Προσαρμογή AR μοντέλου

$x_t = \phi_0 + \phi_1 x_{t-1} + \cdots + \phi_p x_{t-p} + z_t$ τάξη *p*, εκτίμηση παραμέτρων $\phi_1, \ldots, \phi_p, \sigma_z^2$

Προσδιορισμός τάξης AR μοντέλου

- μερική αυτοσυσχέτιση για υστέρηση *τ*: συσχέτιση των x_t, $x_{t-\tau}$ χωρίς τη συσχέτιση με $\underbrace{x_{t-1}, \ldots, x_{t-\tau+1}}$

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

$x_t = \phi_{1,2} x_{t-1} + \phi_{2,2} x_{t-2} + w_t$ εκτίμηση του ϕ_τ για μοντέλο AR(*τ*)

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

Η τάξη είναι *p* αν $\hat{\phi}_{p,p} \neq 0$ και $\hat{\phi}_{k,k} = 0$ για *k*>*p* (πτώση από μη-μηδενική σε μηδενική μερική αυτοσυσχέτιση)

- κριτήρια πληροφορίας

κριτήριο πληροφορίας του Akaike, AIC

διασπορά σφάλματος προσαρμογής

$$\mathrm{AIC}(p) = \ln(s_z^2) + \frac{2p}{N}$$

Εκτίμηση παραμέτρων

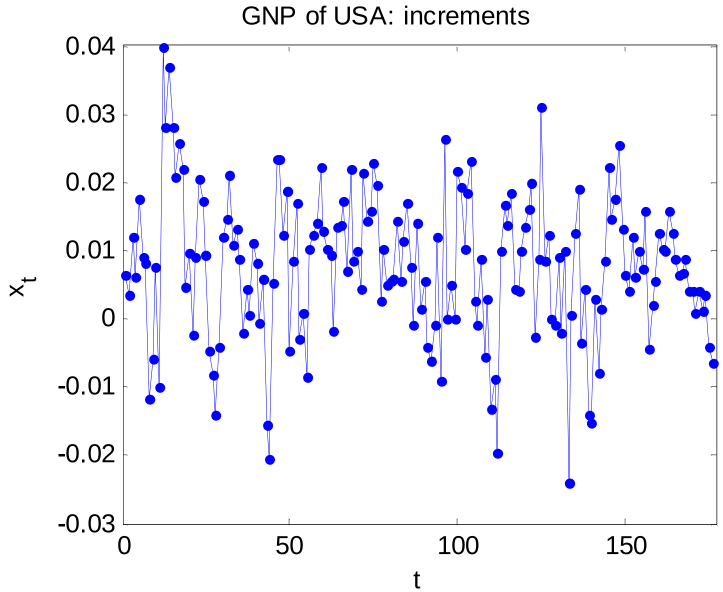
- μέθοδο ελαχίστων τετραγώνων (ordinary least squares, OLS)

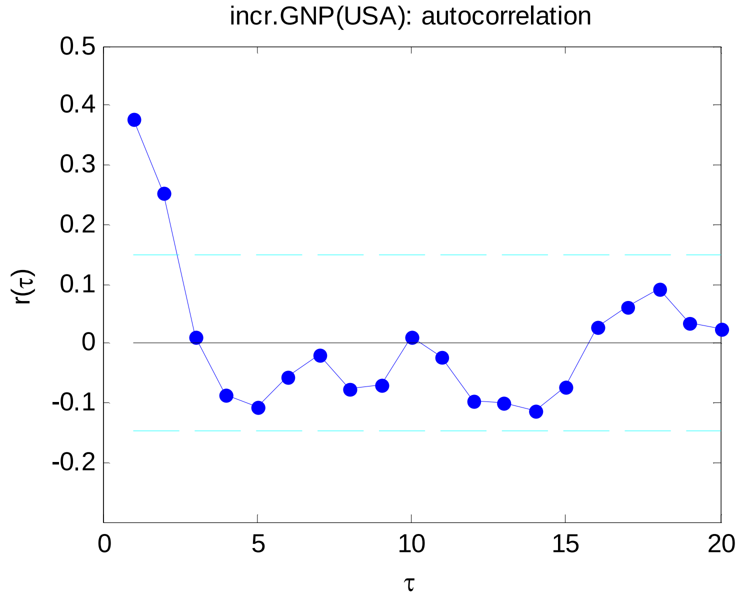
$$\min_{\phi_1,\ldots,\phi_p} \sum_{t=p+1}^{N} \left(y_t - \phi_1 x_{t-1} - \cdots - \phi_p x_{t-p} \right)^2$$

- μέθοδος μέγιστης πιθανοφάνειας (maximum likelihood method)

Παράδειγμα

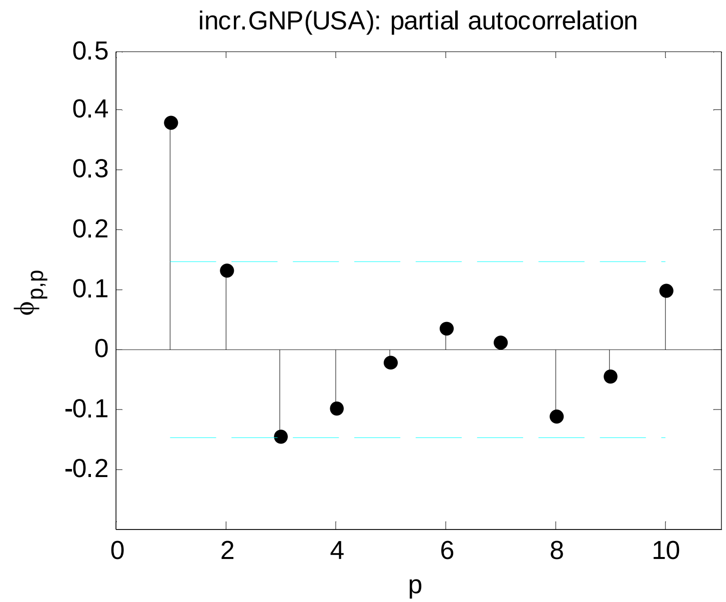
Ρυθμός μεταβολής του ακαθάριστου εθνικού προϊόντος (ΑΕΠ) των ΗΠΑ (τετραμηνιαίες τιμές, 2ο τετράμηνο 1947 – 1ο τετράμηνο 1991). Η εποχικότητα έχει διορθωθεί (αφαιρώντας τον εποχικό κύκλο).

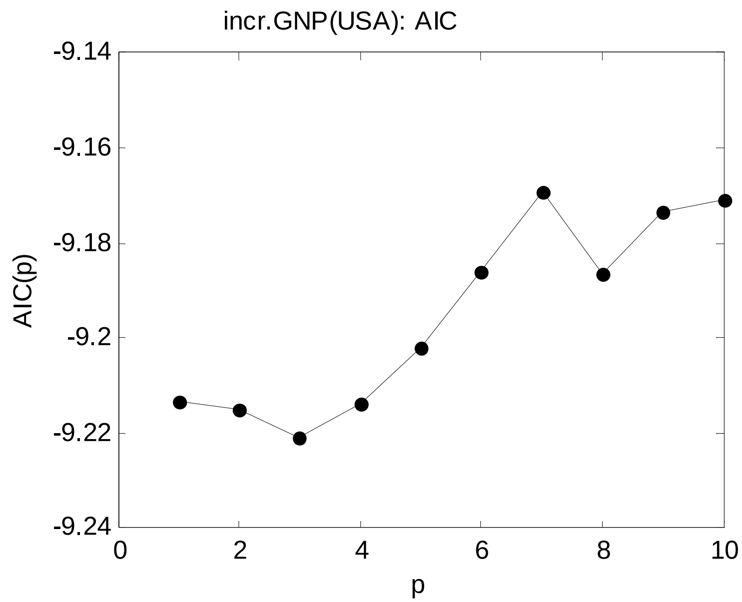




στάσιμη ?

συσχετίσεις μικρής διάρκειας ?





τάξη AR μοντέλου ?

AR(3) ?

εκτίμηση παραμέτρων

$$x_t - \hat{\mu} \to x_t \qquad \hat{\mu} = 0.0077$$

OLS → $\hat{\phi}_1 = 0.35 \qquad \hat{\phi}_2 = 0.18 \qquad \hat{\phi}_3 = -0.14$

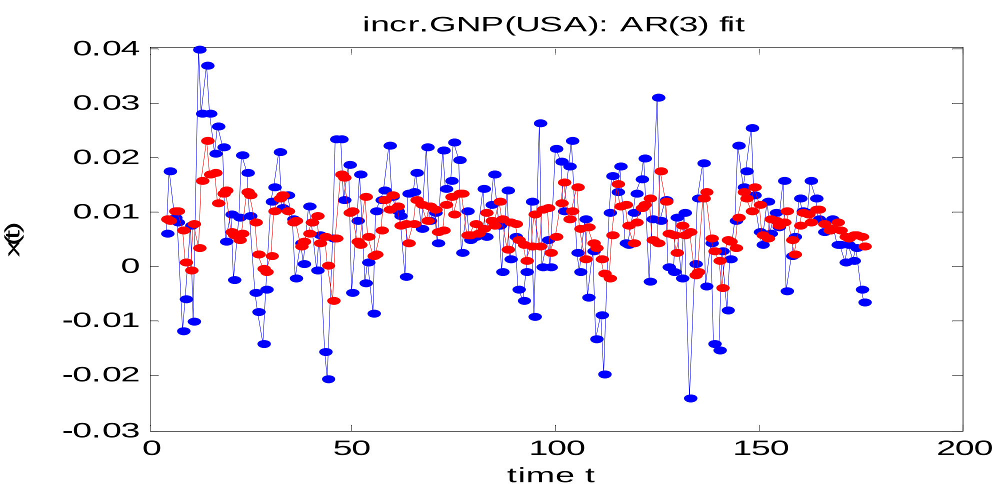
$$\hat{\phi}_0 = \hat{\mu}\left(1 - \hat{\phi}_1 - \hat{\phi}_2 - \hat{\phi}_3\right) = 0.0047$$

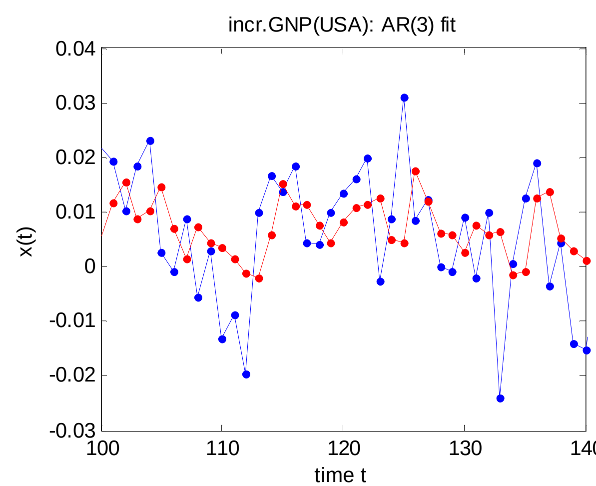
εκτίμηση $\hat{x}_t = 0.0047 + 0.35x_{t-1} + 0.18x_{t-2} - 0.14x_{t-3} \qquad t = 4,\ldots,176$

σφάλματα ή υπόλοιπα (residual) εκτίμησης $\hat{z}_t = x_t - \hat{x}_t$

$$s_z^2 = \hat{\sigma}_z^2 = 0.0000989$$
$$s_z = \hat{\sigma}_z = 0.0098$$

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





Διάγνωση καταλληλότητας

είναι τα υπόλοιπα ανεξάρτητα → έλεγχο ανεξαρτησίας στα $\left\{\hat{z}_t\right\}_{t=p+1}^{N}$

Πρόβλεψη με AR μοντέλα

$$\{x_1, x_2, \ldots, x_n\} \xrightarrow{\quad\text{πρόβλεψη}\quad} \begin{array}{l} x_{n+1}, x_{n+2}, \ldots \\ \hat{x}_{n+1}, \hat{x}_{n+2}, \ldots \\ x_n(1), x_n(2), \ldots \end{array}$$

πρόβλεψης T βημάτων μπροστά
(*T*-step ahead forecast) $\quad x_n(T) = \mathrm{E}\left[x_{n+T} \mid x_n, x_{n-1}, \ldots\right]$

AR(*p*) $\quad x_{n+1} = \phi_0 + \phi_1 x_n + \cdots + \phi_p x_{n+1-p} + z_{n+1}$

$T = 1 \qquad x_n(1) = \mathrm{E}\left[x_{n+1} \mid x_n, x_{n-1}, \ldots, x_{n+1-p}\right] = \phi_0 + \phi_1 x_n + \cdots + \phi_p x_{n+1-p}$

Σφάλμα πρόβλεψης $\quad e_n(1) = x_{n+1} - x_n(1) = z_{n+1} \qquad \mathrm{Var}\left[e_n(1)\right] = \sigma_z^2$

Αν $z_t \sim \mathrm{N}(0, \sigma_z^2)$ → 95% διάστημα πρόβλεψης $\quad x_n(1) \pm 1.96\sigma_w$

$T = 2 \qquad x_n(2) = \phi_0 + \phi_1 x_n(1) + \phi_2 x_n \cdots + \phi_p x_{n+2-p}$

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

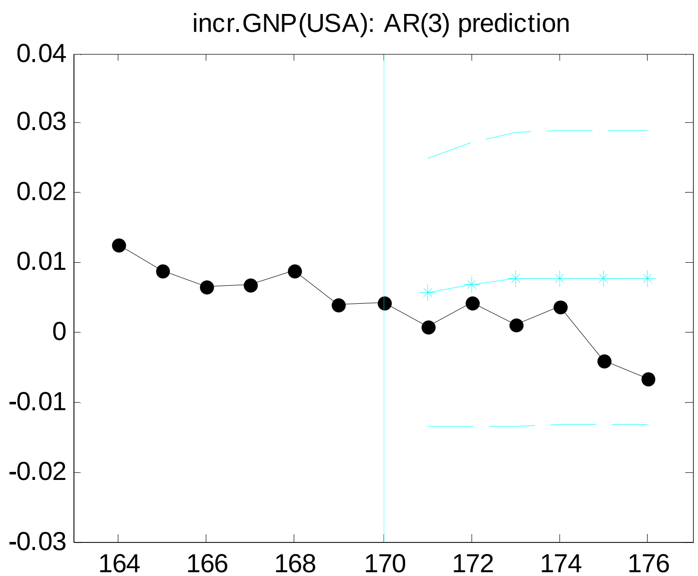
$$\mathrm{Var}\left[e_n(2)\right] = \left(1 + \phi_1^2\right)\sigma_w^2$$

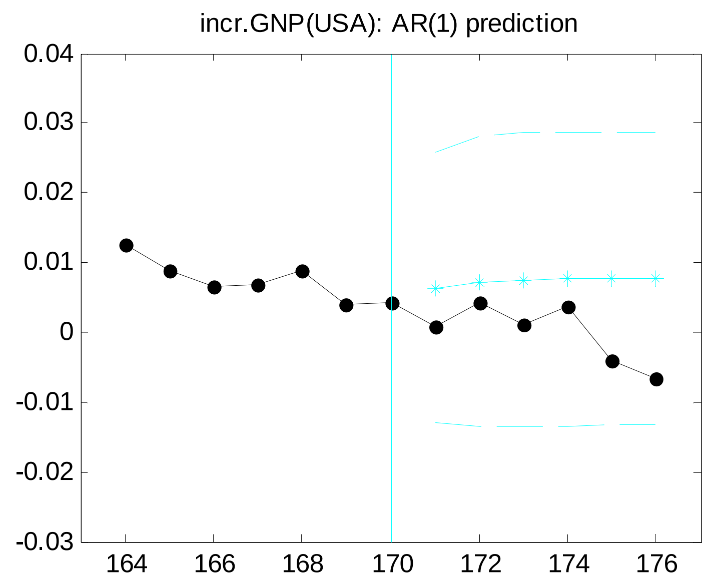
$T \to \infty \qquad x_n(T) \to \mu$

Παράδειγμα μεταβολή ΑΕΠ των ΗΠΑ

$$\{x_1, x_2, \ldots, x_{170}\} \longrightarrow x_{171}, x_{172}, \ldots, x_{176}$$

προσαρμόζουμε το AR(3) $\longrightarrow x_{170}(1), x_{170}(2), \ldots, x_{170}(6)$





Αξιολόγηση της επίδοσης μοντέλου σε πρόβλεψη

- αντεπικύρωση (cross validation)

- *σύνολο εκμάθησης ή εκπαίδευσης* (training ή learning set) $x_1, x_2, \ldots x_{N_1}$

  σύνολο ελέγχου ή *επικύρωσης* (test ή validation set) $x_{N_1+1}, x_{N_2+2}, \ldots x_N$

προβλέψεις $x_{N_1}(T), x_{N_1+1}(T), \ldots, x_{N-T}(T)$

σφάλματα πρόβλεψης *T*-βημάτων

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

μέσο τετραγωνικό σφάλμα
(mean square error, **mse**)

$$\mathrm{mse}(T) = \frac{1}{n-n_1-T+1}\sum_{j=N_1}^{N-T} e_j(T)^2 = \frac{1}{N-N_1-T+1}\sum_{j=N_1}^{N-T}\left(x_{j+T} - x_j(T)\right)^2$$

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

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

κανονοικοποίηση του rmse
(normalized root mean square error, **nrmse**)

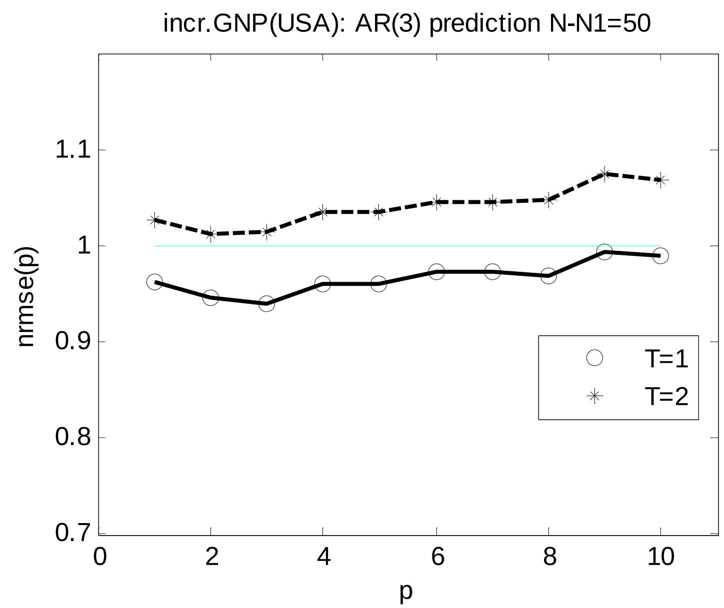
$$\mathrm{nrmse}(T) = \frac{\sqrt{\frac{1}{N-N_1-T+1}\sum_{j=N_1}^{N-T}\left(x_{j+T} - x_j(T)\right)^2}}{\sqrt{\frac{1}{N-N_1-T+1}\sum_{j=N_1}^{N-T}\left(x_{j+T} - \bar{x}\right)^2}}$$

Παράδειγμα μεταβολή ΑΕΠ των ΗΠΑ

σύνολο εκμάθησης $\{x_1, x_2, \ldots, x_{126}\}$

σύνολο ελέγχου $\{x_{127}, x_{128}, \ldots, x_{176}\}$

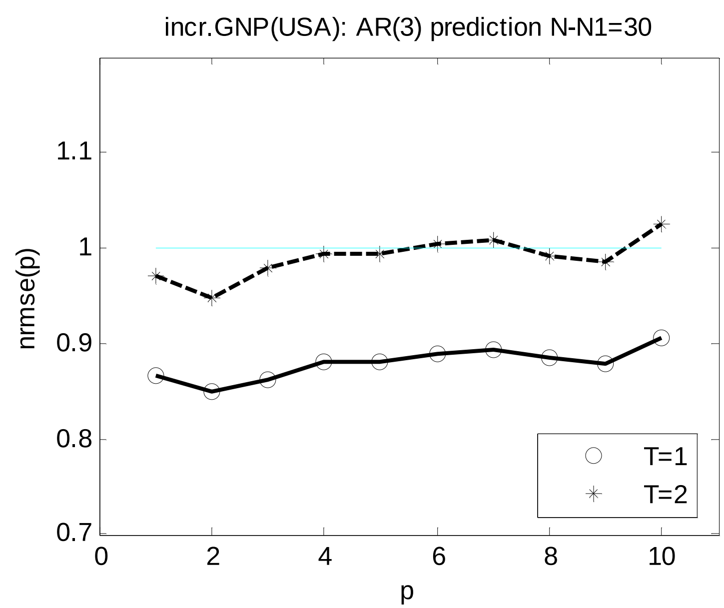
$$N - N_1 = 50$$



σύνολο εκμάθησης $\{x_1, x_2, \ldots, x_{146}\}$

σύνολο ελέγχου $\{x_{147}, x_{128}, \ldots, x_{176}\}$

$$N - N_1 = 30$$



Πρόβλεψη μη-στάσιμων χρονοσειρών με ARIMA

Μεθοδολογία Box-Jenkins

1. γράφημα χρονοσειράς και αυτοσυσχέτισης

→ *ασήμαντες αυτοσυσχετίσεις*
➔ λευκός θόρυβος
➔ κανένα γραμμικό μοντέλο

→ *ισχυρές αυτοσυσχετίσεις που φθίνουν αργά*
➔ μη-στάσιμη χρονοσειρά
➔ πρώτες διαφορές y → x

$$x_t = y_t - y_{t-1}$$

2. Επιλέγεται η τάξη του AR, ARMA μοντέλου για τη $\{x_t\}_{t=1}^{N}$

3. Προσαρμόζεται το μοντέλο AR(*p*) ή ARMA(*p*,*q*) στη $\{x_t\}_{t=1}^{N}$ και ελέγχεται η καταλληλότητα του (αν τα υπόλοιπα είναι ανεξάρτητα)

4. Με το επιλεγμένο μοντέλο γίνονται προβλέψεις της $\{x_t\}_{t=1}^{N}$ και μετασχηματίζονται στην αρχική χρονοσειρά $\{y_t\}_{t=1}^{N}$

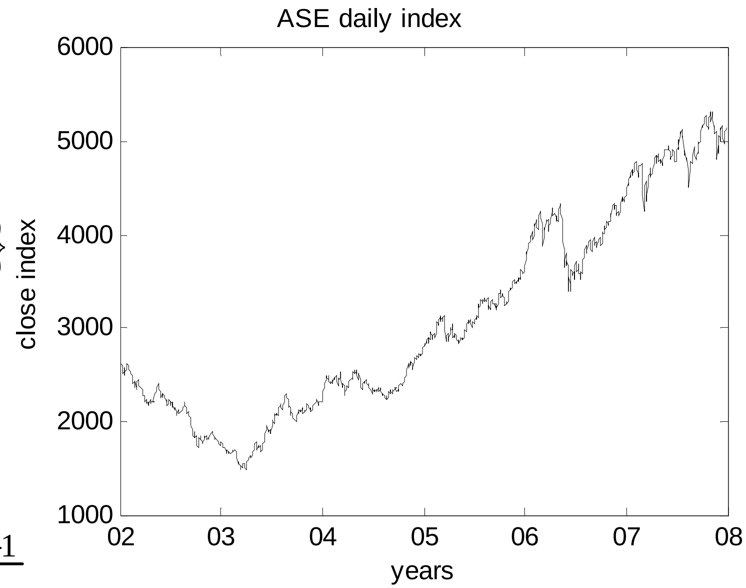
Πρόβλεψη T=1 $\quad x_n(1) \longrightarrow y_n(1) = y_n + x_n(1)$

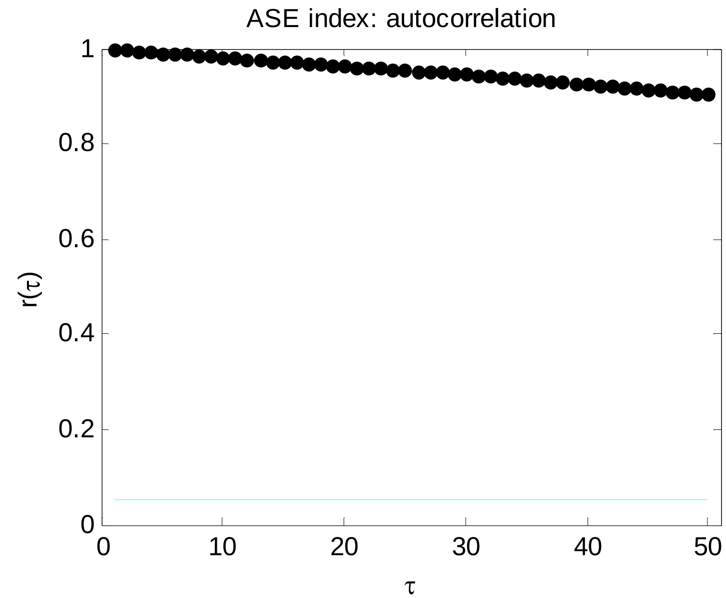
Πρόβλεψη T=2 $\quad x_n(2) \longrightarrow y_n(2) = y_n(1) + x_n(2)$

Πρόβλεψη T $\quad x_n(T) \longrightarrow y_n(T) = y_n(T-1) + x_n(T)$

Παράδειγμα Ημερήσιος δείκτης ΧΑΑ από 2/1/02 – 31/12/07, πρόβλεψη ως 13/3/08

$$\{y_0, \ldots, y_{1496}\}$$

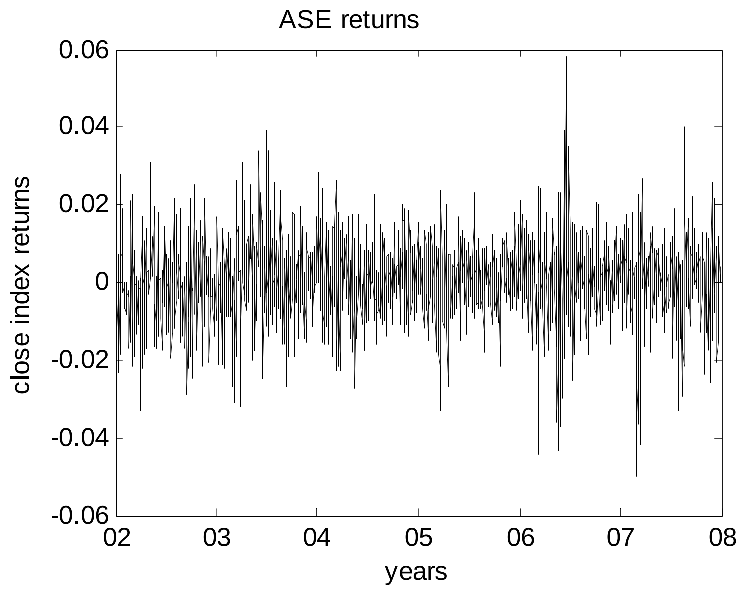


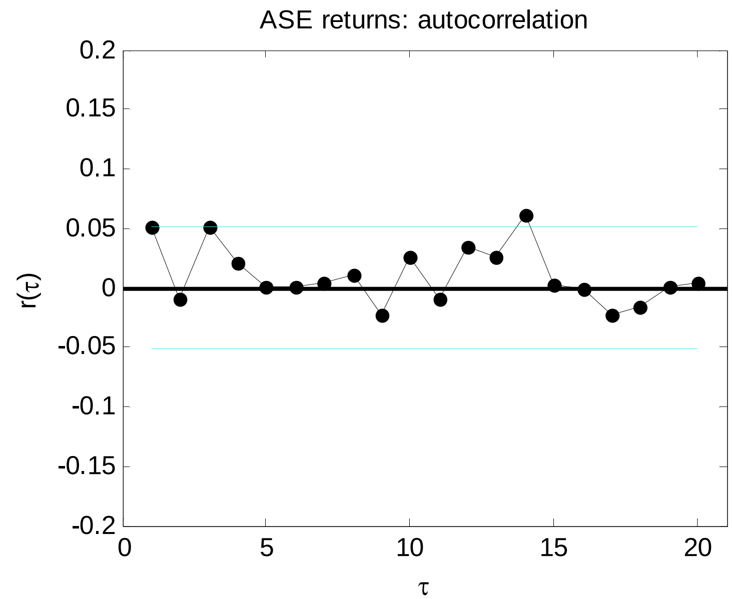


$$x_t = \frac{y_t - y_{t-1}}{y_t}$$

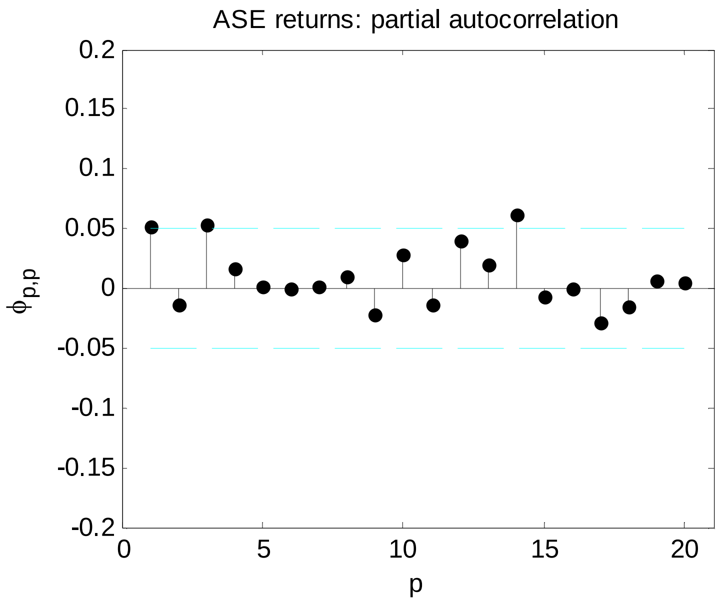
$$\{x_1, \ldots, x_{1496}\}$$

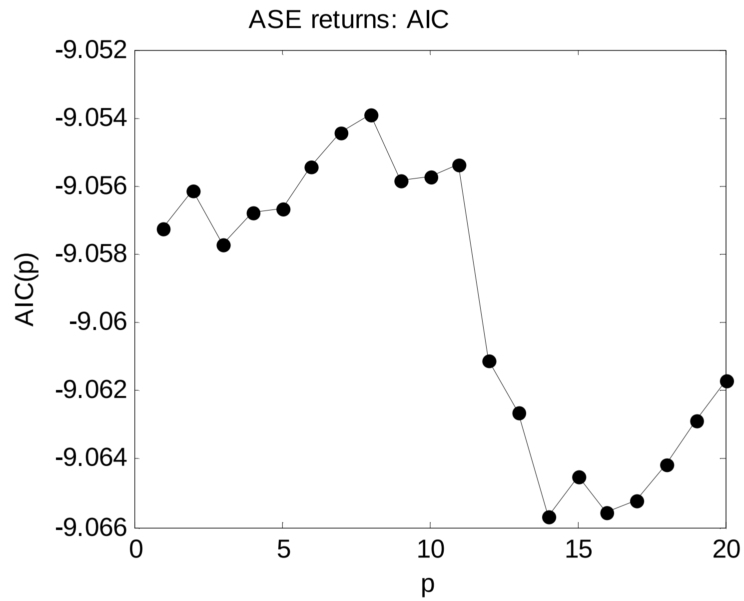
$$N_1 = 1496$$



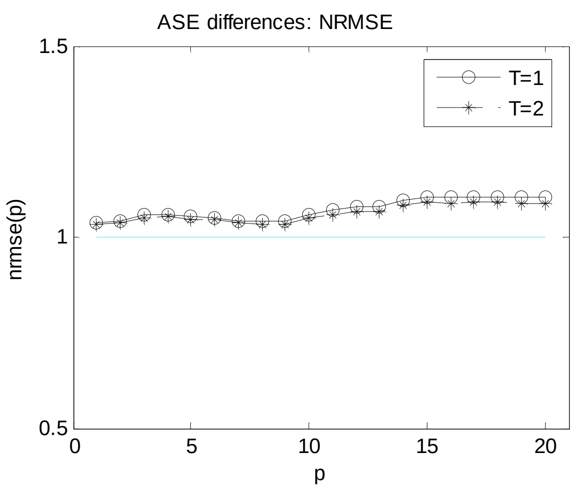


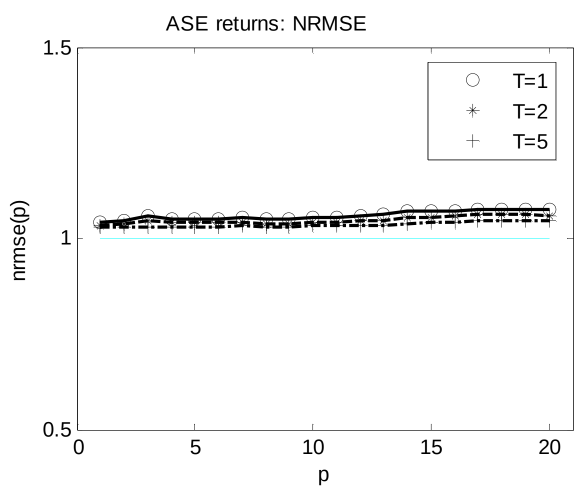
τάξη AR μοντέλου

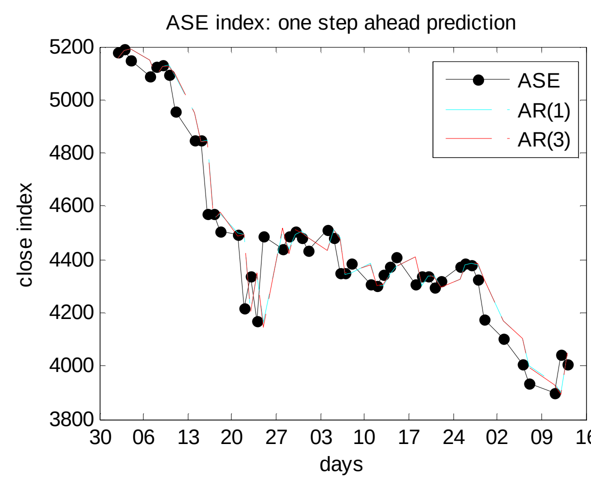




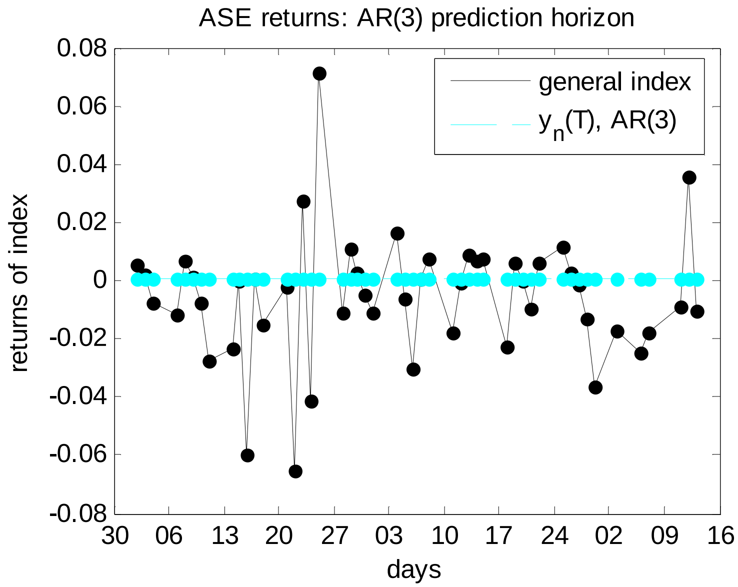
Αξιολόγηση πρόβλεψης, σύνολο ελέγχου 2/1/2008 - 13/3/08 $N = 1544$ $N - N_1 = 48$

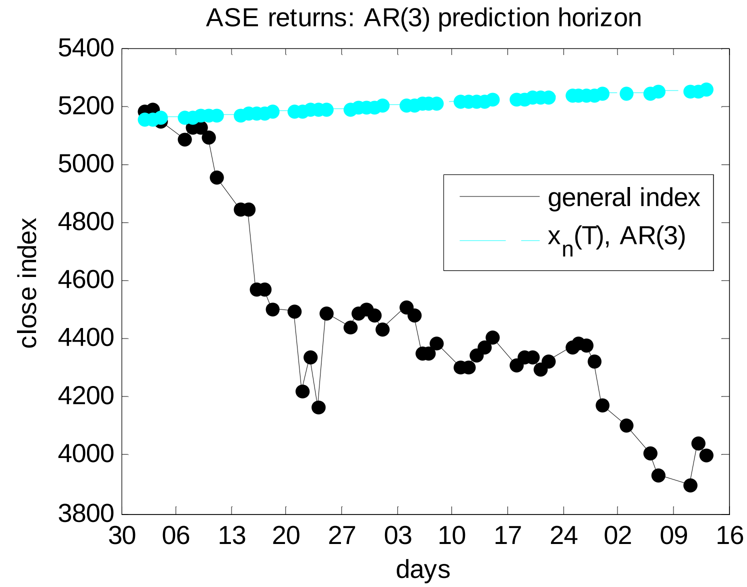






Πρόβλεψη με αφετηρία 31/12/2007 και ως 13/3/08 $n = 1496$ $T = 1, 2, \ldots, 49$





AR(1) $x_t = 0.0004 + 0.052 x_{t-1} + z_t$

AR(3) $x_t = 0.0004 + 0.053 x_{t-1} - 0.015 x_{t-2} + 0.053 x_{t-3} + z_t$

Οι συντελεστές είναι κοντά στο 0 !