβριο και θα προσθέσουμε σε αυτήν την τιμή του ετήσιου κύκλου για το μήνα Σεπτέμβριο. Αναλυτικά, έχοντας τις τιμές GICP από τον Ιανουάριο 2001 ως τον Αύγουστο 2005 ( $n = 56$ ), η τάση επεκτείνεται για το Σεπτέμβριο 2005 ως

$$\mu_{n+1} = 103.9 + 0.31 \cdot (n+1) \Rightarrow \mu_{57} = 103.9 + 0.31 \cdot 57 = 121.70.$$

Η τιμή του ετήσιου κύκλου για το Σεπτέμβριο βρέθηκε ως η μέση τιμή Σεπτεμβρίου για τα έτη 2001 – 2004,  $s_y = 0.16$ , και προστίθεται στην τιμή του  $\mu_{57}$  για να δώσει την πρόβλεψη του GICP για το μήνα Σεπτέμβριο  $\hat{y}_{57} = y_{56}(1) = 121.86$ . Όμοια μπορούμε να κάνουμε προβλέψεις και για τους επόμενους μήνες. Στο Σχήμα 34α δίνονται με ανοιχτούς κύκλους οι προβλέψεις για τους 5 επόμενους μήνες.

#### **4.1.2.2 Προσαρμογή μοντέλου AR και πρόβλεψη**

Πριν προχωρήσουμε στην πρόβλεψη χρονοσειρών με μοντέλα AR θα μελετήσουμε κάποια βασικά στοιχεία για την προσαρμογή κατάλληλου AR μοντέλου σε μια στάσιμη χρονοσειρά  $\{x_t\}_{t=1}^n$ , όπως η χρονοσειρά των υπολοίπων στο προηγούμενο παράδειγμα του δείκτη GICP. Η διαδικασία προσαρμογής του μοντέλου AR περιλαμβάνει την επιλογή της τάξης  $p$  του μοντέλου, την εκτίμηση των παραμέτρων του μοντέλου  $\phi_1, \dots, \phi_p, \sigma_z^2$  και τέλος το διαγνωστικό έλεγχο καταλληλότητας του μοντέλου.

Για το πρώτο βήμα, την επιλογή της τάξης  $p$  του μοντέλου AR, αναφέρουμε δύο κριτήρια. Το πρώτο βασίζεται στη **μερική αυτοσυσχέτιση** (partial autocorrelation). Όπως η συνάρτηση αυτοσυσχέτισης έτσι και η συνάρτηση μερικής αυτοσυσχέτισης ορίζεται για κάθε υστέρηση  $\tau$ , που εδώ τη συμβολίζουμε  $p$  γιατί αναφέρεται στην τάξη του μοντέλου AR. Αν θεωρήσουμε τις μεταβλητές της χρονοσειράς σε παράθυρο μήκους  $p+1$ , δηλαδή  $x_t, x_{t-1}, \dots, x_{t-p+1}, x_{t-p}$ , η μερική αυτοσυσχέτιση μεταξύ των  $x_t$  και  $x_{t-p}$  (για υστέρηση  $p$ ) μετράει τη συσχέτιση τους ουδετεροποιώντας την επίδραση των ενδιάμεσων μεταβλητών, δηλαδή συσχέτιση μεταξύ των  $x_t$  και  $x_{t-p}$  που προκύπτει από τη συσχέτιση της κάθε μιας από τις  $x_t$  και  $x_{t-p}$  με τις ενδιάμεσες μεταβλητές  $x_{t-1}, \dots, x_{t-p+1}$ . Ο υπολογισμός της μερικής αυτοσυσχέτισης γίνεται από την εκτίμηση των παραμέτρων μοντέλων AR με αυξανόμενη τάξη. Συγκεκριμένα για τάξη 1, το μοντέλο είναι

$$x_t = \phi_{1,1} x_{t-1} + z_t,$$

όπου ο δεύτερος δείκτης στο συντελεστή αναφέρεται στην τάξη του μοντέλου. Ο εκτιμώμενος συντελεστής  $\hat{\phi}_{1,1}$  αποτελεί τη μερική αυτοσυσχέτιση υστέρησης 1 (και είναι ίδια με την αυτοσυσχέτιση, δεν υπάρχουν ενδιάμεσες μεταβλητές). Για τάξη 2, το μοντέλο είναι

$$x_t = \phi_{1,2} x_{t-1} + \phi_{2,2} x_{t-2} + z_t$$

και η μερική αυτοσυσχέτιση υστέρησης 2 είναι ο εκτιμώμενος συντελεστής του  $x_{t-2}$ ,  $\hat{\phi}_{2,2}$ . Το  $\hat{\phi}_{2,2}$  δηλώνει την προσφορά του  $x_{t-2}$  που προστίθεται στην προσφορά του μοντέλου AR(1) για τον προσδιορισμό του  $x_t$ . Κατά τον ίδιο τρόπο για τάξη 3, το μοντέλο είναι

$$x_t = \phi_{1,3}x_{t-1} + \phi_{2,3}x_{t-2} + \phi_{3,3}x_{t-3} + z_t$$

και η μερική αυτοσυσχέτιση υστέρησης 3 είναι ο εκτιμώμενος συντελεστής του  $x_{t-3}$ ,  $\hat{\phi}_{3,3}$ . Η εκτίμηση των συντελεστών συνεχίζεται ως κάποια μεγάλη τάξη.

Αν η χρονοσειρά μπορεί πράγματι να προέρχεται από μια στοχαστική διαδικασία τύπου AR( $p$ ), ο συντελεστής  $\hat{\phi}_{p,p}$  δεν είναι μηδενικός, ενώ για υστερήσεις (ή τάξεις μοντέλου)  $k > p$ , οι συντελεστές  $\hat{\phi}_{k,k}$  θα είναι (στατιστικά) μηδενικοί. Άρα χρησιμοποιώντας το κριτήριο της μερικής αυτοσυσχέτισης μπορούμε να προσδιορίσουμε την τάξη του AR μοντέλου από την υστέρηση στην οποία παρατηρείται η δραστική πτώση (cut off) από μη-μηδενική σε μηδενική μερική αυτοσυσχέτιση.

Η τάξη του AR μοντέλου μπορεί να προσδιορισθεί και από διάφορα κριτήρια πληροφoρίας που βασίζονται στην πιθανοφάνεια (likelihood) των δεδομένων με βάση αυτό το μοντέλο. Ως δείκτης πιθανοφάνειας μπορεί να θεωρηθεί η διασπορά των υπολοίπων  $s_z^2$  από την προσαρμογή του AR( $p$ ) μοντέλου. Το πιο γνωστό κριτήριο είναι το **κριτήριο πληροφoρίας του Akaike** (Akaike information criterion, AIC) που ορίζεται ως

$$\text{AIC}(p) = \ln(s_z^2) + \frac{2p}{n}, \quad (34)$$

όπου  $n$  είναι το μήκος της χρονοσειράς. Είναι φανερό ότι όσο μεγαλώνει η τάξη  $p$  του AR μοντέλου τα **σφάλματα** ή **υπόλοιπα** (residual) της προσαρμογής γίνονται μικρότερα, το  $s_z^2$  μικραίνει, και για πολύ μεγάλες τάξεις το AR μοντέλο προσαρμόζεται σε διακυμάνσεις που δεν αντανακλούν πραγματικές συσχετίσεις αλλά το λευκό θόρυβο. Για αυτό στην (34) υπάρχει ο δεύτερος όρος, που λέγεται και **συνάρτηση ποινής** (penalty function), ο οποίος δρα αρνητικά και αυξάνει την τιμή της συνάρτησης AIC όταν η τάξη του μοντέλου αυξάνει. Υπολογίζοντας το κριτήριο AIC για ικανά μεγάλο αριθμό τάξεων μοντέλου επιλέγουμε εκείνη την τάξη  $p$  που δίνει την ελάχιστη τιμή του AIC.

Αφού έχει γίνει η επιλογή της τάξης  $p$  του μοντέλου AR, εκτιμούμε τις παραμέτρους του AR( $p$ ) (το έχουμε ήδη κάνει αν υπολογίσαμε τη μερική αυτοσυσχέτιση). Οι δύο πιο γνωστοί μέθοδοι εκτίμησης των παραμέτρων είναι η **μέθοδος ελαχίστων τετραγώνων** (ordinary least squares, OLS)

$$\min_{\phi_1, \dots, \phi_p} \sum_{t=p+1}^n (x_t - \phi_1 x_{t-1} - \dots - \phi_p x_{t-p})^2, \quad (35)$$

και η **μέθοδος μέγιστης πιθανοφάνειας** (maximum likelihood method).

Το τελευταίο βήμα της προσαρμογής του μοντέλου στη χρονοσειρά είναι ο διαγνωστικός έλεγχος για την επάρκεια του μοντέλου. Αυτό γίνεται ελέγχοντας αν τα σφάλματα προσαρμογής, δηλαδή οι διαφορές πραγματικών και εκτιμωμένων τιμών που αποτελούνε εκτίμηση του  $z_t$ , είναι iid και κυρίως ανεξάρτητα. Αν δεν είναι αυτό σημαίνει ότι υπάρχουν συσχετίσεις στη χρονοσειρά που το μοντέλο δεν έχει

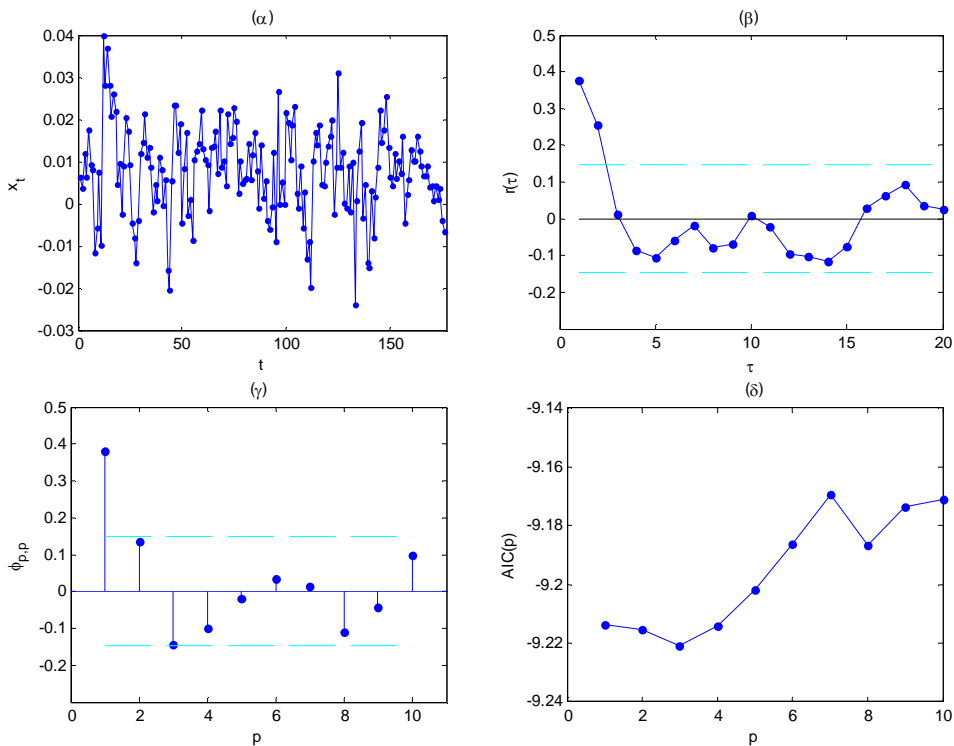


συμπεριλάβει και άρα το μοντέλο είναι ανεπαρκές. Υπάρχουν διάφοροι έλεγχοι ανεξαρτησίας για να εφαρμόσουμε στη χρονοσειρά των σφαλμάτων προσαρμογής αλλά δε θα ασχοληθούμε με αυτό το θέμα εδώ.

Παράδειγμα: Προσαρμογή AR μοντέλου στη χρονοσειρά του ΑΕΠ των ΗΠΑ

Ας δούμε την εφαρμογή της διαδικασίας προσαρμογής σε ένα πραγματικό παράδειγμα, το ρυθμό μεταβολής του ακαθάριστου εθνικού προϊόντος (ΑΕΠ) των ΗΠΑ που δίνεται σε τετραμηνιαίες τιμές από το 2<sup>ο</sup> τετράμηνο 1947 ως το 1<sup>ο</sup> τετράμηνο 1991. Η χρονοσειρά δίνεται στο Σχήμα 35α, όπου η εποχικότητα έχει διορθωθεί (αφαιρώντας τον εποχικό κύκλο). Η αυτοσυσχέτιση της χρονοσειράς που δίνεται στο Σχήμα 35β υποδηλώνει σημαντικές αυτοσυσχετίσεις για τις δύο πρώτες υστερήσεις (τα όρια που φαίνονται στο σχήμα υπολογίζονται ως  $\pm 2/\sqrt{n}$  κάτω από την υπόθεση ότι η χρονοσειρά είναι iid).

Εκτιμήσαμε τα AR μοντέλα για τάξεις  $p = 1, \dots, 10$  και στη συνέχεια υπολογίσαμε την μερική αυτοσυσχέτιση και το AIC κριτήριο για κάθε τάξη  $p$ . Από τη μερική αυτοσυσχέτιση στο Σχήμα 35γ δε μπορούμε να εκτιμήσουμε με ασφάλεια την τάξη. Φαίνεται πως μόνο το  $\hat{\phi}_{1,1}$  είναι στατιστικά μη-μηδενικό, ενώ τα  $\hat{\phi}_{2,2}$  και  $\hat{\phi}_{3,3}$  βρίσκονται στα όρια στατιστικής σημαντικότητας. Το κριτήριο AIC στο Σχήμα 35δ έχει ελάχιστο για  $p = 3$ . Άρα η επιλογή της τάξης 3 υποστηρίζεται από το κριτήριο AIC αλλά όχι από τη μερική αυτοσυσχέτιση. Αυτό φυσικά δεν αποτελεί ένδειξη ότι η μια μέθοδος προσδιορισμού τάξης είναι καλύτερη από την άλλη, αλλά απλά δείχνει ότι διαφορετικές μέθοδοι μπορεί να δώσουν διαφορετικά αποτελέσματα σε πραγματικές εφαρμογές.



Σχήμα 35 (α) χρονοσειρά του ρυθμού μεταβολής του ΑΕΠ των ΗΠΑ. (β) Η αυτοσυσχέτιση της χρονοσειράς. (γ) Η μερική αυτοσυσχέτιση της χρονοσειράς. (δ) Το κριτήριο AIC για την ίδια χρονοσειρά. Στο (β) και (γ) οι οριζόντιες διακεκομμένες γραμμές δηλώνουν τα όρια σημαντικότητας σε στάθμη  $\alpha = 0.05$ .

Προσαρμόζουμε στη χρονοσειρά μοντέλο AR(3) και το εκτιμημένο μοντέλο είναι

$$x_t = 0.0047 + 0.35x_{t-1} + 0.18x_{t-2} - 0.14x_{t-3} + z_t$$

με εκτιμώμενη τυπική απόκλιση του λευκού θορύβου  $s_z = \hat{\sigma}_z = 0.0098$ .

[Για την εκτίμηση, αφαιρέσαμε πρώτα τη μέση τιμή των 176 παρατηρήσεων από κάθε παρατήρηση (βρέθηκε  $\hat{\mu} = 0.0077$ ) και σε αυτήν τη χρονοσειρά εκτιμήσαμε με τη μέθοδο των ελαχίστων τετραγώνων τους συντελεστές  $\hat{\phi}_1 = 0.35$ ,  $\hat{\phi}_2 = 0.18$  και  $\hat{\phi}_3 = -0.14$ , και ο σταθερός όρος υπολογίστηκε ως  $\hat{\phi}_0 = \hat{\mu}(1 - \hat{\phi}_1 - \hat{\phi}_2 - \hat{\phi}_3) = 0.0047$ . Από το μοντέλο αυτό υπολογίστηκε για κάθε χρονική στιγμή  $t = 3, \dots, 176$  ο εκτιμώμενος ρυθμός μεταβολής

$$\hat{x}_t = 0.0047 + 0.35x_{t-1} + 0.18x_{t-2} - 0.14x_{t-3}$$

και στη συνέχεια τα σφάλματα εκτίμησης  $\hat{z}_t = x_t - \hat{x}_t$ . Η διασπορά αυτών των σφαλμάτων είναι η εκτίμηση της διασποράς του λευκού θορύβου στο μοντέλο AR(3),  $s_z^2 = \hat{\sigma}_z^2 = 0.0000989$ .]

Έστω τώρα ότι έχουμε εκτιμήσει ένα μοντέλο AR( $p$ ) σε μια στάσιμη χρονοσειρά  $\{x_t\}_{t=1}^n$ , όπως η χρονοσειρά των υπολοίπων στο προηγούμενο παράδειγμα του ΑΕΠ των ΗΠΑ. Για ένα AR( $p$ ) μοντέλο θεωρούμε ότι η πραγματική τιμή της χρονοσειράς για χρόνο  $n+1$  δίνεται ως

$$x_{n+1} = \phi_0 + \phi_1 x_n + \dots + \phi_p x_{n+1-p} + z_{n+1},$$

(εδώ θεωρούμε ότι η χρονοσειρά έχει μη-μηδενική μέση τιμή). Η καλύτερη πρόβλεψη του  $x_{n+1}$  γνωρίζοντας τις τιμές  $x_n, x_{n-1}, \dots$ , και με βάση το μοντέλο AR( $p$ ) είναι σύμφωνα με τον κανόνα της δεσμευμένης μέσης τιμής στην (32)

$$x_n(1) = E[x_{n+1} | x_n, x_{n-1}, \dots, x_{n+1-p}] = \phi_0 + \phi_1 x_n + \dots + \phi_p x_{n+1-p}, \quad (36)$$

δηλαδή η πρόβλεψη δίνεται από το γραμμικό συνδυασμό των  $p$  τελευταίων (γνωστών) παρατηρήσεων, που είναι το αιτιοκρατικό μέρος του AR( $p$ ). Το *σφάλμα πρόβλεψης* (forecast error) είναι

$$e_n(1) = x_{n+1} - x_n(1) = z_{n+1} \quad (37)$$

και η διασπορά του σφάλματος πρόβλεψης ενός βήματος είναι  $\text{Var}[e_n(1)] = \sigma_z^2$ . Αν θεωρήσουμε ότι ο λευκός θόρυβος στο μοντέλο ακολουθεί κανονική κατανομή, τότε μπορούμε να ορίσουμε 95% διάστημα πρόβλεψης για το ένα βήμα μπροστά ως

$x_n(1) \pm 1.96\sigma_z$ . Στα οικονομικά το σφάλμα πρόβλεψης  $z_{n+1}$  αναφέρεται ως το *σοκ* (shock) της χρονοσειράς στη χρονική στιγμή  $n+1$ .

Στην πράξη οι συντελεστές του AR( $p$ ) μοντέλου, καθώς και η διασπορά του λευκού θορύβου εκτιμώνται από τη χρονοσειρά και οι εκτιμήσεις τους χρησιμοποιούνται για τη σημειακή πρόβλεψη και το διάστημα πρόβλεψης.

Έχοντας υπολογίσει την πρόβλεψη  $x_n(1)$  για το  $x_{n+1}$ , μπορούμε να την αντικαταστήσουμε στη σχέση που μας δίνει την πρόβλεψη για το  $x_{n+2}$

$$x_n(2) = \phi_0 + \phi_1 x_n(1) + \phi_2 x_n \dots + \phi_p x_{n+2-p}.$$

Το σφάλμα πρόβλεψης για δύο βήματα μπροστά είναι

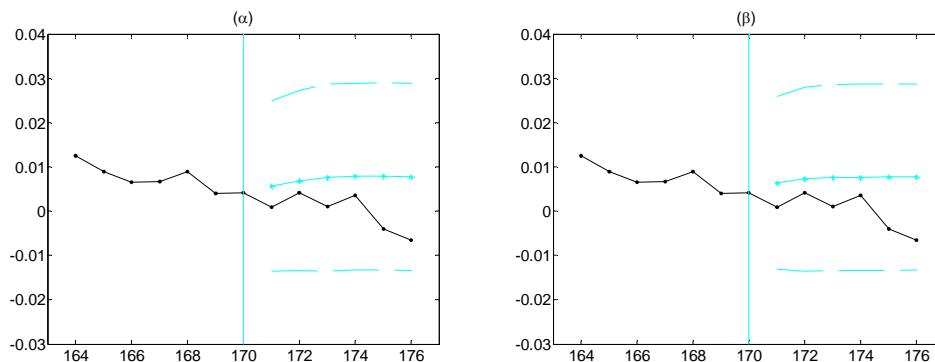
$$e_n(2) = x_{n+2} - x_n(2) = \phi_1 (x_{n+1} - x_n(1)) + z_{n+2} = \phi_1 z_{n+1} + z_{n+2}$$

και η διασπορά του σφάλματος πρόβλεψης δύο βημάτων είναι  $\text{Var}[e_n(2)] = (1 + \phi_1^2) \sigma_\varepsilon^2$ . Το διάστημα πρόβλεψης υπολογίζεται όπως και για ένα βήμα μπροστά.

Αυτή η επαναληπτική διαδικασία συνεχίζεται για να δώσει την πρόβλεψη για 3 βήματα μπροστά και γενικά για ορίζοντα πρόβλεψης  $T$ , όπου κάθε φορά αντικαθιστούμε τις παρατηρήσεις σε χρόνους μεγαλύτερους της αφετηρίας  $n$  (και μικρότερους φυσικά του  $n+T$ ) με τις προβλέψεις που ήδη έγιναν σε προηγούμενα βήματα. Μπορεί να δειχθεί ότι για στάσιμο  $\text{AR}(p)$  μοντέλο καθώς ο ορίζοντας πρόβλεψης αυξάνει ( $T \rightarrow \infty$ ), η πρόβλεψη  $x_n(T)$  τείνει στη μέση τιμή της χρονοσειράς.

Παράδειγμα: Πρόβλεψη με AR μοντέλο της χρονοσειράς του ΑΕΠ των ΗΠΑ

Εφαρμόζουμε τη διαδικασία πρόβλεψης στη χρονοσειρά του ρυθμού μεταβολής του ΑΕΠ των ΗΠΑ. Θεωρούμε ως αφετηρία πρόβλεψης το πρώτο τετράμηνο του 1989 (που αντιστοιχεί στη χρονική στιγμή  $n=170$ ) και ορίζοντα πρόβλεψης τα επόμενα δύο έτη ( $T=6$ ) (ως το πρώτο τετράμηνο του 1991, που είναι και η τελευταία παρατήρηση της χρονοσειράς που μελετήσαμε παραπάνω). Χρησιμοποιώντας το μοντέλο  $\text{AR}(3)$  που εκτιμήσαμε παραπάνω υπολογίζουμε τις σημειακές εκτιμήσεις  $x_{170}(1), \dots, x_{170}(6)$  και τα αντίστοιχα διαστήματα πρόβλεψης. Οι προβλέψεις δίνονται μαζί με τις πραγματικές τιμές του ρυθμού μεταβολής ΑΕΠ στο Σχήμα 36α. Φαίνεται πως οι προβλέψεις συγκλίνουν γρήγορα στη μέση τιμή της χρονοσειράς ( $\hat{\mu} = 0.0077$ ). Τα διαστήματα πρόβλεψης είναι σχετικά μεγάλα και το εύρος τους συγκλίνει μετά τα πρώτα χρονικά βήματα. Αυτά τα στοιχεία δηλώνουν την περιορισμένη δυνατότητα πρόβλεψης για αυτήν τη χρονοσειρά. Όπως φαίνεται στο Σχήμα 36β, οι προβλέψεις είναι ουσιαστικά ίδιες αν χρησιμοποιήσουμε το απλό  $\text{AR}(1)$  μοντέλο, που συνιστάται από μια πιο συντηρητική επιλογή τάξης μοντέλου βασισμένη στη μερική αυτοσυσχέτιση.



Σχήμα 36 Η χρονοσειρά του τετραμηνιαίου ρυθμού μεταβολής ΑΕΠ των ΗΠΑ, όπου η τελευταία τιμή (για χρονική στιγμή 176) αντιστοιχεί στο πρώτο τετράμηνο του 1991. Στο ίδιο διάγραμμα δίνονται οι προβλέψεις με αφετηρία το πρώτο τετράμηνο του 1989 (δίνεται με την κατακόρυφη γραμμή) και ορίζοντα τα 6 τελευταία τετράμηνα. Με διακεκομμένες γκριζες γραμμές δίνονται τα διαστήματα πρόβλεψης. Στο (α) οι προβλέψεις είναι από το  $\text{AR}(3)$  μοντέλο και στο (β) από το  $\text{AR}(1)$ .

**4.1.2.3 Αξιολόγηση της επίδοσης μοντέλου σε πρόβλεψη**

Η αξιολόγηση της επίδοσης ενός μοντέλου πρόβλεψης δε μπορεί να βασίζεται σε μια μεμονωμένη πρόβλεψη από κάποια αφετηρία και για κάποιο ορίζοντα πρόβλεψης  $T$ , όπως στο παραπάνω παράδειγμα, αλλά θα πρέπει να γίνει με κάποιο κριτήριο που

βασίζεται σε στατιστικά αποτελέσματα. Για την αξιολόγηση της πρόβλεψης βήματος  $T = 1, 2, \dots$ , υπολογίζουμε κάποιο μέτρο που συγκεντρώνει τα σφάλματα πρόβλεψης για έναν ικανοποιητικό αριθμό σημείων αφετηρίας.

Υπάρχουν διάφοροι σύνθετοι τρόποι αξιολόγησης, όπως η τεχνική *αντεπικύρωσης* (cross validation). Εδώ απλά θα χωρίσουμε τη χρονοσειρά μήκους  $N$ , σε δύο σύνολα, στο *σύνολο εκμάθησης ή εκπαίδευσης* (training ή learning set)  $x_1, x_2, \dots, x_{N_1}$ , στο οποίο εκτιμούμε τις παραμέτρους του μοντέλου, και στο *σύνολο ελέγχου ή επικύρωσης* (test ή validation set)  $x_{N_1+1}, x_{N_1+2}, \dots, x_N$ , στο οποίο κάνουμε τις προβλέψεις  $x_{N_1}(T), x_{N_1+1}(T), \dots, x_{N-T}(T)$  και υπολογίζουμε τα σφάλματα πρόβλεψης

$$\begin{aligned} e_{N_1}(T) &= x_{N_1+T} - x_{N_1}(T) \\ &\vdots \\ e_{N-T}(T) &= x_N - x_{N-T}(T) \end{aligned}$$

Υπάρχουν διάφορα στατιστικά μέτρα που συγκεντρώνουν τα σφάλματα πρόβλεψης, όπως το **μέσο τετραγωνικό σφάλμα** (mean square error, **mse**)

$$\text{mse}(T) = \frac{1}{N - N_1 - T + 1} \sum_{j=N_1}^{N-T} e_j(T)^2 = \frac{1}{N - N_1 - T + 1} \sum_{j=N_1}^{N-T} (x_{j+T} - x_j(T))^2.$$

Συχνά χρησιμοποιείται η **ρίζα** του **μέσου τετραγωνικού σφάλματος** (root mean square error, **rmse**)

$$\text{rmse}(T) = \sqrt{\frac{1}{N - N_1 - T + 1} \sum_{j=N_1}^{N-T} e_j(T)^2} = \sqrt{\frac{1}{N - N_1 - T + 1} \sum_{j=N_1}^{N-T} (x_{j+T} - x_j(T))^2}.$$

Ένα χρήσιμο μέτρο σφάλματος πρόβλεψης όταν θέλουμε να συγκρίνουμε μοντέλα πρόβλεψης σε διαφορετικές χρονοσειρές είναι η **κανονικοποίηση του rmse** (normalized root mean square error, **normse**) διαιρώντας το rmse με τη δειγματική τυπική απόκλιση των παρατηρήσεων της χρονοσειράς (ή πιο σωστά των παρατηρήσεων που χρησιμοποιούνται στο σχηματισμό των σφαλμάτων)

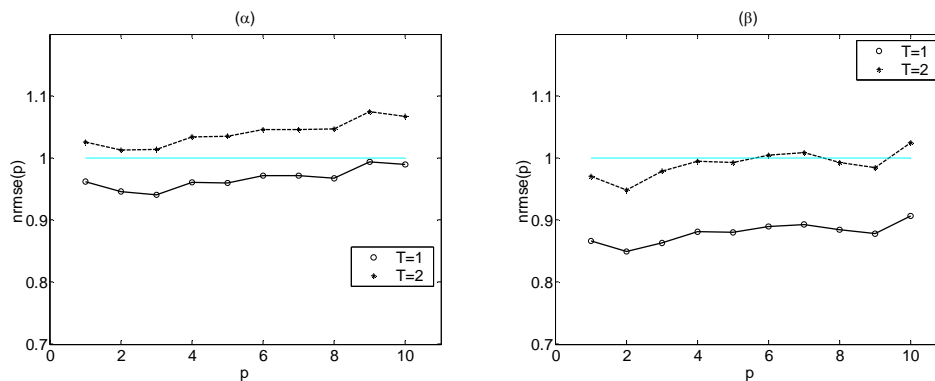
$$\text{normse}(T) = \frac{\sqrt{\frac{1}{N - N_1 - T + 1} \sum_{j=N_1}^{N-T} (x_{j+T} - x_j(T))^2}}{\sqrt{\frac{1}{N - N_1 - T + 1} \sum_{j=N_1}^{N-T} (x_{j+T} - \bar{x})^2}}, \quad (38)$$

όπου  $\bar{x}$  είναι η δειγματική μέση τιμή των  $x_{N_1+T}, x_{N_1+1+T}, \dots, x_N$ . Τιμές του normse κοντά στο 0 δηλώνουν πολύ καλή πρόβλεψη ενώ τιμές του normse κοντά στο 1 δηλώνουν ότι η πρόβλεψη είναι τόσο καλή όσο αν προβλέπαμε με τη μέση τιμή.

#### Παράδειγμα: Αξιολόγησης AR πρόβλεψης στη χρονοσειρά του ΑΕΠ των ΗΠΑ

Συγκρίνουμε τη δυνατότητα πρόβλεψης των μοντέλων AR για τάξεις  $p = 1, \dots, 10$  στη χρονοσειρά του ρυθμού μεταβολής του ΑΕΠ των ΗΠΑ, από το πρώτο τετράμηνο του 1947 ως το δεύτερο τετράμηνο του 1991 ( $n = 176$ ). Εξετάζουμε την πρόβλεψη για ένα και δύο χρονικά βήματα μπροστά, δηλαδή για  $T = 1$  και  $T = 2$ , με μοντέλα AR τάξης  $p = 1, \dots, 10$ . Ως σύνολο ελέγχου παίρνουμε τα τελευταία 50 τετράμηνα (περίπου 17 έτη), δηλαδή οι παράμετροι του κάθε μοντέλου εκτιμώνται στη χρονοσειρά εκμάθησης των πρώτων 126 παρατηρήσεων ( $N_1 = 126$ ) και οι προβλέψεις και τα σφάλματα πρόβλεψης υπολογίζονται στις υπόλοιπες 50

παρατηρήσεις ( $N - N_1 = 50$ ). Στο Σχήμα 37α δίνεται το  $\text{nmse}$  ως συνάρτηση της τάξης  $p$  του AR μοντέλου. Παρατηρούμε πως η πρόβλεψη για ένα βήμα μπροστά είναι γενικά λίγο καλύτερη από τη μέση τιμή, ενώ για  $T=2$  λίγο χειρότερη. Το μικρότερο σφάλμα πρόβλεψης επιτυγχάνεται για μοντέλο AR(3) και η μείωση είναι μικρή. Οι διαφορές δε φαίνεται να είναι σημαντικές και αυτό φαίνεται αν αλλάξουμε το σύνολο ελέγχου. Στο Σχήμα 37β δίνονται τα ίδια αποτελέσματα για σύνολο ελέγχου τις τελευταίες 30 παρατηρήσεις. Παρατηρούμε ότι η πρόβλεψη γενικά βελτιώνεται και η καλύτερη πρόβλεψη για  $T=1$  επιτυγχάνεται τώρα με το μοντέλο AR(2). Γενικά με μικρές χρονοσειρές, η αξιολόγηση της πρόβλεψης είναι δύσκολη καθώς τα στατιστικά μέτρα, όπως το  $\text{nmse}$ , δεν είναι σταθερά σε αλλαγές στο δείγμα αξιολόγησης. Επίσης τα κριτήρια προσδιορισμού τάξης μοντέλου δε δίνουν πάντα τα πιο κατάλληλα μοντέλα για πρόβλεψη.



Σχήμα 37 (α) Σφάλμα πρόβλεψης με AR μοντέλα τάξης  $p=1, \dots, 10$  για τα 50 τελευταία τετράμηνα του τετραμηνιαίου ρυθμού μεταβολής ΑΕΠ των ΗΠΑ την περίοδο 1947 – 1991. Το σφάλμα πρόβλεψης μετρείται με το  $\text{nmse}$  για ένα και δύο χρονικά βήματα μπροστά, όπως δίνεται στο ένθετο. (β) Το ίδιο όπως στο (α) για τα τελευταία 30 τετράμηνα.

#### 4.1.2.4 Πρόβλεψη μη-στάσιμων χρονοσειρών με μοντέλα ARIMA

Τα μοντέλα τύπου ARMA (που συμπεριλαμβάνουν τα AR και MA) είναι κατάλληλα για στάσιμες χρονοσειρές. Όταν η χρονοσειρά δεν είναι στάσιμη μπορούμε να την κάνουμε στάσιμη και να εφαρμόσουμε τα μοντέλα τύπου ARMA.

Αν η μη-στασιμότητα οφείλεται σε περιοδικότητα ή τάση που μπορεί να προσδιοριστεί ως συνάρτηση του χρόνου, στην παράγραφο 4.1.2.1 περιγράψαμε τη διαδικασία για να κάνουμε στάσιμη τη χρονοσειρά. Γενικότερα απαλείφουμε την τάση μιας χρονοσειράς παίρνοντας τις πρώτες διαφορές, ή πιο συχνά στα χρηματοοικονομικά, τις αποδόσεις. Οι πρώτες διαφορές σε χρονοσειρά τυχαίου περιπάτου δίνουν λευκό θόρυβο. Στη γενική περίπτωση όμως οι πρώτες διαφορές δίνουν μια στάσιμη χρονοσειρά  $\{x_t\}_{t=1}^n$  με συσχετίσεις που μπορούν να περιγραφτούν από μοντέλα τύπου ARMA( $p, q$ ). Αν η  $\{x_t\}_{t=1}^n$  που προκύπτει από τις πρώτες διαφορές δεν είναι στάσιμη, επαναλαμβάνουμε το μετασχηματισμό των πρώτων διαφορών στη  $\{x_t\}_{t=1}^n$  μέχρι να προκύψει στάσιμη χρονοσειρά.

Ο συνδυασμός μετασχηματισμού πρώτων διαφορών σε μια μη-στάσιμη χρονοσειρά  $\{y_t\}_{t=0}^n$  και μοντέλου ARMA στην  $\{x_t\}_{t=1}^n$  είναι το μοντέλο ARIMA. Επιτρέποντας  $d$  επαναλήψεις των πρώτων διαφορών για να φτάσουμε σε στάσιμη χρονοσειρά και να προσαρμόσουμε κάποιο ARMA( $p, q$ ) μοντέλο, τα μοντέλα ARIMA συμβολίζονται ως ARIMA( $p, d, q$ ). Τα μοντέλα αυτά λέγονται και **Box-Jenkins**

**μοντέλα** γιατί οι Box και Jenkins πρότειναν πρώτοι μια ολοκληρωμένη διαδικασία εκτίμησης γραμμικών μοντέλων τύπου ARMA σε μη-στάσιμες χρονοσειρές. Η διαδικασία αυτή έχει περιγραφεί παραπάνω στην παρουσίαση των AR μοντέλων και περιλαμβάνει την επιλογή τάξης μοντέλου, την εκτίμηση των παραμέτρων του μοντέλου και τη διάγνωση της καταλληλότητας του μοντέλου. Παρακάτω δίνεται η ολοκληρωμένη διαδικασία στα πλαίσια της γραμμικής ανάλυσης χρονοσειρών.

1. Κατασκευάζουμε το γράφημα της χρονοσειράς και της αυτοσυσχέτισης.
2. Αν η αυτοσυσχέτιση για κάθε υστέρηση είναι στατιστικά ασήμαντη, η προσαρμογή γραμμικού μοντέλου δεν έχει νόημα. Ενδεχομένως ένα μη γραμμικό μοντέλο να προσαρμόζεται στα δεδομένα. (Για αυτό προτείνεται να γίνει έλεγχος ανεξαρτησίας με μη γραμμικές στατιστικές). Αν η αυτοσυσχέτιση είναι ισχυρή και φθίνει αργά με την υστέρηση τότε η χρονοσειρά δεν είναι στάσιμη και παίρνουμε τις πρώτες διαφορές. (Αν η αυτοσυσχέτιση συνεχίζει να μη φθίνει παίρνουμε πάλι πρώτες διαφορές.)
3. Στη χρονοσειρά των πρώτων διαφορών προσαρμόζεται μοντέλο AR ή ARMA τάξης που έχει επιλεχτεί κατάλληλα χρησιμοποιώντας τη μερική αυτοσυσχέτιση ή κάποιο κριτήριο πληροφορίας.
4. Προσαρμόζουμε το μοντέλο στη χρονοσειρά και ελέγχουμε την καταλληλότητα του. Κυρίως ελέγχουμε αν η χρονοσειρά των υπολοίπων είναι iid.
5. Έχοντας καταλήξει στο πιο κατάλληλο μοντέλο, το χρησιμοποιούμε για να κάνουμε πρόβλεψη και μετασχηματίζουμε τις προβλέψεις πίσω στην αρχική χρονοσειρά.

Το τελευταίο σημείο χρειάζεται επεξήγηση. Οι γενικοί τύποι για τις προβλέψεις με ARIMA( $p, d, q$ ) είναι πολύπλοκοι, αλλά κάποιος μπορεί να καταλάβει πως γίνεται η πρόβλεψη με ARIMA μοντέλο θεωρώντας το ARIMA( $p, 1, q$ ). Από την αρχική χρονοσειρά  $\{y_0, y_1, \dots, y_n\}$  προκύπτει η χρονοσειρά των πρώτων διαφορών  $\{x_1, x_2, \dots, x_n\}$ , όπου  $x_t = y_t - y_{t-1}$ . Εφαρμόζοντας το μοντέλο ARMA( $p, q$ ) στην  $\{x_1, x_2, \dots, x_n\}$  βρίσκουμε την πρόβλεψη για ένα χρονικό βήμα  $x_n(1)$  και η πρόβλεψη για την αρχική χρονική σειρά είναι

$$y_n(1) = y_n + x_n(1). \quad (39)$$

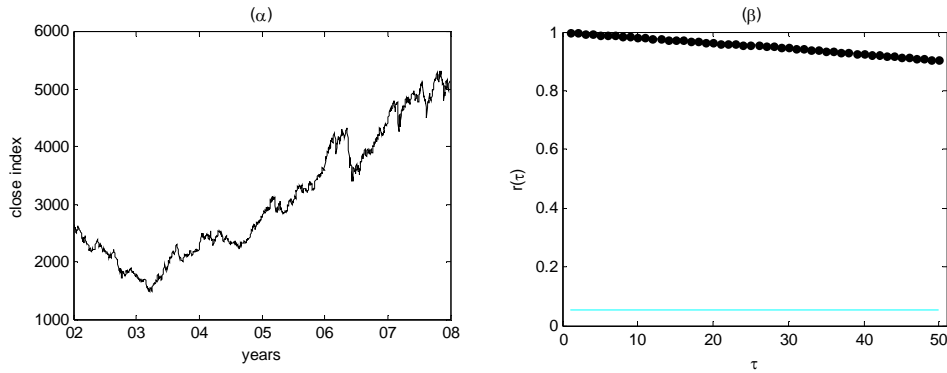
Το σφάλμα πρόβλεψης  $e_n(1)$  του  $y_{n+1}$  είναι το ίδιο με το σφάλμα πρόβλεψης του  $x_{n+1}$ . Γενικά η πρόβλεψη για  $T$  χρονικά βήματα μπροστά είναι

$$y_n(T) = y_n(T-1) + x_n(T), \quad (40)$$

όπου  $x_n(T)$  είναι η πρόβλεψη του  $x_{n+T}$  με το μοντέλο ARMA( $p, q$ ) και το  $y_n(T-1)$  είναι γνωστό από την πρόβλεψη του  $y_{n+T-1}$  στο προηγούμενο βήμα.

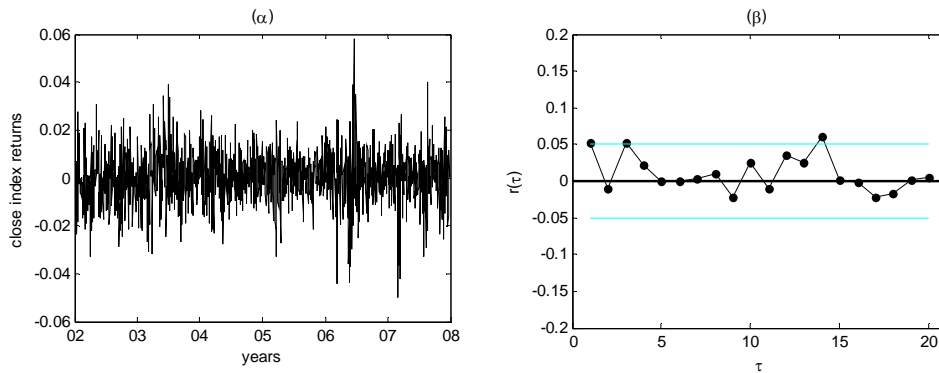
#### Παράδειγμα: Box Jenkins διαδικασία στη χρονοσειρά του XAA

Ας δούμε τη διαδικασία Box-Jenkins στην πράξη στη χρονοσειρά του ημερήσιου γενικού δείκτη χρηματιστηρίου αξιών Αθηνών (XAA). Επιλέγουμε μια αρκετά μεγάλη χρονοσειρά με μήκος  $n=1496$  που καλύπτει την περίοδο 1/1/2002 – 31/12/2007. Η χρονοσειρά  $\{y_t\}_{t=1}^n$  σε αυτήν την περίοδο φαίνεται να έχει ομοσκεδαστικότητα, δηλαδή η διακύμανση των τιμών της δε διαφέρει σημαντικά σε διαφορετικές περιόδους. Φυσικά περιέχει τάσεις όπως εύκολα φαίνεται από το διάγραμμα της χρονοσειράς και της αυτοσυσχέτισης, η οποία φθίνει πολύ αργά από τη μονάδα προς το μηδέν (δες Σχήμα 38).



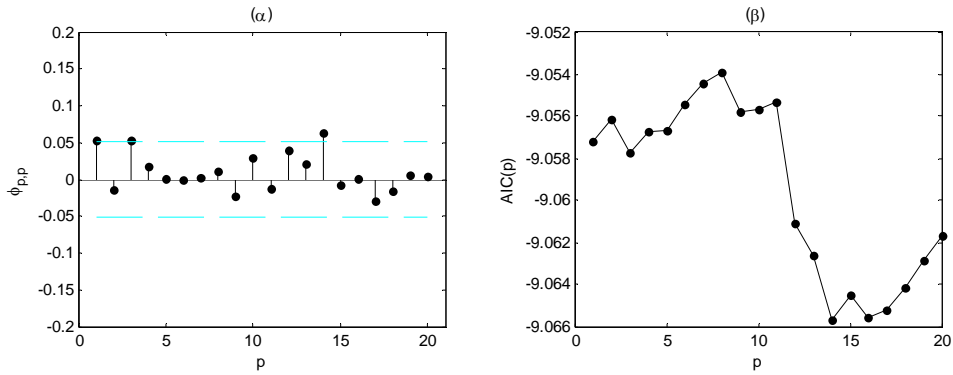
**Σχήμα 38 (α) Η χρονοσειρά του γενικού δείκτη ΧΑΑ την περίοδο 1/1/2002 – 31/13/2007. (β) Η αυτοσυσχέτιση της χρονοσειράς στο (α).**

Ως πρώτο βήμα της ανάλυσης παίρνουμε τις πρώτες διαφορές του γενικού δείκτη, ή αντίστοιχα τις αποδόσεις ή τις λογαριθμικές αποδόσεις του γενικού δείκτη. Η χρονοσειρά των αποδόσεων  $\{x_t\}_{t=1}^n$  φαίνεται να είναι στάσιμη (δες Σχήμα 39α). Το διάγραμμα της αυτοσυσχέτισης στο Σχήμα 39β δε δείχνει να υπάρχουν στατιστικά σημαντικές αυτοσυσχετίσεις, παρά μόνο οριακά για υστερήσεις 1, 3 και 14. Το αποτέλεσμα αυτό συνιστούν πολύ ασθενείς συσχετίσεις στα δεδομένα, που μπορεί απλά να οφείλονται σε κάποια μικρή ετεροσκεδαστικότητα ή και ασθενείς τάσεις (π.χ. όπως στην περίοδο του 2003 στο Σχήμα 39α). Μπορεί όμως και να οφείλονται στη δυναμική του συστήματος του ελληνικού χρηματιστηρίου, που φυσικά θα θέλαμε να εξηγήσουμε. Στα πλαίσια της γραμμικής ανάλυσης η περιγραφή της δυναμικής γίνεται με μοντέλα τύπου ARMA που περιγράφουν την εξάρτηση της απόδοσης του γενικού δείκτη από τις αποδόσεις σε προηγούμενες μέρες.



**Σχήμα 39 (α) Η χρονοσειρά των αποδόσεων του γενικού δείκτη ΧΑΑ την περίοδο 1/1/2002 – 31/12/2007. (β) Η αυτοσυσχέτιση της χρονοσειράς στο (α). Τα όρια της στατιστικής σημαντικότητας δίνονται με τις οριζόντιες γκριζες γραμμές.**

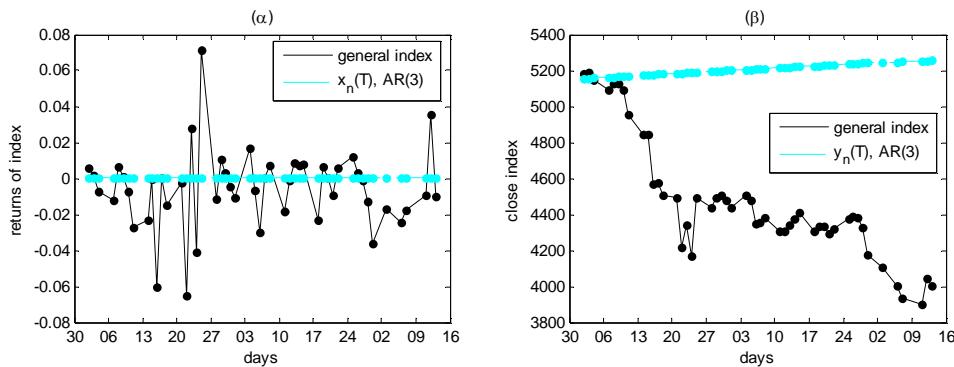
Επειδή ενδιαφερόμαστε να κάνουμε προβλέψεις θέλουμε να προσαρμόσουμε κάποιο AR μοντέλο, παρά ARMA, στη χρονοσειρά των αποδόσεων  $\{x_1, \dots, x_{1495}\}$ . Για τον προσδιορισμό της τάξης του AR μοντέλου υπολογίζουμε τη μερική αυτοσυσχέτιση και τη συνάρτηση AIC για τάξεις  $p = 1, \dots, 20$  (δες Σχήμα 40).



**Σχήμα 40 (α)** Η μερική αυτοσυσχέτιση για τη χρονοσειρά των αποδόσεων του γενικού δείκτη XAA την περίοδο 1/1/2002 – 31/13/2007. Τα όρια της στατιστικής σημαντικότητας δίνονται με τις οριζόντιες γκριζες γραμμές (β) Το AIC κριτήριο για την ίδια χρονοσειρά και AR μοντέλα τάξης  $p = 1, \dots, 20$ .

Η μερική αυτοσυσχέτιση έχει την ίδια μορφή με την αυτοσυσχέτιση και δε μας παρέχει ασφαλή εκτίμηση της κατάλληλης τάξης. Η συνάρτηση AIC φαίνεται να έχει τοπικά ελάχιστα στο 3 και στο 9 και το σφαιρικό ελάχιστο είναι για  $p = 14$ . Αυτή η εκτίμηση της τάξης είναι ύποπτα υψηλή για δεδομένα που οριακά φαίνονται να έχουν συσχετίσεις. Τα κριτήρια της μερικής αυτοσυσχέτισης και AIC δεν φαίνεται λοιπόν να προσδιορίζουν με κάποια ασφάλεια την πιο κατάλληλη τάξη του AR μοντέλου.

Δοκιμάσαμε όλα τα μοντέλα AR,  $p = 1, \dots, 20$ , για να κάνουμε αυθεντικές προβλέψεις σε τιμές αποδόσεων του γενικού δείκτη πέρα από τη χρονική περίοδο που καλύπτει η χρονοσειρά, και συγκεκριμένα για τις επόμενες μέρες και μέχρι τις 13/3/2008. Με αφετηρία την 2/1/2008 προβλέψαμε τις τιμές απόδοσης του ημερήσιου γενικού δείκτη για αυξανόμενο ορίζοντα πρόβλεψης  $T$  από 2/1/2008 ως 13/3/2008 (δηλαδή τα βήματα πρόβλεψης είναι  $T = 1, \dots, 49$  και οι προβλέψεις είναι  $y_{293}(1), \dots, y_{293}(16)$ ). Οι προβλέψεις με όλα τα AR μοντέλα συνέκλιναν γρήγορα με το βήμα  $T$  στη μέση απόδοση του γενικού δείκτη (περίπου 0). Στο Σχήμα 41α φαίνεται αυτό για το AR(3). Αντίστοιχα η πρόβλεψη του γενικού δείκτη τείνει προς τη μέση τιμή του δείκτη και δεν καταφέρνει να προσεγγίσει την πορεία τιμών του δείκτη XAA (δες Σχήμα 41β).



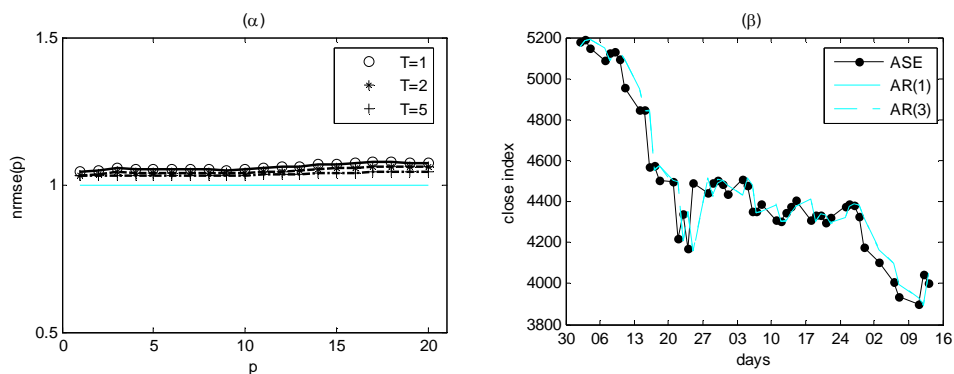
**Σχήμα 41 (α)** Πρόβλεψη με AR(3) των αποδόσεων του γενικού δείκτη με αφετηρία την 31/12/2007 και για 1 μέχρι 49 μέρες μπροστά. Ο τύπος γραμμών για τις πραγματικές και προβλεπόμενες τιμές δίνονται στο ένθετο. Στον οριζόντιο άξονα δίνονται οι μέρες του μήνα. **(β)** Το ίδιο όπως στο (α) αλλά για τις τιμές του γενικού δείκτη XAA.



Υπολογίσαμε σε αυτήν την περίοδο και για μικρά βήματα ( $T = 1, \dots, 5$ ) το  $\text{nmse}$ , ως στατιστικό μέτρο για το σφάλμα πρόβλεψης, και βρέθηκε να είναι λίγο μεγαλύτερο του 1 για όλα τα βήματα  $T$ , δηλαδή η πρόβλεψη είναι λίγο χειρότερη από την πρόβλεψη με τη μέση απόδοση, όπως φαίνεται στο Σχήμα 42α. Οι προβλέψεις των αποδόσεων με AR μοντέλα είναι κοντά στη μέση απόδοση, που είναι σχεδόν 0. Η πρόβλεψη με τη μέση απόδοση αντιστοιχεί στη διατήρηση της τελευταίας τιμής του γενικού δείκτη για την πρόβλεψη μελλοντικών τιμών του γενικού δείκτη. Αυτό φαίνεται από τη σχέση απόδοσης και τιμής

$$x_t = \frac{y_t - y_{t-1}}{y_{t-1}} \quad \text{και} \quad y_t = y_{t-1}(1 + x_t),$$

αν θεωρήσουμε ότι  $x_t \cong 0$ . Πράγματι παρατηρούμε Σχήμα 42β ότι οι προβλέψεις του γενικού δείκτη για την επόμενη μέρα ( $T=1$ ) με μοντέλο των αποδόσεων AR(1) και AR(3) ακολουθούν τις τιμές του γενικού δείκτη με υστέρηση 1, δηλαδή  $y_n(1) = y_n$ .



Σχήμα 42 (α) Σφάλμα  $\text{nmse}$  για την πρόβλεψη ενός, δύο και πέντε βημάτων μπροστά του ημερήσιου γενικού δείκτη XAA για την περίοδο 2/1/2007 – 13/3/2007 (όπως δίνεται στο ένθετο) με μοντέλα  $\text{AR}(p)$  για  $p = 1, \dots, 20$ . (β) Πρόβλεψη ενός βήματος για την ίδια περίοδο (στον οριζόντιο άξονα δίνονται οι μέρες του μήνα). Οι προβλέψεις έγιναν στις αποδόσεις με  $\text{AR}(1)$  και  $\text{AR}(3)$  μοντέλο, όπως δίνεται στο ένθετο.

Συμπερασματικά θα λέγαμε ότι η πρόβλεψη του ημερήσιου γενικού δείκτη XAA ακολουθώντας τη διαδικασία Box – Jenkins αποτυγχάνει. Πράγματι η χρονοσειρά του γενικού δείκτη XAA φαίνεται να είναι συνεπής με την υπόθεση του τυχαίου περιπάτου. Τα αποτελέσματα αυτά αφήνουν δύο ανοιχτά ενδεχόμενα: (α) δεν υπάρχει κάποια δυναμική της χρηματαγοράς που μπορεί να εκτιμηθεί από τις τιμές του γενικού δείκτη, (β) υπάρχει αλλά η γραμμική ανάλυση και τα γραμμικά μοντέλα δεν καταφέρνουν να αναδείξουν. Στη συνέχεια θα μελετήσουμε άλλες μεθόδους ανάλυσης και πρόβλεψης χρονοσειρών που βασίζονται στην υπόθεση ύπαρξης μη-γραμμικής δυναμικής, δηλαδή υποθέτουμε ότι η χρονοσειρά παράγεται από κάποιο μη-γραμμικό δυναμικό σύστημα που έχει τη δυνατότητα να παράγει τροχιές (όπως αυτή που μετράμε) με φαινομενικά τυχαία μορφή.

## 4.2 Μη-γραμμικά μοντέλα πρόβλεψης χρονοσειρών

Μέχρι τώρα μελετήσαμε γραμμικά χαρακτηριστικά των χρονοσειρών, κυρίως την αυτοσυσχέτιση, και γραμμικά μοντέλα χρονοσειρών, τα μοντέλα τύπου ARMA και ARIMA. Τα στατιστικά χαρακτηριστικά της χρονοσειράς που μπορεί να μιμηθεί ένα γραμμικό μοντέλο περιορίζονται στη μέση τιμή, στη διασπορά και στην αυτοσυσχέτιση. Αυτά τα χαρακτηριστικά είναι αρκετά για να ορίσουν πλήρως μια κανονική (Gaussian) διαδικασία, αλλά δεν αποτελούν ικανοποιητική περιγραφή μιας

μη-κανονικής διαδικασίας από την οποία μπορεί να προκύπτει η χρονοσειρά ενός χρηματο-οικονομικού δείκτη. Οι χρηματιστηριακές και οικονομετρικές χρονοσειρές δε χαρακτηρίζονται από κανονική κατανομή και τα γραμμικά μοντέλα τύπου ARIMA δεν περιγράφουν ικανοποιητικά αυτές τις χρονοσειρές, με αποτέλεσμα να μην έχουν ικανότητα πρόβλεψης. Υπάρχει ένα πλήθος άλλων προσεγγίσεων με βάση τη μη-γραμμικότητα που αναφέρεται ως **μη-γραμμική ανάλυση χρονοσειρών** (nonlinear time series analysis) και αφορά τη διερεύνηση και εκτίμηση χαρακτηριστικών που αντιστοιχούν σε σύνθετες (μη-γραμμικές) δομές και δίνουν περισσότερη πληροφορία για το υπό μελέτη σύστημα. Για το σκοπό αυτό επιστρατεύονται μέθοδοι που πηγάζουν από τη θεωρία των μη-γραμμικών δυναμικών συστημάτων και του χάους. Με αυτά τα θέματα ασχολήθηκαν κατά κύριο λόγο φυσικοί, αλλά τα τελευταία χρόνια αυτή η προσέγγιση μη-γραμμικής ανάλυσης χρονοσειρών έχει συμπεριλάβει πολλά θέματα στατιστικής.

Τα μη-γραμμικά δυναμικά συστήματα έχουν τη δυνατότητα να δημιουργούν ποικίλες καταστάσεις χωρίς την επίδραση του θορύβου,  $z_t$ , τις οποίες τα γραμμικά μοντέλα αδυνατούν να περιγράψουν. Αντίστροφα, αυτό σημαίνει ότι μπορεί τα δεδομένα να κρύβουν μεγαλύτερη δομή από αυτήν που μπορεί να ανακαλύψει ένα γραμμικό μοντέλο. Επειδή σε οικονομετρικά προβλήματα η κλασική στατιστική ανάλυση χρονοσειρών δεν έχει αποδώσει ικανοποιητικά αποτελέσματα, οι μέθοδοι της μη-γραμμικής δυναμικής έχουν αποκτήσει μεγάλο ενδιαφέρον.

Η μεθοδολογία της μη-γραμμικής ανάλυσης περιλαμβάνει την εκτίμηση μη-γραμμικών χαρακτηριστικών, όπως τη *μορφοκλασματική διάσταση* (fractal dimension) και τους *εκθέτες Lyapunov* (Lyapunov exponents). Επίσης υπάρχουν διάφορες κλάσεις μη-γραμμικών μοντέλων που χρησιμοποιούνται για προβλέψεις, όπως τα *νευρωνικά δίκτυα* (neural networks), τα *μη-γραμμικά προσθετικά μοντέλα* (nonlinear additive models) και τα *τοπικά γραμμικά μοντέλα* (local linear models). Εδώ δε θα μελετήσουμε όλη αυτή τη μεθοδολογία παρά μόνο τα τοπικά μοντέλα πρόβλεψης. Πρώτα όμως θα κάνουμε μια σύντομη ανασκόπηση των δυναμικών συστημάτων.

#### 4.2.1 Χρονοσειρές από δυναμικά συστήματα

Θεωρούμε πως το υπό μελέτη σύστημα που παρατηρούμε μέσω μιας χρονοσειράς είναι ένα μη-γραμμικό δυναμικό σύστημα. Αυτό ορίζεται μαθηματικά από εξισώσεις διαφορών, αν ο χρόνος είναι διακριτός (π.χ. ο ημερήσιος γενικός δείκτης), ή από διαφορικές εξισώσεις, αν ο χρόνος είναι συνεχής, που πρακτικά σημαίνει ότι η δειγματοληψία είναι πολύ πυκνή (π.χ. συναλλαγματική αξία ανά δευτερόλεπτο). Συνήθως στα οικονομετρικά προβλήματα ο χρόνος θεωρείται διακριτός. Σ' αυτήν την περίπτωση το δυναμικό σύστημα ορίζεται ως

$$s_{t+1} = f(s_t), \quad (41)$$

όπου  $f$  η συνάρτηση του συστήματος και  $s_t = [s_{t,1}, \dots, s_{t,d}]' \in \mathbb{R}^d$  το σημείο ή διάνυσμα θέσης (κατάστασης) του συστήματος τη χρονική στιγμή  $t$  στον Ευκλείδειο χώρο διάστασης  $d$ . Αρχίζοντας από ένα αρχικό σημείο (αρχική συνθήκη)  $s_0$  και με επαναλήψεις της παραπάνω εξίσωσης ορίζεται η τροχιά του συστήματος  $\{s_0, s_1, \dots\}$ .

Θεωρούμε πως η τροχιά του συστήματος περιορίζεται στο  $\mathbb{R}^d$  και καταλήγει ασυμπτωτικά σε ένα αναλλοίωτο σύνολο σημείων που λέγεται **ελκυστής** (attractor). Γενικά ο ελκυστής μπορεί να είναι (υποθέτουμε εδώ διακριτά συστήματα):

- ένα *ευσταθές σημείο ισορροπίας* του συστήματος (stable equilibrium point),
- ένα *πεπερασμένο σύνολο σημείων* (για περιοδικές τροχιές),

- κάποιο άλλο μη-πεπερασμένο σύνολο σημείων.

Για την τελευταία κατηγορία, ιδιαίτερο ενδιαφέρον έχει ο παράξενος ελκυστής (strange attractor), που παρουσιάζει την ιδιότητα της αυτο-ομοιότητας (self-similarity) σε διαφορετικές κλίμακες του χώρου, είναι δηλαδή ένα μορφοκλασματικό σύνολο (fractal). Τα μορφοκλασματικά σύνολα χαρακτηρίζονται από τη μορφοκλασματική διάσταση (fractal dimension), που είναι ένας μη-ακέραιος αριθμός και δηλώνει το βαθμό αυτό-ομοιότητας. Η μορφοκλασματική διάσταση είναι πάντα μικρότερη της Ευκλείδειας διάστασης στην οποία βρίσκεται ο ελκυστής. Οι παράξενοι ελκυστές σχηματίζονται από τις τροχιές δυναμικών συστημάτων που παρουσιάζουν ευαίσθητη εξάρτηση στις αρχικές συνθήκες (sensitive dependence on initial conditions) και λέγονται χαοτικά δυναμικά συστήματα (chaotic dynamical systems).

Τα χαοτικά δυναμικά συστήματα παρουσιάζουν ιδιαίτερο ενδιαφέρον γιατί λόγω της ιδιότητας της ευαισθησίας στις αρχικές συνθήκες, κοντινές τροχιές τους αποκλίνουν πολύ γρήγορα με αποτέλεσμα να παρουσιάζουν στοχαστική συμπεριφορά. Αντίστροφα, πολλά πραγματικά συστήματα που φαίνονται τυχαία ίσως να έχουν αιτιοκρατική μη-γραμμική και ενδεχομένως χαοτική δομή. Αυτό σημαίνει ότι χρησιμοποιώντας κατάλληλες μεθόδους που βασίζονται στη θεωρία των δυναμικών συστημάτων και του χάους έχουμε τη δυνατότητα να εξηγήσουμε και να προβλέψουμε τέτοια συστήματα (σε μικρό χρονικό ορίζοντα).

Μια μονοδιάστατη χρονοσειρά  $\{x_t\}$  για  $t=1, \dots, n$ , μπορεί να θεωρηθεί ότι προέρχεται από ένα δυναμικό σύστημα ως η προβολή  $x_t$  κάθε σημείου  $s_t$  της τροχιάς του συστήματος σε κάποιο άξονα του  $\mathbb{R}^d$ , δηλαδή

$$x_t = h(s_t), \quad (42)$$

όπου η συνάρτηση προβολής  $h: \mathbb{R}^d \mapsto \mathbb{R}$  λέγεται και συνάρτηση παρατήρησης (observation function). Για παράδειγμα η προβολή του  $s_t$  μπορεί να είναι πάνω σε έναν από τους κύριους άξονες του  $\mathbb{R}^d$ , δηλαδή το  $x_t$  να είναι μια από τις συνιστώσες  $s_{1,t}, \dots, s_{d,t}$ .

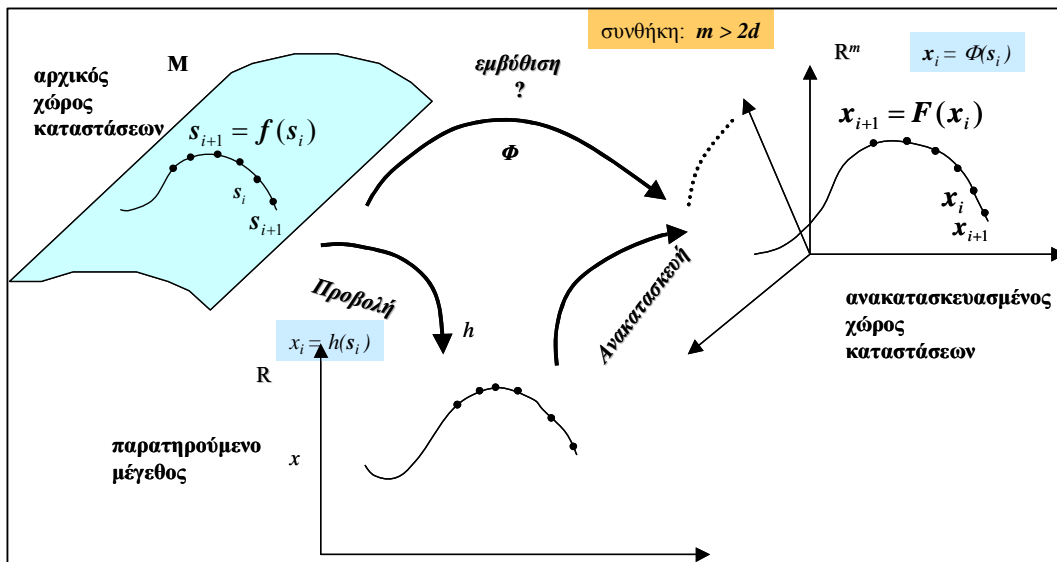
Σε μια πρώτη προσέγγιση θα θέλαμε ίσως να εκτιμήσουμε τη συνάρτηση του συστήματος  $f$ , τη διάσταση του Ευκλείδειου χώρου καταστάσεων  $d$  στο οποίο βρίσκονται οι τροχιές που παράγει το σύστημα, καθώς και τη συνάρτηση παρατήρησης  $h$ . Δεν είναι όμως δυνατόν από την παρατήρηση μιας μονοδιάστατης μεταβλητής  $x_t$  που αφορά το υπό μελέτη σύστημα να αντλήσουμε όλες αυτές τις πληροφορίες, δηλαδή να πετύχουμε αντιστροφή και από την προβολή στο  $\mathbb{R}$  να γυρίσουμε στο χώρο των καταστάσεων  $\mathbb{R}^d$ . Μπορούμε όμως να φτιάξουμε ένα είδωλο του αρχικού ελκυστή όπως θα δούμε παρακάτω.

#### 4.2.2 Ανακατασκευή χώρου καταστάσεων

Η δυναμική ανάλυση χρονοσειρών βασίζεται κατά κύριο ρόλο στη δημιουργία ενός ειδώλου του αρχικού ελκυστή. Ο σχηματισμός ενός νέου «ισοδύναμου» ελκυστή είναι δυνατός από το θεώρημα του Takens, που είναι θεμελιώδης για τη δυναμική ανάλυση χρονοσειρών.

Το **θεώρημα του Takens** (Takens theorem) επιτρέπει κάτω από κάποιες συνθήκες να κατασκευάσουμε ένα νέο χώρο καταστάσεων διάστασης  $m$  στον οποίο οι ανακατασκευασμένες τροχιές  $\{\mathbf{x}_t\} \in \mathbb{R}^m$  από τη χρονοσειρά  $\{x_t\}$ , δηλαδή ο ανακατασκευασμένος ελκυστής, διατηρεί τις χαρακτηριστικές ιδιότητες του αρχικού

ελκυστή. Αυτό σημαίνει πως το ανακατασκευασμένο δυναμικό σύστημα,  $x_{t+1} = F(x_t)$ , έχει τα ίδια δυναμικά χαρακτηριστικά με το αρχικό σύστημα,  $s_{t+1} = f(s_t)$ . Αυτή η ιδιότητα αναφέρεται και ως **εμβύθιση** (embedding). Το θεώρημα του Takens δίνει τη συνθήκη  $m \geq 2D+1$  για την ανακατασκευή του χώρου καταστάσεων, όπου  $D$  είναι η μορφοκλασματική διάσταση του ελκυστή (που είναι όμως άγνωστη!). Στο σχεδιάγραμμα στο Σχήμα 43 παρουσιάζεται το πρόβλημα της ανακατασκευής του χώρου καταστάσεων από τη χρονοσειρά. Θα πρέπει να σημειωθεί ότι το θεώρημα του Takens αναφέρεται σε χρονοσειρές με άπειρο μήκος και χωρίς θόρυβο. Στην πράξη αυτές οι συνθήκες φυσικά δε συναντώνται και γι αυτό τα αποτελέσματα της ανακατασκευής μπορούν να θεωρηθούν μόνο προσεγγιστικά.



Σχήμα 43 Σχεδιάγραμμα ανακατασκευής χώρου καταστάσεων.

Η ανακατασκευή του χώρου καταστάσεων γίνεται απλά με τη δημιουργία σημείων  $x_t \in \mathbb{R}^m$  από τις μονοδιάστατες παρατηρήσεις  $\{x_1, x_2, \dots, x_n\}$ . Η πιο απλή μέθοδος ανακατασκευής του χώρου καταστάσεων είναι η **μέθοδος των υστερήσεων** (method of delays) και τα σημεία  $x_t$  σχηματίζονται απλά ως

$$x_t = [x_t, x_{t-\tau}, \dots, x_{t-(m-1)\tau}]'. \quad (43)$$

Οι παράμετροι της ανακατασκευής είναι:

1. Η **διάσταση εμβύθισης** (embedding dimension)  $m$  που ορίζει τον αριθμό των παρατηρήσεων που γίνονται συνιστώσες του ανακατασκευασμένου διανύσματος.
2. Η **υστέρηση** (delay)  $\tau$  που ορίζει με ποια χρονική διαφορά επιλέγονται οι  $m$  παρατηρήσεις για χρόνους μικρότερους της χρονικής στιγμής  $t$ .

Θα πρέπει να σημειωθεί ότι η πληροφορία που διοχετεύεται από τη χρονοσειρά στο κάθε ανακατασκευασμένο διάνυσμα  $x_t$  καλύπτει το χρονικό παράθυρο  $\tau_w = (m-1)\tau$ . Για να επιτευχθεί σωστή ανακατασκευή θα πρέπει το παράθυρο αυτό να μην είναι ούτε πολύ μικρό, αλλιώς δε θα έχει χρησιμοποιηθεί αρκετή πληροφορία, ούτε πολύ μεγάλο, αλλιώς θα υπάρχει περιττή πληροφορία που θα περιπλέξει την ανακατασκευή.

Για την ανακατασκευή του χώρου καταστάσεων πρέπει να επιλέξουμε κατάλληλες τιμές για τις παραμέτρους  $m$  και  $\tau$ . Για τις οικονομικές χρονοσειρές που μας ενδιαφέρουν, συνήθως δε μας απασχολεί η εκτίμηση της υστέρησης και θέτουμε  $\tau = 1$ . Για την εκτίμηση της διάστασης εμπύθισης  $m$  έχουν προταθεί κάποιες μέθοδοι, όπως η *μέθοδος των ψευδών γειτόνων* (false nearest neighbors), αλλά δε φαίνεται να υπάρχει κάποια μέθοδος που να εκτιμά πάντα κατάλληλες τιμές του  $m$  σε πρακτικές εφαρμογές.

### 4.2.3 Τοπικά μοντέλα πρόβλεψης

Θεωρώντας την ανακατασκευή του χώρου των καταστάσεων και την υπόθεση ότι η χρονοσειρά είναι η προβολή κάποιου δυναμικού συστήματος, τα σημεία του ανακατασκευασμένου ελκυστή ορίζονται από τη συνάρτηση (διανυσματικό πεδίο)  $F: \mathbb{R}^m \mapsto \mathbb{R}^m$ ,  $\mathbf{x}_{i+1} = F(\mathbf{x}_i)$ , ή γενικά για κάποιο χρονικό βήμα  $T$ ,  $\mathbf{x}_{i+T} = F^T(\mathbf{x}_i)$ . Σε προβλήματα πρόβλεψης της χρονοσειράς δε μας ενδιαφέρει να εκτιμήσουμε την  $F^T$  αλλά μόνο τη συνιστώσα της  $F^T: \mathbb{R}^m \mapsto \mathbb{R}$  που ορίζει το στοιχείο  $x_{i+T}$  της χρονοσειράς,  $x_{i+T} = F^T(\mathbf{x}_i)$ . Είναι φανερό ότι ένα ικανοποιητικό μοντέλο για την συνάρτηση  $F^T$  θα πρέπει να είναι μη-γραμμικό για να μπορεί να κατέχει τις ιδιότητες του αρχικού συστήματος όπως αυτές διατηρούνται μέσω της ανακατασκευής. Σημειώνεται ότι αν το σύστημα είναι χαοτικό τότε έχει ευαισθησία στις αρχικές συνθήκες κι άρα ανεξάρτητα από την καταλληλότητα του μοντέλου οι προβλέψεις περιορίζονται σε μικρό χρονικό ορίζοντα και η αβεβαιότητα της πρόβλεψης αυξάνει εκθετικά σύμφωνα με τον μέγιστο εκθέτη Lyapunov.

Το πρόβλημα της πρόβλεψης είναι, έχοντας τις παρατηρήσεις  $\{x_1, \dots, x_n\}$  μιας τυχαίας μεταβλητής, να προβλέψουμε τις τιμές της στις επόμενες χρονικές στιγμές,  $x_{n+1}, \dots, x_{n+T}$  για κάποιο χρονικό ορίζοντα  $T$ . Η βασική ιδέα με τα **τοπικά μοντέλα πρόβλεψης** (local prediction models ή local state space models) είναι ότι εντοπίζουμε στην ιστορία της χρονοσειράς τμήματα ή παράθυρα μεγέθους  $m$  (έστω  $\{x_{j-m+1}, \dots, x_j\}$  ένα τέτοιο τμήμα) που μοιάζουν με το τελευταίο τμήμα μεγέθους  $m$  που έχουμε στη διάθεση μας,  $\{x_{n-m+1}, \dots, x_n\}$ . Αν η χρονοσειρά έχει δημιουργηθεί από κάποιο αιτιοκρατικό δυναμικό σύστημα τότε τα όμοια τμήματα θα έχουν και όμοιες εικόνες, δηλαδή το  $x_{j+1}$  (που είναι γνωστό) θα είναι κοντά στο  $x_{n+1}$  (που θέλουμε να εκτιμήσουμε). Οι εικόνες των όμοιων τμημάτων από την ιστορία της χρονοσειράς συνθέτουν την πρόβλεψη του  $x_{n+1}$  και αυτό μπορεί να επεκταθεί και για μεγαλύτερο χρονικό ορίζοντα  $T > 1$ .

Στον ανακατασκευασμένο χώρο καταστάσεων  $\mathbb{R}^m$  αυτά τα τμήματα της χρονοσειράς αντιστοιχούν σε γειτονικά σημεία του  $\mathbf{x}_n = [x_n, x_{n-1}, \dots, x_{n-m+1}]'$ , δηλαδή του τελευταίου ανακατασκευασμένου σημείου από τη χρονοσειρά  $\{x_1, \dots, x_n\}$  που αντιστοιχεί στη χρονική στιγμή  $n$  (ο χρόνος αφετηρίας). Έστω  $\mathbf{x}_{n(1)}$  το κοντινότερο σημείο στο  $\mathbf{x}_n$  που βρέθηκε για χρόνους μικρότερους του  $n$ . Η πιο απλή πρόβλεψη του  $x_{n+1}$  είναι από την εικόνα του  $\mathbf{x}_{n(1)}$ , δηλαδή

$$\hat{x}_{n+1} \equiv x_n(1) = x_{n(1)+1}. \quad (44)$$

Αυτή είναι η **πρόβλεψη μηδενικής τάξης** (zeroth order prediction). Αν συμπεριλάβουμε και άλλα γειτονικά σημεία, έστω  $K$  γειτονικά σημεία

$\mathbf{x}_{n(1)}, \mathbf{x}_{n(2)}, \dots, \mathbf{x}_{n(K)}$ , μπορούμε να κάνουμε πρόβλεψη παίρνοντας το μέσο όρο των εικόνων των γειτονικών σημείων και είναι

$$x_n(1) = \frac{1}{K} \sum_{j=1}^K x_{n(j)+1}. \quad (45)$$

Αυτή είναι η **τοπική πρόβλεψη του μέσου όρου απεικονίσεων** (local average mapping, LAM). Μπορεί εναλλακτικά να οριστεί από το σταθμισμένο μέσο όρο, σταθμίζοντας με την απόσταση του κάθε γειτονικού σημείου από το  $\mathbf{x}_n$ .

Πιο σύνθετο τοπικό μοντέλο από την πρόβλεψη μέσου όρου απεικονίσεων είναι το **τοπικό γραμμικό μοντέλο** (local linear model, LLM) που ορίζει την εικόνα  $x_{n+1}$  ως γραμμικό συνδυασμό των τελευταίων  $m$  στοιχείων  $x_{n-m+1}, \dots, x_n$

$$x_{n+1} = \phi_0^{(n)} + \phi_1^{(n)} x_n + \phi_2^{(n)} x_{n-1} + \dots + \phi_m^{(n)} x_{n-m+1}, \quad (46)$$

όπου οι παράμετροι  $\phi_0^{(n)}, \phi_1^{(n)}, \dots, \phi_m^{(n)}$  δεν είναι σταθεροί αλλά ορίζονται (και εκτιμώνται) για κάθε χρονική στιγμή  $n$ . Το μοντέλο αυτό μοιάζει με το σκελετό ενός  $AR(m)$  μοντέλου (χωρίς το λευκό θόρυβο). Η βασική διαφορά από το κλασικό AR μοντέλο είναι ότι για την εκτίμηση των παραμέτρων  $\phi_0^{(n)}, \phi_1^{(n)}, \dots, \phi_m^{(n)}$  δε χρησιμοποιούνται όλα τα στοιχεία της χρονοσειράς αλλά μόνο  $K$  τμήματα που μοιάζουν στο τμήμα  $\{x_{n-m+1}, \dots, x_n\}$ , δηλαδή τα  $K$  γειτονικά σημεία  $\mathbf{x}_{n(1)}, \mathbf{x}_{n(2)}, \dots, \mathbf{x}_{n(K)}$  του  $\mathbf{x}_n = [x_n, x_{n-1}, \dots, x_{n-m+1}]'$  στον ανακατασκευασμένο χώρο καταστάσεων  $\mathbb{R}^m$ . Υποθέτουμε ότι το τοπικό γραμμικό μοντέλο ισχύει για μια μικρή περιοχή γύρω από το  $\mathbf{x}_n$  και άρα και για τα σημεία  $\mathbf{x}_{n(1)}, \mathbf{x}_{n(2)}, \dots, \mathbf{x}_{n(K)}$ .

Η εκτίμηση των παραμέτρων δίνεται εύκολα από τη μέθοδο ελαχίστων τετραγώνων. Η πρόβλεψη του  $x_{n+1}$  δίνεται από το γραμμικό μοντέλο με τις εκτιμήσεις των παραμέτρων  $\phi_0^{(n)}, \phi_1^{(n)}, \dots, \phi_m^{(n)}$ . Για πρόβλεψη  $T$  χρονικών βημάτων μπορούμε απλά να αλλάξουμε στα μοντέλα πρόβλεψης τις εικόνες των γειτονικών σημείων για ένα χρονικό βήμα,  $x_{n(j)+1}$ , με τις εικόνες για  $T$  χρονικά βήματα,  $x_{n(j)+T}$ . Μπορούμε επίσης να κάνουμε την πρόβλεψη επαναληπτικά προβλέποντας κάθε φορά ένα χρονικό βήμα μπροστά. Για παράδειγμα, έχοντας υπολογίσει το  $x_n(1)$  σχηματίζουμε το διάνυσμα  $\mathbf{x}_{n+1}$  αντικαθιστώντας τη συνιστώσα  $x_{n+1}$  με το  $x_n(1)$ , βρίσκουμε τα γειτονικά σημεία του  $\mathbf{x}_{n+1}$  και προχωράμε στον υπολογισμό του  $x_n(2)$ .

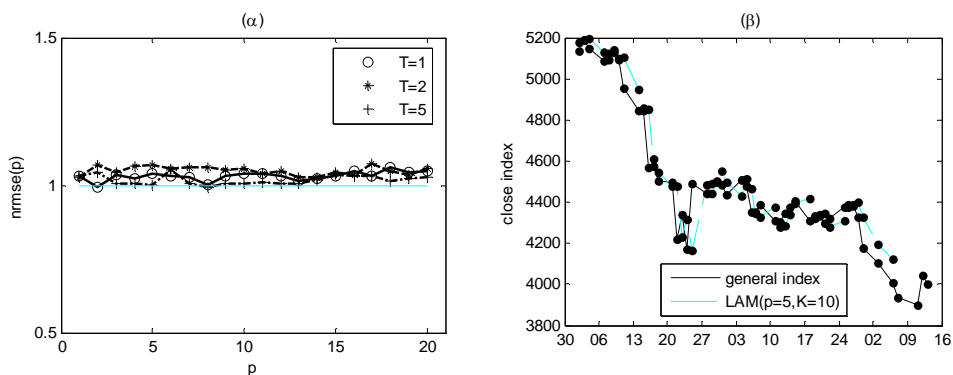
Για την αξιολόγηση της απόδοσης των μοντέλων πρόβλεψης χρησιμοποιείται η ίδια διαδικασία όπως και για τα γραμμικά μοντέλα. Ειδικότερα το λάθος πρόβλεψης μετριέται από το `rmse`. Αν η χρονοσειρά  $\{x_t\}$  είναι μη-στάσιμη, γίνεται πρώτα στάσιμη με τις διαδικασίες που περιγράφηκαν στην Παράγραφο 4.1.2, όπως παίρνοντας τις πρώτες διαφορές ή τις αποδόσεις. Η πρόβλεψη σε αυτήν την περίπτωση γίνεται όπως στη διαδικασία Box-Jenkins. Για την πρόβλεψη της μη-στάσιμης  $\{y_t\}$  γίνεται πρώτα η πρόβλεψη στη στάσιμη χρονοσειρά  $\{x_t\}$ , π.χ. στη χρονοσειρά των πρώτων διαφορών. Η πρόβλεψη για χρονική στιγμή  $n$  και ένα βήμα μπροστά σε αυτή τη χρονοσειρά,  $x_n(1)$ , προστίθεται στην τιμή της αρχικής χρονοσειράς για χρόνο  $n$ ,  $y_n$ , για να πάρουμε την πρόβλεψη  $y_n(1) = y_n + x_n(1)$ .

Η πρόβλεψη με τα τοπικά γραμμικά μοντέλα και με τα μη-γραμμικά μοντέλα γενικότερα επηρεάζεται από την παράμετρο της διάστασης εμπύθισης  $m$ . Η επίδραση

του  $m$  είναι πιο σημαντική από την επίδραση της τάξης  $p$  στα μοντέλα AR, όπου είδαμε ότι για μεγαλύτερη τάξη οι προβλέψεις δεν αλλάζουν σημαντικά. Για την τοπική πρόβλεψη μέσου όρου και την πρόβλεψη με τοπικά γραμμικά μοντέλα η απόδοση του μοντέλου επηρεάζεται και από τον αριθμό των γειτονικών σημείων  $K$ . Ανεξάρτητα από την καταλληλότητα του μοντέλου και την επιλογή των παραμέτρων η πρόβλεψη χασοτικών χρονοσειρών έχει νόημα για μικρό χρονικό ορίζοντα που ορίζεται από τον μέγιστο εκθέτη Lyapunov.

Παράδειγμα: Μη-γραμμική πρόβλεψη του δείκτη XAA

Κάνουμε προβλέψεις με τα τοπικά μοντέλα για τη χρονοσειρά του γενικού δείκτη, για την οποία κάναμε προβλέψεις με τα γραμμικά μοντέλα. Η χρονοσειρά του γενικού δείκτη  $\{y_t\}$  είναι μη-στάσιμη, γι αυτό ακολουθώντας τη διαδικασία Box-Jenkins για ARIMA μοντέλα, πρώτα κάνουμε πρόβλεψη για τη χρονοσειρά των αποδόσεων  $\{x_t\}$  (ή ισοδύναμα των πρώτων διαφορών). Τα γραμμικά μοντέλα που εξετάσαμε δεν έδειξαν να έχουν ικανότητα πρόβλεψης. Τα τοπικά μοντέλα δε φαίνεται να δίνουν καλύτερες προβλέψεις, όπως φαίνεται στο Σχήμα 44.



Σχήμα 44 (α) Σφάλμα nmse για την πρόβλεψη ενός, δύο και πέντε βημάτων μπροστά του ημερήσιου γενικού δείκτη XAA για την περίοδο 2/1/2007 – 13/3/2007 (όπως δίνεται στο ένθετο) με μοντέλα  $AR(p)$  για  $p = 1, \dots, 20$ . (β) Πρόβλεψη ενός βήματος για την ίδια περίοδο (στον οριζόντιο άξονα δίνονται οι μέρες του μήνα). Οι προβλέψεις έγιναν στις αποδόσεις με  $AR(1)$  και  $AR(3)$  μοντέλο, όπως δίνεται στο ένθετο..

Οι προβλέψεις ενός βήματος μπροστά ( $T = 1$ ) του ημερήσιου γενικού δείκτη XAA για την περίοδο 2/1/2008 – 13/5/2008 δεν παρουσιάζουν σημαντικές διαφορές στο γραμμικό και στο μη-γραμμικό μοντέλο (δες Σχήμα 42). Τα αποτελέσματα αυτά επιβεβαιώνονται για διαφορετικές τιμές της διάστασης του χώρου ανακατασκευής  $m$ , που για τα AR μοντέλα είναι ισοδύναμη με την τάξη  $p$ .

Γενικά η πρόβλεψη σε προβλήματα οικονομίας και χρηματο-οικονομίας είναι ένα δύσκολο εγχείρημα. Οι χρονοσειρές σε τέτοια προβλήματα παρουσιάζουν μεγάλη στοχαστικότητα καθώς επιδρούν πολλοί εξωτερικοί και τυχαίοι παράγοντες στη διαμόρφωση των παρατηρούμενων δεικτών. Στη συνέχεια θα μελετήσουμε ένα χαρακτηριστικό των χρηματο-οικονομιών δεικτών που έχει επίσης ενδιαφέρον να προβλέψουμε, που αναφέρεται ως μεταβλητότητα ή αστάθεια (volatility).

### 4.3 Αστάθεια χρονοσειράς

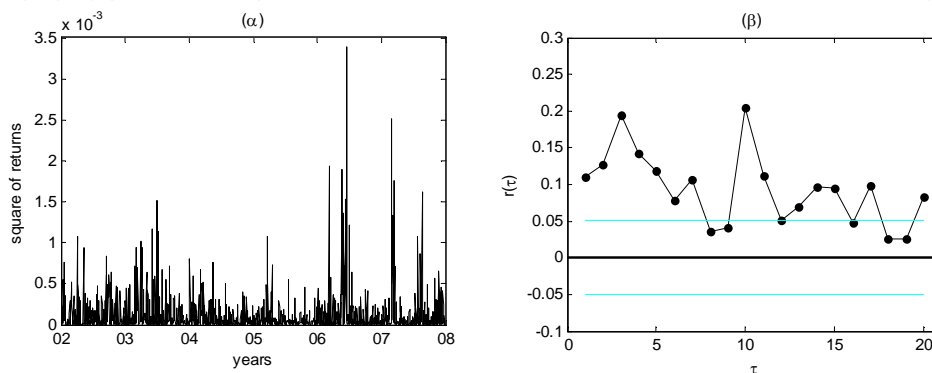
Η μεταβλητότητα ή **αστάθεια** (volatility) δηλώνει την αυξομείωση των τιμών της χρονοσειράς και μπορεί απλά να ορισθεί ως η διασπορά της απόδοσης της παρατηρούμενης μεταβλητής σε κάθε χρονική στιγμή  $t$ ,  $\sigma_t^2$  (ή αντίστοιχα η τυπική

απόκλιση  $\sigma_t$ ). Φυσικά μια στάσιμη χρονοσειρά έχει σταθερή διασπορά και άρα δε μπορούμε να μιλάμε για αστάθεια. Υπάρχουν όμως χρονοσειρές, όπως οι αποδόσεις χρηματο-οικονομικών δεικτών, με μεταβαλλόμενη διασπορά, παρ' όλο που ασυμπτωτικά (δηλαδή για μεγάλο χρονικό ορίζοντα) η διασπορά φαίνεται να συγκλίνει σε κάποια τιμή (δεν αυξάνει με τη χρονική κλίμακα).

Υποθέτοντας πως η μέση τιμή της χρονοσειράς απόδοσης ενός δείκτη  $\{x_t\}_{t=1}^N$  είναι 0, η αστάθεια σε κάθε χρόνο  $t$  μπορεί να εκτιμηθεί στιγμιαία από το τετράγωνο της απόδοσης ως  $s_t^2 = x_t^2$  ή με αναφορά στην τυπική απόκλιση από την απόλυτη τιμή της ως  $s_t = |x_t|$ . Η στιγμιαία αστάθεια της χρονοσειράς μπορεί να μην εμφανίζεται τυχαία. Οι αποδόσεις χρηματιστηριακών δεικτών μπορεί να μην έχουν συσχετίσεις αλλά τα τετράγωνα τους να έχουν.

Παράδειγμα: Στιγμαία αστάθεια στη χρονοσειρά του ΧΑΑ

Ας πάρουμε ως παράδειγμα την απόδοση του γενικού δείκτη ΧΑΑ. Είδαμε στη προηγούμενη παράγραφο πως η διαδικασία Box-Jenkins δεν απέδωσε γιατί οι συσχετίσεις στη χρονοσειρά  $\{y_t\}_{t=0}^N$  την περίοδο 2/1/2002 – 13/3/2008 που εξετάσαμε είναι πολύ ασθενείς. Η αυτοσυσχέτιση της χρονοσειράς βρέθηκε να είναι οριακά σημαντική για κάποιες υστερήσεις και τα γραμμικά και τοπικά μοντέλα δε μπορούσαν να δώσουν προβλέψεις. Τα αποτελέσματα αυτά δείχνουν την αδυναμία πρόβλεψης της πορείας του δείκτη, δηλαδή αν θα ανέβει ή θα κατέβει και πόσο. Ένα άλλο θέμα είναι η το πόσο αυξομείωσης του δείκτη που στηρίζεται στην ανάλυση των τετραγώνων των αποδόσεων  $\{x_t^2\}$ . Στο Σχήμα 45 δίνεται η χρονοσειρά των τετραγώνων των αποδόσεων του γενικού δείκτη καθώς και η αυτοσυσχέτιση της. Παρατηρούμε ότι υπάρχουν σημαντικές αυτοσυσχετίσεις για πολλές υστερήσεις. Η περιγραφή τέτοιων συσχετίσεων γίνεται με μοντέλα τύπου GARCH που θα δούμε αργότερα.



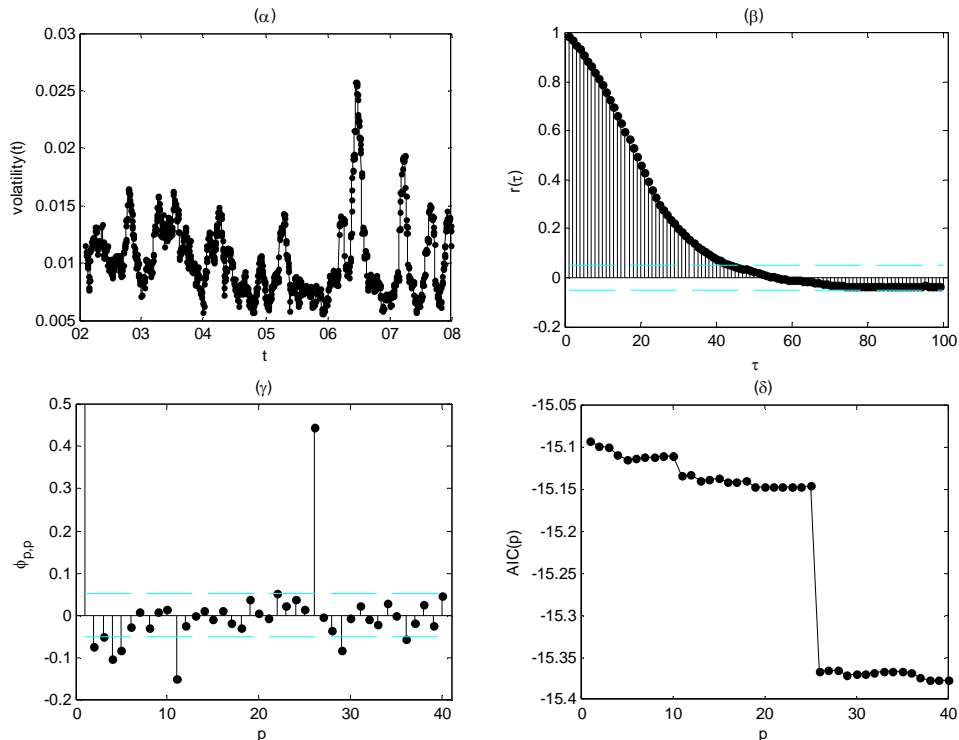
**Σχήμα 45 (α) Η χρονοσειρά των τετραγώνων των αποδόσεων του γενικού δείκτη ΧΑΑ την περίοδο 2/1/2002 – 13/3/2008. (β) Η αυτοσυσχέτιση της χρονοσειράς στο (α). Τα όρια της στατιστικής σημαντικότητας δίνονται με τις οριζόντιες γκριζες γραμμές.**

Η αστάθεια συνήθως δεν εκτιμάται από μια χρονική στιγμή, όπως παραπάνω, αλλά σε ένα παράθυρο χρόνου. Για παράδειγμα αν ένας δείκτης μετριέται ανά λεπτό, η αστάθεια σε κάθε λεπτό μπορεί να εκτιμάται από τη διασπορά την τελευταία ώρα, ενώ αν ο δείκτης είναι ημερήσιος, η αστάθεια για κάθε μέρα μπορεί να ορίζεται με τη διασπορά την τελευταία βδομάδα (5 εργάσιμες μέρες) ή τον τελευταίο μήνα (προσεγγιστικά τις τελευταίες 25 μέρες). Με αυτόν τον τρόπο σχηματίζεται η χρονοσειρά της αστάθειας του δείκτη και θα θέλαμε να προβλέψουμε τη χρονοσειρά αυτή για να εκτιμήσουμε την αστάθεια του δείκτη που αναμένουμε τις επόμενες χρονικές στιγμές.

Παράδειγμα: Πρόβλεψη αστάθειας του δείκτη ΧΑΑ με γραμμικά και τοπικά μοντέλα



Υπολογίζουμε την αστάθεια του δείκτη ΧΑΑ σε παράθυρο 25 ημερών. Από τη χρονοσειρά των αποδόσεων  $\{x_t\}_{t=1}^{N_1}$  για  $N_1=1495$ , η χρονοσειρά της αστάθειας υπολογίζεται για κάθε χρονική στιγμή  $t$  από τη διασπορά των  $\{x_{t-24}, x_{t-23}, \dots, x_t\}$  και έχει μήκος  $N_1=1471$ . Στο Σχήμα 46α δίνεται η χρονοσειρά αστάθειας και στο Σχήμα 46β η αυτοσυσχέτιση της. Παρατηρούμε ότι υπάρχει ισχυρή συσχέτιση για μεγάλο εύρος υστερήσεων που προκύπτει από τον τρόπο που υπολογίζεται η αστάθεια. Αντίστοιχα η μερική αυτοσυσχέτιση και η συνάρτηση AIC δίνουν εκτιμήσεις μεγάλης τάξης για το μοντέλο AR, ως και  $p=26$  (η μεγάλη κορυφή στη μερική αυτοσυσχέτιση και πτώση στο AIC για αυτήν την τιμή εξηγείται από το χρονικό παράθυρο που χρησιμοποιήθηκε στον υπολογισμό της αστάθειας).

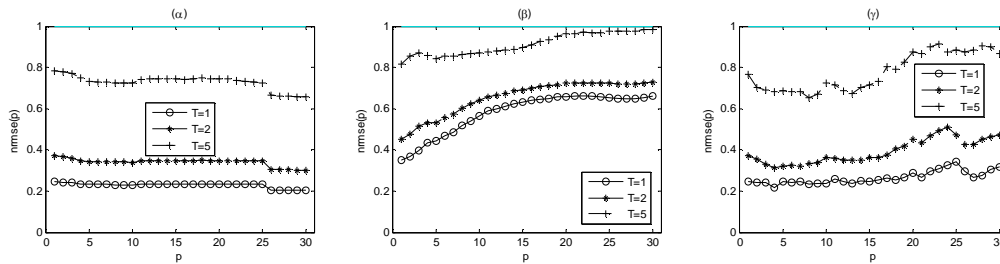


Σχήμα 46 (α) Η χρονοσειρά αστάθειας (σε παράθυρο 25 ημερών) των αποδόσεων του γενικού δείκτη ΧΑΑ την περίοδο 2/1/2002 – 13/3/2008. (β) Η αυτοσυσχέτιση της χρονοσειράς στο (α). Τα όρια της στατιστικής σημαντικότητας δίνονται με τις οριζόντιες γκρίζες γραμμές. (γ) Η μερική αυτοσυσχέτιση της χρονοσειράς στο (α). (δ) Η συνάρτηση AIC της χρονοσειράς στο (α).

Δοκιμάσαμε τα μοντέλα AR για τάξεις  $p=1, \dots, 30$ , για να κάνουμε προβλέψεις για  $T=1, \dots, 5$  σε τιμές αστάθειας του γενικού δείκτη την περίοδο από 2/1/2008 ως 13/3/2008. Όπως δίνεται στο Σχήμα 47α, το σφάλμα πρόβλεψης είναι μικρό για ένα βήμα μπροστά και αυξάνει για περισσότερα βήματα. Οι διαφορές στην πρόβλεψη είναι μικρές ως προς την τάξη και φαίνεται μια μικρή πτώση του σφάλματος όταν η τάξη γίνεται 26, όπως θα περιμέναμε και από τα αποτελέσματα για τη μερική αυτοσυσχέτιση και τη συνάρτηση AIC.

Η πρόβλεψη με το τοπικό μοντέλο μέσου όρου (LAM) είναι χειρότερη από την πρόβλεψη με AR μοντέλο, όπως φαίνεται στο Σχήμα 47β. Χρησιμοποιώντας τοπικό γραμμικό μοντέλο (LLM), η πρόβλεψη βελτιώνεται αλλά δε φαίνεται να ξεπερνά τα επίπεδα των γραμμικών μοντέλων. Στα τοπικά μοντέλα θα πρέπει να επιλεγεί κατάλληλα και το πλήθος των γειτονικών σημείων  $K$  και ίσως με κατάλληλη επιλογή

του  $K$  να μπορούμε να πετύχουμε καλύτερες προβλέψεις της αστάθειας του δείκτη από τα μοντέλα AR.



Σχήμα 47 (α) Σφάλμα  $rmse$  για την πρόβλεψη ενός, δύο και πέντε βημάτων μπροστά της αστάθειας του ημερήσιου γενικού δείκτη ΧΑΑ για την περίοδο 2/1/2007 – 13/3/2007 (όπως δίνεται στο ένθετο) με μοντέλα  $AR(p)$  για  $p = 1, \dots, 30$ . (β) Το ίδιο όπως στο (α) αλλά με τοπικά μοντέλα μέσω όρου από 10 γειτονικά σημεία. (γ) Το ίδιο όπως στο (α) αλλά με τοπικά γραμμικά μοντέλα χρησιμοποιώντας 50 γειτονικά σημεία.

#### 4.4 Γραμμικά μοντέλα πρόβλεψης της διασποράς

Ένας βασικός διαχωρισμός που γίνεται στην ανάλυση οικονομικών χρονοσειρών είναι η ανάλυση με έμφαση στη μέση τιμή (mean) και η ανάλυση με έμφαση στη μεταβλητότητα (variance) ή αστάθεια (volatility). Τα γραμμικά και μη-γραμμικά μοντέλα που μελετήσαμε στις Παραγράφους 4.1 και 4.2 αναφέρονται στην πρόβλεψη της μέσης τιμής. Για την αστάθεια είδαμε στην προηγούμενη παράγραφο πως μπορούμε να δημιουργήσουμε μια νέα χρονοσειρά αστάθειας για να τη μελετήσουμε και προβλέψουμε. Εδώ θα εξετάσουμε μια άλλη προσέγγιση, όπου η αστάθεια ή μεταβλητότητα ορίζεται στιγμιαία και αναπτύσσονται μοντέλα που περιγράφουν τη χρονική ανάπτυξη της από προηγούμενες τιμές των τετραγώνων της χρονοσειράς ή και προηγούμενες τιμές της αστάθειας.

Οι αποδόσεις χρηματο-οικονομικών δεικτών θεωρούνται ότι έχουν μεταβαλλόμενη διασπορά που μπορεί να παίρνει υψηλές τιμές σε κάποιες χρονικές στιγμές ή περιόδους. Θα μπορούσε κάποιος να προσδώσει αυτή τη συμπεριφορά αποκλειστικά στο γεγονός ότι η χρονοσειρά των αποδόσεων ακολουθεί μη-Γκαουσιανή κατανομή με παχιές ουρές, αλλά οι μεγάλες ή μικρές αυξομειώσεις των αποδόσεων δεν εμφανίζονται τυχαία. Παρατηρείται όταν οι αυξομειώσεις αλλάζουν, να διατηρούνται σε αυτό το επίπεδο για κάποια χρονική περίοδο πριν αλλάξουν πάλι, δημιουργούνται δηλαδή συστάδες ίδιας διασπορά (volatility clustering) (δες Σχήμα 4 για τη χρονοσειρά S&P500). Αυτό οδηγεί στο συμπέρασμα πως η διασπορά σε κάποια χρονική στιγμή  $t$  μπορεί να εξηγηθεί από τις τιμές του δείκτη (ή καλύτερα της απόδοσης του δείκτη) σε προηγούμενες χρονικές στιγμές. Τα μοντέλα που θα μελετήσουμε παρακάτω έχουν δημιουργηθεί για αυτόν τον σκοπό.

##### 4.4.1 Μοντέλα ARCH

Το πρώτο μοντέλο που θα μελετήσουμε ονομάζεται **αυτοπαλινδρομούμενο μοντέλο με δεσμευμένη ετεροσκεδαστικότητα** (autoregressive model with conditional heteroscedasticity, ARCH). Οι υποθέσεις για αυτό το μοντέλο είναι:

1. Η χρονοσειρά είναι ασυσχέτιστη.
2. Η μεταβαλλόμενη διασπορά είναι δεσμευμένη σε προηγούμενες τιμές της χρονοσειράς.
3. Η ασυμπτωτική διασπορά είναι σταθερή.

Η πρώτη συνθήκη μπορεί πάντα να ισχύει αν για μια χρονοσειρά με αυτοσυσχετίσεις προσαρμόσουμε ένα μοντέλο AR και θεωρήσουμε τη χρονοσειρά των υπολοίπων. Συνήθως αυτό το βήμα δε χρειάζεται σε χρονοσειρές χρηματοοικονομικών δεικτών γιατί όπως είδαμε δεν έχουν σημαντικές συσχετίσεις. Η δεύτερη συνθήκη δηλώνει και την υπόθεση στην οποία βασίζεται το μοντέλο ARCH. Αν και σύμφωνα με τη συνθήκη / υπόθεση αυτή η χρονοσειρά δεν είναι τοπικά (χρονικά) στάσιμη, η τρίτη συνθήκη δηλώνει τη στασιμότητα της χρονοσειράς καθώς σε μεγάλες χρονικές κλίμακες η διασπορά είναι σταθερή.

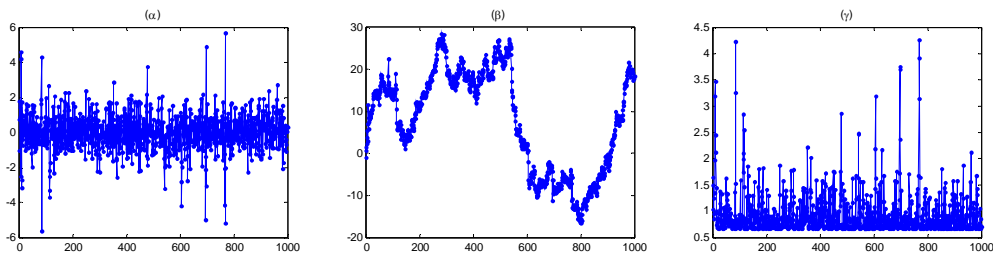
Μαθηματικά μια στοχαστική διαδικασία  $\{x_t\}$  είναι τύπου ARCH όταν

$$x_t = \sigma_t z_t,$$

όπου  $z_t$  είναι iid με μέση τιμή 0 και διασπορά 1, και

$$\sigma_t^2 = \alpha_0 + \alpha_1 x_{t-1}^2 + \dots + \alpha_q x_{t-q}^2, \quad (47)$$

όπου  $\alpha_i > 0$ ,  $i = 0, 1, \dots, q$ . Η παράμετρος  $q$  είναι η τάξη του ARCH και δηλώνει τη μνήμη της εξάρτησης της δεσμευμένης διασποράς. Το  $z_t$  θεωρείται συνήθως να έχει κανονική κατανομή, αλλά αυτό δεν σημαίνει πως και η  $x_t$  έχει κανονική κατανομή. Αυτό φαίνεται από το παράδειγμα της ARCH(1) που δίνεται στο Σχήμα 48. Παρατηρούμε ότι η χρονοσειρά  $\{x_t\}$  που παράγεται από την ARCH(1) έχει διακυμάνσεις που σε κάποιες στιγμές αυξάνουν πολύ και δεν αντιστοιχούν σε κανονική κατανομή (Σχήμα 48α). Η στιγμιαία τυπική απόκλιση  $\{\sigma_t\}$  που δίνεται στο Σχήμα 48γ μεταβάλλεται αλλά όχι εντελώς τυχαία αλλά σύμφωνα με τη σχέση  $\sigma_t^2 = \alpha_0 + \alpha_1 x_{t-1}^2$ . Η χρονοσειρά  $\{y_t\}$  του προφίλ που παίρνουμε με την ολοκλήρωση (άθροισμα) των τιμών της  $\{x_t\}$  σε κάθε χρονική στιγμή, φαίνεται να έχει τη μορφή που συναντάμε σε χρονοσειρές χρηματοοικονομικών δεικτών.



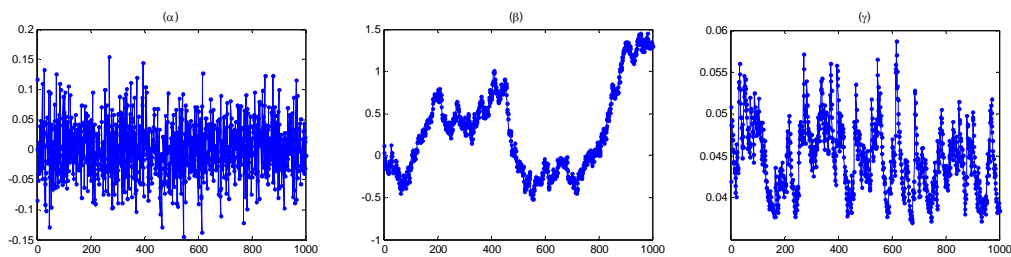
**Σχήμα 48 (α)** Χρονοσειρά  $\{x_t\}_{t=1}^{1000}$  από στοχαστική διαδικασία ARCH(1)  $x_t = \sigma_t z_t$  με  $\sigma_t^2 = 0.45 + 0.55x_{t-1}^2$  και  $z_t \sim N(0, 1)$ . **(β)** Χρονοσειρά του προφίλ  $\{y_t\}_{t=1}^{1000}$  της χρονοσειράς στο (α). **(γ)** Χρονοσειρά του  $\{\sigma_t\}_{t=1}^{1000}$  για τη χρονοσειρά στο (α).

Η στιγμιαία διασπορά ή τυπική απόκλιση για μια ARCH(1) χρονοσειρά, όπως αυτή στο Σχήμα 48γ, φαίνεται να αλλάζει ραγδαία. Αν θέλουμε να πετύχουμε πιο ομαλές αλλαγές της διασποράς θα πρέπει να αυξήσουμε σημαντικά την τάξη  $p$  του ARCH( $p$ ). Αντί για αυτό μπορούμε να θεωρήσουμε εκτός από την εξάρτηση της στιγμιαίας (ή δεσμευμένης) διασποράς  $\sigma_t^2$  στις  $q$  τελευταίες τετραγωνικές τιμές της  $\{x_t\}$  και την εξάρτηση της σε  $p$  προηγούμενες στιγμιαίες διασπορές. Έτσι προκύπτει το **γενικευμένο αυτοπαλινδρομούμενο μοντέλο με δεσμευμένη ετεροσκεδαστικότητα** (generalized ARCH, GARCH). Η στοχαστική διαδικασία τύπου GARCH δίνεται όπως για την ARCH αλλά η διασπορά ορίζεται ως

$$\sigma_t^2 = \alpha_0 + \alpha_1 x_{t-1}^2 + \dots + \alpha_q x_{t-q}^2 + \beta_1 \sigma_{t-1}^2 + \dots + \beta_p \sigma_{t-p}^2, \quad (48)$$

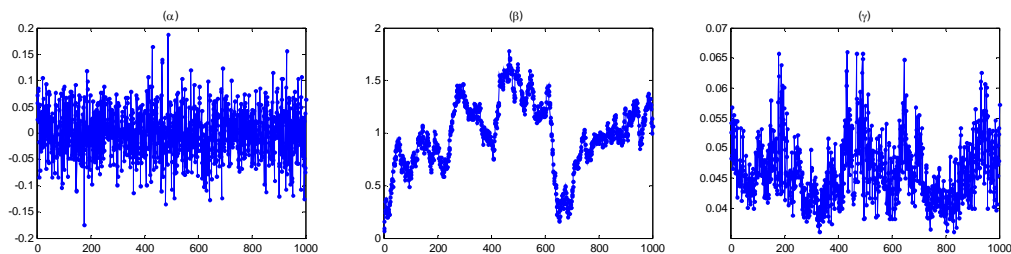
και συμβολίζεται GARCH( $p,q$ ). Αν για το ARCH χρειάζεται μεγάλη τάξη  $q$ , εισάγοντας την εξάρτηση της  $\sigma_t^2$  σε προηγούμενη(ες) δεσμευμένες διασπορές (μικρό  $p$ ) ελαττώνει σημαντικά την τάξη  $q$ . Πιο συνήθεις στα χρηματο-οικονομικά είναι το σύστημα GARCH(1,1),  $\sigma_t^2 = \alpha_0 + \alpha_1 x_{t-1}^2 + \beta_1 \sigma_{t-1}^2$ .

Η αντιστοιχία ARCH( $q$ ) για μεγάλο  $q$  με GARCH( $p,q$ ) για μικρά  $p$  και  $q$  φαίνεται στο παρακάτω παράδειγμα. Στο Σχήμα 49 δίνονται οι χρονοσειρές  $\{x_t\}$ ,  $\{y_t\}$  και  $\{\sigma_t\}$  που δημιουργήθηκαν από ένα μοντέλο GARCH(1,1). Παρατηρούμε ότι η διασπορά αλλάζει με κάποια ομαλότητα που οφείλεται στην ισχυρή εξάρτηση του  $\sigma_t^2$  στο  $\sigma_{t-1}^2$ . Για να πετύχουμε τέτοια ομαλότητα με μοντέλο ARCH θα πρέπει να επιλέξουμε μεγάλη τάξη  $q$ . Προσαρμόσαμε στη GARCH(1,1) χρονοσειρά μοντέλο τύπου ARCH. Για να πετύχουμε ισχυρά συσχετισμένη δεσμευμένη διασπορά, έπρεπε να αυξήσουμε αρκετά την τάξη του ARCH. Στο Σχήμα 50 δίνονται οι χρονοσειρές  $\{x_t\}$ ,  $\{y_t\}$  και  $\{\sigma_t\}$  που δημιουργήθηκαν από το προσαρμοσμένο μοντέλο ARCH(25).



**Σχήμα 49 (α)** Χρονοσειρά  $\{x_t\}_{t=1}^{1000}$  από στοχαστική διαδικασία GARCH(1,1)

$x_t = \sigma_t z_t$  με  $\sigma_t^2 = 0.0001 + 0.05x_{t-1}^2 + 0.9\sigma_{t-1}^2$  και  $z_t \sim N(0,1)$ . **(β)** Χρονοσειρά του προφίλ  $\{y_t\}_{t=1}^{1000}$  της χρονοσειράς στο (α). **(γ)** Χρονοσειρά του  $\{\sigma_t\}_{t=1}^{1000}$  για τη χρονοσειρά στο (α).



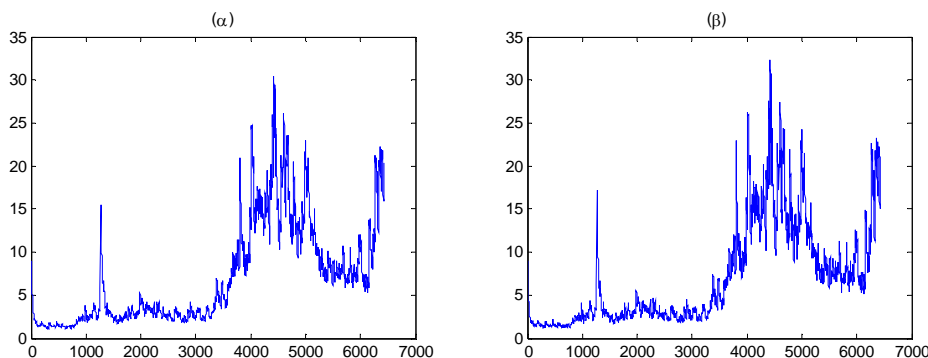
**Σχήμα 50 (α)** Χρονοσειρά  $\{x_t\}_{t=1}^{1000}$  από στοχαστική διαδικασία ARCH(25) με

συντελεστές που επιλέχτηκαν από προσαρμογή στη GARCH(1,1) χρονοσειρά στο Σχήμα 49. **(β)** Χρονοσειρά του προφίλ  $\{y_t\}_{t=1}^{1000}$  της χρονοσειράς στο (α). **(γ)** Χρονοσειρά του  $\{\sigma_t\}_{t=1}^{1000}$  για τη χρονοσειρά στο (α).

Δε θα αναφερθούμε στη διαδικασία προσαρμογής μοντέλων ARCH και GARCH σε χρονοσειρές, αλλά θα ολοκληρώσουμε την παρουσίαση των μοντέλων αυτών με την εφαρμογή τους στο δείκτη S&P500.

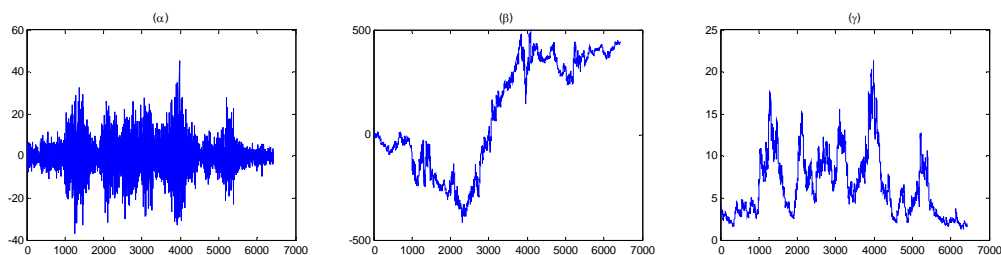
Παράδειγμα: Προσαρμογή GARCH μοντέλου στο δείκτη S&P500

Όπως φαίνεται στο Σχήμα 4 η χρονοσειρά των μεταβολών του δείκτη S&P500 παρουσιάζει αυξομειώσεις με συσταδοποίηση της διασποράς. Αυτό ενισχύει την υπόθεση ότι η δεσμευμένη διασπορά εξαρτάται και μάλιστα σημαντικά από προηγούμενες τιμές της απόδοσης (τετραγωνικές) αλλά και της διασποράς. Για αυτό προσαρμόζουμε μοντέλο GARCH(1,1) στη χρονοσειρά  $\{x_t\}$  των πρώτων διαφορών του δείκτη S&P500. Η εκτίμηση της στιγμιαίας διασποράς από το GARCH(1,1) δίνει εξηγεί τις αυξομειώσεις στις τιμές των πρώτων διαφορών, όπως φαίνεται από τη σύγκριση των διαγραμμάτων στο Σχήμα 51 και στο Σχήμα 4α. Μάλιστα η εκτίμηση αυτή δεν αλλάζει με την κατανομή της iid μεταβλητή  $z_t$ , όπως φαίνεται στο Σχήμα 51, όπου δίνεται η χρονοσειρά της εκτίμησης της στιγμιαίας διασποράς υποθέτοντας Γκαουσιανή κατανομή και κατανομή Student για την iid μεταβλητή  $z_t$ .



**Σχήμα 51 (α)** Χρονοσειρά της εκτίμησης του  $\{\sigma_t\}_{t=1}^N$  ( $N=6423$ ) των πρώτων διαφορών του δείκτη S&P500 την περίοδο 20/10/1982 – 13/3/2008 από μοντέλο GARCH(1,1) υποθέτοντας Γκαουσιανή κατανομή για την τυχαία μεταβλητή  $z_t$ . **(β)** Το ίδιο όπως στο (α) αλλά όταν το  $z_t$  ακολουθεί κατανομή student με 5 βαθμούς ελευθερίας.

Το εκτιμώμενο μοντέλο GARCH(1,1) με Γκαουσιανό θόρυβο έχει τους συντελεστές  $\alpha_0=0.0282$ ,  $\alpha_1=0.0595$  και  $\beta_1=0.9405$ , δηλαδή έχει ισχυρή εξάρτηση της διασποράς σε υστέρηση ένα. Το ίδιο ισχύει όταν το  $z_t$  ακολουθεί Student κατανομή. Ενδεικτικά δημιουργούμε πραγματοποιήσεις με το παραπάνω μοντέλο και οι χρονοσειρές  $\{x_t\}$ ,  $\{y_t\}$  και  $\{\sigma_t\}$  δίνονται στο Σχήμα 52. Η χρονοσειρά προσομοίωσης του δείκτη  $\{y_t\}$  καθώς και των πρώτων διαφορών  $\{x_t\}$  φαίνεται να έχουν παρόμοια μορφή με τις αντίστοιχες χρονοσειρές του δείκτη S&P 500.



**Σχήμα 52 (α)** Χρονοσειρά  $\{x_t\}_{t=1}^{6423}$  από το προσαρμόσμενο GARCH(1,1) στις πρώτες διαφορές του δείκτη S&P 500. **(β)** Χρονοσειρά του προφίλ  $\{y_t\}_{t=1}^{6423}$  της χρονοσειράς στο (α). **(γ)** Χρονοσειρά του  $\{\sigma_t\}_{t=1}^{6423}$  για τη χρονοσειρά στο (α).

Με τα μοντέλα τύπου GARCH ολοκληρώνεται η παρουσίαση των μοντέλων πρόβλεψης χρηματο-οικονομικών δεικτών.

## Ασκήσεις Κεφαλαίου 4

1. Δημιουργείστε μια συνάρτηση που θα υπολογίζει τη μερική αυτοσυσχέτιση σε μια χρονοσειρά. Ως είσοδο η συνάρτηση θα παίρνει το διάνυσμα της χρονοσειράς και την τιμή της μέγιστης τάξης (υστέρησης). Ως έξοδο θα δίνει ένα διάνυσμα με τις τιμές της μερικής αυτοσυσχέτισης για κάθε τάξη ως τη μέγιστη τάξη. Δοκιμάστε τη συνάρτηση αυτή σε χρονοσειρές από AR διαδικασία για διαφορετικές τάξεις.

Βοήθεια matlab: Η μερική αυτοσυσχέτιση για κάθε τάξη  $p$  είναι ο συντελεστής του τελευταίου όρου του  $AR(p)$ . Για την εκτίμηση των παραμέτρων του  $AR(p)$  μπορείτε να χρησιμοποιήσετε τις συναρτήσεις `ar` για τον υπολογισμό του μοντέλου και `polydata` για την εξαγωγή του πολυωνύμου (διανύσματος) από τη δομή του μοντέλου που δίνει η συνάρτηση `ar`. Για το σχηματισμό χρονοσειρών από AR διαδικασία, χρησιμοποιείστε τη συνάρτηση `AR.m` που δίνεται στην ιστοσελίδα του μαθήματος.

2. Ένας έλεγχος για την ανεξαρτησία χρονοσειράς (τυχαία σειρά παρατηρήσεων) είναι από την συνάρτηση αυτοσυσχέτισης, δηλαδή ελέγχοντας αν η αυτοσυσχέτιση  $r(\tau)$  για κάθε υστέρηση  $\tau$ , από ένα ως κάποια μέγιστη υστέρηση  $\tau_{\max}$ , είναι στατιστικά ασήμαντη, δηλαδή είναι μέσα στα όρια  $\pm 1.96/\sqrt{N}$  (για επίπεδο εμπιστοσύνης 95%), όπου  $N$  είναι το μήκος της χρονοσειράς. Έχει βρεθεί από τους Ljung και Box ότι το παρακάτω άθροισμα των τετραγώνων των αυτοσυσχετίσεων

$$Q = N(N+2) \sum_{\tau=1}^{\tau_{\max}} \frac{r(\tau)^2}{N-\tau}$$

ακολουθεί κατανομή  $\chi^2$  με  $\tau_{\max}$  βαθμούς ελευθερίας. Άρα μπορούμε να κάνουμε έλεγχο υπόθεσης με μηδενική υπόθεση ότι η χρονοσειρά είναι τυχαία (έχει μηδενική αυτοσυσχέτιση για όλες τις υστερήσεις). Υπολογίζουμε το  $Q$  από τη χρονοσειρά και απορρίπτουμε αυτή τη μηδενική υπόθεση σε επίπεδο σημαντικότητας  $\alpha$  (δηλαδή σε  $1-\alpha$  επίπεδο εμπιστοσύνης) αν  $Q > \chi_{1-\alpha, \tau_{\max}}^2$ , όπου  $\chi_{1-\alpha, \tau_{\max}}^2$  είναι η αντίστοιχη κρίσιμη τιμή (το όριο της ουράς της κατανομής) για πιθανότητα  $1-\alpha$  και  $\tau_{\max}$  βαθμούς ελευθερίας. Αντίστοιχα μπορεί κάποιος να υπολογίσει την  $p$ -τιμή του ελέγχου ως το  $\alpha$  που δίνει ως κρίσιμη τιμή  $\chi_{1-\alpha, \tau_{\max}}^2$  την τιμή της  $Q$ .

Δημιουργείστε μια συνάρτηση που κάνει αυτόν τον έλεγχο. Η συνάρτηση θα πρέπει να παίρνει ως είσοδο τη χρονοσειρά και τη μέγιστη υστέρηση και να δίνει ως έξοδο την  $p$ -τιμή του ελέγχου Ljung-Box. Δοκιμάστε τη συνάρτηση αυτή σε χρονοσειρές από AR διαδικασία κάποιας τάξης  $p$  εφαρμόζοντας μοντέλα AR μικρότερης τάξης και τάξης  $p$ .

Βοήθεια matlab: Για τον υπολογισμό της κρίσιμης τιμής της  $\chi^2$  με  $\tau_{\max}$  βαθμούς ελευθερίας χρησιμοποίησε τη συνάρτηση `chi2inv` και για την  $p$ -τιμή τη συνάρτηση `chi2cdf`.

3. Δίνεται η χρονοσειρά της τιμής κλεισίματος του δείκτη S&P500 για την περίοδο 20/10/1982 – 13/3/2008. Δημιουργείστε τη χρονοσειρά των αποδόσεων. Ελέγξτε αν η χρονοσειρά αυτή είναι ανεξάρτητη χρησιμοποιώντας τον έλεγχο Ljung-Box (δες Άσκηση 2). Αν δεν προκύπτει ανεξαρτησία, κάνετε προσαρμογή κατάλληλου AR μοντέλου. Επιλέξτε την τάξη του AR μοντέλου από τη μερική αυτοσυσχέτιση (δες Άσκηση 1) και το κριτήριο AIC.

Βοήθεια matlab: Για τον υπολογισμό του κριτηρίου AIC μπορείς να χρησιμοποιήσεις τη συνάρτηση `aic` που παίρνει ως όρισμα τη δομή του μοντέλου που υπολογίζει η συνάρτηση `ar` και δίνει την τιμή της συνάρτησης AIC για το συγκεκριμένο μοντέλο. Η χρονοσειρά του SP500 δίνεται στην ιστοσελίδα του μαθήματος.

4. Κάνετε την ίδια διαδικασία όπως στην Άσκηση 3 για τη χρονοσειρά των πρώτων διαφορών που θα δημιουργήσετε από τη χρονοσειρά του δείκτη της μηνιαίας τιμής συναλλάγματος βρετανικής λίρας σε δολλάριο (GBR/USD) για την περίοδο από Ιανουάριο 1971 ως Αύγουστο 2000. Η χρονοσειρά αυτού του δείκτη δίνεται στην ιστοσελίδα του μαθήματος (`gbpusdmonthly.dat`).