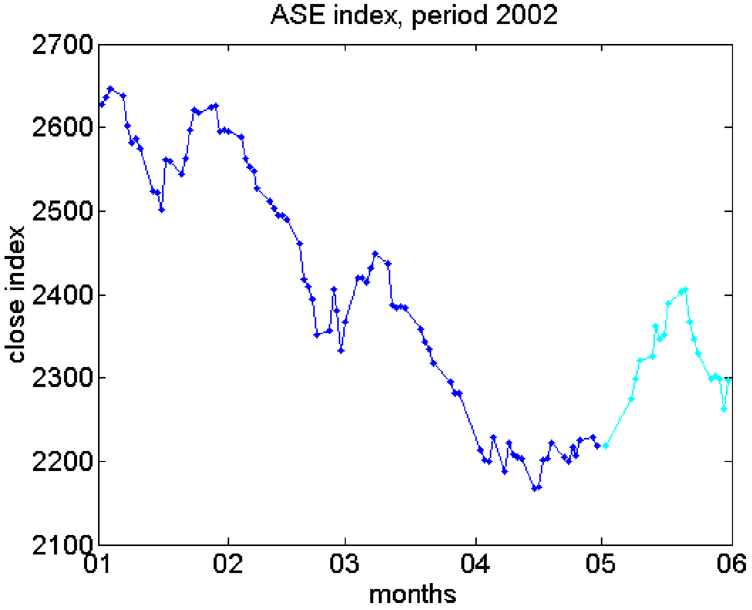
Μάθημα 4:
Πρόβλεψη χρονοσειρών

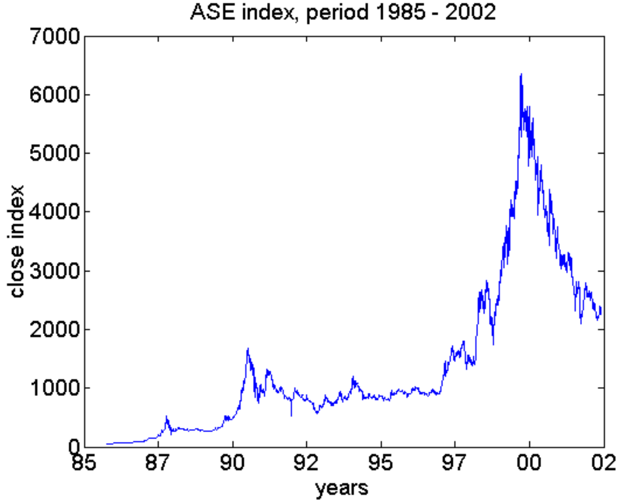
- Απλές τεχνικές πρόβλεψης
- Πρόβλεψη στάσιμων χρονοσειρών με γραμμικά μοντέλα
- Πρόβλεψη μη-στάσιμων χρονοσειρών
- Ασκήσεις

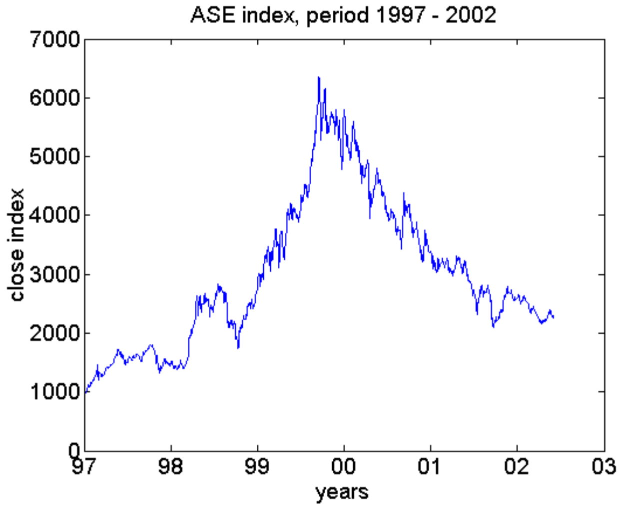


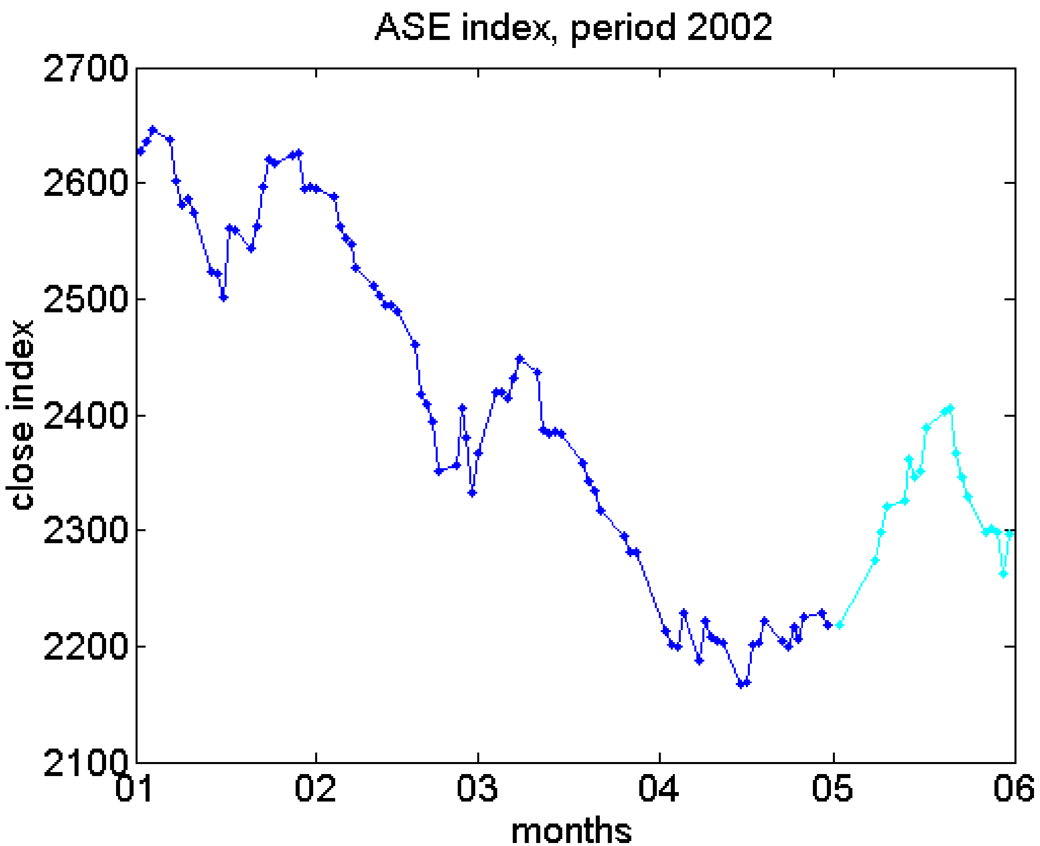
Πρόβλεψη Χρονοσειρών

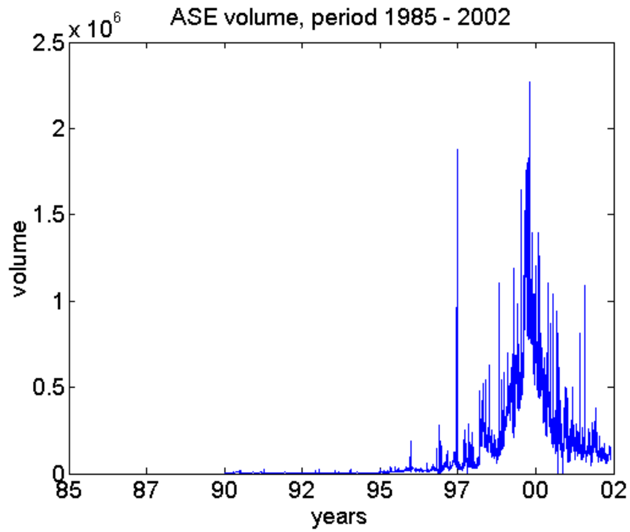
Μοντέλα για χρονικές σειρές (AR, MA, ARMA, ARIMA, SARIMA) ➔ **πρόβλεψη**
Πολλές εφαρμογές

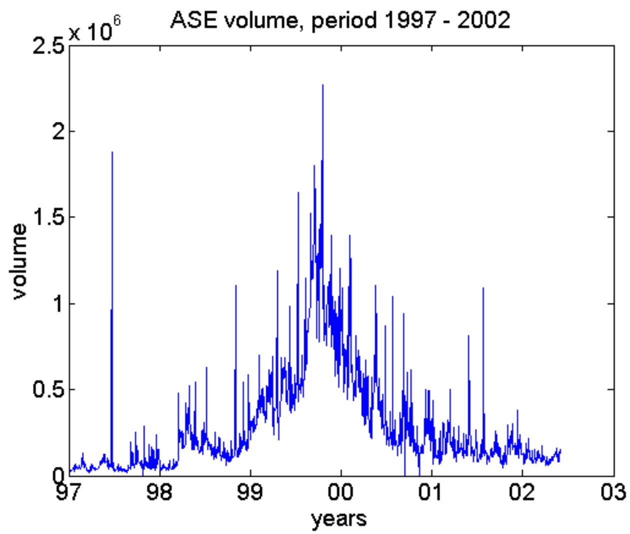
Δείκτης και όγκος συναλλαγών Χρηματιστηρίου Αξιών Αθηνών (ΧΑΑ)

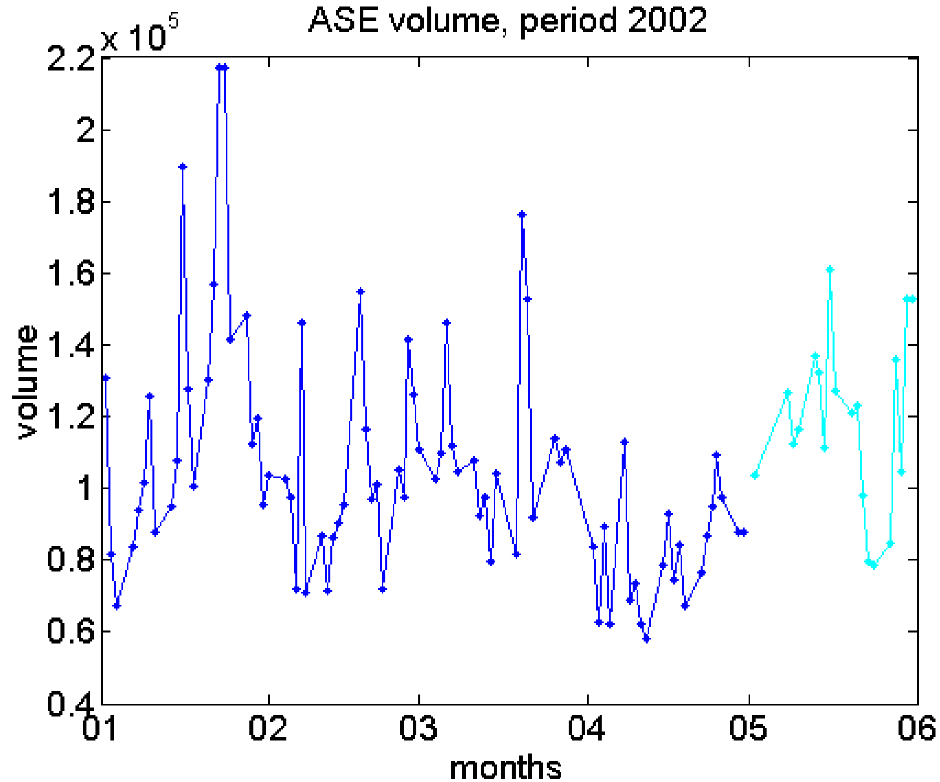






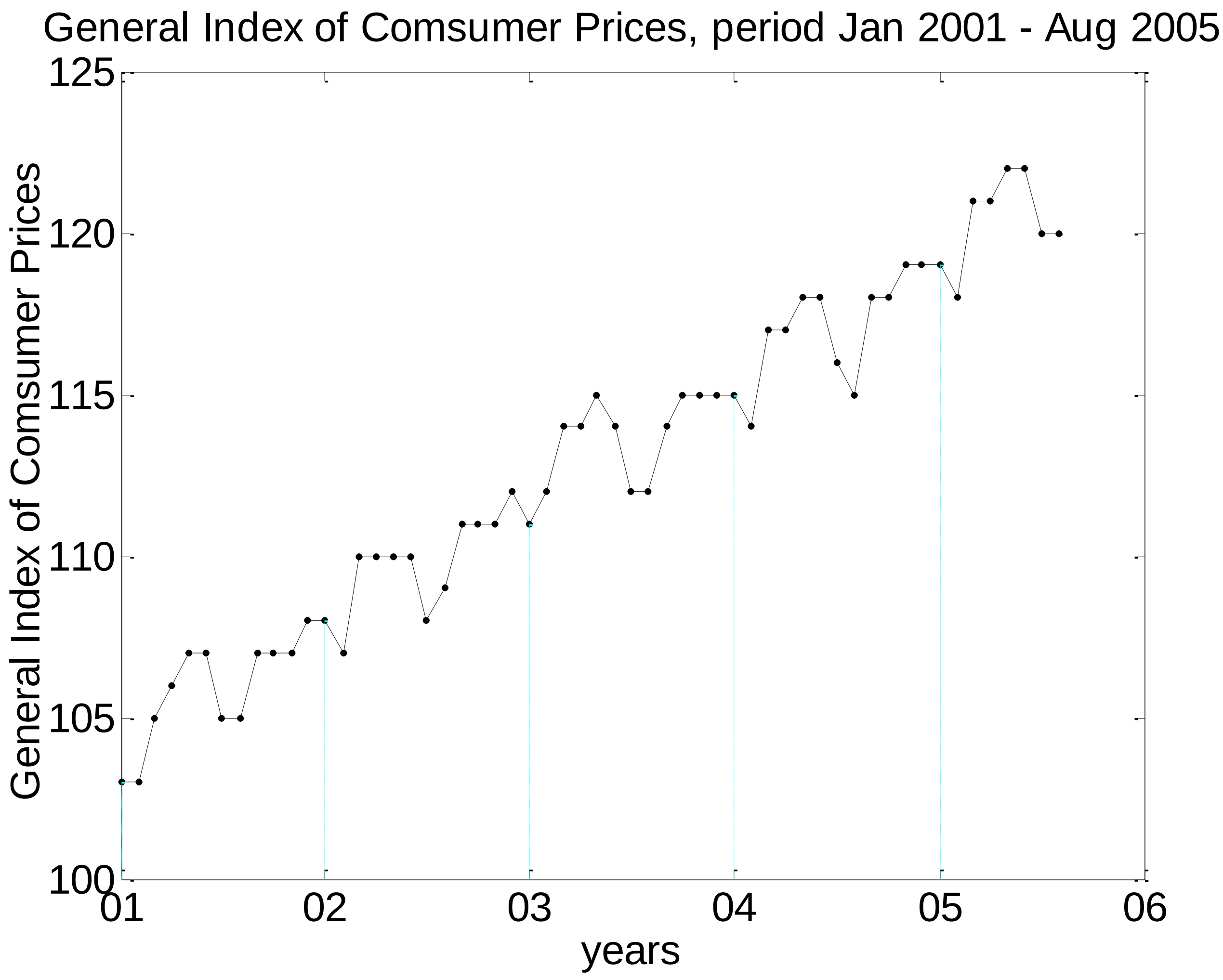






Θα μπορούσαμε να προβλέψουμε τους δύο δείκτες για το Μάιο 2002 (έστω τις πρώτες μέρες του μήνα) γνωρίζοντας τα δεδομένα ως και το τέλος Απριλίου 2002?

Γενικός δείκτης τιμών καταναλωτή (GICP)



Πως θα μεταβληθεί δείκτης GICP τους επόμενους μήνες?

Πρόβλημα πρόβλεψης

- Γνωρίζουμε τη χρονική σειρά $\{x_1, x_2, \ldots, x_n\}$
- Θέλουμε να εκτιμήσουμε το x_{n+k}

Πρόβλεψη $x_n(k)$

Σφάλμα πρόβλεψης: $e_n(k) = x_{n+k} - x_n(k)$

Στοχαστική διαδικασία $\{X_n\}$

πρόβλεψη $X_n(k)$ είναι η εκτίμηση του στοιχείου X_{n+k} της $\{X_n\}$

Βέλτιστη πρόβλεψη : $X_n(k) = \mathrm{E}[X_{n+k} \mid X_n, X_{n-1}, \ldots]$

Ιδιότητες καλής πρόβλεψης:

- η *αμεροληψία* (unbiasedness) : $\mathrm{E}[X_n(k)] = X_{n+k}$
- η *αποτελεσματικότητα* (efficiency), δηλαδή μικρή διασπορά σφάλματος

$$\mathrm{Var}[\varepsilon_n(k)] = \mathrm{Var}[X_{n+k} - X_n(k)]$$

Συνδυασμός αμεροληψίας και αποτελεσματικότητας ➔
ελαχιστοποίηση του **μέσου τετραγωνικού σφάλματος πρόβλεψης**

$$\mathrm{E}\left[\left(X_{n+k} - X_n(k)\right)^2\right]$$

Αξιολόγηση απόδοσης μοντέλου πρόβλεψης:

Γνωρίζουμε $\{x_1, x_2, \ldots, x_n\}$ σύνολο εκμάθησης learning /training set

γνωρίζουμε επίσης $\{x_{n+1}, x_{n+2}, \ldots, x_{n+l}\}$ σύνολο αξιολόγησης test / validation set

μοντέλο πρόβλεψης

$$x_n(k), x_{n+1}(k), \ldots, x_{n+l-k}(k)$$

Σφάλματα πρόβλεψης *k* χρονικά βήματα μπροστά

$$e_n(k), e_{n+1}(k), \ldots, e_{n+l-k}(k)$$

$$e_j(k) = x_{j+k} - x_j(k) \quad j = n, n+1, \ldots, n+l-k$$

Στατιστικά μέτρα σφάλματος

Εκτίμηση μέσου τετραγωνικού σφάλματος (*mean square error*, **mse**)

$$\mathrm{mse}(k) = \frac{1}{l-k+1}\sum_{j=n}^{n+l-k} e_j(k)^2 = \frac{1}{l-k+1}\sum_{j=n}^{n+l-k}\left(x_{j+k} - x_j(k)\right)^2$$

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

$$\mathrm{rmse}(k) = \sqrt{\frac{1}{l-k+1}\sum_{j=n}^{n+l-k} e_j(k)^2} = \sqrt{\frac{1}{l-k+1}\sum_{j=n}^{n+l-k}\left(x_{j+k} - x_j(k)\right)^2}$$

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

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

nrmse → 0
πολύ καλή πρόβλεψη

nrmse ≈ 1
πρόβλεψη ισοδύναμη με μέση τιμή

Προβλέψεις:

Πρόβλεψη πολλών βημάτων μπροστά για μια χρονική στιγμή

Δίνεται $\{x_1, x_2, \ldots, x_n\}$, προβλέπουμε $x_n(1), x_n(2), \ldots, x_n(k)$

Πρόβλεψη για κάποιο βήμα μπροστά για πολλές χρονικές στιγμές

Δίνεται $\{x_1, x_2, \ldots, x_n, x_{n+1}, \ldots, x_{n+l}\}$, θέλουμε να μετρήσουμε την απόδοση ενός μοντέλου πρόβλεψης.

1. Εκτιμούμε τις παραμέτρους του μοντέλου από τα $\{x_1, x_2, \ldots, x_n\}$

2. Κάνουμε τις προβλέψεις για κάποιο βήμα k

$$x_n(k), x_{n+1}(k), \ldots, x_{n+l-k}(k)$$

3. Υπολογίζουμε κάποιο στατιστικό μέτρο των σφαλμάτων πρόβλεψης

$$\mathrm{rmse}(k) = \sqrt{\frac{1}{l-k+1}\sum_{j=n}^{n+l-k} e_j(k)^2} = \sqrt{\frac{1}{l-k+1}\sum_{j=n}^{n+l-k} \left(x_{j+k} - x_j(k)\right)^2}$$

όρια πρόβλεψης $\boxed{x_n(k) \pm c_{1-\alpha/2}\sqrt{\mathrm{Var}\left(e_n(k)\right)}}$ $z_t \sim \mathrm{N}(0, \sigma_z^2) \longrightarrow c_{1-\alpha/2} \equiv z_{1-\alpha/2}$

Απλές τεχνικές πρόβλεψης

Αιτιοκρατική τάση

$x_t = \mu_t + z_t$ → Λευκός θόρυβος $z_t \sim \mathrm{WN}(0, \sigma_z^2)$

μ_t → Τάση, συνάρτηση του χρόνου

$x_{n+k} = \mu_{n+k} + z_{n+k}$

Πρόβλεψη: $x_n(k) = \mathrm{E}\left[\mu_{n+k} + z_{n+k} \mid x_n, x_{n-1}, \ldots, x_1\right] = \ldots = \mu_{n+k}$

Λύση: επέκταση της συνάρτησης μ_t για χρόνους > n

Σφάλμα πρόβλεψης: $e_n(k) = z_{n+k}$

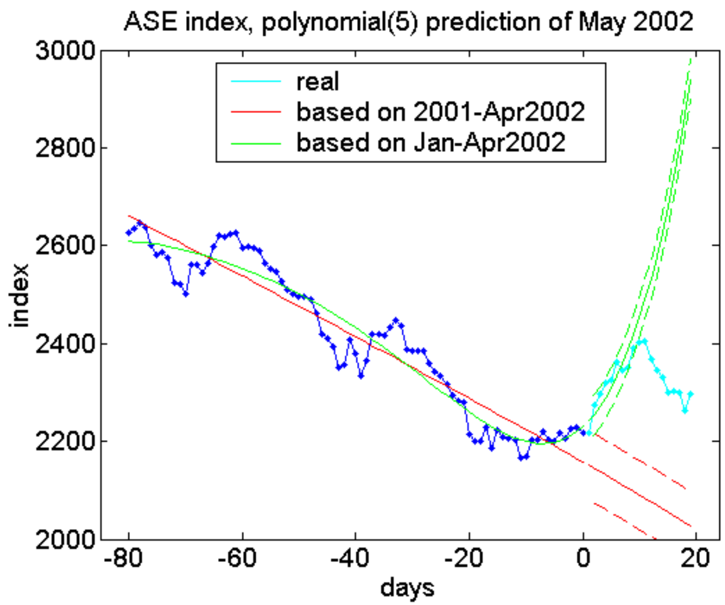
μ_t = ?

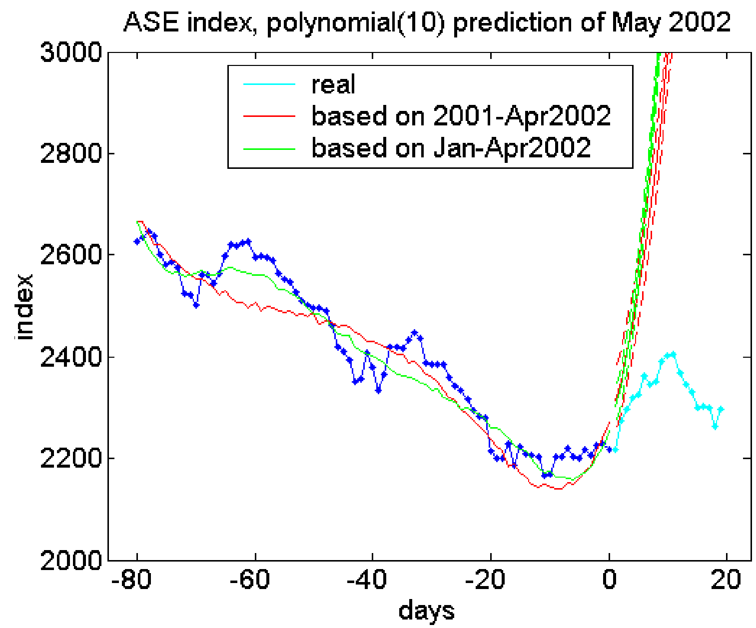
- γνωστό ➔ απλή αντικατάσταση
- άγνωστο ➔ εκτίμηση
  - καθολικά (προσαρμογή σε $\{x_1, x_2, \ldots, x_n\}$)
  - τοπικά (προσαρμογή στις m τελευταίες παρατηρήσεις $\{x_{n-m+1}, x_{n-m+2}, \ldots, x_n\}$)

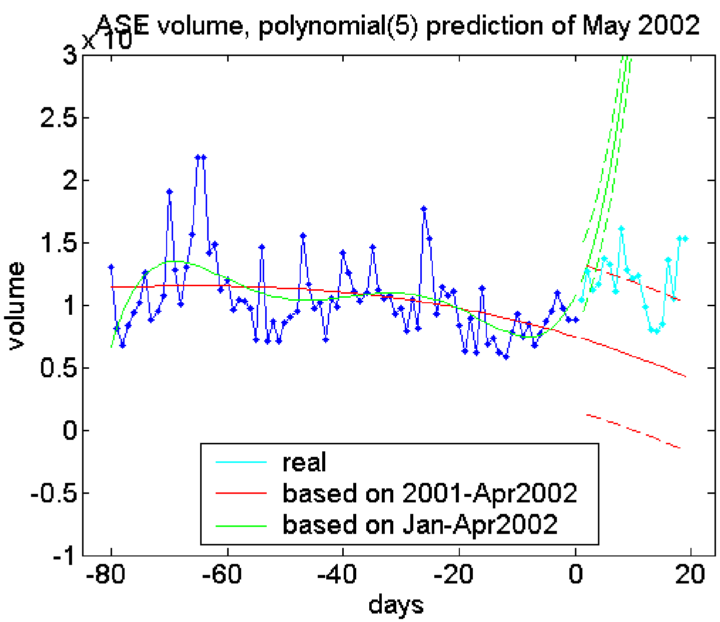
π.χ. με πολυώνυμο

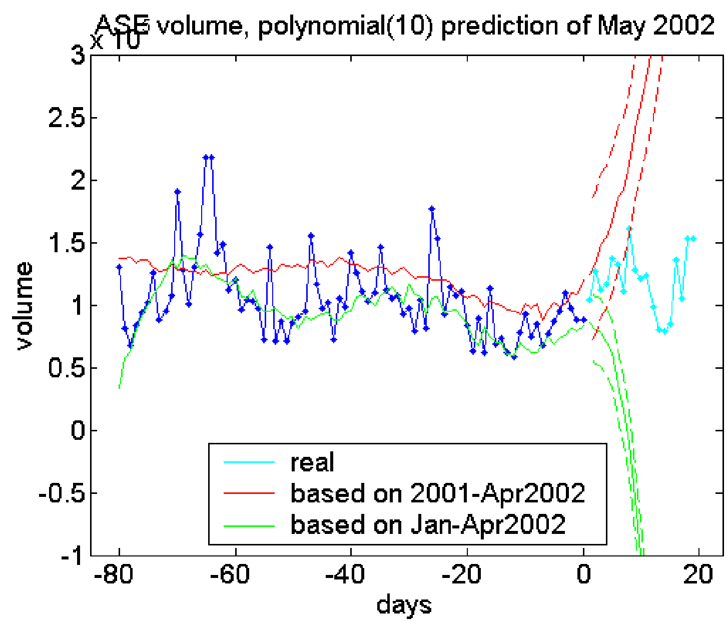
$$p_m(t) = c_0 + c_1 t + \cdots + c_m t^m$$

Δείκτης και όγκος ΧΑΑ, πρόβλεψη με επέκταση τάσεων (εκτίμηση με πολυώνυμα)









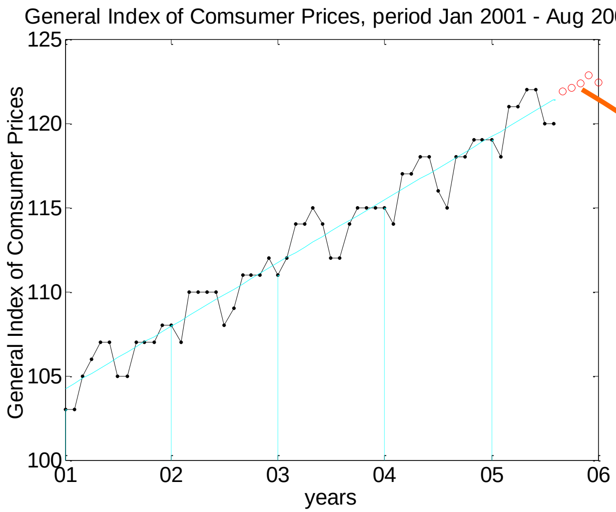
Αιτιοκρατικός εποχικός όρος $x_t = s_t + z_t$

Αιτιοκρατικός εποχικός όρος και αιτιοκρατική τάση $x_t = \mu_t + s_t + z_t$

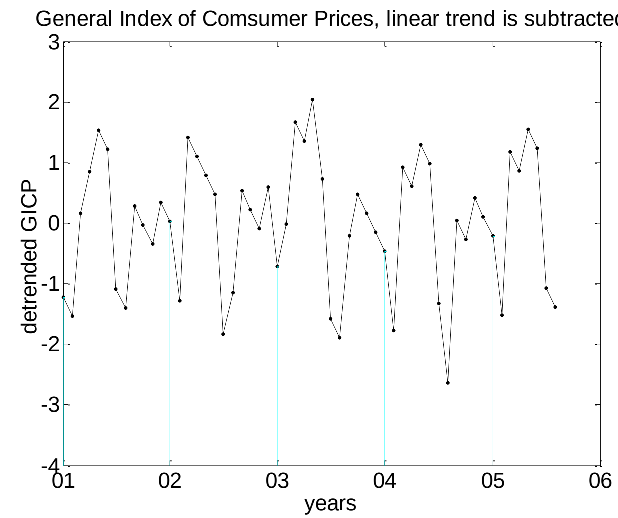
Ίδια προσέγγιση: εκτίμηση του αιτιοκρατικού μέρους

GICP, Ιανουάριος 2001 – Αύγουστος 2005

$\{x_t\}_{t=1}^{56}$



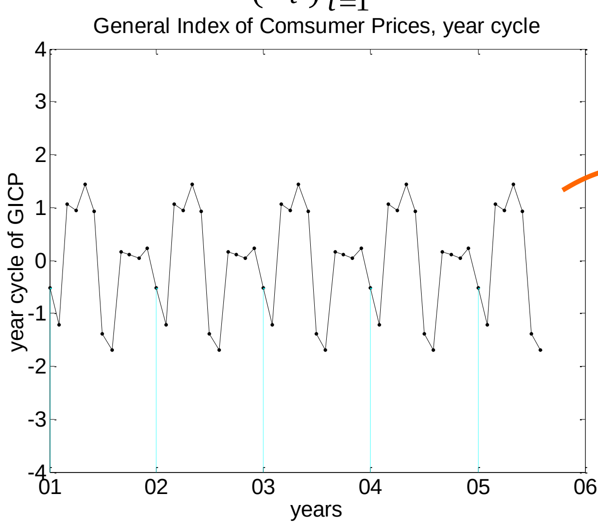
$x'_t = x_t - \mu_t$



$$\mu_t = 103.9 + 0.31t$$

$$x_t = \mu_t + s_t + z_t$$

$\{s_t\}_{t=1}^{n}$



$$z_t = x'_t - s_t = x_t - \mu_t - s_t$$

$$x_n(k) = \mu_{n+k} + s_{n+k}$$

Πρόβλεψη Σεπτεμβρίου 2005

$$\mu_{57} = 103.9 + 0.31 * 57 = 121.70$$

$$s_9 = 0.16$$

$$x_{56}(1) = 121.86$$

Εκθετική ομαλοποίηση

Εκτίμηση του x_{n+k} ως σταθμισμένο άθροισμα των προηγούμενων παρατηρήσεων

$$x_n(k) = c_0 x_n + c_1 x_{n-1} + \cdots + c_{n-1} x_1 = \sum_{j=0}^{n-1} c_j x_{n-j} \qquad \sum_{j=0}^{n-1} c_j = 1$$

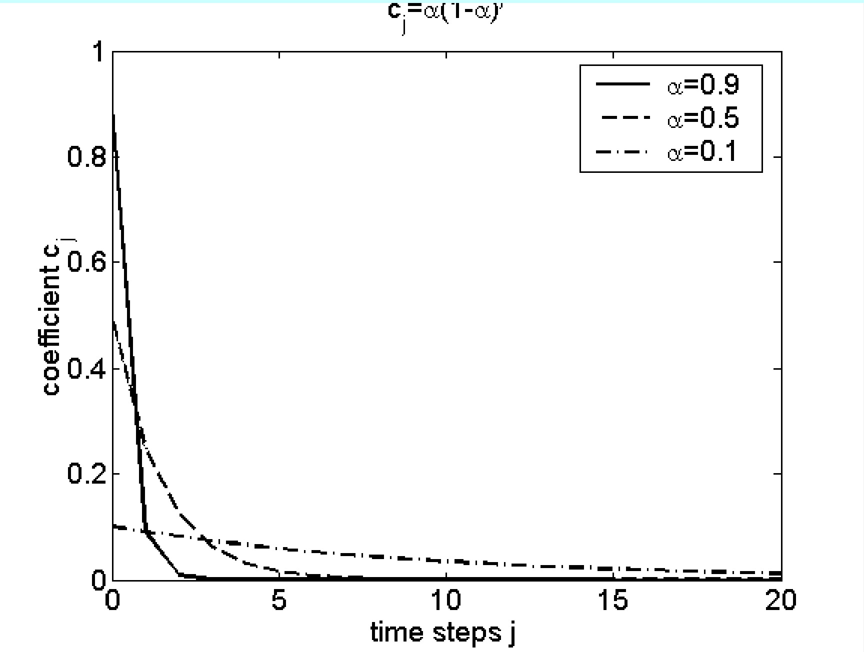
Επιθυμητή συνθήκη για τα βάρη: $c_0 \geq c_1 \geq \cdots \geq c_{n-1}$

Ορισμός βαρών με μια παράμετρο α :

$$c_j = \alpha(1-\alpha)^j, \quad j = 0,1,\ldots,n-1, \quad 0 < \alpha < 1$$

Αναδρομική σχέση :

$$x_n(k) = \alpha x_n + (1-\alpha) x_{n-1}(k)$$

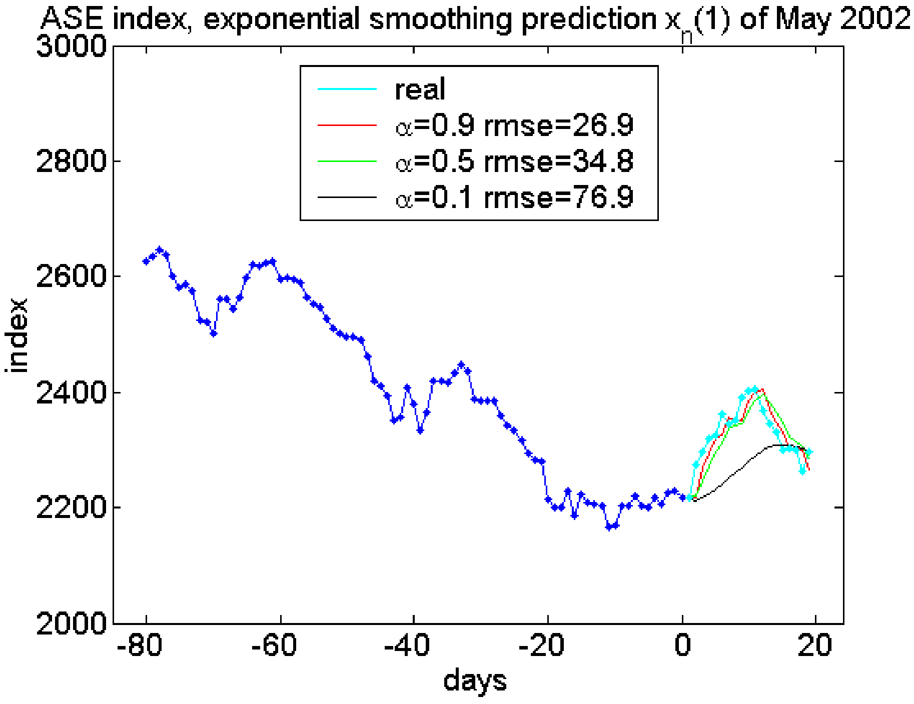


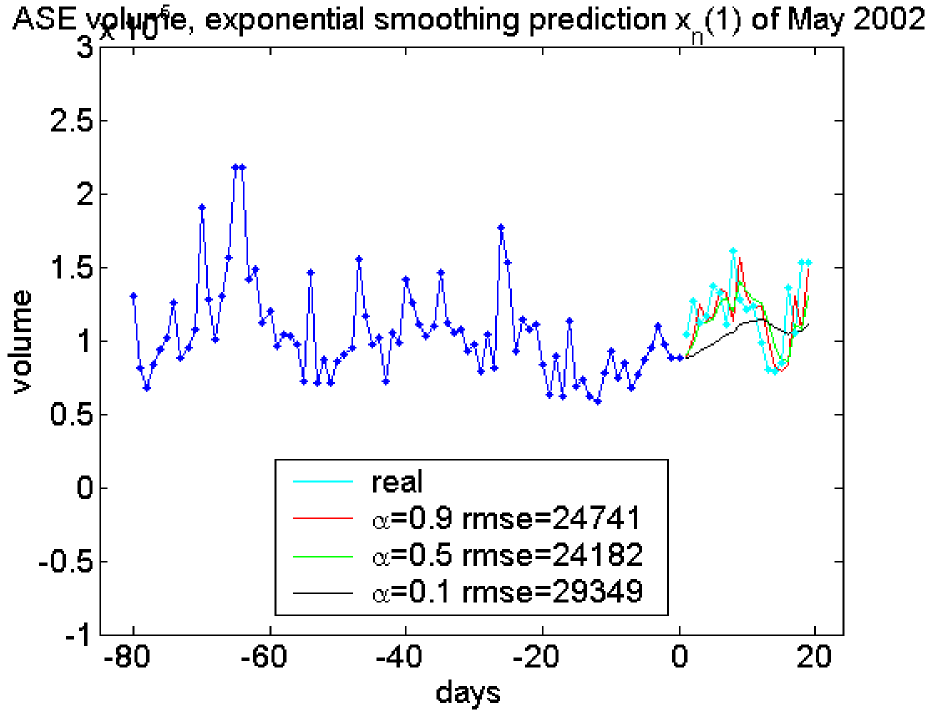
Πρόβλεψη με εκθετική ομαλοποίηση: Παραδείγματα

<u>*Δείκτης και όγκος ΧΑΑ*</u>

Πρόβλεψη με χρονικό βήμα 1 για όλες τις μέρες του Μαΐου 2002

Σύγκριση της απόδοσης του μοντέλου εκθετικής ομαλοποίησης για διάφορα α





Μεγάλο α (βάρος στις πολύ πρόσφατες παρατηρήσεις) δίνει την καλύτερη πρόβλεψη

Πρόβλεψη στάσιμων χρονοσειρών με γραμμικά μοντέλα

Προβλέψεις με AR, MA και ARMA

Πρόβλεψη με αυτοπαλινδρομούμενα μοντέλα (AR)

Δίνεται η χρονική σειρά $\{x_1, x_2, \ldots, x_n\}$

AR(1) μοντέλο $x_t = \phi\, x_{t-1} + z_t \longrightarrow$ Λευκός θόρυβος $z_t \sim \mathrm{WN}(0, \sigma_z^2)$

$t \to n+1 \qquad x_{n+1} = \phi\, x_n + z_{n+1}$

Βέλτιστη πρόβλεψη για βήμα 1: $x_n(1) = \phi\, x_n$

Σφάλμα πρόβλεψης :

$$e_n(1) = z_{n+1}$$

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

$t \to n+2 \qquad x_{n+2} = \phi\, x_{n+1} + z_{n+2}$

Βέλτιστη πρόβλεψη για βήμα 2: $x_n(2) = \phi\, x_n(1) = \phi^2 x_n$

Σφάλμα πρόβλεψης :

$$e_n(2) = \phi\, z_{n+1} + z_{n+2}$$

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

$t \to n+k$

Βέλτιστη πρόβλεψη για βήμα *k*: $x_n(k) = \phi^k x_n$

Σφάλμα πρόβλεψης :

$$e_n(k) = \phi^{k-1} z_{n+1} + \cdots + \phi\, z_{n+k-1} + z_{n+k}$$

$$\mathrm{Var}\left[e_n(k)\right] = \sigma_z^2 \frac{1-\varphi^{2k}}{1-\varphi^2}$$

Δίνεται η χρονική σειρά $\{x_1, x_2, \ldots, x_n\}$

AR(p) μοντέλο $x_t = \phi_1 x_{t-1} + \cdots + \phi_p x_{t-p} + z_t$

$t \rightarrow n+1 \qquad x_{n+1} = \phi_1 x_n + \cdots + \phi_p x_{n-p+1} + z_{n+1}$

Βέλτιστη πρόβλεψη για βήμα 1:

$$x_n(1) = \phi_1 x_n + \cdots + \phi_p x_{n-p+1}$$

Σφάλμα πρόβλεψης :

$$e_n(1) = z_{n+1}$$

$$\mathrm{Var}\big(e_n(1)\big) = \sigma_z^2$$

$t \rightarrow n+2 \qquad x_{n+2} = \phi_1 x_{n+1} + \cdots + \phi_p x_{n-p+2} + z_{n+2}$

Βέλτιστη πρόβλεψη για βήμα 2:

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

Σφάλμα πρόβλεψης :

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

$$\mathrm{Var}\big(e_n(2)\big) = (\phi^2 + 1)\sigma_z^2$$

$t \rightarrow n+k$

$$x_{n+k} = \phi_1 x_{n-1+k} + \cdots + \phi_p x_{n-p+k} + z_{n+k}$$

Βέλτιστη πρόβλεψη για βήμα k:

$$x_n(k) = \phi_1 x_n(k-1) + \cdots + \phi_p x_n(k-p)$$

$$\text{όπου} \quad x_n(j) = \begin{cases} \text{πρόβλεψη} \quad x_n(j) & j > 0 \\ \text{παρατήρηση} \quad x_{n+j} & j \le 0 \end{cases}$$

Σφάλμα πρόβλεψης :

$$e_n(k) = \phi_1 e_n(k-1) + \cdots + \phi_k e_n(1) + z_{n+k}$$

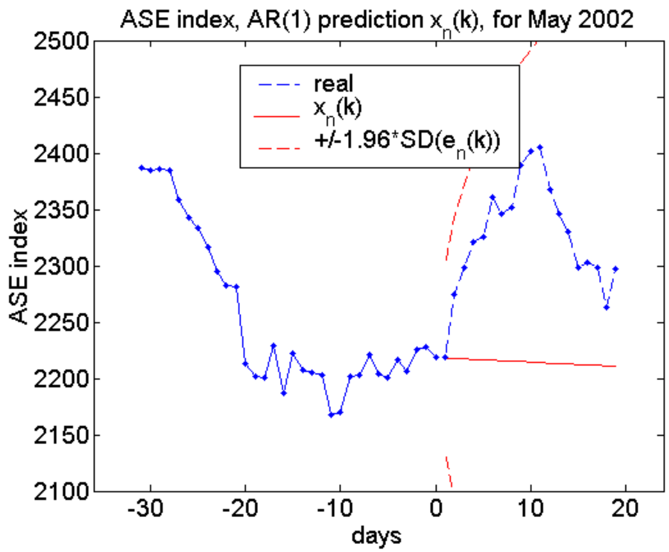
$$e_n(k) = \sum_{j=0}^{k-1} b_j z_{n+k-j}$$

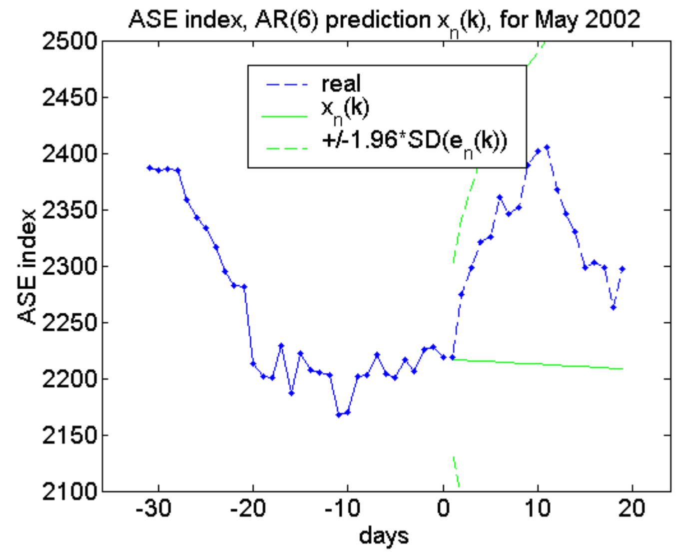
$$\mathrm{Var}\big(e_n(k)\big) = \sigma_z^2 \sum_{j=0}^{k-1} b_j^2$$

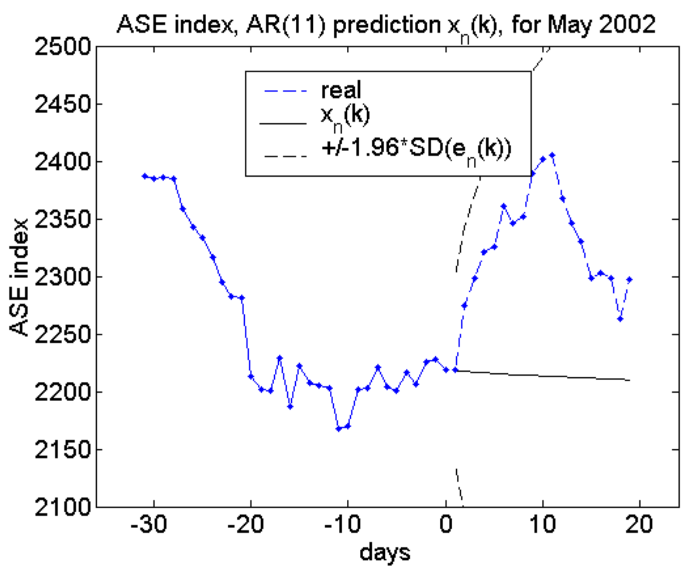
Δείκτης ΧΑΑ, πρόβλεψη για Μάιο 2002

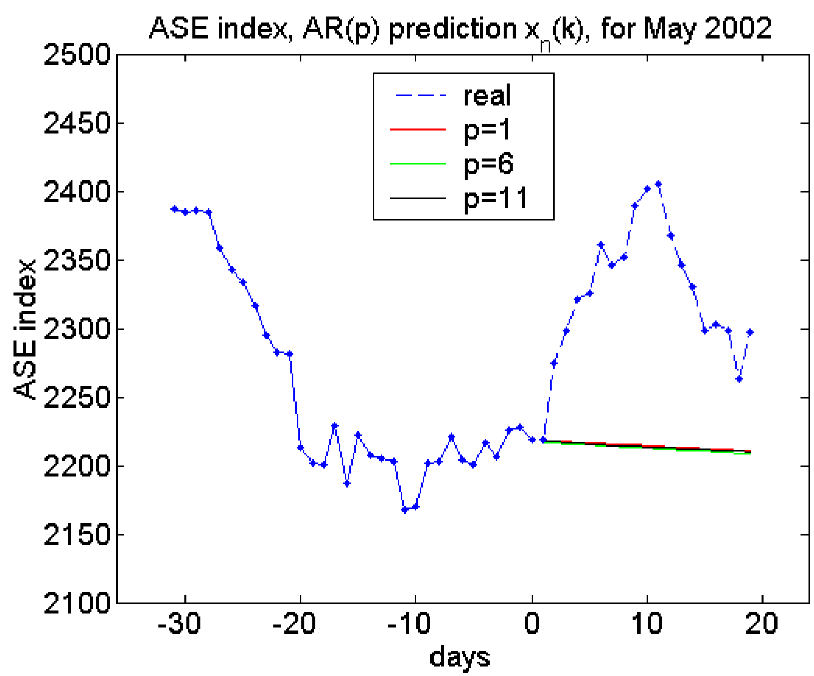
$$x_n(k),\ \ n = 30.04.2002,\ \ k = 1,\ldots,20$$

AR(1)	AR(6)	AR(11)
0.9995	1.1535	1.1523
	-0.2126	-0.2131
	0.0944	0.0961
	-0.0655	-0.0663
	0.0103	0.0135
	0.0194	-0.0031
		-0.0058
		0.0190
		-0.0175
		0.0458
		-0.0213



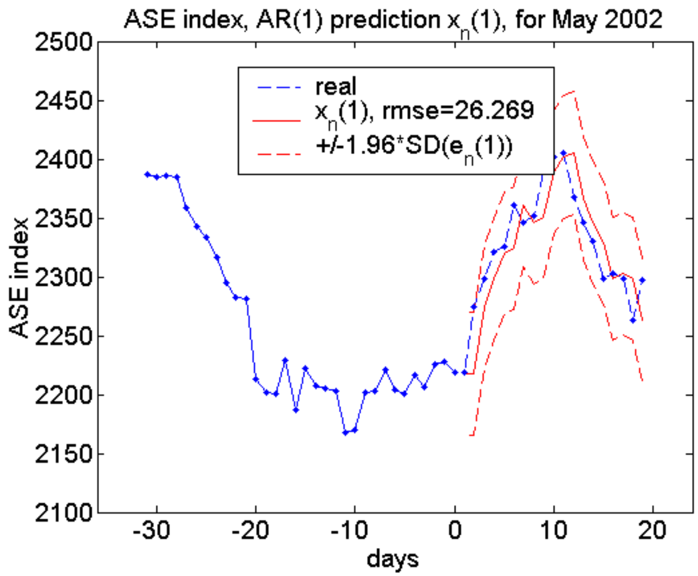


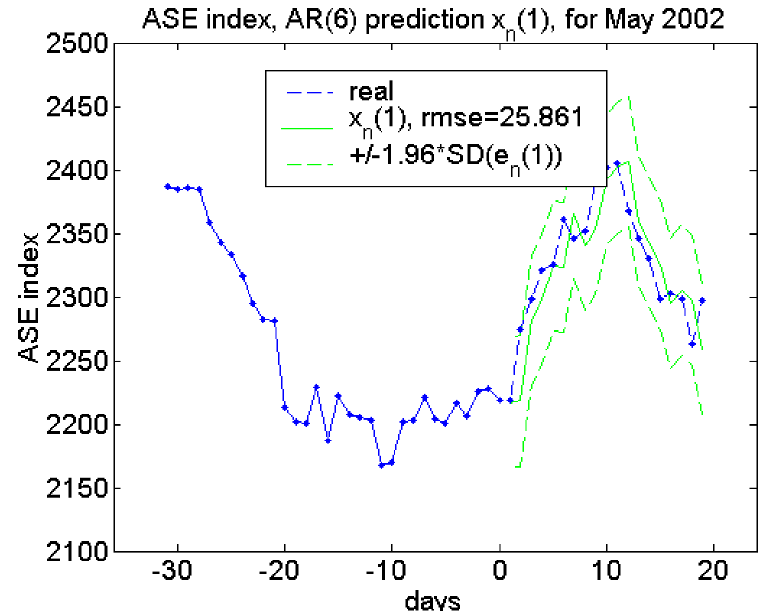


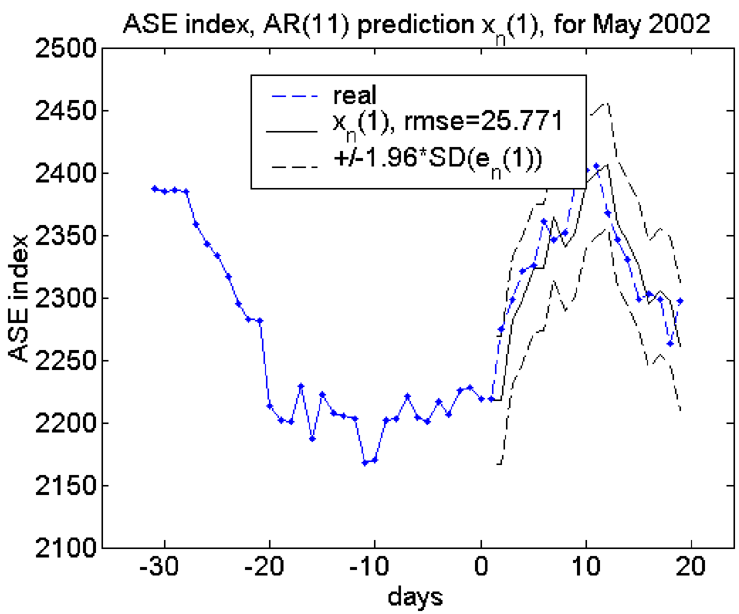


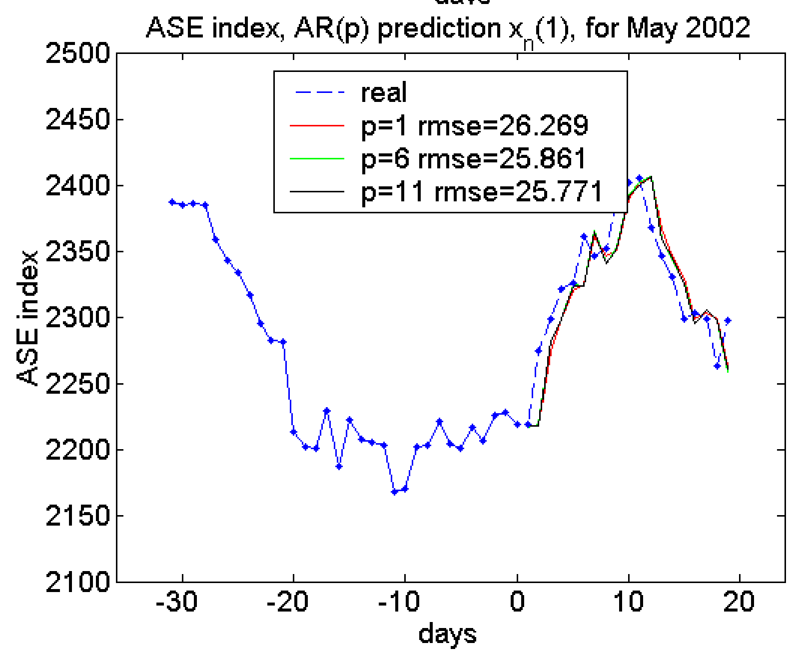
Δείκτης ΧΑΑ, ημερήσιες προβλέψεις για Μάιο 2002

$$x_n(1),\ \ n = 2.05.2002 - 31.05.2002$$





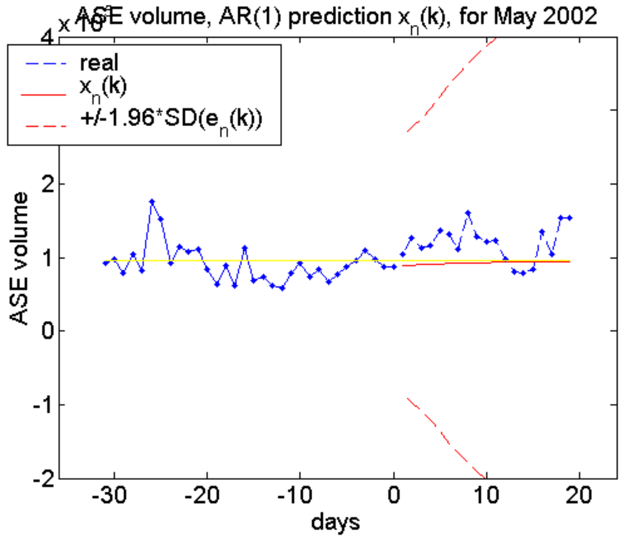


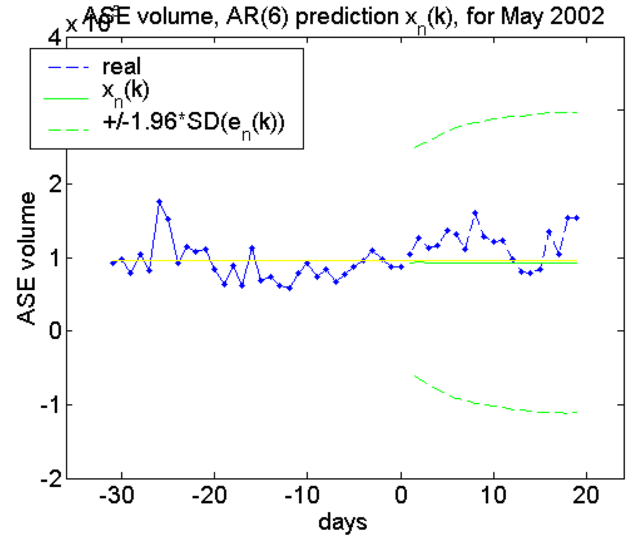


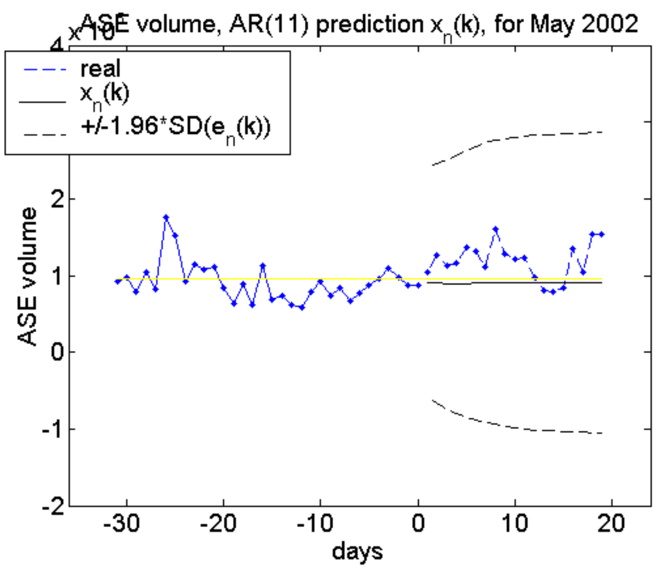
Όγκος ΧΑΑ, πρόβλεψη για Μάιο 2002

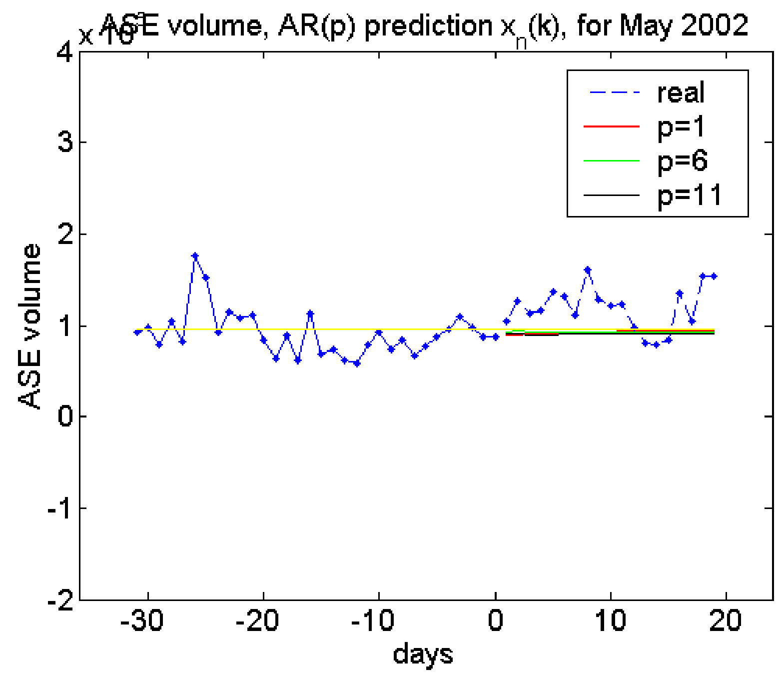
$$x_n(k), \quad n = 30.04.2002, \quad k = 1, \ldots, 20$$

AR(1)	AR(6)	AR(11)
0.9097	0.3412	0.3251
	0.2092	0.1955
	0.1557	0.1380
	0.1369	0.1138
	0.0773	0.0455
	0.0528	0.0009
		0.0350
		0.0068
		0.0249
		0.0420
		0.0527



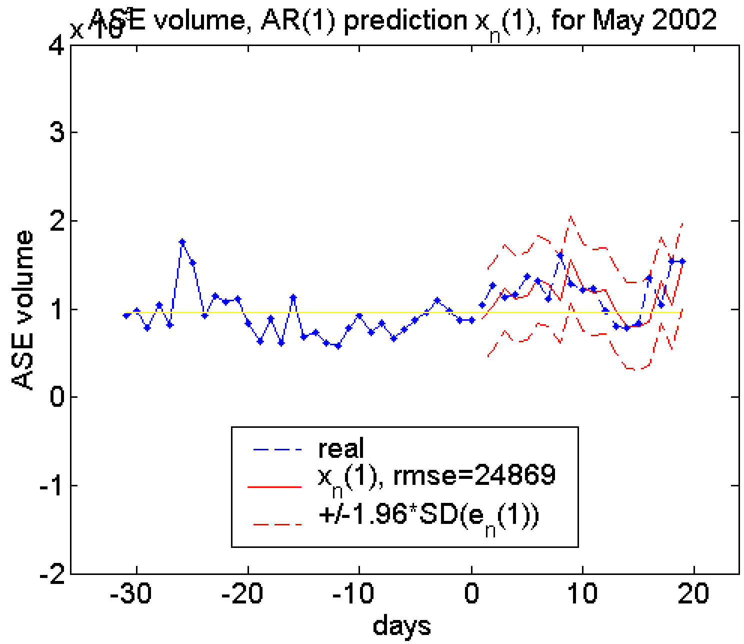


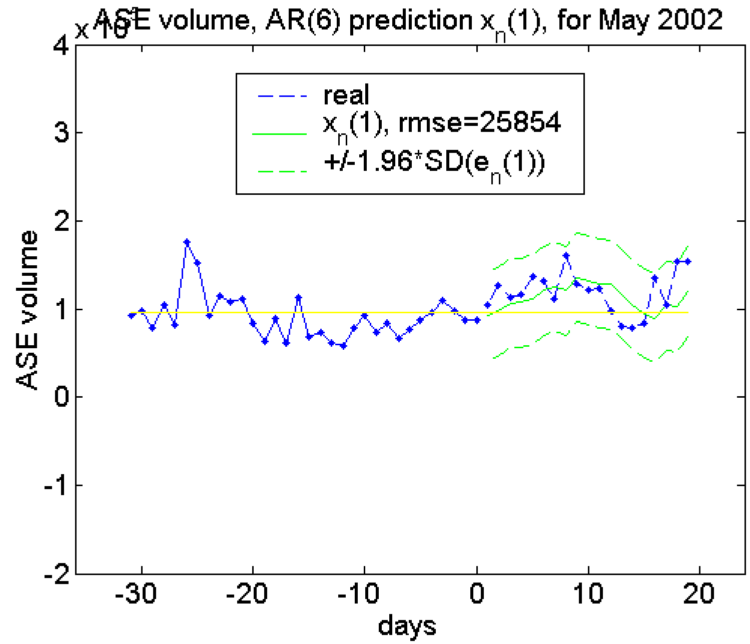


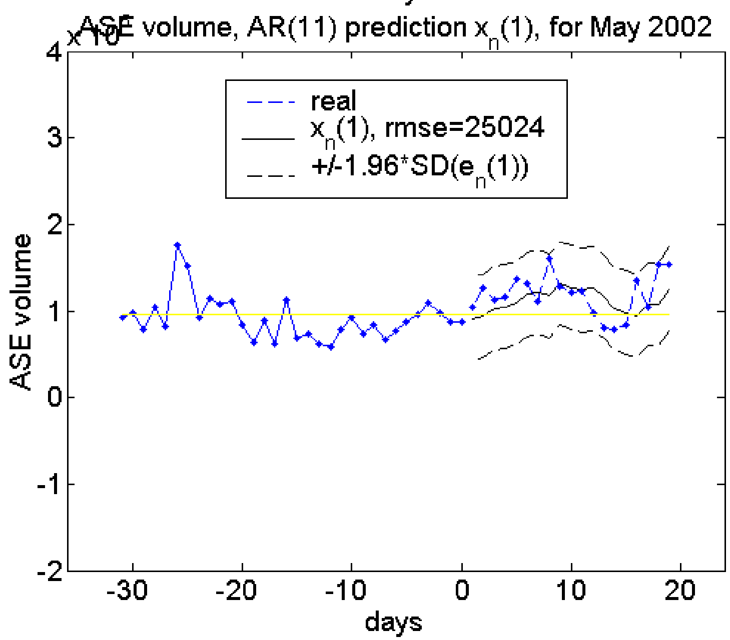


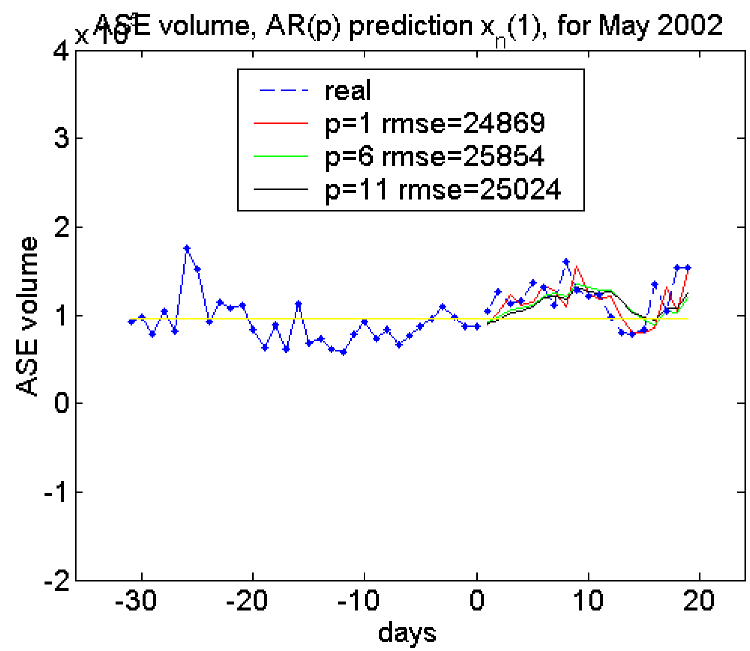
Όγκος ΧΑΑ, ημερήσιες προβλέψεις για Μάιο 2002

$x_n(1),\ \ n = 2.05.2002 - 31.05.2002$



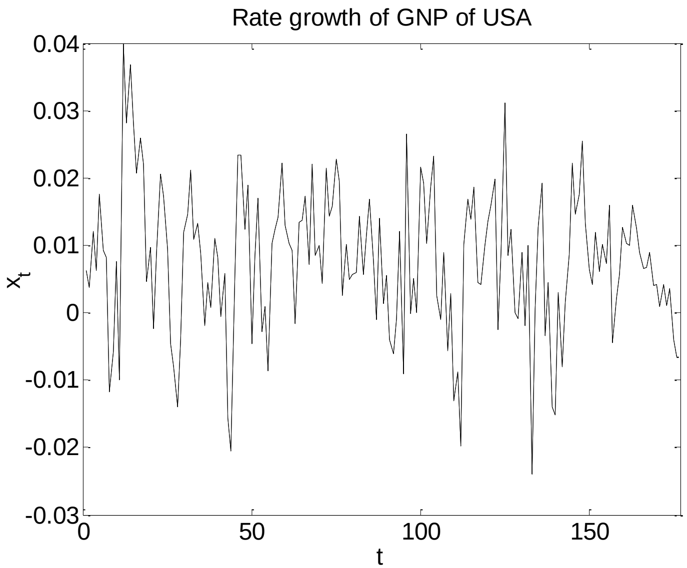


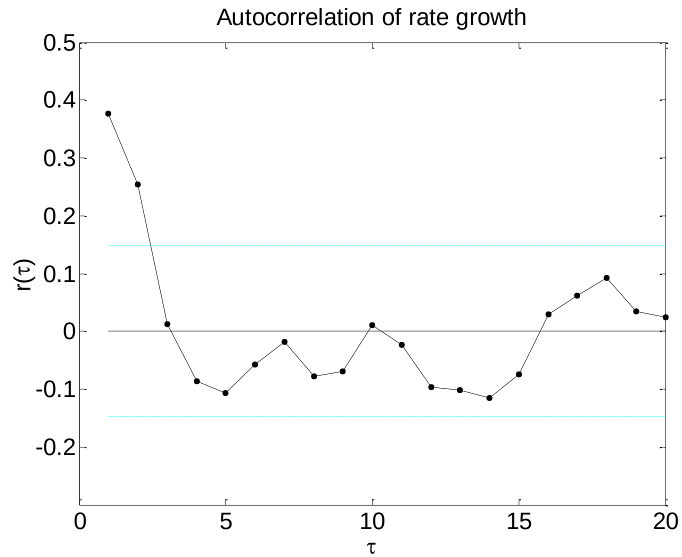


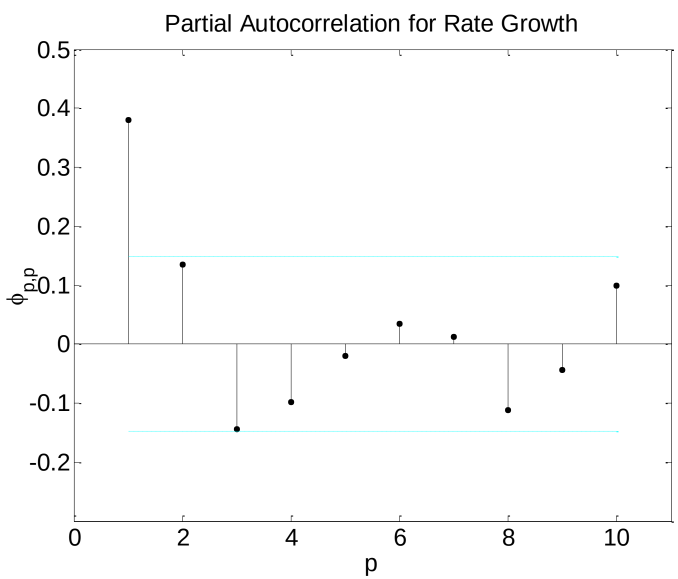


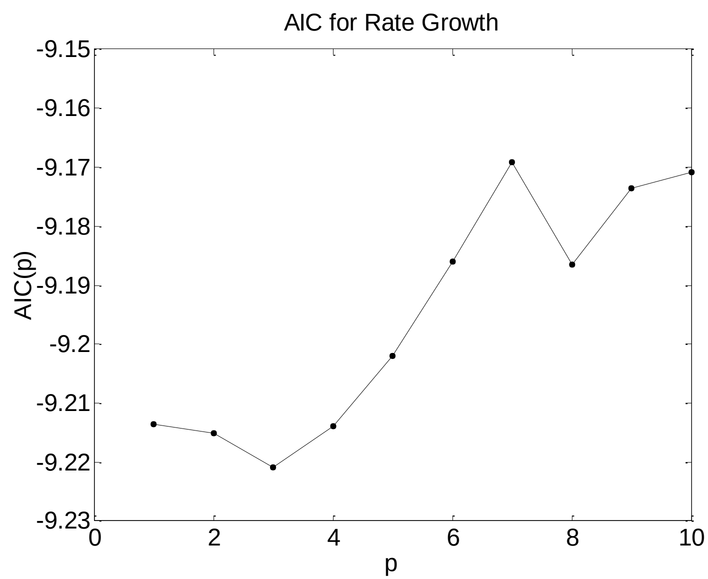
Ρυθμός μεταβολής του ΑΕΠ των ΗΠΑ

Οι τιμές είναι τετραμηνιαίες, από το δεύτερο τετράμηνο του 1947 ως το πρώτο τετράμηνο του 1991 (n=176)









Ρυθμός μεταβολής του ΑΕΠ των ΗΠΑ

AR(3) $x_t = \phi_0 + \phi_1 x_{t-1} + \phi_2 x_{t-2} + \phi_3 x_{t-3} + z_t$

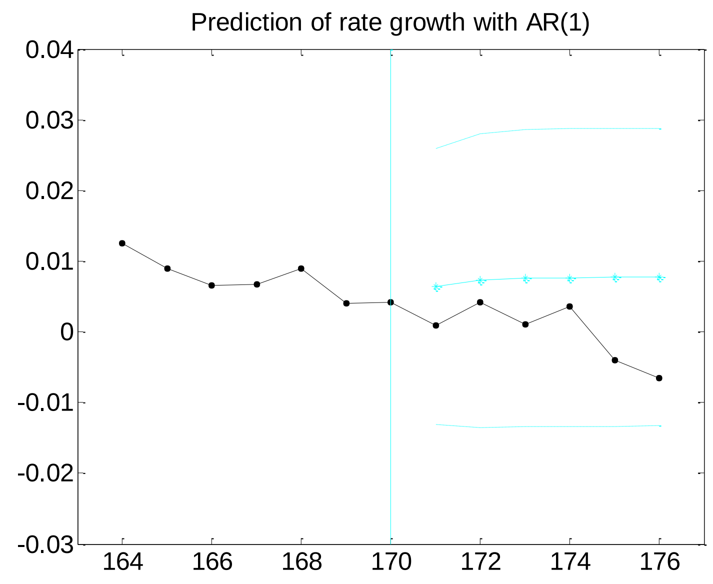
$$x_n(k), \; n = 170, \; k = 1, \ldots, 6$$

$$\hat{\mu} = 0.0077 \qquad \hat{\phi}_1 = 0.35 \qquad \hat{\phi}_2 = 0.18 \qquad \hat{\phi}_3 = -0.14$$

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

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

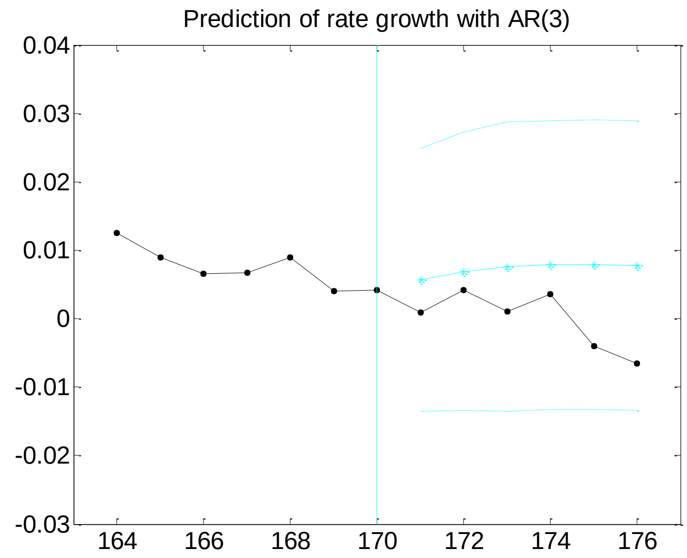
$$s_z = \hat{\sigma}_z = 0.0098$$



AR(1)

$$x_t = 0.0047 + 0.38 x_{t-1} + z_t$$

$$s_z = \hat{\sigma}_z = 0.0099$$

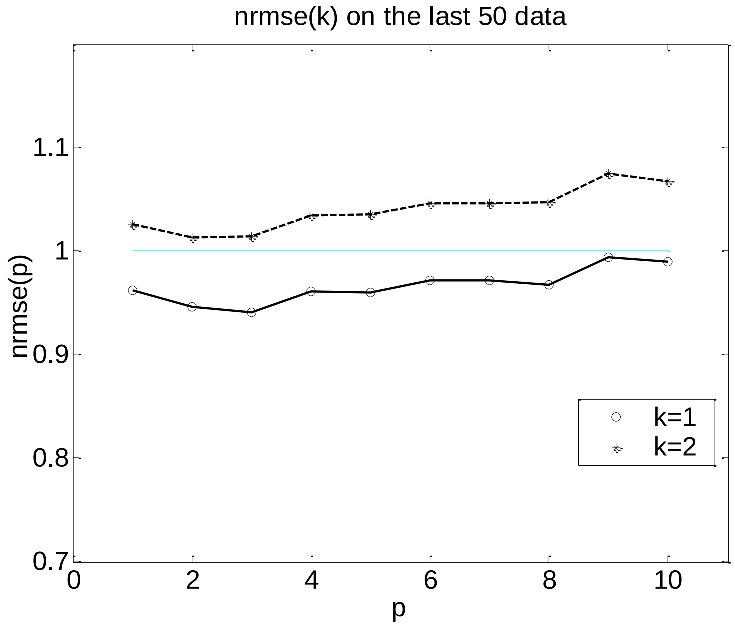


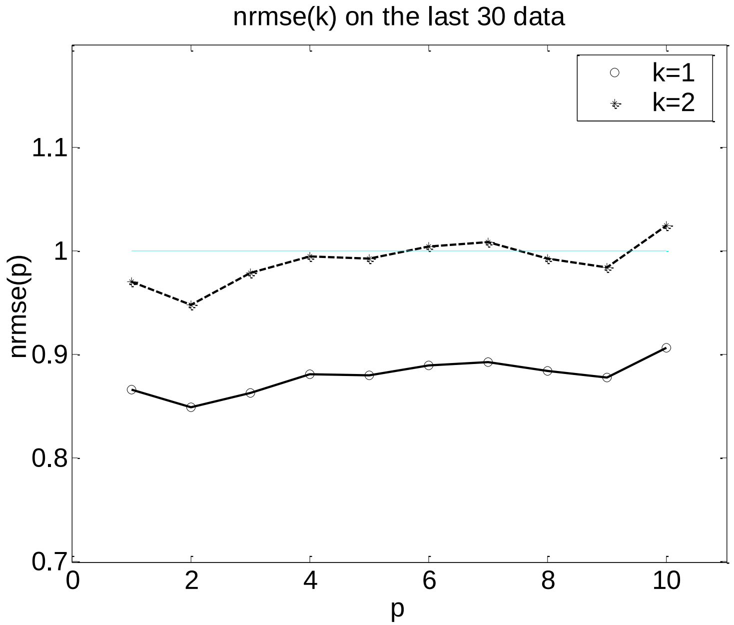
Ρυθμός μεταβολής του ΑΕΠ των ΗΠΑ

Απόδοση πρόβλεψης *k βημάτων* AR(*p*), *p*=1,…,10

$$\boxed{x_n(1),\ \ n = 126 - 176}$$

$$\boxed{x_n(1),\ \ n = 146 - 176}$$





$$z_t \sim \mathrm{WN}(0, \sigma_z^2) \qquad \mathrm{E}\left(z_{n+j} \mid x_n, x_{n-1}, \ldots\right) = \begin{cases} 0 & \text{αν } j > 0 \\ z_{n+j} & \text{αν } j \leq 0 \end{cases}$$

MA(1) μοντέλο $x_t = z_t + \theta\, z_{t-1}$

$t \to n+1 \qquad x_{n+1} = z_{n+1} + \theta\, z_n$

Βέλτιστη πρόβλεψη για βήμα 1: $x_n(1) = \theta\, z_n$

Σφάλμα πρόβλεψης :

$$e_n(1) = z_{n+1}$$

$$\mathrm{Var}\left(e_n(1)\right) = \sigma_z^2$$

$t \to n+2 \qquad x_{n+2} = z_{n+2} + \theta\, z_{n+1}$

Βέλτιστη πρόβλεψη για βήμα 2: $x_n(2) = 0$

Σφάλμα πρόβλεψης :

$$e_n(2) = x_{n+2}$$

$$\mathrm{Var}\left(e_n(2)\right) = \mathrm{Var}\left(x_{n+2}\right) = \sigma_x^2$$

Γενικά για βήμα πρόβλεψης *k*:

$$x_n(k) = \begin{cases} \theta\, z_n & \text{για } k = 1 \\ 0 & \text{για } k > 1 \end{cases} \qquad e_n(k) = \begin{cases} z_{n+1} & \text{για } k = 1 \\ x_{n+k} & \text{για } k > 1 \end{cases}$$

MA(q) μοντέλο $x_t = z_t + \theta_1 z_{t-1} + \cdots + \theta_q z_{t-q}$

$t \to n+1$ $\quad x_{n+1} = z_{n+1} + \theta_1 z_n + \cdots + \theta_q z_{n-q+1}$

Βέλτιστη πρόβλεψη για βήμα 1:

$x_n(1) = \theta_1 z_n + \cdots + \theta_q z_{n-q+1}$

Σφάλμα πρόβλεψης :

$e_n(1) = z_{n+1}$

$\mathrm{Var}\big(e_n(1)\big) = \sigma_z^2$

$t \to n+2$ $\quad x_{n+2} = z_{n+2} + \theta_1 z_{n+1} + \theta_2 z_n + \cdots + \theta_q z_{n-q+2}$

Βέλτιστη πρόβλεψη για βήμα 2:

$x_n(2) = \theta_2 z_n + \cdots + \theta_q z_{n-q+2}$

Σφάλμα πρόβλεψης :

$e_n(2) = z_{n+2} + \theta_1 z_{n+1}$

$\mathrm{Var}\big(e_n(2)\big) = (\theta^2 + 1)\sigma_z^2$

$t \to n+k$ $\quad x_{n+k} = z_{n+k} + \theta_1 z_{n+k-1} + \cdots + \theta_q z_{n+k-q}$

Βέλτιστη πρόβλεψη για βήμα k:

$$x_n(k) = \begin{cases} \theta_k z_n + \theta_{k+1} z_{n-1} + \cdots + \theta_q z_{n-q+k} & \text{αν } k \le q \\ 0 & \text{αν } k > q \end{cases}$$

Σφάλμα πρόβλεψης :

$$e_n(k) = z_{n+k} + \theta_1 z_{n+k-1} + \cdots + \theta_{k-1} z_{n+1}$$

$$e_n(k) = \sum_{j=0}^{k-1} \theta_j z_{n+k-j}$$

$$\mathrm{Var}\big(e_n(k)\big) = \sigma_z^2 \sum_{j=0}^{k-1} \theta_j^2$$

Ρυθμός μεταβολής του ΑΕΠ των ΗΠΑ

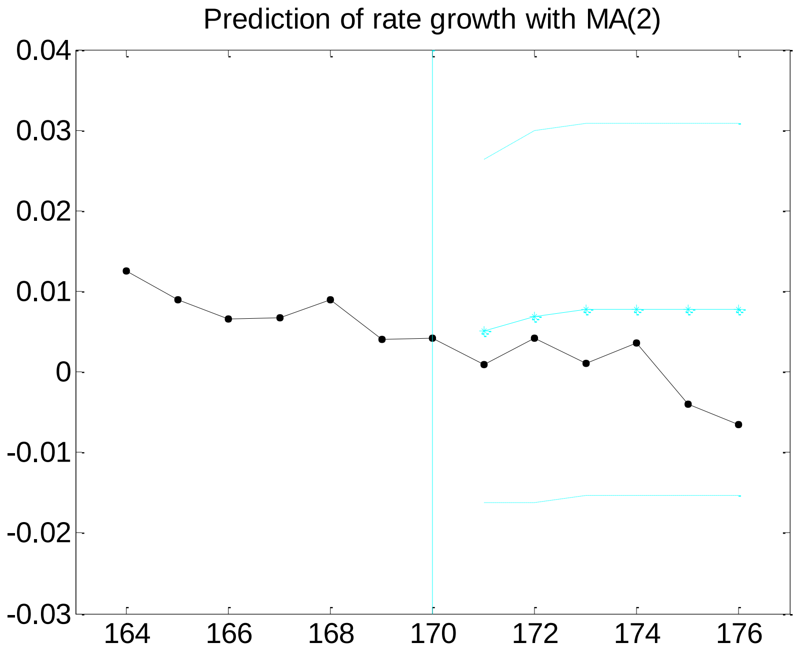
Οι τιμές είναι τετραμηνιαίες, από το δεύτερο τετράμηνο του 1947 ως το πρώτο τετράμηνο του 1991 (n=176)

MA(2)

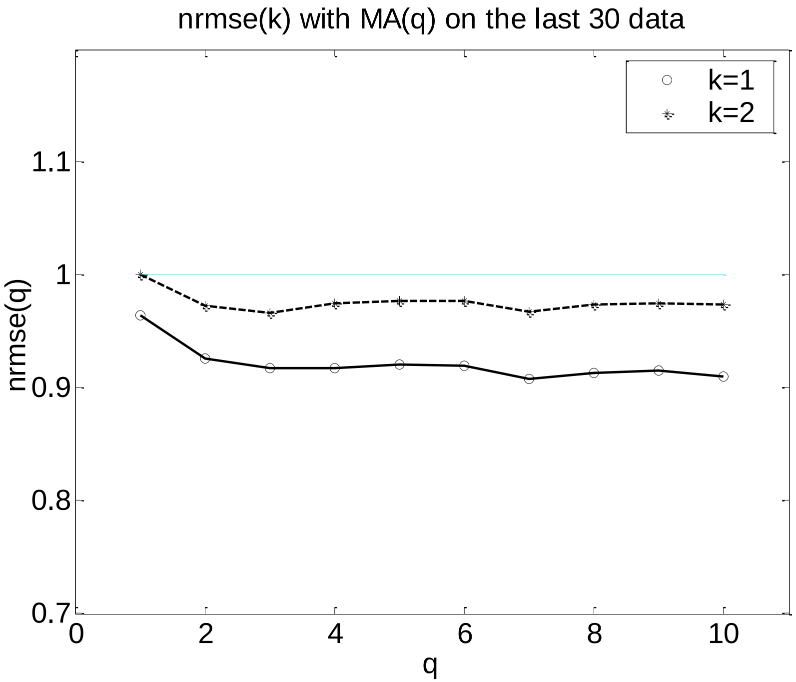
$$x_t = 0.0077 + z_t - 0.41 z_{t-1} - 0.40 z_{t-2}$$

$$s_z = \hat{\sigma}_z = 0.0109$$

$\boxed{x_n(k),\ n=170,\ k=1,\dots,6}$



$\boxed{x_n(1),\ n=146-176}$



ARMA(p,q) μοντέλο

$$x_t = \phi_1 x_{t-1} + \cdots + \phi_p x_{t-p} + z_t + \theta_1 z_{t-1} + \cdots + \theta_q z_{t-q}$$

$$t \to n+1$$

$$x_{n+1} = \phi_1 x_n + \cdots + \phi_p x_{n-p+1} + z_{n+1} + \theta_1 z_n + \cdots + \theta_q z_{n-q+1}$$

Βέλτιστη πρόβλεψη για βήμα 1:

$$x_n(1) = \phi_1 x_n + \cdots + \phi_p x_{n-p+1} + \theta_1 z_n + \cdots + \theta_q z_{n-q+1}$$

Σφάλμα πρόβλεψης :

$$e_n(1) = z_{n+1}$$

$$\mathrm{Var}\left(e_n(1)\right) = \sigma_z^2$$

Γενικά για βήμα πρόβλεψης *k*:

$$x_n(k) = \begin{cases} \phi_1 x_n(k-1) + \cdots + \phi_p x_n(k-p) + \theta_k z_n + \cdots + \theta_q z_{n-q+k} & \text{αν } k \le q \\ \phi_1 x_n(k-1) + \cdots + \phi_p x_n(k-p) & \text{αν } k > q \end{cases}$$

Πρόβλεψη με ARMA: σύνθεση των προβλέψεων με AR και MA

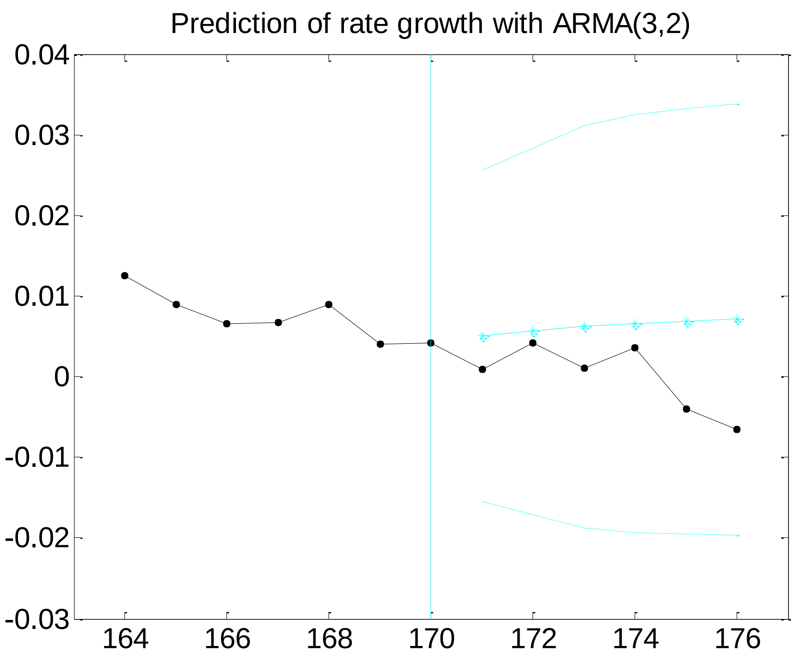
Ρυθμός μεταβολής του ΑΕΠ των ΗΠΑ

ARMA(3,2)

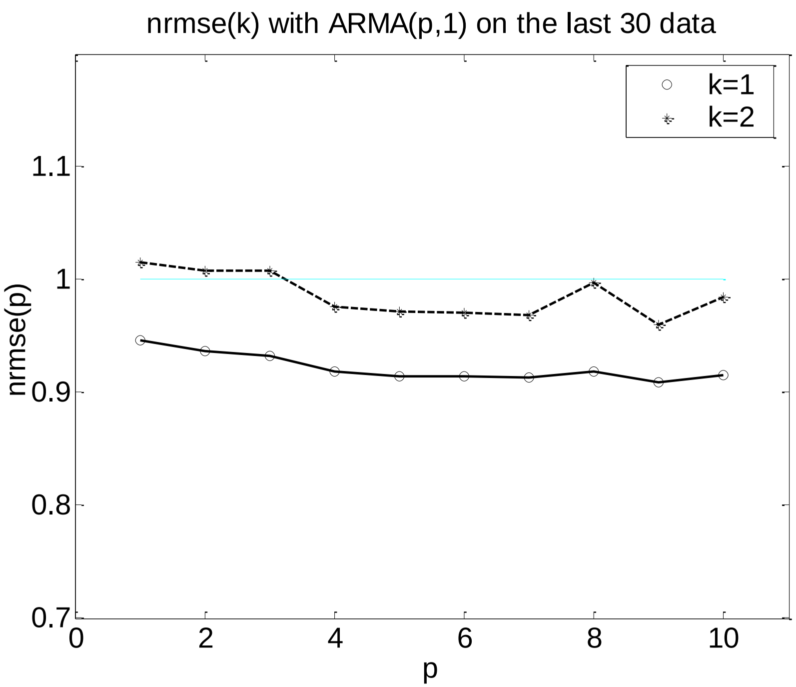
$$x_t = 0.0034 + 0.15x_{t-1} + 0.29x_{t-2} + 0.12x_{t-3} + z_t - 0.33z_{t-1} - 0.13z_{t-2}$$

$$s_z = \hat{\sigma}_z = 0.0105$$

$\boxed{x_n(k),\ n=170,\ k=1,\dots,6}$



$\boxed{x_n(1),\ n=146-176}$



Πρόβλεψη μη-στάσιμων χρονοσειρών

Μη-στάσιμη χρονική σειρά $\left\{y_1, y_2, \ldots, y_n\right\}$

<u>Βήματα πρόβλεψης:</u>

1. μετασχηματισμός σε στάσιμη
2. πρόβλεψη του x_{n+k} με κάποιο μοντέλο
3. αντίστροφος μετασχηματισμός για την πρόβλεψη

$$\begin{tikzcd}
\left\{y_1, y_2, \ldots, y_n\right\} \arrow[d, "1"'] \arrow[rr, dashed] & & y_n(k) \\
\left\{x_1, x_2, \ldots, x_n\right\} \arrow[rr, "2"] & & x_n(k) \arrow[u, "3"']
\end{tikzcd}$$

Κλασικό μοντέλο για το y_t: $y_t = \mu_t + s_t + x_t$

$t \to n+k$ Πρόβλεψη του y_{n+k} : $y_{n+k} = \mu_{n+k} + s_{n+k} + x_{n+k}$

Εκτίμηση των μ_t και s_t
ως συναρτήσεις του χρόνου t

1. $x_t = y_t - \mu_t - s_t$
2. $x_n(k)$: πρόβλεψη (τύπου ARMA) του x_{n+k}
3. $y_n(k) = \mu_{n+k} + s_{n+k} + x_n(k)$

Απαλοιφή των μ_t και s_t
(χρησιμοποιώντας διαφορές)
➔ πρόβλεψη με μοντέλα ARIMA

ARIMA(*p,1,q*)

Βήματα πρόβλεψης του $y_n(1)$ *:*

1. μετασχηματισμός: $\{y_1, y_2, \ldots, y_n\} \rightarrow \{x_2, x_3 \ldots, x_n\}$ στάσιμη

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

2. πρόβλεψη του x_{n+1} με ARMA(*p,q*) ➔ $x_n(1)$
3. αντίστροφος μετασχηματισμός: $y_n(1) \leftarrow x_n(1)$

$$y_n(1) = y_n + x_n(1)$$

Σφάλμα πρόβλεψης :

$$e_n(1) = \tilde{e}_n(1)$$

σφάλμα πρόβλεψης του $x_n(1)$

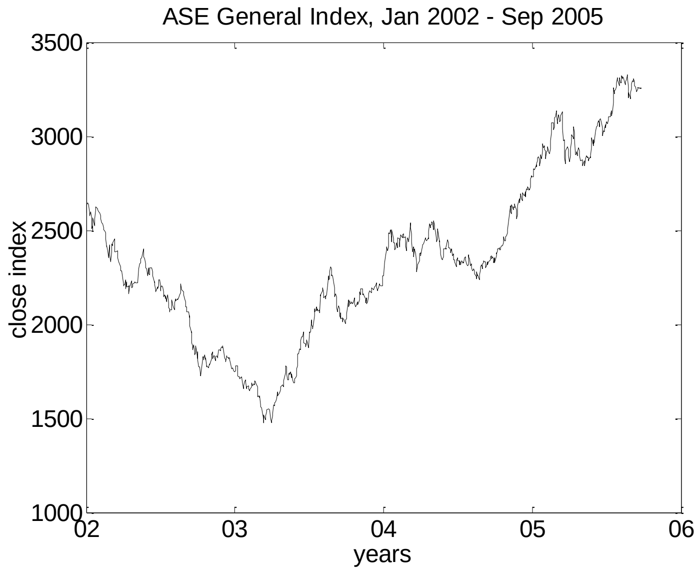
Γενικά για βήμα πρόβλεψης *k* :

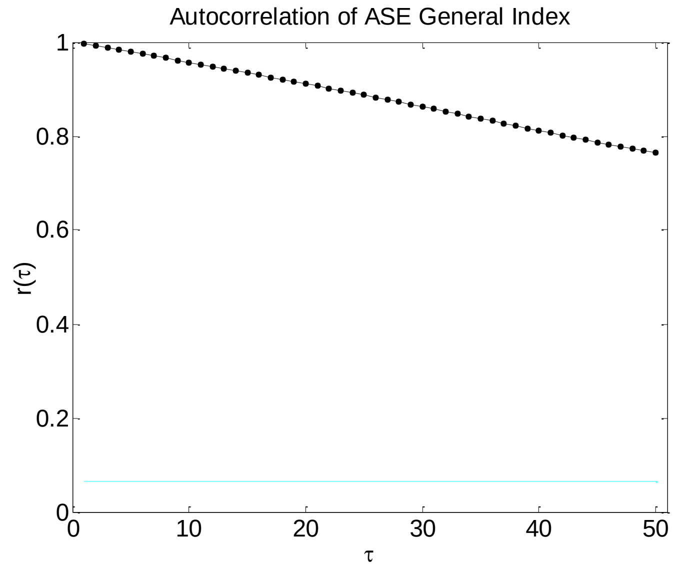
$$y_n(k) = y_n(k-1) + x_n(k)$$

γνωστό από την πρόβλεψη του y_{n+k-1} ARMA(*p,q*) πρόβλεψη του x_{n+k}

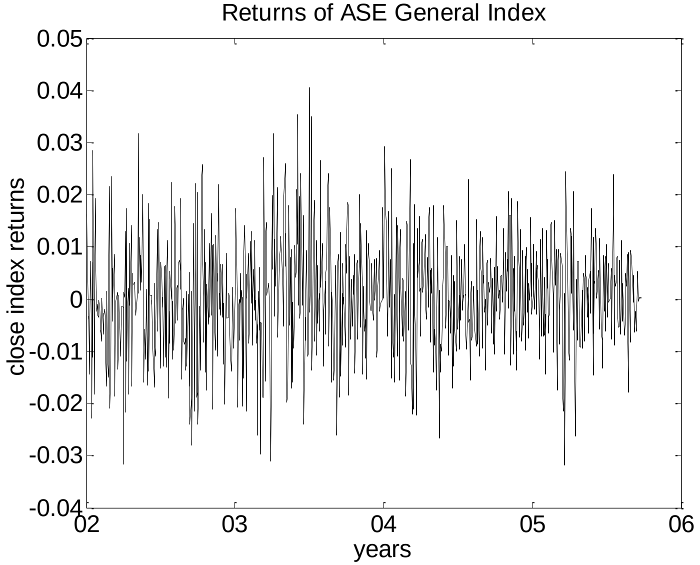
Παρόμοια διαδικασία πρόβλεψης για ARIMA(*p,d,q*)

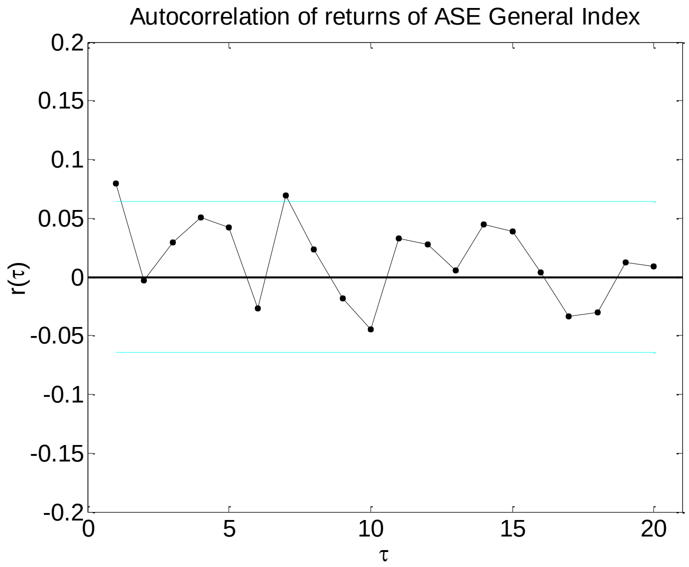
Γενικός δείκτης Χρηματιστηρίου Αθηνών Περίοδος Ιανουάριος 2002 – Σεπτέμβριος 2005





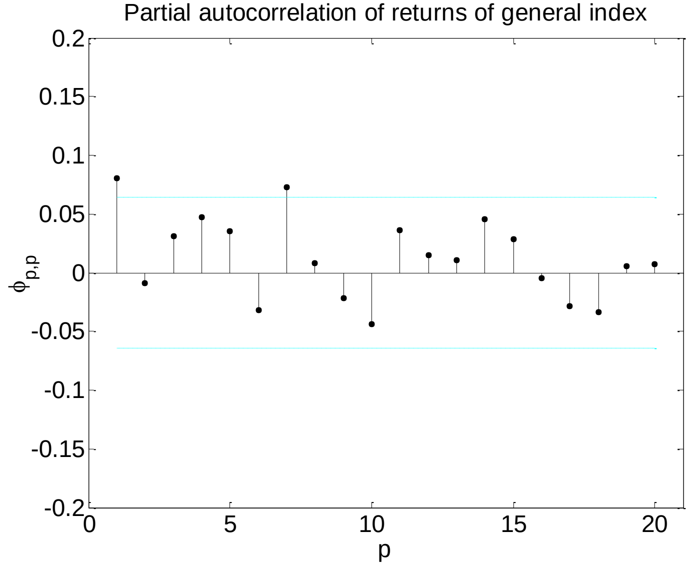
Απόδοση $x_t = \dfrac{y_t - y_{t-1}}{y_{t-1}}$



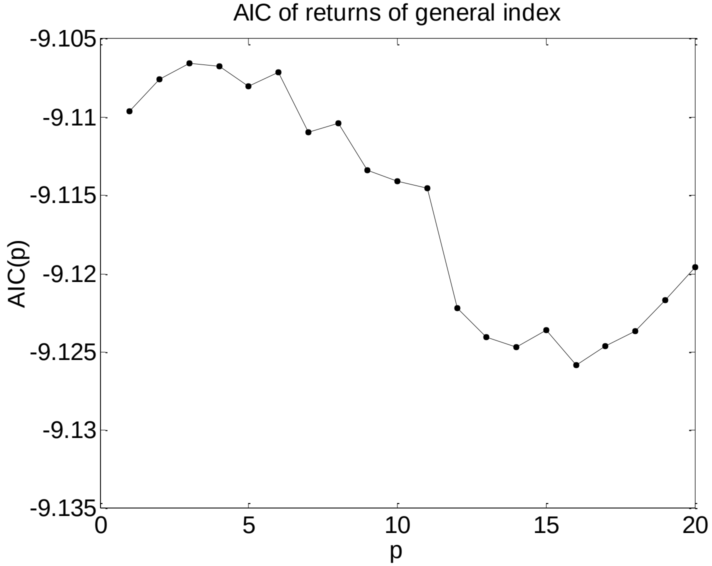


Γενικός δείκτης Χρηματιστηρίου Αθηνών Περίοδος Ιανουάριος 2002 – Σεπτέμβριος 2005

Τάξη AR μοντέλου

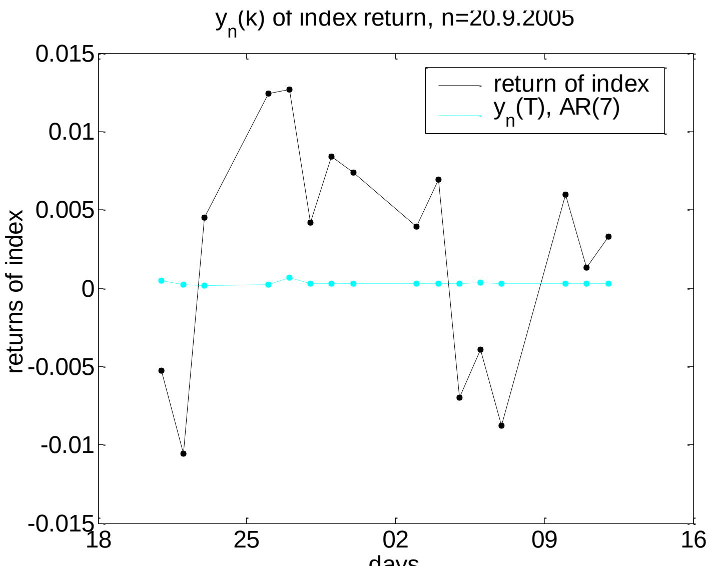


$$x_t = \frac{y_t - y_{t-1}}{y_{t-1}}$$



Πρόβλεψη πολλών βημάτων με αφετηρία 20.9.2005

αποδόσεις γενικού δείκτη



$$y_t = y_{t-1}(1 + x_t)$$

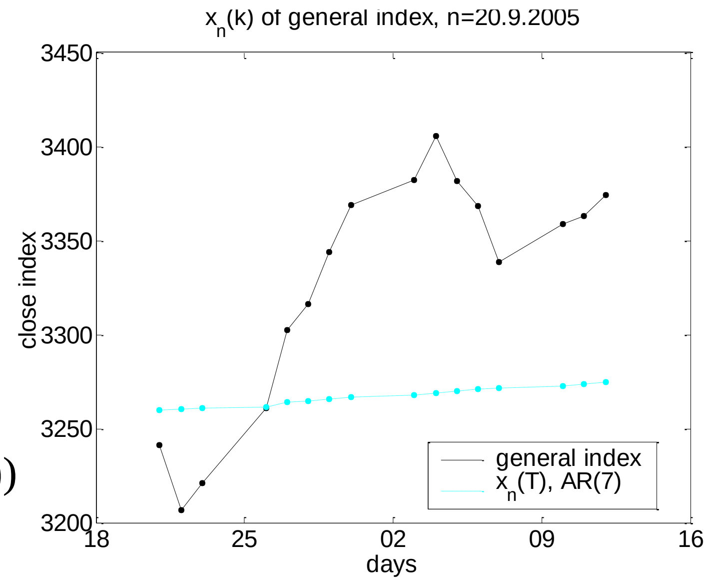
$$y_{n+1} = y_n(1 + x_{n+1})$$

Πρόβλεψη

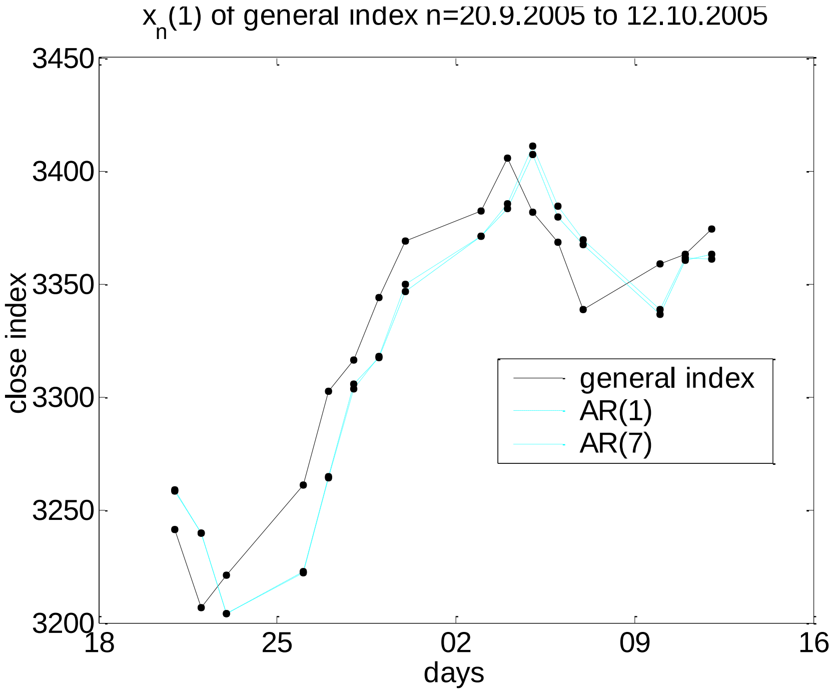
$$y_n(1) = y_n(1 + x_n(1))$$

$$y_n(k) = y_n(k-1)(1 + x_n(k))$$

γενικός δείκτης



Πρόβλεψη ενός βήματος την περίοδο 20.9.2005 – 12.10.2005



Εκτίμηση του σφάλματος πρόβλεψης με AR(*p*) μοντέλα για την περίοδο 20.9.2005 – 12.10.2005