



Διατμηματικό πρόγραμμα μεταπτυχιακών σπουδών “ΥΔΡΑΥΛΙΚΗ ΜΗΧΑΝΙΚΗ”

ΕΦΑΡΜΟΣΜΕΝΗ ΑΝΑΛΥΣΗ ΥΔΡΟΛΟΓΙΚΩΝ ΧΡΟΝΙΚΩΝ ΣΕΙΡΩΝ

Σπηλιώτης Μιχάλης και Κιτσικούδης Β.

ARMA, ARIMA, SARIMA (θεωρία και
εφαρμογές)

ΞΑΝΘΗ, 2020

Μη αιτιατά υποδείγματα πρόβλεψης

2

- Πρόβλεψη αποκλειστικά με βάση τις προηγούμενες τιμές:
 - Καθοριστικά υποδείγματα: ο τυχαίος παράγοντας εμφανίζεται ως κατάλοιπο λάθους (πρώτη εφαρμογή)
 - **Στοχαστικά υποδείγματα: Τυχαίος παράγοντας αποτελεί μηχανισμό μέσα από τον οποίο δημιουργείται η χρονοσειρά (Δημέλη, 2002)**

Λευκός Θόρυβος

3

- Κανένα ευκρινές σχήμα, ε_t

$$E(\varepsilon_t) = 0$$

$$\gamma_0 = E(\varepsilon_t^2) = \sigma^2 (\neq 0)$$

$$\gamma_k = E(\varepsilon_t \varepsilon_{t-k}) = 0, \text{ για } k \neq 0$$

- Χρήση τυχαίων αριθμών

Ανασκόπηση 1^{ου} μαθήματος

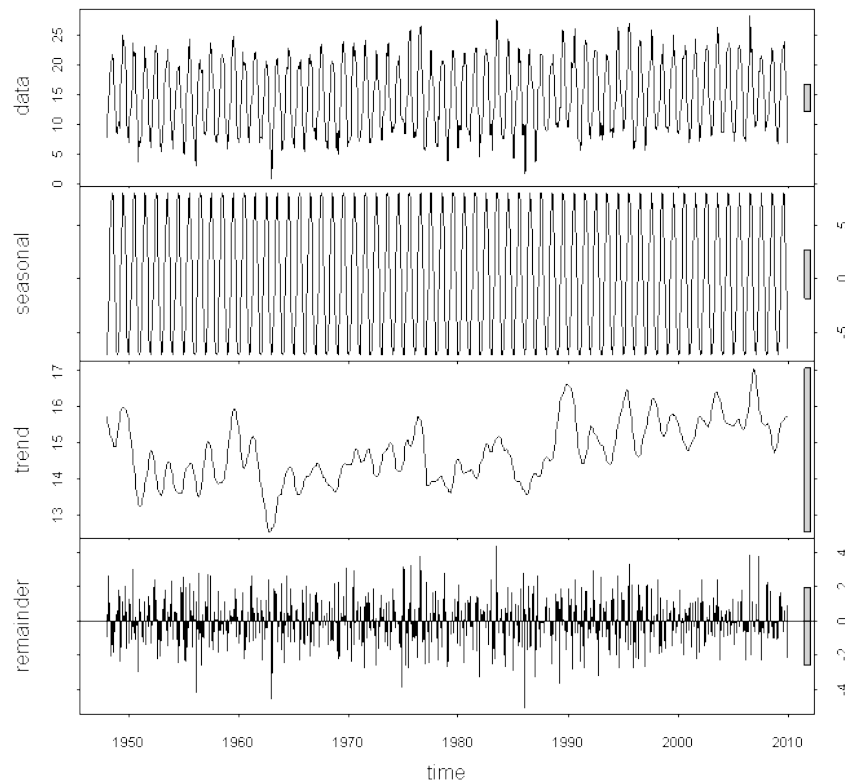
4

- Μοντελοποίηση μονομεταβλητών χρονικών σειρών με τη χρήση του κινητού μέσου όρου για την εξομάλυνσή τους.
- Προσδιορισμός της τάσης με πολλαπλή γραμμική παλινδρόμηση και αποδόμηση της χρονικής σειράς σε συνιστώσες τάσης, εποχικότητας και θορύβου.
- Μοντελοποίηση των παραπάνω συνιστωσών και επανασύνθεση σε μία χρονική σειρά που θα χρησιμοποιηθεί για μελλοντικές προβλέψεις.



Ανασκόπηση 1^{ου} μαθήματος (συνέχεια)

5



- Αποσύνθεση μιας χρονικής σειράς, που παρουσιάζει την θερμοκρασία στο Heathrow, στις συνιστώσες της.

6

Βασικές γνώσεις

Μερικοί απλοί τελεστές

7

- Ένας τελεστής που χρησιμοποιείται ευρύτατα στη μελέτη των χρονικών σειρών είναι ο “οπισθοδρομικός” (backward shift) τελεστής (B) ο οποίος ορίζεται ως:

$$Bz_t = z_{t-1}$$

και γενικότερα:

$$B^j z_t = z_{t-j}$$

- Η αντίστροφη διαδικασία υλοποιείται μέσω ενός αντίστοιχου τελεστή, του “πρόσω” (forward shift) τελεστή (F) που ορίζεται αντίστοιχα ως:

$$Fz_t = z_{t+1}$$

και γενικότερα:

$$F^j z_t = z_{t+j}$$

όπου $z_1, z_2, \dots, z_t, \dots, z_N$ η χρονική σειρά.



Μερικοί απλοί τελεστές (συνέχεια)

8

- Ένας ακόμα τελεστής είναι ο “οπισθοδρομικός τελεστής διαφοράς” (backward difference operator) ο οποίος ορίζεται ως:

$$\nabla z_t = z_t - z_{t-1} = (1 - \mathbf{B})z_t$$

Αυτοδιακύμανση

9

γ_k = αυτοδιακύμανση μεταξύ παρατηρήσεων που απέχουν χρόνο k :

$$\begin{aligned}\gamma_k &= \text{cov}(y_t, y_{t+k}) = \\ &= E[(y_t - \bar{y}_t)(y_{t+k} - \bar{y}_{t+k})]\end{aligned}$$

- για στατική χρονοσειρά

$$\bar{y}_t = \bar{y}_{t+k} (= E(y_{t+k})) = \mu_y$$

άρα
$$\gamma_k = \text{cov}(y_t, y_{t+k}) = E[(y_t - \mu_t)(y_{t+k} - \mu_t)]$$

Για $k=0$, τετράγωνο τυπική απόκλιση
 $\gamma_0 = \sigma_y^2$

Συντελεστής αυτοσυσχέτισης

10

$$\rho_k = \frac{\text{COV}(y_t, y_{t+k})}{\sqrt{\text{Var}(y_t)} \sqrt{\text{Var}(y_{t+k})}} =$$
$$= \frac{\text{COV}(y_t, y_{t+k})}{\text{Var}(y_t)} \quad \left(\begin{array}{l} \text{"} \\ \sigma_y \end{array} \quad \begin{array}{l} \text{"} \\ \sigma_y \end{array} \text{ σταθερά} \right)$$

$$\Rightarrow \rho_k = \frac{\gamma_k}{\gamma_0}$$

Επομένως υπολογίζεται ως:

$$\rho_k = \frac{\gamma_k}{\gamma_0} = \frac{\sum_{t=1}^{N-k} (y_t - \bar{y})(y_{t+k} - \bar{y})}{\sum_{t=1}^N (y_t - \bar{y})^2}$$

ΣΤΑΣΙΜΗ ΧΡΟΝΟΣΕΙΡΑ

Αυτοσυσχετίσεις και μερικές αυτοσυσχετίσεις

- Η εκτίμηση της (k) τάξης υστέρησης (lag) αυτοσυσχέτισης είναι:

$$r_k = \frac{c_k}{c_0} \quad c_k = \frac{1}{N} \sum_{t=1}^{N-k} (z_t - \bar{z})(z_{t+k} - \bar{z})$$

Για μια ικανοποιητική προσέγγιση των r_k χρειάζονται τουλάχιστον $N=50$ παρατηρήσεις, όταν δεν υπάρχει εποχικότητα, και το κατάλληλο πλήθος των αυτοσυσχετίσεων είναι $\approx N/4$

- Η έννοια των μερικών αυτοσυσχετίσεων είναι εντελώς αντίστοιχη εκείνης που χρησιμοποιείται στην πολ/πλή παλινδρόμηση. Το ερώτημα της τάξης του αυτοπαλινδρομικού μοντέλου (το p) είναι ανάλογο προς εκείνο του πλήθους των ανεξάρτητων μεταβλητών.

$$\bar{z}_t = \phi_1 \bar{z}_{t-1} + \phi_2 \bar{z}_{t-2} + \dots + \phi_p \bar{z}_{t-p}$$

Μερική Αυτοσυσχέτιση

2.5 Μερική Αυτοσυσχέτιση

Μια άλλη συνάρτηση που χρησιμοποιείται στη μελέτη των χαρακτηριστικών μιας χρονολογικής σειράς είναι η συνάρτηση μερικής αυτοσυσχέτισης. Γενικά ο συντελεστής μερικής συσχέτισης $r_{xy; z_1, \dots, z_v}$ μετρά τη συσχέτιση μεταξύ των δύο μεταβλητών x και y όταν έχει αφαιρεθεί η επίδραση που ασκούν άλλες μεταβλητές έστω οι z_1, \dots, z_v πάνω σ' αυτές. Στην περίπτωση μιας χρονολογικής σειράς, ο συντελεστής μερικής αυτοσυσχέτισης μεταξύ y_t και y_{t+k} ορίζεται ως ο συντελεστής συσχέτισης μεταξύ τους όταν έχουν ληφθεί υπόψη οι συσχετίσεις όλων των ενδιάμεσων τιμών $y_{t+1}, y_{t+2}, \dots, y_{t+k-1}$.

Η έννοια της μερικής αυτοσυσχέτισης μπορεί να εξηγηθεί καλύτερα μέσα από την παλινδρόμηση. Ας θεωρήσουμε την παλινδρόμηση της y_t πάνω στις y_{t-1} και y_{t-2} :

$$y_t = \phi_{12}y_{t-1} + \phi_{22}y_{t-2} + \varepsilon_t \quad (2.7)$$

όπου ο πρώτος υποδείκτης του ϕ δηλώνει τη χρονική υστέρηση της μεταβλητής (1 για την y_{t-1}) και ο δεύτερος τη μέγιστη τάξη της παλινδρόμησης (2 στη συγκεκριμένη περίπτωση). Τότε ο συντελεστής της y_{t-2} μετρά το συντελεστή μερικής αυτοσυσχέτισης δευτέρας τάξεως (ϕ_{22}) καθόσον αυτός δείχνει τη συσχέτιση μεταξύ των y_t και y_{t-2} όταν έχει συμπεριληφθεί στην παλινδρόμηση η ενδιάμεση y_{t-1} .

Γενικότερα, ο συντελεστής μερικής αυτοσυσχέτισης p -οστής τάξεως συμβολίζεται με το γράμμα Φ_{pp} και είναι ο συντελεστής του y_{t-p} στην παλινδρόμηση

Δημέλη, 2002

$$y_t = \Phi_{1p}y_{t-1} + \Phi_{2p}y_{t-2} + \dots + \Phi_{pp}y_{t-p} + \varepsilon_t \quad (2.8)$$

Είναι προφανές ότι ο συντελεστής πρώτης τάξης Φ_{11} ταυτίζεται με τον απλό συντελεστή αυτοσυσχέτισης ρ_1 .

Στην πράξη εκτιμούμε υποδείγματα της μορφής (2.8), ξεκινώντας έχοντας μια χρονική υστέρηση του y_t και προσθέτοντας διαδοχικά από μια υστέρηση κάθε φορά. Οι εκτιμήσεις ελαχίστων τετραγώνων των συντελεστών Φ_{ss} σε κάθε τέτοιο υπόδειγμα για $s = 1, 2, \dots, p$ μας δίνει τη σειρά των μερικών αυτοσυσχετίσεων $\hat{\Phi}_{11}, \hat{\Phi}_{22}, \hat{\Phi}_{33}, \dots, \hat{\Phi}_{pp}$.

Έτσι, για παράδειγμα, οι τρεις πρώτοι συντελεστές προκύπτουν από τις εκτιμήσεις των υποδειγμάτων:

$$\begin{aligned} y_t &= \Phi_{11}y_{t-1} + \varepsilon_{1t} && \rightarrow \hat{\Phi}_{11} \\ y_t &= \Phi_{12}y_{t-1} + \Phi_{22}y_{t-2} + \varepsilon_{2t} && \rightarrow \hat{\Phi}_{22} \\ y_t &= \Phi_{13}y_{t-1} + \Phi_{23}y_{t-2} + \Phi_{33}y_{t-3} + \varepsilon_{3t} && \rightarrow \hat{\Phi}_{33} \end{aligned} \quad (2.9)$$

Οι τιμές των Φ_{ss} για διάφορες τιμές του $s = 1, 2, \dots$ αποτελούν τη **συνάρτηση μερικής αυτοσυσχέτισης (partial autocorrelation function)** και συμβολίζεται με τα γράμματα PACF ή απλώς PAC. Η μορφή της συνάρτησης αυτής καθώς και της απλής συνάρτησης αυτοσυσχέτισης (ACF) είναι εξαιρετικά χρήσιμες καθόσον μας καθοδηγούν στον προσδιορισμό της μορφής της стоχαστικής διαδικασίας που δημιούργησε τη δεδομένη χρονολογική σειρά όπως θα δούμε παρακάτω.

Στο σημείο αυτό είναι χρήσιμο να μελετήσουμε μερικές απλές μορφές χρονολογικών σειρών πριν προχωρήσουμε στη μελέτη των πιο πολύπλοκων μορφών.

Αυτοσυσχετίσεις και μερικές αυτοσυσχετίσεις (συνέχεια)

14

- Η εκτίμηση της (k) τάξης υστέρησης (lag) αυτοσυσχέτισης είναι:

$$r_k = \frac{c_k}{c_0} \quad c_k = \frac{1}{N} \sum_{t=1}^{N-k} (z_t - \bar{z})(z_{t+k} - \bar{z})$$

- Παράδειγμα: Εκτίμηση αυτοσυσχέτισης πρώτης τάξης (r_1) των 10 πρώτων παρατηρήσεων μιας χρονικής σειράς:

$$z_t: 47,64,23,71,38,64,55,41,59,48 \rightarrow \bar{z} = 51$$

$$z_t - \bar{z}: -4,13,-28,20,-13,13,4,-10,8,-3$$

$$\sum_{i=1}^9 (z_t - \bar{z})(z_{t+1} - \bar{z}) = (-4)(13) + (13)(28) + \dots + (8)(-3)$$

$$= -1497 \rightarrow c_1 = \frac{-1497}{10} = -149.7$$

$$c_0 = 189.6$$



$$r_1 = \frac{-149.7}{189.6} = -0.79$$



Μονιμότητα χρονοσειράς

15

• Μονιμότητα Χρονοσειράς:

$$\bar{X}_t = \mu = \frac{1}{N} \sum_{i=1}^N X_i(t) \xrightarrow{\text{σειρά}} \bar{X}_{t+T} = \frac{1}{N} \sum_{i=1}^N X_i(t+T)$$

ή ισοδύναμα:

αν η υστέρηση καθορίζει τη διακύμανση και δx
ο χρόνος t :

$$\text{Cov}(X_t, X_{t+T}) = \frac{1}{N} \sum_{i=1}^N X_i(t) X_i(t+T) - \mu \bar{X}_{t+T}$$

- Συνήθως, οι χρονοσειρές μου, αναφέρονται οφεί έχει αφαιρεθεί ο μείκτος όρος.
- AR, MA, ARMA, στατικές χρονοσειρές
- Για μη στατικές χρονοσειρές.
Διάφοροι τρόποι μονιμοποίησης. Ένας από αυτούς είναι η διαφορά τάξης d :

Τάξης 1:

$$\nabla Z_t = Z_t - Z_{t-1} = (1-B)Z_t.$$

Τάξης 2:

$$\nabla^2 Z = \nabla \cdot \nabla Z_t = \nabla (Z_t - Z_{t-1}) =$$

$$= \nabla Z_t - \nabla Z_{t-1} = Z_t - Z_{t-1} - Z_{t-1} + Z_{t-2} =$$

$$= Z_t - 2Z_{t-1} + Z_{t-2}$$

Άρα: $\nabla^2 Z_t = (1-B)^2 Z_t$

Τρόπος
μονιμοποίησης
της
χρονοσειρών

Μόνιμη χρονοσειρά (έως και ARMA)

Μόνιμη χρονοσειρά

18

- Αυτοπαλίνδρομα μοντέλα, AR
- Μοντέλα κινητού μέσου όρου, MA
- Μεικτά ARMA
- Μονιμοποίηση χρονοσειρών με διαφορές (μετάβαση στην επομένη ενότητα)

Αυτοπαλίνδρομο μοντέλο AR

Το αυτοπαλινδρομικό μοντέλο

- Ένα στοχαστικό μοντέλο που αποδεικνύεται εξαιρετικά χρήσιμο στην απεικόνιση πλείστων περιπτώσεων χρονικών σειρών που παρουσιάζονται στην πράξη, είναι το λεγόμενο **αυτοπαλινδρομικό μοντέλο** (autoregressive model).

$$\bar{z}_t = \varphi_1 \bar{z}_{t-1} + \varphi_2 \bar{z}_{t-2} + \dots + \varphi_p \bar{z}_{t-p} + \alpha_t$$

όπου \bar{z}_t είναι η απόκλιση της αντίστοιχης τιμής από τη μέση τιμή (μ), π.χ. $\bar{z}_t = z_t - \mu$.

- Η τιμή \bar{z}_t μιας εξαρτημένης μεταβλητής συνδέεται γραμμικά με τις αντίστοιχες τιμές p ανεξάρτητων μεταβλητών και ενός όρου σφάλματος α_t .

Το αυτοπαλινδρομικό μοντέλο (συνέχεια)

21

- Αν θεωρηθεί ο αυτοπαλινδρομικός τελεστής τάξης p που ορίζεται ως:

$$\varphi(\mathbf{B}) = 1 - \varphi_1 \mathbf{B} - \varphi_2 \mathbf{B}^2 - \dots - \varphi_p \mathbf{B}^p$$

τότε το αυτοπαλινδρομικό μοντέλο μπορεί συνεπτυγμένα να γραφεί ως:

$$\varphi(\mathbf{B})\bar{z}_t = \alpha_t$$

και είναι ισοδύναμο με:

$$\bar{z}_t = \varphi_1 \bar{z}_{t-1} + \varphi_2 \bar{z}_{t-2} + \dots + \varphi_p \bar{z}_{t-p} + \alpha_t$$

από τα προηγούμενα



AR(1)

$$Y_t = \phi_1 Y_{t-1} + \delta + \varepsilon_t.$$

$$E(Y_{t+1}) = \phi_1 E(Y_{t+1}) + \delta$$

$$\Leftrightarrow E(Y_{t+1}) = \frac{\delta}{1 - \phi_1} = \mu.$$

$$\text{άρα όταν } E(Y_{t+1}) = 0 \rightarrow \delta = 0$$

Συνήθως επιδίωξη

$$\tilde{Y}_t = Y_t - E(Y_{t+1}) = Y_t - \frac{\delta}{1 - \phi_1}$$

Απο αυτό. Αφαιρώντας το μέσο όρο, δεν θα υπάρξει σταθερός όρος, $\delta = 0$.

$$\tilde{Y}_t = \phi_1 \tilde{Y}_{t-1} + \varepsilon_t.$$

Έχοντας αφαιρέσει το μέσο όρο (όπως συμβαίνει δεν θα εμφανιστεί ο όρος δ)

Αυτοπαλίνδρομο μοντέλο δεύτερης τάξης AR(2)

23

Το **Αυτοπαλίνδρομο υπόδειγμα δεύτερης τάξης AR(2)** είναι της μορφής:

$$Y_t = \phi_0 + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \varepsilon_t$$

ή

$$y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + \varepsilon_t$$

Δεν το πολυχρησιμοποιούμε, συνήθως αφαιρούμε το μέσο όρο και πάμε στο παρακάτω μοντέλο

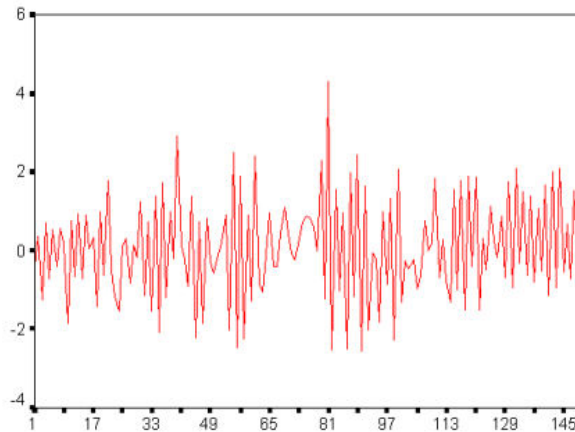
Αν εφαρμοσθεί ο **τελεστής υστερήσεως (lag operator) B** η σχέση γράφεται και ως εξής:

$$(1 - \phi_1 B - \phi_2 B^2) y_t = \varepsilon_t$$

Λουκάς, 2015

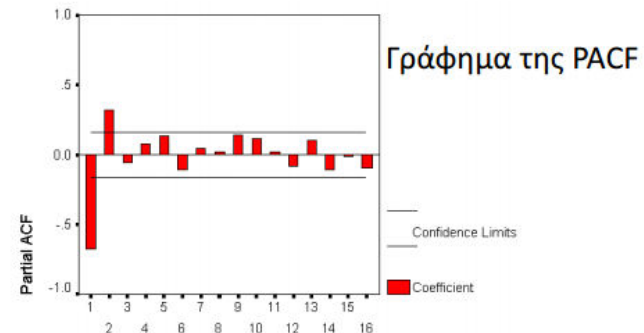
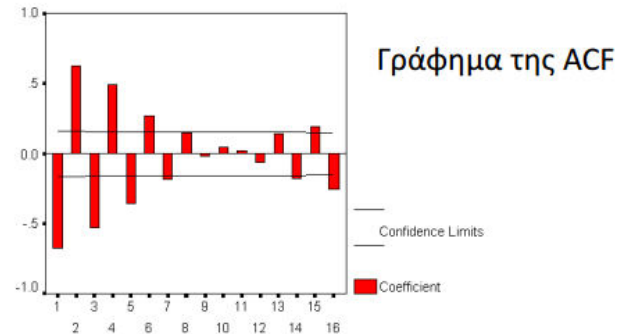
Αυτοπαλίνδρομο μοντέλο δεύτερης τάξης AR(2)

Παράδειγμα 4: Με προσομοίωση φτιάξαμε 150 παρατηρήσεις από μια AR(2) χ.σ. $(1+0.5B-0.3B^2)z_t = a_t$. Η σειρά έχει το παρακάτω γράφημα:



Παρατηρήσεις:

- το correlogram της σειράς παρουσιάζει μείωση εκθετική ενώ τα πρόσυμα των αυτοσυσχετίσεων πηγαίνουν εναλλάξ
- στο γράφημα των ϕ_{kk} σημαντικές είναι μόνο οι ϕ_{11} και ϕ_{22}



Αυτοπαλίνδρομα μοντέλα

Εξ. Yule-Walker

25

Εξίσωση Yule-Walker:
(μόνο για AR(p)):

$$\tilde{Y}_t = \phi_1 \tilde{Y}_{t-1} + \phi_2 \tilde{Y}_{t-2} + \dots + \phi_p \tilde{Y}_{t-p} + \varepsilon_t$$

$$\tilde{Y}_t \cdot \tilde{Y}_{t-k} = \phi_1 \tilde{Y}_{t-1} \cdot \tilde{Y}_{t-k} + \dots + \phi_p \tilde{Y}_{t-p} \cdot \tilde{Y}_{t-k} + \varepsilon_t \tilde{Y}_{t-k}$$

↙ Μέσοι όροι:

$$E(\tilde{Y}_t \cdot \tilde{Y}_{t-k}) = \phi_1 E(\tilde{Y}_{t-1} \cdot \tilde{Y}_{t-k}) + \phi_p E(\tilde{Y}_{t-p} \cdot \tilde{Y}_{t-k}) + O(\varepsilon)$$

Σχόλιο

$$E(\varepsilon_t \tilde{Y}_{t-k}) = 0$$

\tilde{Y}_{t-k} , είναι συνάρτηση των λευκών θορύβων
έως $t-k$, ενώ οι επόμενες διατεταγμένες ~~ε~~
δεν συχίζονται.

Σχόλιο

28

- Ο λευκός θόρυβος έχει μέσο όρο μηδέν αλλά τυπική απόκλιση που εξαρτάται από τη διασπορά της στάσιμης χρονοσειράς
- Για παράδειγμα σε AR(1) ισχύει:

$$\begin{aligned} \text{Var}(\tilde{y}_t^2) &= E(\varepsilon_t + \phi_1 \varepsilon_{t-1} + \phi_1^2 \varepsilon_{t-2} + \dots)^2 \\ &= \sigma_\varepsilon^2 (1 + \phi_1^2 + \phi_1^4 + \dots) \\ &\Rightarrow \gamma_0 = \text{Var}(y_t) = \frac{\sigma_\varepsilon^2}{1 - \phi_1^2} \end{aligned}$$

$$\rightarrow \sigma_\varepsilon^2 = \text{Var}(y_t) \cdot (1 - \phi_1^2)$$

Πιο γενική απόδειξη τυπική απόκλιση σε αυτοπαλίνδρομα μοντέλα

29

Σχέση τυπική απόκλισης σ_{y_t} του y_t και σ_{a_t} της a_t .

Ισχύει

$$\tilde{y}_t = a_t + \phi_1 \tilde{y}_{t-1} + \dots + \phi_p \tilde{y}_{t-p}$$

Πολλαπλασιάζουμε με \tilde{y}_t και παίρνουμε το μέσο όρο:

$$E(\tilde{y}_t^2) = E(a_t \tilde{y}_t) + \phi_1 E(\tilde{y}_t \tilde{y}_{t-1}) + \dots + \phi_p E(\tilde{y}_t \tilde{y}_{t-p})$$

$$\Leftrightarrow \sigma_{y_t}^2 = \phi_1 \gamma_1 + \dots + \phi_p \gamma_p + E(a_t \tilde{y}_t)$$

$$E(a_t \tilde{y}_t) = \sigma_{a_t}^2, \text{ εφόσον η } \tilde{y}_t$$

συνδέεται μόνο με \tilde{a}_t

$$E(y_t a_t) = E((\phi_1 y_{t-1} + \dots + \phi_p y_{t-p} + a_t) a_t) = E(a_t^2) = \sigma_{a_t}^2$$

Τυπική απόκλιση λευκού θορύβου

30

□ Διαιρώντας με $\sigma_y^2 = \gamma_0$

$$\sigma_y^2 = \varphi_1 \gamma_1 + \dots \varphi_p \gamma_p + \sigma_a^2 \Leftrightarrow$$

$$1 = \varphi_1 \frac{\gamma_1}{\gamma_0} + \dots \varphi_p \frac{\gamma_p}{\gamma_0} + \frac{\sigma_a^2}{\sigma_y^2} \left(\sigma_y^2 = \gamma_0 \text{ συμβολισμός} \right) \Leftrightarrow$$

$$1 = \varphi_1 \rho_1 + \dots \varphi_p \rho_p + \frac{\sigma_a^2}{\sigma_y^2} \Leftrightarrow$$

$$\sigma_a^2 = \sigma_y^2 \left(1 - \varphi_1 \rho_1 - \dots \varphi_p \rho_p \right)$$

Αυτοπαλίνδρομα μοντέλα πρώτης τάξεως

Αυτοπαλίνδρομα μοντέλα πρώτης τάξεως

32

Μόνο για $AR(L)$

$$\rho_1 = \phi_L$$

$AR(1)$

$\rho(1)$ σημαντική

εκθ. μίσθω

$\phi_1 \neq 0$

$\phi_2 = \phi_3 = \dots = 0$

Από rule-walkover:

$$\rho_2 = \phi_1 \rho_1 + \cancel{\phi_2} \rho_1^2 = \rho_1^2$$

$$\rho_3 = \phi_1 \rho_2 + 0 + \dots = \phi_1 \rho_1^2 = \rho_1^3$$

Ομοίως:

$$\rho_k = \phi_1^k$$

$$\sigma_z^2 = \frac{\sigma_a^2}{1 - \phi_1^2}$$

Μόνιμα Μοντέλα AR(q)

γενικό υπόδειγμα:

$$Y_t = \delta + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \dots + \phi_q Y_{t-q} + \varepsilon_t$$

↑
-

υπόδειγμα ανανδρόμησης αν αριθμό τιμών + δυνάμεις

Θόρυβος.

(Η εφεπτηρήνη τιμή Y_t ανανδρόμείται, αν- τις προηγούμενες τιμές της ίδιας μεταβλητής + τυχαίο όρος σφάλματος.

Παράδειγμα (Δηρίδης, 2002)

$$y_t = 0.8y_{t-1} + \varepsilon_t,$$

$$\sigma_\varepsilon^2 = 4.$$

Εστω η
χρονοσειρά.

Συμπερίληψη: $AR(1)$

$$\phi_1 = 0.8.$$

$$\text{var}(y_t) = \gamma_0 = \frac{\sigma_\varepsilon^2}{1 - \phi_1^2} = \frac{4}{1 - 0.8^2} = 11.11.$$

$$\gamma_1 = \phi_1 \gamma_0$$

$$\gamma_2 = \phi_1^2 \gamma_0 = 0.8^2 \cdot 11.11 = 7.11 \dots$$

γ_{k+1}

$$\gamma_k = \phi_1^k \cdot \gamma_0 \text{ (Μόνο για } AR(1)\text{)}$$

Μιμίκου, επαλήθευση με χρήση τυχαίων αριθμών

35

ΠΙΝΑΚΑΣ 10. Έφαρμογή μοντέλου AR(1)

i	q_i	$q_i - \bar{x}$	$r_1(q_i - \bar{x})$	$\bar{x} + r_1(q_i - \bar{x})$	t_i	$t_i s \sqrt{1 - r_1^2}$	\hat{q}_{i+1}
0	588.80	-5.32	-2.01	592.11	-0.523	-62.61	529.50
1	505.20	-88.92	-33.63	560.49	0.611	73.25	633.74
2	654.85	60.73	22.97	617.09	-0.359	-43.04	574.05
3	556.40	-37.72	-14.27	579.85	-0.393	-47.12	532.73
4	513.73	-80.39	-30.40	563.72	0.084	10.07	573.79
5	573.83	-20.29	-7.67	586.45	-0.931	-111.61	474.84
6	434.33	-159.79	-60.43	533.69	-0.027	-3.24	530.45
7	526.06	-68.06	-25.74	568.38	0.798	95.66	664.04
8	692.63	98.51	37.26	631.38	1.672	200.45	831.83
9	895.32	301.20	113.91	708.03	-1.077	-129.11	578.92

όπου \bar{x} η εκτίμηση από το δείγμα του μέσου όρου μ , $\bar{x}=594.12$
 r_1 η εκτίμηση του ρ_1 , $r_1=0.37819$, και S η εκτίμηση του σ ,
 $S=129.32$, t_i είναι η κανονική τυχαία μεταβλητή $[0,1]$ και η
 τελευταία στήλη του πίνακα δίνει τις συνθετικές ροές, που
 φαίνονται να είναι πολύ κοντά στις πραγματικές με περίπου ίδιο
 μέσο όρο, τυπική απόκλιση και συντελεστή αυτοσυσχετίσεως πρώτης
 τάξεως.

Εφαρμοσμένη ανάλυση υδρολογικών χρονικών σειρών

Fiering, πάνω σε αυτοπαίνδρομα μοντέλα,
διάλλειμα, χρήση τυχαίων αριθμών

Για ετήσια μεγέθη

Έστω τη σειρά ανεξαρτήτων
Τυχαίων αριθμών ϵ_i με μέσο όρο μηδέν
και συνάρτηση από κεντρική $Z_i \sim N(0,1)$

Τότε το $t_t \sigma \sqrt{1-\rho_1^2}$ είναι

μετεβλητή ανεξάρτητη ϵ_i με μέσο όρο μηδέν
και συνάρτηση από κεντρική $\sigma \sqrt{1-\rho_1^2}$.

όσο t_t είναι ο t -όρος t -όρος.

$$q_i - \mu = \rho_1 (q_{i-1} - \mu) + t_i \sigma \sqrt{1-\rho_1^2}$$

συνολική χρονοσυνάρτηση

Χρήση τυχαίων
αριθμών

Ετήσιες χρονοσειρές

38

Μοντέλο Fienberg. \rightarrow AR(2)

\rightarrow Μήσοι όροι

\rightarrow Ετήσιο

\rightarrow Μηνιαίο.

$$\begin{aligned}
 & \text{Ετήσιο:} \\
 & \tilde{y}_t - \mu = \rho_L \cdot (\tilde{y}_{t-1} - \mu) + \epsilon_t \cdot \sigma \sqrt{(1-\rho_L^2)} \\
 & \quad \quad \quad \uparrow \quad \quad \quad \uparrow \\
 & \quad \quad \quad \text{τυχαίος} \quad \quad \quad \sigma \epsilon. \\
 & \quad \quad \quad \text{αριθμός} \quad \quad \quad \text{σφάλμα}
 \end{aligned}$$

$\Rightarrow \epsilon_t = \text{τυχαίος αριθμός } N(0,1).$

\Rightarrow ε να δατασει ωπονολη γύηηη
 $(\gamma_{12} = \frac{1 - \rho^3}{(1 - \rho^2)^{1.5}} \gamma^*).$

Πρόβλεψη με
 τυχαίους αριθμούς

Μηνιαίες χρονοσειρές

39

$$X_{i,j+1} = \mu_{j+1} + \rho_j \frac{\sigma_{j+1}}{\sigma_j} (X_{ij} - \mu_j) + t_{i,j+1} \sigma_{j+1} \sqrt{1 - \rho_j^2}$$

ρ_j is serial correlation between flows of j^{th} month and $j+1^{\text{th}}$ month.

$t_{i,j+1} \sim N(0, 1)$

Πρόβλεψη με
τυχαίους αριθμούς

Όχι ακριβώς AR

Μοντέλα κινητού μέσου όρου

Το μοντέλο κινητού μέσου όρου

41

- Ένα άλλο ιδιαίτερα πρακτικού ενδιαφέροντος μοντέλο για διάφορες χρονικές σειρές της πράξης είναι το μοντέλο που απεικονίζει τη σειρά μέσα από μια πεπερασμένη ακολουθία (q) προηγούμενων τιμών λευκών θορύβων (α_t).

- Η έκφραση της μορφής

$$z_t = \alpha_t - \theta_1 \alpha_{t-1} - \theta_2 \alpha_{t-2} - \dots - \theta_q \alpha_{t-q}$$

καλείται διαδικασία κινητού μέσου όρου (q) τάξης.

- Αν όπως και προηγουμένως οριστεί ένας κινητού μέσου όρου τελεστής τάξης (q):

$$\theta(\mathbf{B}) = 1 - \theta_1 \mathbf{B} - \theta_2 \mathbf{B}^2 - \dots - \theta_q \mathbf{B}^q$$

τότε το αντίστοιχο μοντέλο μπορεί να γραφεί ως: $\bar{z}_t = \theta(\mathbf{B})\alpha_t$



Υπόδειγμα κινητού μέσου όρου πρώτης τάξης MA(1)

Η μορφή του υποδείγματος
MA(1) είναι η ακόλουθη:

$$Y_t = \mu + \varepsilon_t + \theta_1 \varepsilon_{t-1}$$

ή

$$Y_t = \mu + \varepsilon_t - \theta_1 \varepsilon_{t-1}$$

ή

$$Y_t - \mu = y_t = \varepsilon_t + \theta_1 \varepsilon_{t-1}$$

Δεν το πολυχρησιμοποιούμε,
συνήθως αφαιρούμε το μέσο όρο
και πάμε στο παρακάτω μοντέλο

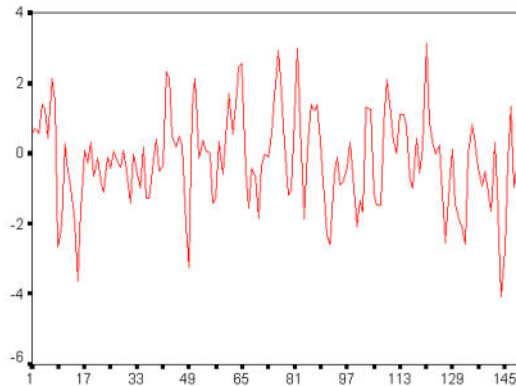
Αν χρησιμοποιήσουμε τον **τελεστή υστερήσεως (Lag operator) L** η ανωτέρω σχέση γίνεται:

$$y_t = (1 + \theta_1 L) \varepsilon_t$$



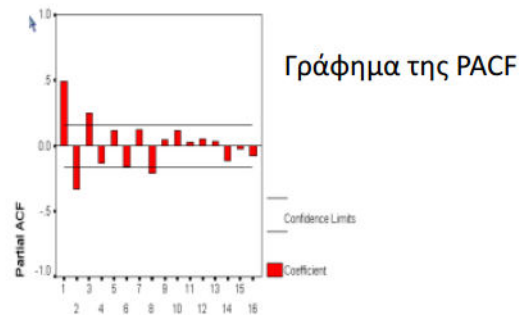
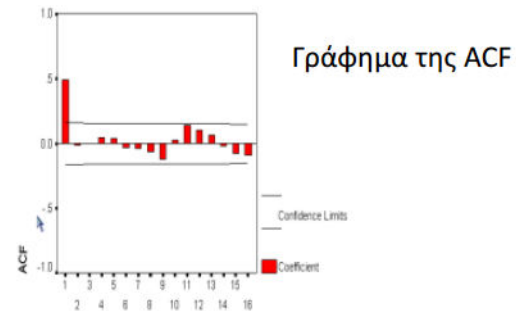
Κινούμενου μέσου μοντέλα τάξης q (MA(q))

Παράδειγμα 8: Με προσομοίωση φτιάξαμε 150 παρατηρήσεις από την MA(1)
χ.σ. $z_t = (1 + 0.8B)a_t$. Η σειρά έχει το παρακάτω γράφημα:



Παρατηρήσεις:

- ρ_1 σημαντική, ρ_k για $k > 1$ ασήμαντες
- φ_{kk} μειώνονται εκθετικά, με εναλλασσόμενο πρόσημο ($\theta_1 = -0.8 < 0$) ξεκινώντας από θετική τιμή



Συμπεριφορές

44

Πίνακας: Μορφές της συνάρτησης αυτοσυσχέτισης και μερικής αυτοσυσχέτισης

Διαδικασία	Αυτοσυσχετίσεις ACF	Μερικές αυτοσυσχετίσεις PACF
AR(1)	Φθίνουν γεωμετρικά από ρ_1	Μηδέν μετά το ρ_{11}
AR(2)	Φθίνουν γεωμετρικά από ρ_2	Μηδέν μετά το ρ_{22}
AR (p)	Φθίνουν γεωμετρικά ή φθίνουν με ημιτονοειδή συμπεριφορά από ρ_p	Μηδέν μετά το ρ_{pp}



MA(1)	Μηδέν μετά το ρ_1	Φθίνουν γεωμετρικά μετά από ρ_{11}
MA(2)	Μηδέν μετά το ρ_2	Φθίνουν γεωμετρικά μετά από ρ_{22}
MA(q)	Μηδέν μετά το ρ_q	Φθίνουν γεωμετρικά μετά από ρ_{qq}
ARMA (1,1)	Φθίνουν γεωμετρικά από ρ_1	Φθίνουν γεωμετρικά μετά από ρ_{11}
ARMA (p,q)	Φθίνουν γεωμετρικά από ρ_q	Φθίνουν γεωμετρικά μετά από ρ_{pp}



MA(1)

$$y_t = \varepsilon_t - \theta \varepsilon_{t-1}$$

$$\gamma_0 = \sigma_\varepsilon^2 (1 + \theta^2)$$

$$\gamma_1 = E(y_t y_{t-1}) =$$

$$= E((\varepsilon_t - \theta \varepsilon_{t-1})(\varepsilon_{t-1} - \theta \varepsilon_{t-2})) =$$

$$= \dots - \theta E(\varepsilon_{t-1}^2) = -\theta \sigma_\varepsilon^2$$

$$\rho_1 = \frac{\gamma_1}{\gamma_0} = \frac{-\theta \sigma_\varepsilon^2}{\sigma_\varepsilon^2 (1 + \theta^2)} =$$

$$\rho_1 = \frac{-\theta}{1 + \theta^2}$$

$$\rho_2, \rho_3, \dots = 0 \text{ since MA}(1).$$

ARMA

Μικτά αυτοπαλινδρομικά – κινητού μέσου
όρου μοντέλα

Μικτά αυτοπαλινδρομικά – κινητού μέσου όρου μοντέλα

48

- Προκειμένου να επιτευχθεί μεγαλύτερη ευελιξία στην προσαρμογή των παρατηρούμενων χρονικών σειρών είναι μερικές φορές σκόπιμο να περιληφθούν τόσο αυτοπαλινδρομικοί, όσο και κινητού μέσου όρου παράγοντες στο μοντέλο, ταυτόχρονα:

$$\bar{z}_t = \varphi_1 \bar{z}_{t-1} + \varphi_2 \bar{z}_{t-2} + \dots + \varphi_p \bar{z}_{t-p} + \alpha_t - \theta_1 \alpha_{t-1} - \theta_2 \alpha_{t-2} - \dots - \theta_q \alpha_{t-q}$$

ή ισοδύναμα:

$$\varphi(\mathbf{B})\bar{z}_t = \theta(\mathbf{B})\alpha_t$$

- Στην πράξη αποδεικνύεται ότι η παρουσίαση μόνιμων (stationary) χρονικών σειρών είτε με αυτοπαλινδρομικά μοντέλα μόνο, είτε κινητού μέσου όρου, είτε μικτής μορφής γίνεται με μοντέλα όπου τα (p) και (q) σπάνια είναι μεγαλύτερα του 2.

Σχόλιο:

49

- Σε μοντέλα ARMA, κατά τη συνήθη μορφή, χωρίς σταθερούς όρους πρέπει να έχω στάσιμες χρονοσειρές, συνήθως θεωρώ ότι επιτυγχάνεται αφαιρώντας το μέσο όσο (και διαιρώντας με την τυπική απόκλιση πολλές φορές). Επομένως δεν θεωρώ τις «πραγματικές χρονοσειρές» αλλά αφαιρώντας το μέσο όρο κλπ.
- Σε μοντέλα APIMA η μονιμοποίηση επιτυγχάνεται με τη d τάξεως διαφορά

- Στην στατιστική ανάλυση χρονικών σειρών το αυτοπαλινδρομικό κινητού μέσου όρου μοντέλο (**A**uto**R**egressive **M**oving **A**verage - **ARMA**) παρέχει την περιγραφή μίας μόνιμης (stationary) στοχαστικής διαδικασίας.
- Για μία δεδομένη χρονική σειρά, ένα μοντέλο **ARMA** αποτελεί εργαλείο για την κατανόηση της σειράς και ενδεχομένως για την πρόβλεψη μελλοντικών τιμών.
- Το μοντέλο **ARMA** πρωτοεμφανίστηκε από τον Whittle (1951) και έγινε ευρέως γνωστό από τους Box and Jenkins (1970).
- Το **ARMA** αποτελείται από δύο πολυώνυμα, ένα για την αυτοπαλινδρόμηση (**AR**) και ένα για τον κινητό μέσο όρο (**MA**).
- Το μοντέλο συνήθως αναφέρεται ως **ARMA** (p,q), όπου p είναι η τάξη της αυτοπαλινδρόμησης και q η τάξη του κινητού μέσου όρου.

Whittle, P. (1951). *Hypothesis Testing in Time Series Analysis*. Almqvist and Wicksell.

Box, George and Jenkins, Gwilym (1970) *Time series analysis: Forecasting and control*, San Francisco: Holden-Day.



ARIMA

ARIMA (μετάβαση στην επομένη ενότητα)

52

- Ένα μοντέλο **ARMA**(p,q) μπορεί να εφαρμοστεί σε μία χρονική σειρά που χαρακτηρίζεται ως στάσιμη.
- Σχεδόν πάντα οι φυσικές χρονικές σειρές παρουσιάζουν μία αξιολογη τάση, περιοδικότητα ή και τα δύο.



- Η απομάκρυνσή τους επιτυγχάνεται λαμβάνοντας τις διαδοχικές διαφορές της χρονικής σειρά (d) 1^{ης}, 2^{ης} ή 3^{ης} τάξης, μέχρι, δηλαδή, η χρονική σειρά να είναι κατά προσέγγιση στάσιμη (να μην εμφανίζει δηλαδή αξιολογη τάση και περιοδικότητα). ARIMA(p,d,q)

In general, an **ARIMA** (p, d, q) process can be written as

$$\phi(B)(1 - B)^d Y_t = \theta(B)e_t.$$

- $AR(p) \longleftrightarrow ARIMA(p, 0, 0)$
- $MA(q) \longleftrightarrow ARIMA(0, 0, q)$
- $ARMA(p, q) \longleftrightarrow ARIMA(p, 0, q)$
- $ARI(p, d) \longleftrightarrow ARIMA(p, d, 0)$
- $IMA(d, q) \longleftrightarrow ARIMA(0, d, q).$

ΠΙΝΑΚΑΣ 5.1

Σχήματα	Αυτοσυσχέτιση	Μερική Αυτοσυσχέτιση
MA	Αιχμές σε χρονικές υστερήσεις από 1 ως q και στη συνέχεια μηδενίζεται	Σταδιακή μείωση
AR	Σταδιακή μείωση σύμφωνα με το $\rho_j = \phi_1\rho_{j-1} + \dots + \phi_p\rho_{j-p}$	Αιχμές σε χρονικές υστερήσεις από 1 ως p και στη συνέχεια μηδενίζεται
ARMA	Ακανόνιστο σχήμα σε χρονικές υστερήσεις από 1 ως q και στη συνέχεια σταδιακή μείωση σύμφωνα με το $\rho_j = \phi_1\rho_{j-1} + \dots + \phi_p\rho_{j-p}$	Σταδιακή μείωση

Εποχικότητα

56

- Εντός του έτους με βήμα μηνιαίο (π.χ. θέμα μηνιαία τιμή: μέση ημερήσια παροχή)
- Εντός της εβδομάδας με βήμα 7 ή 5 (σε οικονομικά προβλήματα)
- Εβδομαδιαίο κλπ
- Τότε ορίζουμε και ένα μοντέλο εποχικότητας S

Μονο εποχικά μοντέλα

Καθαρά εποχικό μοντέλο

58

□ SARIMA(P, D, Q)_s

$$\Phi_P(B^s)(1-B^s)^D Y_t = \Theta_Q(B^s) a_t$$

$$\Phi_P(B^s) = 1 - \Phi_1 B^s - \Phi_2 B^{2s} - \dots - \Phi_P B^{sP}$$

$$\Theta_Q(B^s) = 1 - \Theta_1 B^s - \Theta_2 B^{2s} - \dots - \Theta_Q B^{sQ}$$

Εποχικό μοντέλο κινούμενου μέσου

59

όρου (μόνο)

**Tebbs , 2013. STAT 520
FORECASTING AND TIME
SERIES, Lecture Notes**

10.2.1 $MA(Q)_s$

TERMINOLOGY: Suppose $\{e_t\}$ is a zero mean white noise process with $\text{var}(e_t) = \sigma_e^2$. A seasonal moving average (MA) model of order Q with seasonal period s , denoted by $MA(Q)_s$, is

$$Y_t = e_t - \Theta_1 e_{t-s} - \Theta_2 e_{t-2s} - \cdots - \Theta_Q e_{t-Qs}.$$

A nonzero mean μ could be added for flexibility (as with nonseasonal models), but we take $\mu = 0$ for simplicity.

$MA(1)_{12}$: When $Q = 1$ and $s = 12$, we have

$$Y_t = e_t - \Theta e_{t-12}.$$



$MA(1)_{12}$

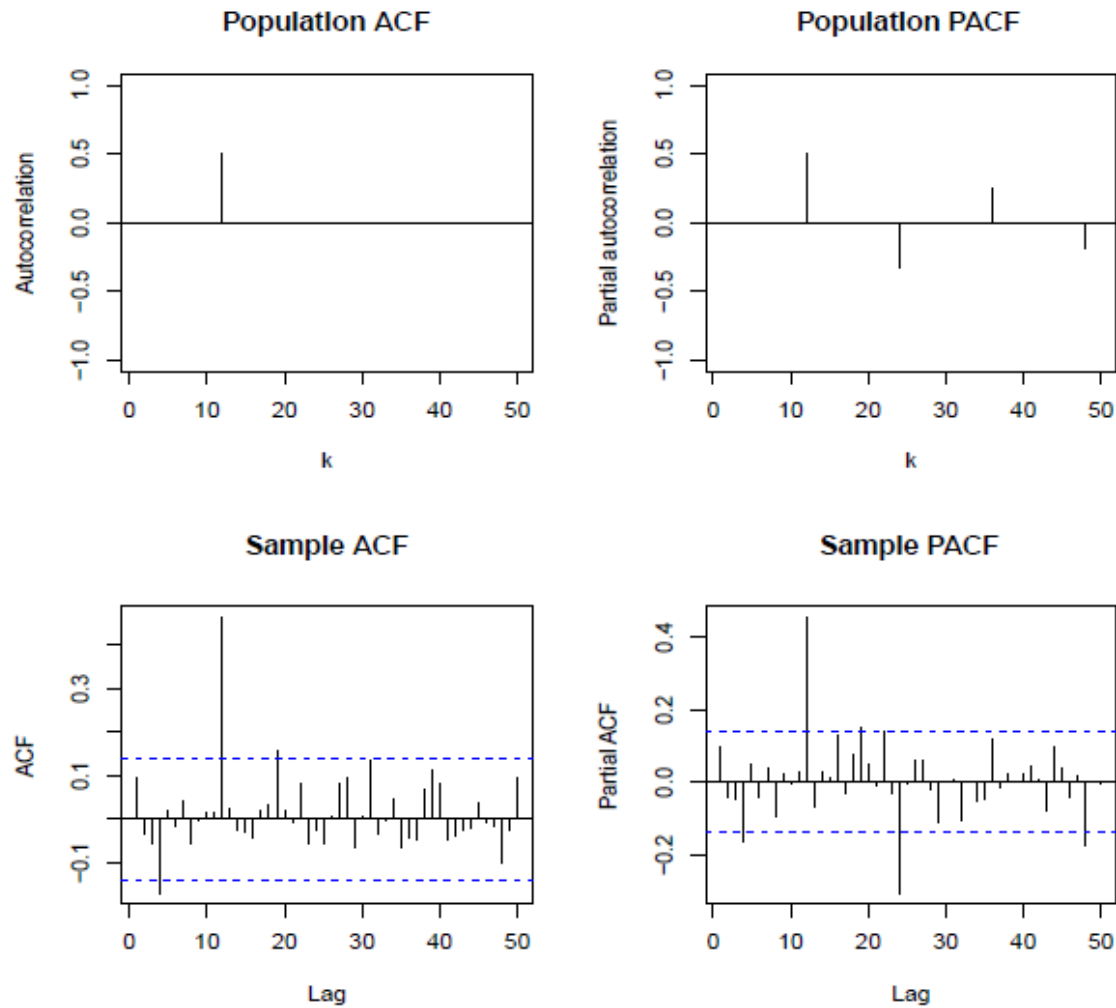


Figure 10.4: $MA(1)_{12}$ with $\Theta = -0.9$. Upper left: Population ACF. Upper right: Population PACF. Lower left (right): Sample ACF (PACF) using data in Figure 10.3.

Εποχικό αυτόπαλίνδρομο μοντέλο

61

10.2.2 $AR(P)_s$

TERMINOLOGY: Suppose $\{e_t\}$ is a zero mean white noise process with $\text{var}(e_t) = \sigma_e^2$. A seasonal autoregressive (AR) model of order P with seasonal period s , denoted by $AR(P)_s$, is

$$Y_t = \Phi_1 Y_{t-s} + \Phi_2 Y_{t-2s} + \cdots + \Phi_P Y_{t-Ps} + e_t.$$

A nonzero mean μ could be added for flexibility (as with nonseasonal models), but we take $\mu = 0$ for simplicity.

$AR(1)_{12}$: When $P = 1$ and $s = 12$, we have

$$Y_t = \Phi Y_{t-12} + e_t.$$

**Tebbs , 2013. STAT 520
FORECASTING AND TIME
SERIES, Lecture Notes**

- Similar to a nonseasonal $AR(1)$ process, a seasonal $AR(1)_{12}$ process is stationary if and only if $-1 < \Phi < 1$. An $AR(1)_{12}$ process is automatically invertible.

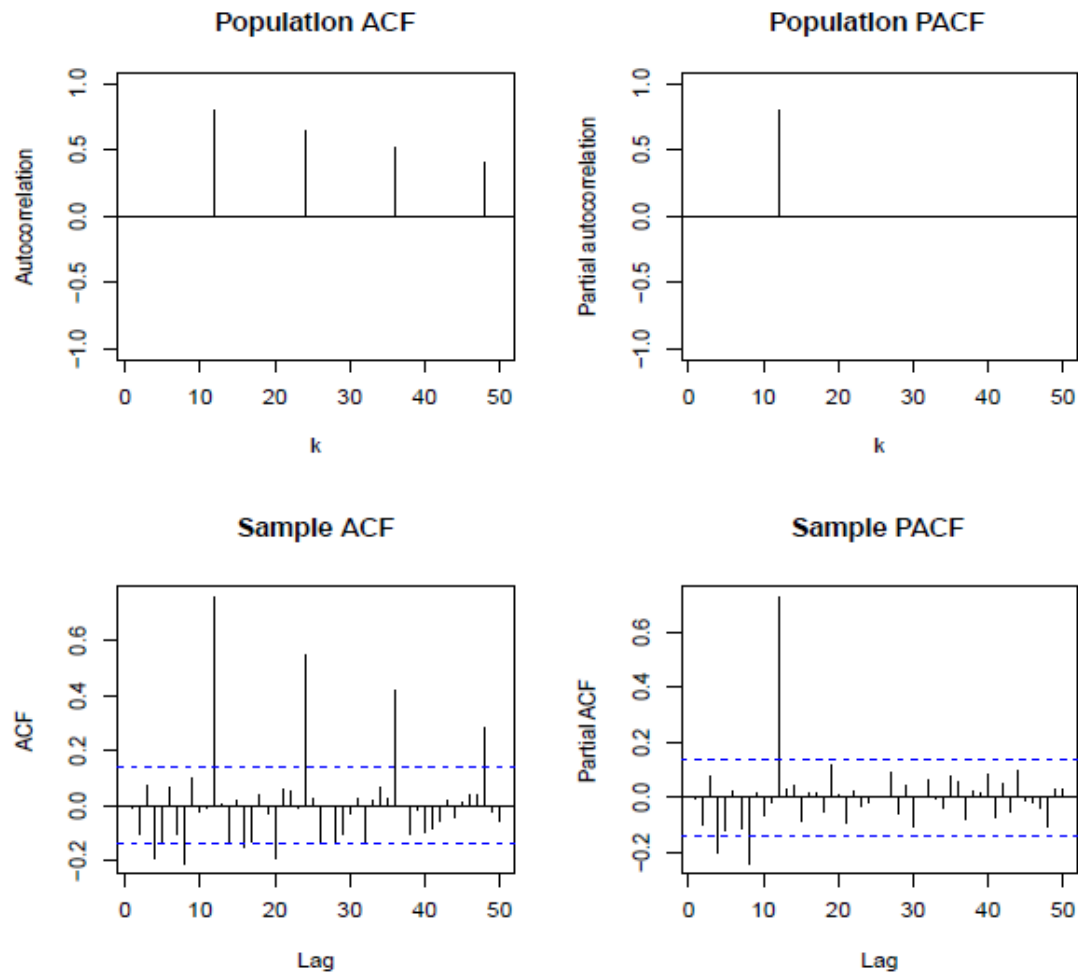


Figure 10.6: $AR(1)_{12}$ with $\Phi = -0.9$. Upper left: Population ACF. Upper right: Population PACF. Lower left (right): Sample ACF (PACF) using data in Figure 10.5.

Μεικτό μοντέλο (όπως στο θέμα)

Πολλαπλασιαστική εποχική σειρά

64

- Συνδυασμός μοντέλων κλασικού *ARIMA* και εποχικότητας
- $ARIMA(p, d, q)(P, D, Q)_s$.

$$\varphi_p(B)\Phi_P(B^s)(1-B)^d(1-B^s)^D Y_t = \theta_q(B)\Theta_Q(B^s)a_t$$

Παράδειγμα (χωρίς διαφορές)

65

$\text{MA}(1) \times \text{AR}(1)_{12}$: Suppose $\{e_t\}$ is a zero mean white noise process with $\text{var}(e_t) = \sigma_e^2$.

Consider the two models

$$Y_t = e_t - \theta e_{t-1} \iff Y_t = (1 - \theta B)e_t$$

and

$$Y_t = \Phi Y_{t-12} + e_t \iff (1 - \Phi B^{12})Y_t = e_t,$$

**Tebbs , 2013. STAT 520
FORECASTING AND TIME
SERIES, Lecture Notes**

a nonseasonal $\text{MA}(1)$ and a seasonal $\text{AR}(1)_{12}$, respectively.

- The defining characteristic of the nonseasonal $\text{MA}(1)$ is that the only nonzero autocorrelation occurs at lag $k = 1$.
- The defining characteristic of the seasonal $\text{AR}(1)_{12}$ is that the autocorrelation decays across seasonal lags $k = 12, 24, 36, \dots$.

COMBINING THE MODELS: Consider combining these two models to form

$$(1 - \Phi B^{12})Y_t = (1 - \theta B)e_t,$$

Παράδειγμα, χωρίς διαφορές

$\text{MA}(1) \times \text{MA}(1)_{12}$: Suppose $\{e_t\}$ is a zero mean white noise process with $\text{var}(e_t) = \sigma_e^2$.

Consider the **nonseasonal** $\text{MA}(1)$ model

$$Y_t = e_t - \theta e_{t-1} \iff Y_t = (1 - \theta B)e_t$$

and the **seasonal** $\text{MA}(1)_{12}$ model

$$Y_t = e_t - \Theta e_{t-12} \iff Y_t = (1 - \Theta B^{12})e_t.$$

- The defining characteristic of the nonseasonal $\text{MA}(1)$ process is that the only nonzero autocorrelation occurs at lag $k = 1$.
- The defining characteristic of the seasonal $\text{MA}(1)_{12}$ process is that the only nonzero autocorrelation occurs at lag $k = 12$.

MA(1) × MA₁₂(1)

67

**Tebbs , 2013. STAT 520
FORECASTING AND TIME
SERIES. Lecture Notes**

COMBINING THE MODELS: Consider taking the nonseasonal MA characteristic operator $\theta(B) = 1 - \theta B$ and the nonseasonal one $\Theta(B) = 1 - \Theta B^{12}$ and multiplying them together to get the new model

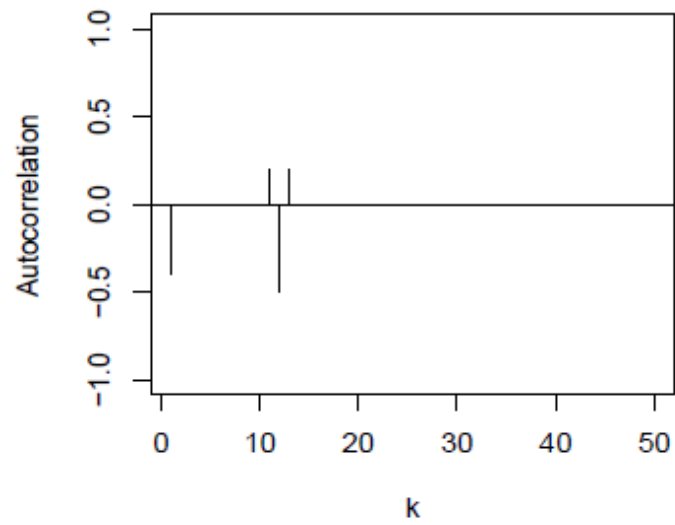
$$\begin{aligned} Y_t &= (1 - \theta B)(1 - \Theta B^{12})e_t \\ &= (1 - \theta B - \Theta B^{12} + \theta\Theta B^{13})e_t, \end{aligned}$$

or, equivalently,

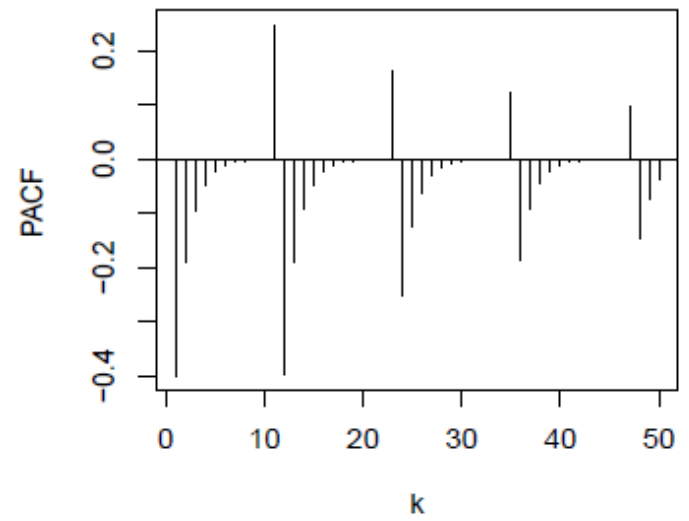
$$Y_t = e_t - \theta e_{t-1} - \Theta e_{t-12} + \theta\Theta e_{t-13}.$$

Λόγω του MA(1)
εμπεριέχεται όχι
μόνο ο όρος 12
μήνες πίσω αλλά
και 13 μήνες
πίσω

MA1*MA1(12) ACF



MA1*MA1(12) PACF



AIRLINE MODEL

69

$$W_t = (1 - \theta B)(1 - \Theta B^{12})a_t$$

$$W_t = a_t - \theta a_{t-1} - \Theta a_{t-12} + \theta\Theta a_{t-13}$$

$$W_t \sim I(0)$$

Αν έχει κανονική υστέρηση αυτή μεταφ. Και στο μοντέλο εποχικότητας

$$\gamma_k = \begin{cases} (1 + \theta^2)(1 + \Theta^2)\sigma_a^2, & k = 0 \\ -\theta(1 + \Theta^2)\sigma_a^2, & |k| = 1 \\ -\Theta(1 + \theta^2)\sigma_a^2, & |k| = 12 \\ \theta\Theta\sigma_a^2, & |k| = 11, 13 \\ 0, & o.w. \end{cases} \quad \rho_k = \begin{cases} \frac{-\theta}{(1 + \theta^2)}, & |k| = 1 \\ \frac{-\Theta}{(1 + \Theta^2)}, & |k| = 12 \\ \frac{\theta\Theta}{(1 + \theta^2)(1 + \Theta^2)}, & |k| = 11, 13 \\ 0, & o.w. \end{cases}$$

...με κινητούς μέσους όρους

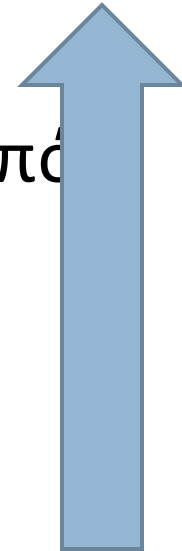
70

- SARIMA(0,1,1)(0,1,1)₁₂

$$(1 - B)(1 - B^{12})Y_t = (1 - \theta B)(1 - \Theta B^{12})a_t$$

όπου $|\theta| < 1$ και $|\Theta| < 1$.

- Box and Jenkins (1976) μηνιαίες σειρές από επιβάτες αεροπλάνων



Διαφορές στους εποχικούς όρους

TERMINOLOGY: The seasonal difference operator ∇_s is defined by

$$\nabla_s Y_t = Y_t - Y_{t-s} = (1 - B^s)Y_t,$$

for a seasonal period s . For example, with $s = 12$ and monthly data, the **first seasonal differences** are

$$\nabla_{12} Y_t = Y_t - Y_{t-12} = (1 - B^{12})Y_t,$$

that is, the first differences of the January observations, the first differences of the February observations, and so on.

Μοντέλα: $(p, 1, q) \times (P, 1, Q)_{12}$

72

$$\begin{aligned}\nabla\nabla_{12}Y_t &= (1 - B)(1 - B^{12})Y_t \\ &= (1 - B - B^{12} + B^{13})Y_t \\ &= Y_t - Y_{t-1} - Y_{t-12} + Y_{t-13}.\end{aligned}$$



Οι όροι των διαφορών

Γενικό μοντέλο

73

$$\varphi_p(\mathbf{B})\Phi_P(\mathbf{B}^s)(1-\mathbf{B})^d(1-\mathbf{B}^s)^D Y_t = \theta_q(\mathbf{B})\Theta_Q(\mathbf{B}^s)a_t$$

TERMINOLOGY: Suppose that $\{e_t\}$ is zero mean white noise with $\text{var}(e_t) = \sigma_e^2$. The multiplicative seasonal autoregressive integrated moving average (SARIMA) model with seasonal period s , denoted by $\text{ARIMA}(p, d, q) \times \text{ARIMA}(P, D, Q)_s$, is

$$\phi(B)\Phi_P(B^s)\nabla^d\nabla_s^D Y_t = \theta(B)\Theta_Q(B^s)e_t,$$

where the nonseasonal AR and MA characteristic operators are

$$\phi(B) = (1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p)$$

$$\theta(B) = (1 - \theta_1 B - \theta_2 B^2 - \dots - \theta_q B^q),$$

the seasonal AR and MA characteristic operators are

$$\Phi_P(B^s) = 1 - \Phi_1 B^s - \Phi_2 B^{2s} - \dots - \Phi_P B^{Ps}$$

$$\Theta_Q(B^s) = 1 - \Theta_1 B^s - \Theta_2 B^{2s} - \dots - \Theta_Q B^{Qs},$$

and

$$\nabla^d\nabla_s^D Y_t = (1 - B)^d(1 - B^s)^D Y_t.$$

Γενικό
μοντέλο

Συμβολική γραφή

$$\phi(B)\Phi_P(B^s)\nabla^d\nabla_s^D Y_t = \theta(B)\Theta_Q(B^s)e_t,$$

where the nonseasonal AR and MA characteristic operators are

$$\phi(B) = (1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p)$$

$$\theta(B) = (1 - \theta_1 B - \theta_2 B^2 - \dots - \theta_q B^q),$$

the seasonal AR and MA characteristic operators are

$$\Phi_P(B^s) = 1 - \Phi_1 B^s - \Phi_2 B^{2s} - \dots - \Phi_P B^{Ps}$$

$$\Theta_Q(B^s) = 1 - \Theta_1 B^s - \Theta_2 B^{2s} - \dots - \Theta_Q B^{Qs},$$

and

$$\nabla^d\nabla_s^D Y_t = (1 - B)^d(1 - B^s)^D Y_t.$$

MODEL FITTING: We use R to fit this $ARIMA(0, 1, 0) \times ARIMA(3, 1, 0)_{12}$ model using maximum likelihood. Here is the output:

```
> milk.arima010.arima310 =  
  arima(milk,order=c(0,1,0),method='ML',seasonal=list(order=c(3,1,0),period=12))  
> milk.arima010.arima310  
Coefficients:  
      sar1      sar2      sar3  
 -0.9133 -0.8146 -0.6002  
s.e.    0.0696    0.0776    0.0688  
sigma^2 estimated as 121.4:  log likelihood = -512.03,  aic = 1030.05
```

Παράδειγμα: $(0, 1, 0) \times (3, 1, 0)_{12}$
 $\Phi_1 = -0.9133, \Phi_2 = -0.8146, \Phi_3 = -0.6002$

The fitted model is

$$(1 + 0.9133B^{12} + 0.8146B^{24} + 0.6002B^{36}) \underbrace{(1 - B)(1 - B^{12})}_{= \nabla \nabla_{12} Y_t} Y_t = e_t.$$

ΧΕΙΡΙΣΜΟΣ ΠΡΟΓΡΑΜΜΑΤΟΣ

	C	var												
1	84,87													
2	111,23													
3	84,61													
4	24,71													
5	32,48													
6	93,74													
7	478,13													
8	305,16													
9	2342,40													
10	8892,10													
11	7983,39													
12	3791,20													
13	19276,13													
14	2325,80													
15	878,45													
16	124,23													
17	97,46													
18	9969,06													
19	59175,67													
20	4196,77													
21	4190,33													
22	6589,03													
23	659,74													
24	114,90													
25	75,61													
26	82,20													
27	57,39													
28	49,32													
29	74,29													
30	108,19													
31	1565,80													

- Reports
- Descriptive Statistics
- Tables
- RFM Analysis
- Compare Means
- General Linear Model
- Generalized Linear Models
- Mixed Models
- Correlate
- Regression
- Loglinear
- Neural Networks
- Classify
- Dimension Reduction
- Scale
- Nonparametric Tests
- Forecasting
 - Create Models...
 - Apply Models...
- Survival
- Multiple Response
 - Seasonal Decomposition...
 - Spectral Analysis...
- Missing Value Analysis...
- Multiple Imputation
- Complex Samples
 - Sequence Charts...
 - Autocorrelations...
 - Cross-Correlations...
- Quality Control
- ROC Curve...



SPSS Statistics Data Editor - [DataSet1] - SPSS Statistics Data Editor

Visible: 1 of 1 Variables

	C	var	var	var	var	var	var	var	var	var	var	var	var	var	var	var	var	var	var	var
1	84,87																			
2	111,23																			
3	84,61																			
4	24,71																			
5	32,48																			
6	93,74																			
7	478,13																			
8	305,16																			
9	2342,40																			
10	8892,10																			
11	7983,39																			
12	3791,20																			
13	19276,13																			
14	2325,80																			
15	878,45																			
16	124,23																			
17	97,46																			
18	9969,06																			
19	59175,67																			
20	4196,77																			
21	4190,33																			
22	6589,03																			
23	659,74																			
24	114,90																			
25	75,61																			
26	82,20																			
27	57,39																			
28	49,32																			
29	74,29																			
30	108,19																			
31	4566,80																			

Time Series Modeler

Variables: daily mean sediment transport rate tons/day [C]

Dependent Variables:

Estimation Period

Start: First case

End: Last case

Start: First case after end of estimation period

End: Last case in active dataset

Time Series Modeler

Before you use this dialog, starting time and time interval should be defined for your time series. This will ensure that the output is properly labeled and will allow you to obtain seasonal models, if applicable.

Press OK to define a time series model, treating cases as unlabeled time periods.

Press Define Dates to set the starting time and time interval.

Don't show this message again

OK Define Dates...

OK Paste Reset Cancel Help

SPSS Statistics Processor is ready

11:50 μμ
3/2/2015

Εισαγωγή δεδομένων (από υπάρχον αρχείο ή με πληκτρολόγηση) και χρονικού βήματος της χρονοσειράς.

Analyze → Forecasting → Create Models → Define Dates (Από το αναδυόμενο πλαίσιο) → Επιλογή χρονικού βήματος που αντιστοιχεί στο υπό εξέταση πρόβλημα και εισαγωγή αρχής της χρονοσειράς.



	C	var	var	var	var	var	var	var	var	var	var	var	var	var	var	var	var	var	var	var	
1	84,87																				
2	111,23																				
3	84,61																				
4	24,71																				
5	32,48																				
6	93,74																				
7	478,13																				
8	305,16																				
9	2342,40																				
10	8892,10																				
11	7983,39																				
12	3791,20																				
13	19276,13																				
14	2325,80																				
15	878,45																				
16	124,23																				
17	97,46																				
18	9969,06																				
19	59175,67																				
20	4196,77																				
21	4190,33																				
22	6589,03																				
23	659,74																				
24	114,90																				
25	75,61																				
26	82,20																				
27	57,39																				
28	49,32																				
29	74,29																				
30	108,19																				
31	1565,80																				

Define Dates

Cases Are:

- Years
- Years, quarters
- Years, months**
- Years, quarters, months
- Days
- Weeks, days
- Weeks, work days(5)
- Weeks, work days(6)
- Hours
- Days, hours

Current Dates: None

First Case Is:

Year: 1900

Month: 1 Periodicity at higher level: 12

OK Reset Cancel Help



	C	YEAR_	MONTH_	DATE_	var	var	var	var	var	var	var	var	var	var	var	var	var	var
1	84,87	1900		1 JAN 1900														
2	111,23	1900		2 FEB 1900														
3	84,61	1900		3 MAR 1900														
4	24,71	1900		4 APR 1900														
5	32,48	1900		5 MAY 1900														
6	93,74	1900		6 JUN 1900														
7	478,13	1900		7 JUL 1900														
8	305,16	1900		8 AUG 1900														
9	2342,40	1900		9 SEP 1900														
10	8892,10	1900		10 OCT 1900														
11	7983,39	1900		11 NOV 1900														
12	3791,20	1900		12 DEC 1900														
13	19276,13	1901		1 JAN 1901														
14	2325,80	1901		2 FEB 1901														
15	878,45	1901		3 MAR 1901														
16	124,23	1901		4 APR 1901														
17	97,46	1901		5 MAY 1901														
18	9969,06	1901		6 JUN 1901														
19	59175,67	1901		7 JUL 1901														
20	4196,77	1901		8 AUG 1901														
21	4190,33	1901		9 SEP 1901														
22	6589,03	1901		10 OCT 1901														
23	659,74	1901		11 NOV 1901														
24	114,90	1901		12 DEC 1901														
25	75,61	1902		1 JAN 1902														
26	82,20	1902		2 FEB 1902														
27	57,39	1902		3 MAR 1902														
28	49,32	1902		4 APR 1902														
29	74,29	1902		5 MAY 1902														
30	108,19	1902		6 JUN 1902														
31	4565,80	1902		7 JUL 1902														

Time Series Modeler

Variables Statistics Plots Output Filter Save Options

Variables:
YEAR, not periodic [YEAR_]
MONTH, period 12 [MONTH_]

Dependent Variables:
daily mean sediment transport rate tons/day [C]

Independent Variables:

Method: Expert Modeler Criteria...

Model Type: All models

Estimation Period:
Start: First case
End: Last case

Forecast Period:
Start: First case after end of estimation period
End: Last case in active dataset

OK Paste Reset Cancel Help

05325000.sav [DataSet1] - SPSS Statistics Data Editor

File Edit View Data Transform Analyze Graphs Utilities Add-ons Window Help

1 : C 84,8709677419355 Visible: 4 of 4 Variables

	C	YEAR_	MONTH_	DATE_	var	var	var	var	var	var	var	var	var	var	var	var	var	var	var	
1	84,87	1900	1	JAN 1900																
2	111,23	1900	2	FEB 1900																
3	84,61	1900	3	MAR 1900																
4	24,71	1900	4	APR 1900																
5	32,48	1900	5	MAY 1900																
6	93,74	1900	6	JUN 1900																
7	478,13	1900	7	JUL 1900																
8	305,16	1900	8	AUG 1900																
9	2342,40	1900	9	SEP 1900																
10	8892,10	1900	10	OCT 1900																
11	7983,39	1900	11	NOV 1900																
12	3791,20	1900	12	DEC 1900																
13	19276,13	1901	1	JAN 1901																
14	2325,80	1901	2	FEB 1901																
15	878,45	1901	3	MAR 1901																
16	124,23	1901	4	APR 1901																
17	97,46	1901	5	MAY 1901																
18	9969,06	1901	6	JUN 1901																
19	59175,67	1901	7	JUL 1901																
20	4196,77	1901	8	AUG 1901																
21	4190,33	1901	9	SEP 1901																
22	6589,03	1901	10	OCT 1901																
23	659,74	1901	11	NOV 1901																
24	114,90	1901	12	DEC 1901																
25	75,61	1902	1	JAN 1902																
26	82,20	1902	2	FEB 1902																
27	57,39	1902	3	MAR 1902																
28	49,32	1902	4	APR 1902																
29	74,29	1902	5	MAY 1902																
30	108,19	1902	6	JUN 1902																
31	1565,80	1902	7	JUL 1902																

Data View Variable View

Time Series Modeler

Variables Statistics Plots Output Filter Save Options

Display fit measures, Ljung-Box statistic, and number of outliers by model

Fit Measures

Stationary R square Mean absolute error

R square Maximum absolute percentage error

Root mean square error Maximum absolute error

Mean absolute percentage error Normalized BIC

Statistics for Comparing Models

Goodness of fit

Residual autocorrelation function (ACF)

Residual partial autocorrelation function (PACF)

Display forecasts

Statistics for Individual Models

Parameter estimates

Residual autocorrelation function (ACF)

Residual partial autocorrelation function (PACF)

OK Paste Reset Cancel Help

Καρτέλα Statistics: Εμφανίζει στα αποτελέσματα ορισμένα στατιστικά μεγέθη. Προτείνονται (επιπρόσθετα από τα τικαρισμένα από προεπιλογή του προγράμματος) τα

- Root mean square error
- Normalized BIC
- Τα 3 από το πλαίσιο Statistics for Individual Models (Parameter estimates, Residual autocorrelation function (ACF), Residual partial autocorrelation function (PACF))

12:03 πμ 4/2/2015

05325000.sav [DataSet1] - SPSS Statistics Data Editor

File Edit View Data Transform Analyze Graphs Utilities Add-ons Window Help

1 : C 84,8709677419355 Visible: 4 of 4 Variables

	C	YEAR	MONTH	DATE	var	var	var	var	var	var	var	var	var	var	var	var	var	var
1	84,87	1900	1	JAN 1900														
2	111,23	1900	2	FEB 1900														
3	84,61	1900	3	MAR 1900														
4	24,71	1900	4	APR 1900														
5	32,48	1900	5	MAY 1900														
6	93,74	1900	6	JUN 1900														
7	478,13	1900	7	JUL 1900														
8	305,16	1900	8	AUG 1900														
9	2342,40	1900	9	SEP 1900														
10	8892,10	1900	10	OCT 1900														
11	7983,39	1900	11	NOV 1900														
12	3791,20	1900	12	DEC 1900														
13	19276,13	1901	1	JAN 1901														
14	2325,80	1901	2	FEB 1901														
15	878,45	1901	3	MAR 1901														
16	124,23	1901	4	APR 1901														
17	97,46	1901	5	MAY 1901														
18	9969,06	1901	6	JUN 1901														
19	59175,67	1901	7	JUL 1901														
20	4196,77	1901	8	AUG 1901														
21	4190,33	1901	9	SEP 1901														
22	6589,03	1901	10	OCT 1901														
23	659,74	1901	11	NOV 1901														
24	114,90	1901	12	DEC 1901														
25	75,61	1902	1	JAN 1902														
26	82,20	1902	2	FEB 1902														
27	57,39	1902	3	MAR 1902														
28	49,32	1902	4	APR 1902														
29	74,29	1902	5	MAY 1902														
30	108,19	1902	6	JUN 1902														
31	1566,80	1902	7	JUL 1902														

Time Series Modeler

Variables Statistics **Plots** Output Filter Save Options

Plots for Comparing Models

- Stationary R square
- Maximum absolute percentage error
- R square
- Maximum absolute error
- Root mean square error
- Normalized BIC
- Mean absolute percentage error
- Residual autocorrelation function (ACF)
- Mean absolute error
- Residual partial autocorrelation function (PACF)

Plots for Individual Models

- Series
- Residual autocorrelation function (ACF)

Each Plot Displays

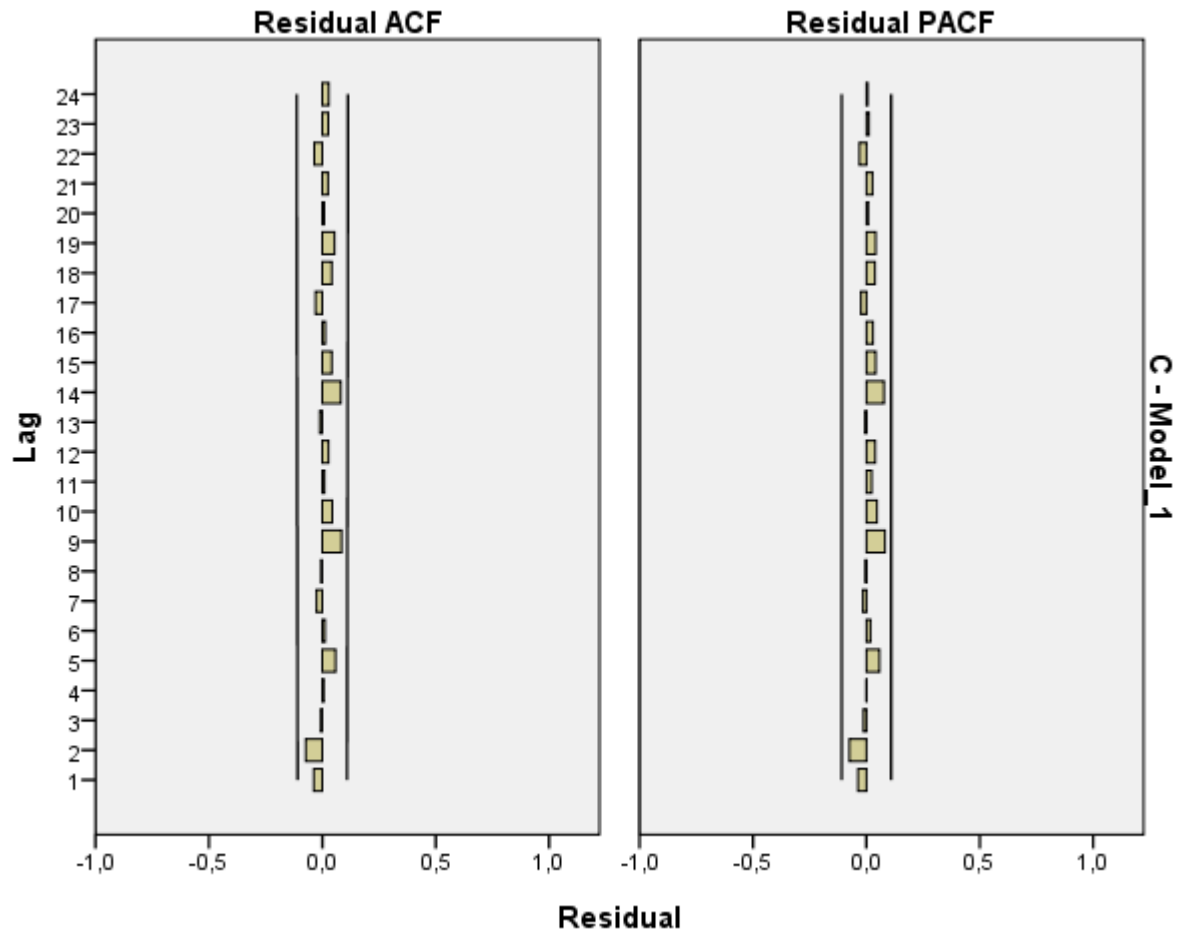
- Observed values
- Forecasts
- Fit values
- Confidence intervals for forecasts
- Confidence intervals for fit values

OK Paste Reset Cancel Help

Data View Variable View

SPSS Statistics Processor is ready

12:08 πμ 4/2/2015



Model Statistics

Model	Number of Predictors	Model Fit statistics				Ljung-Box Q(18)			Number of Outliers
		Stationary R-squared	RMSE	MAE	Normalized BIC	Statistics	DF	Sig.	
daily mean sediment transport rate tons/day-Model_1	0	,662	8674,619	3325,209	18,172	10,839	16	,819	0

ARIMA Model Parameters

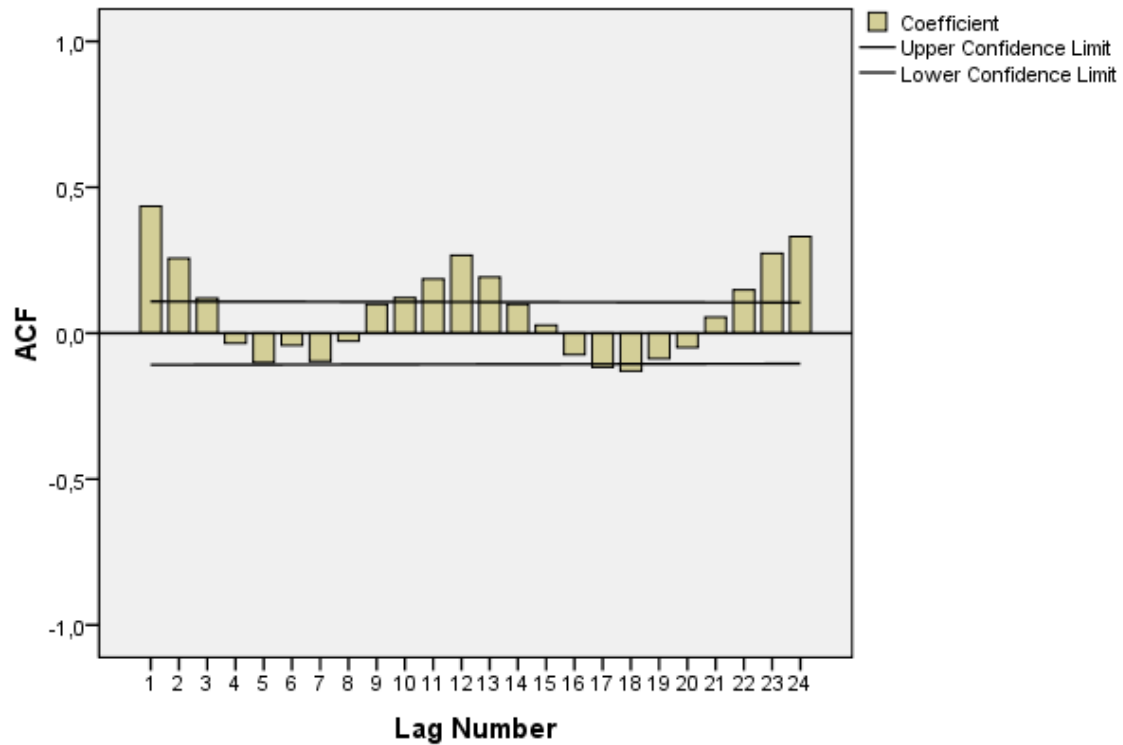
				Estimate	SE	t	Sig.
daily mean sediment transport rate tons/day-Model_1	daily mean sediment transport rate tons/day	Natural Log	AR Lag 1	,795	,034	23,073	,000
			Seasonal Difference	1			
			MA, Seasonal Lag 1	,933	,048	19,599	,000

- Έλεγχος του διαγράμματος Residual ACF. Δεν πρέπει να υπάρχουν μεγάλες αυτοσυσχετίσεις (δηλ. να υπάρχει μόνο θόρυβος).
- Έλεγχος του Model Statistics. Η τιμή του Ljung-Box Q(18) πρέπει να είναι $\geq 0,05$. Το τεστ Ljung-Box είναι ένα στατιστικό τεστ που δείχνει κατά πόσο οι αυτοσυσχετίσεις μιας χρSig. στο ονική σειράς είναι διάφορες του μηδενός και καθορίζει εάν το μοντέλο είναι ορθά ορισμένο. Σε αντίθετη περίπτωση (Sig. $\leq 0,05$) το μοντέλο δεν είναι αποδεκτό επειδή υπάρχει κάποια δομή μέσα στο μοντέλο που δεν λήφθηκε υπόψη.
- Έλεγχος του ARIMA Model Parameters. Η τιμή του Sig. στους συντελεστές πρέπει να είναι $\leq 0,01$ (ή $\leq 0,05$) για να είναι στατιστικά σημαντικοί, διαφορετικά το μοντέλο απορρίπτεται. Σε περίπτωση που η τιμή της σταθεράς δεν είναι $\leq 0,01$ (ή $\leq 0,05$) αυτή αγνοείται και το υπόλοιπο μοντέλο γίνεται αποδεκτό.

Original autocorrelation

87

daily mean sediment transport rate tons/day



Προσοχή στα υπερόπλα

88

- Spss
- Οικονομικά μεγέθη με εποχικότητα: ημέρα μήνα (π.χ. μισθοδοσία δημοσίων υπαλλήλων) -μήνας – χρόνος
- Υδρολογικά με εποχικότητα: μήνας-χρόνος