



National and Kapodistrian
UNIVERSITY OF ATHENS

Τεχνικές Ανάλυσης και Πρόβλεψης Τηλεπικοινωνιακών Αγορών

BOX-JENKINS (ARIMA)

ARIMA

AR AutoRegressive

I Integrated

MA Moving Average

- › Είναι μία κατηγορία γραμμικών μοντέλων, που είναι ικανά να αναπαραστήσουν στάσιμες και μη στάσιμες χρονοσειρές.
- › Η μεθοδολογία για αναγνώριση, προσαρμογή και έλεγχο καταλοίπων των ARIMA μοντέλων εξελίχθηκε ιδιαίτερα από τη δουλειά των G.E.P. Box και G.M. Jenkins. Για αυτό το λόγο, η μοντελοποίηση και πρόβλεψη ARIMA συχνά αναφέρεται ως Box-Jenkins μεθοδολογία.

ARIMA

- › Τα ARIMA μοντέλα συνδυάζουν τις ιδιότητες τριών διαφορετικών υπο-μοντέλων (υπο-συστημάτων):
 - Αυτοπαλινδρόμησης (autoregression)
 - Ολοκλήρωσης (integration)
 - Εξομάλυνσης με κινητό μέσο (moving average)

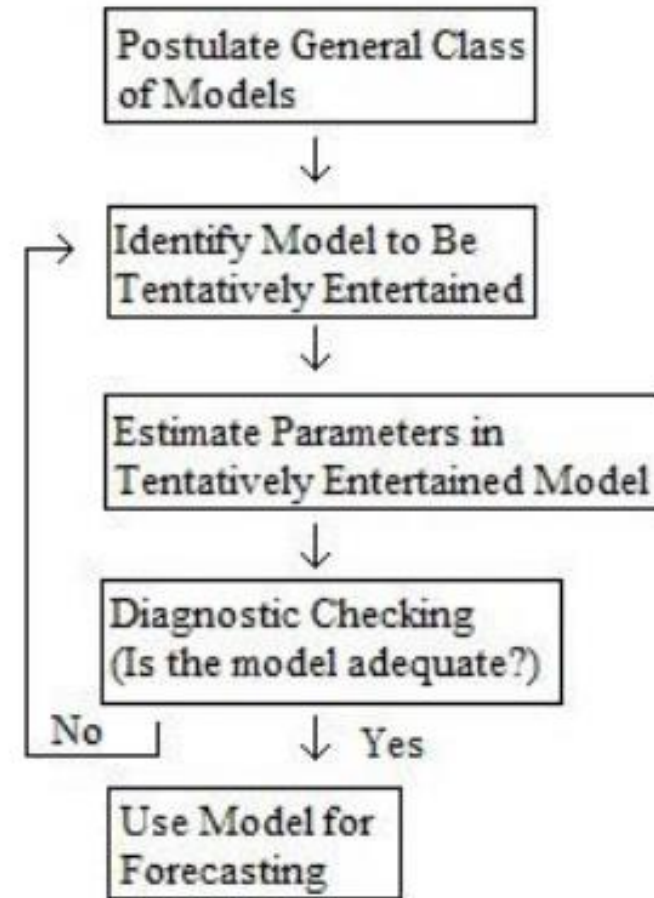
- › Ένα ARIMA μοντέλο μπορεί να διαθέτει και τα τρία υποσυστήματα.

ARIMA

- › Τα ARIMA μοντέλα δεν περιλαμβάνουν ανεξάρτητες μεταβλητές στην κατασκευή τους.
- › Αντί για αυτό κάνουν χρήση της διαθέσιμης πληροφορίας, που κρύβουν οι χρονοσειρές, για τη δημιουργία προβλέψεων.

Μεθοδολογία Box – Jenkins

- › Η Box-Jenkins μεθοδολογία αναφέρεται σε ένα σύνολο διαδικασιών για αναγνώριση, προσαρμογή και έλεγχο ARIMA μοντέλων με δεδομένα χρονοσειρών.
- › Οι προβλέψεις έπονται απευθείας από τη μορφή του προσαρμοσμένου μοντέλου.



Μεθοδολογία Box – Jenkins

- › Η αρχική επιλογή ενός ARIMA μοντέλου βασίζεται στην εξέταση ενός διαγράμματος των χρονοσειρών σαν συνάρτηση του χρόνου και στην εξέταση της αυτοσυσχέτισής τους για μερικές χρονικές καθυστερήσεις (lags).
- › Το μοτίβο των αυτοσυσχετίσεων δείγματος, το οποίο υπολογίζεται από τις χρονοσειρές, συνδυάζεται με το γνωστό μοτίβο αυτοσυσχέτισης, που αντιστοιχεί σε ένα συγκεκριμένο ARIMA μοντέλο.
- › Αυτός ο συνδυασμός γίνεται και για τις αυτοσυσχετίσεις και για τις μερικές αυτοσυσχετίσεις.

Μεθοδολογία Box – Jenkins

Στην επιλογή του μοντέλου, να έχουμε υπόψη, ότι:

- › Οι αυτοσυσχετίσεις, οι οποίες υπολογίζονται από τα δεδομένα, δεν ταιριάζουν ακριβώς με κάποιο μοτίβο από το σύνολο των θεωρητικών αυτοσυσχετίσεων ARIMA μοντέλων και
- › Οι αυτοσυσχετίσεις υπόκεινται στη μεταβολή του δείγματος.

Μερική Αυτοσυσχέτιση

- › Η μερική αυτοσυσχέτιση ανάμεσα στα Y_t και Y_{t-s} αναφέρεται στη συσχέτιση ανάμεσα στην Y_t και στην Y_{t-s} , όταν έχουν αφαιρεθεί οι γραμμικές επιδράσεις όλων των ενδιάμεσων μεταβλητών $Y_{t-1}, Y_{t-2}, \dots, Y_{t-(s-1)}$

Μερική Αυτοσυσχέτιση

- › Αν παραστήσουμε με ρ_{ss} το συντελεστή μερικής αυτοσυσχέτισης s τάξης, δηλαδή το συντελεστή αυτοσυσχέτισης ανάμεσα στην Y_t και στην Y_{t-s} , για $s=1, 2, \dots$, τότε:

$$y_t = \rho_{1s} y_{t-1} + \rho_{2s} y_{t-2} + \rho_{3s} y_{t-3} + \dots + \rho_{ss} y_{t-s} + \varepsilon_t$$

όπου:

$$y_t = Y_t - \mu$$

- › Είναι προφανές, ότι ο μερικός συντελεστής αυτοσυσχέτισης ρ_{11} συμπίπτει με τον συντελεστή αυτοσυσχέτισης r_1 .
Δηλαδή $r_1 = \rho_{11}$.

Αυτοπαλίνδρομα Μοντέλα

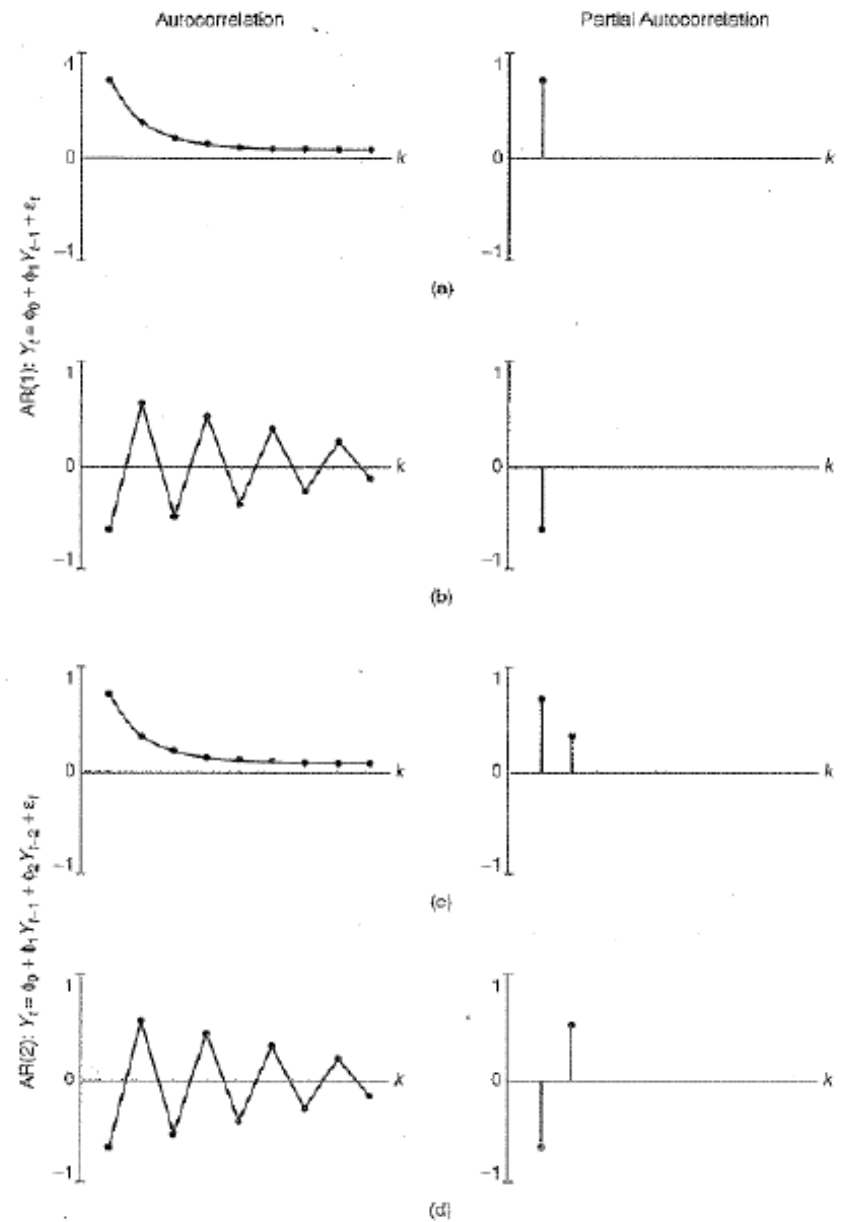
› Ένα αυτοπαλινδρομικό μοντέλο p τάξης – AR(p) – παίρνει τη μορφή:

$$Y_t = \varphi_0 + \varphi_1 Y_{t-1} + \varphi_2 Y_{t-2} + \dots + \varphi_p Y_{t-p} + \varepsilon_t$$

όπου,

- $Y_t = H$ εξαρτημένη μεταβλητή στον χρόνο t .
- $Y_{t-1}, Y_{t-2}, \dots, Y_{t-p} = H$ εξαρτημένη μεταβλητή στους χρόνους για lag $t-1, t-2, \dots, t-p$.
- $\varphi_0, \varphi_1, \varphi_2, \dots, \varphi_p =$ οι συντελεστές, που πρέπει να εκτιμηθούν.
- $\varepsilon_t =$ ο όρος σφάλματος στο χρόνο t , ο οποίος αναπαριστά τις επιδράσεις των μεταβλητών, που δεν ερμηνεύονται από το μοντέλο. Οι υποθέσεις σχετικά με τον όρο σφάλματος είναι ίδιες με εκείνες για το τυπικό μοντέλο παλινδρόμησης.

Αυτοπαλίνδρομα Μοντέλα



Γράφημα 1: Συντελεστές αυτοσυσχέτισης και μερικής αυτοσυσχέτισης των AR(1) και AR(2) μοντέλων

Αυτοπαλίνδρομα Μοντέλα

- › Η Εξίσωση του αυτοπαλίνδρομου μοντέλου έχει την εμφάνιση ενός μοντέλου παλινδρόμησης με καθυστερημένες (lagged) τιμές της εξαρτημένης μεταβλητής στις θέσεις της ανεξάρτητης μεταβλητής, ως εκ τούτου και το όνομα αυτοπαλινδρομικό μοντέλο.
- › Τα αυτοπαλινδρομικά μοντέλα είναι κατάλληλα και για στάσιμες χρονοσειρές και ο συντελεστής ϕ_0 σχετίζεται με το σταθερό επίπεδο των σειρών.

Μοντέλα Κινητού Μέσου

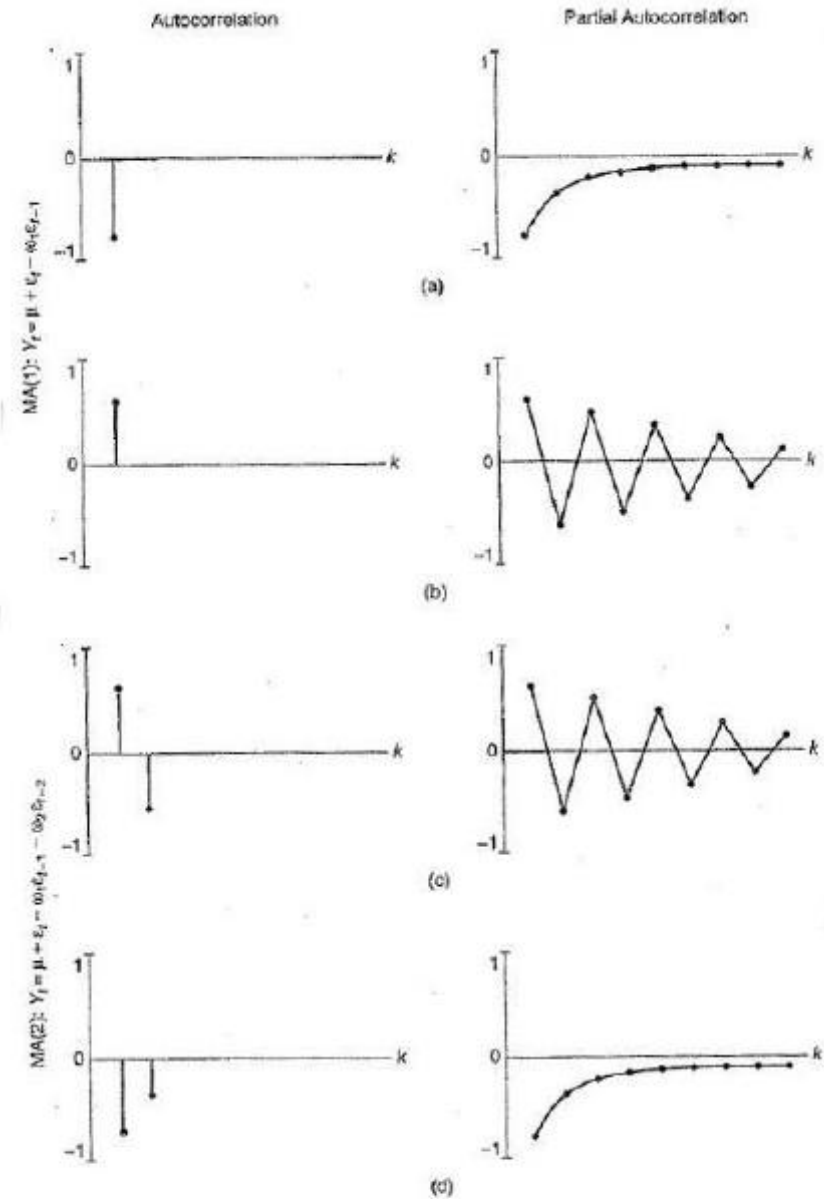
- › Ένα μοντέλο κινούμενου μέσου (moving average) q τάξης – MA(q) – παίρνει τη μορφή:

$$Y_t = \mu + \varepsilon_t - \omega_1 \varepsilon_{t-1} - \omega_2 \varepsilon_{t-2} - \dots - \omega_q \varepsilon_{t-q}$$

όπου:

- › $Y_t = Y$ εξαρτημένη μεταβλητή στον χρόνο t
- › $\mu = 0$ σταθερός μέσος της διαδικασίας
- › $\omega_0, \omega_1, \omega_2, \dots, \omega_p,$ = Οι συντελεστές που πρέπει να εκτιμηθούν.
- › $\varepsilon_t = 0$ όρος σφάλματος στο χρόνο t , ο οποίος αναπαριστά τις επιδράσεις των μεταβλητών, που δεν ερμηνεύονται από το μοντέλο.
- › $\varepsilon_0, \varepsilon_1, \varepsilon_2, \dots, \varepsilon_{t-q} = \varepsilon$ Σφάλματα από προηγούμενες χρονικές περιόδους.

Μοντέλα Κινητού Μέσου



Γράφημα 2: Συντελεστές αυτοσυσχέτισης και μερικής αυτοσυσχέτισης των MA(1) και MA(2) μοντέλων

Μοντέλα Κινητού Μέσου

- › Η εξίσωση είναι παρόμοια με την Εξίσωση του AR μοντέλου, εκτός από το ότι η εξαρτημένη μεταβλητή Y_t εξαρτάται από τις προηγούμενες τιμές των σφαλμάτων, παρά από την ίδια τη μεταβλητή.
- › Τα μοντέλα κινούμενου μέσου (MA) παρέχουν προβλέψεις βασιζόμενα σε ένα γραμμικό συνδυασμό ενός πεπερασμένου αριθμού παρελθοντικών σφαλμάτων, ενώ τα αυτοπαλινδρομικά μοντέλα (AR) προβλέπουν το Y_t σαν μία γραμμική συνάρτηση ενός πεπερασμένου αριθμού παρελθοντικών τιμών του Y_t .
- › Είναι κατάλληλα για στάσιμες χρονοσειρές.

Μοντέλα Κινητού Μέσου

- › Ο όρος κινούμενος μέσος για το μοντέλο είναι ιστορικός και δεν θα πρέπει να μπερδεύεται με τις διαδικασίες κινούμενου μέσου.
- › Ο όρος κινούμενος μέσος αναφέρεται στο γεγονός, ότι η απόκλιση της απόκρισης από το μέσο, $Y_t - \mu$, είναι ένας γραμμικός συνδυασμός των τωρινών και παρελθοντικών σφαλμάτων και ότι καθώς ο χρόνος προχωράει μπροστά, τα σφάλματα, που περιλαμβάνονται σε αυτό το γραμμικό συνδυασμό, προχωράνε επίσης μπροστά.

Αυτοπαλινδρομικά Μοντέλα Κινητού Μέσου

- › Ένα μοντέλο με αυτοπαλινδρομικούς όρους μπορεί να συνδυαστεί με ένα μοντέλο, που έχει όρους κινούμενου μέσου και να οδηγήσει σε ένα μεικτό αυτοπαλινδρομικό μοντέλο κινούμενου μέσου (autoregressive-moving average model).
- › Γίνεται χρήση του συμβολισμού $ARMA(p,q)$, όπου το p είναι η τάξη του αυτοπαλινδρομικού μέρους και q είναι η τάξη του μέρους κινούμενου μέσου.

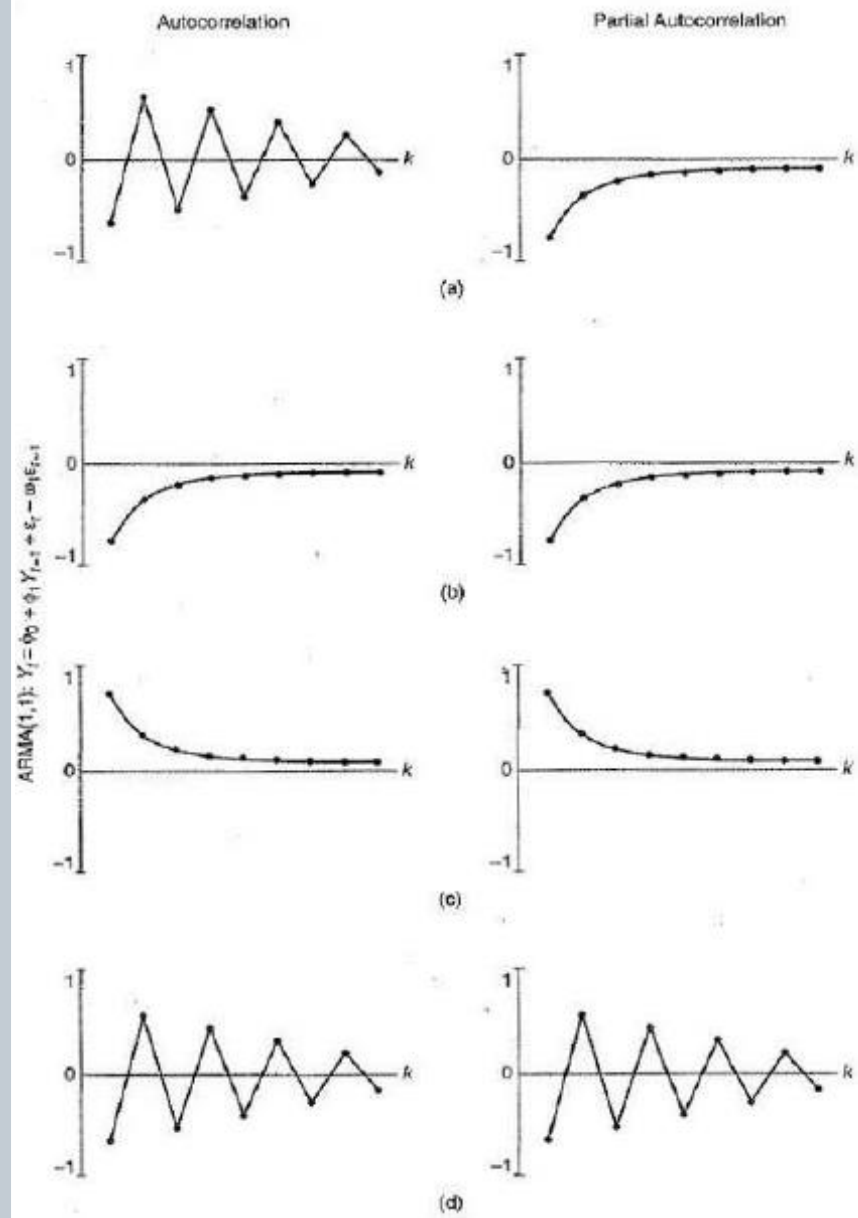
Αυτοπαλινδρομικά Μοντέλα Κινητού Μέσου

› Ένα ARMA(p,q) μοντέλο έχει τη γενική μορφή:

$$Y_t = \varphi_0 + \varphi_1 Y_{t-1} + \varphi_2 Y_{t-2} + \dots + \varphi_p Y_{t-p} + \varepsilon_t - \omega_1 \varepsilon_{t-1} - \omega_2 \varepsilon_{t-2} - \dots - \omega_q \varepsilon_{t-q}$$

- › Τα ARMA(p,q) μοντέλα μπορούν να περιγράψουν μία ευρεία ποικιλία συμπεριφορών για στάσιμες χρονοσειρές.
- › Προβλέψεις που παράγονται από ένα ARMA(p,q) μοντέλο, θα εξαρτώνται από τις πρόσφατες και περασμένες τιμές της απόκρισης, Y , καθώς επίσης από τις πρόσφατες και περασμένες τιμές των σφαλμάτων (καταλοίπων), e .

Αυτοπαλινδρομικά Μοντέλα Κινητού Μέσου



Γράφημα 3: Συντελεστές αυτοσυσχέτισης και μερικής αυτοσυσχέτισης ενός mixed ARMA(1,1) μοντέλου

Σύνοψη ARMA

| ΣΥΜΒΟΛΟ | ΑΥΤΟΣΥΣΧΕΤΙΣΕΙΣ | ΜΕΡΙΚΕΣ ΑΥΤΟΣΥΣΧΕΤΙΣΕΙΣ |
|----------------|--|--|
| MA(q) | Διακόπτονται μετά την τάξη q της διαδικασίας | Φθίνουν |
| AR(p) | Φθίνουν | Διακόπτονται μετά την τάξη p της διαδικασίας |
| ARMA(p,q) | Φθίνουν | Φθίνουν |

Ολοκληρωμένη Διαδικασία

- › Όταν μία σειρά μετατρέπεται σε στάσιμη παίρνοντας τις πρώτες διαφορές, λέμε ότι η σειρά είναι ολοκληρωμένη πρώτης τάξης (integrated first order) και παριστάνεται με $I(1)$.
- › Αν πάρουμε τις δεύτερες διαφορές τότε η σειρά είναι ολοκληρωμένη δεύτερης τάξης και παριστάνεται με $I(2)$.
- › Αν πάρουμε d διαφορές, τότε η σειρά είναι ολοκληρωμένη d τάξης.

Περιοριστικοί Όροι

- › Είναι καλή πρακτική να ξεκινάμε με ένα μικρό αριθμό ξεκάθαρα αιτιολογημένων παραμέτρων και να προσθέτουμε μία παράμετρο τη φορά, όταν χρειάζεται.
- › Από την άλλη πλευρά, αν οι παράμετροι σε ένα προσαρμοσμένο ARIMA μοντέλο δεν είναι σημαντικές (όπως αυτό κρίνεται από τους t λόγους), διαγράφουμε μία παράμετρο τη φορά και αναπροσαρμόζουμε το μοντέλο. Εξαιτίας της υψηλής συσχέτισης μεταξύ των εκτιμώμενων παραμέτρων, μπορεί να έχουμε την περίπτωση, κατά την οποία μία προηγουμένως μη σημαντική παράμετρος γίνεται σημαντική.

Κριτήρια Επιλογής Μοντέλου

- › Τα ARIMA μοντέλα αναγνωρίζονται (επιλέγονται) κοιτάζοντας ένα διάγραμμα των σειρών και ταιριάζοντας μοτίβα αυτοσυσχέτισης δείγματος και μερικής αυτοσυσχέτισης δείγματος με τα γνωστά θεωρητικά μοτίβα των ARIMA διαδικασιών.
- › Έχει αναπτυχθεί μία προσέγγιση στην επιλογή μοντέλου, που θεωρεί και την προσαρμογή μοντέλου και τον αριθμό των παραμέτρων.

AIC

- › Το κριτήριο πληροφόρησης του Akaike (1974), ή AIC, επιλέγει το καλύτερο μοντέλο από ένα σύνολο υποψήφίων μοντέλων σαν αυτό που ελαχιστοποιεί:

$$AIC = \ln \hat{\sigma}^2 + \frac{2}{n} r$$

όπου:

- › σ^2 = το άθροισμα των τετραγώνων των καταλοίπων διαιρεμένο με τον αριθμό των παρατηρήσεων n και όχι με τους βαθμούς ελευθερίας
- › n = ο αριθμός των παρατηρήσεων
- › r = ο συνολικός αριθμός των παραμέτρων μαζί με τον σταθερό όρο στο μοντέλο ARIMA.

BIC

- › Το Bayesian κριτήριο πληροφόρησης, που αναπτύχθηκε από τον Schwarz (1978), ή BIC, επιλέγει το μοντέλο, που ελαχιστοποιεί το:

$$BIC = \ln \hat{\sigma}^2 + \frac{\ln n}{n} r$$

Κριτήρια Επιλογής Μοντέλου

- › Ο δεύτερος όρος στο AIC και στο BIC είναι ένας παράγοντας που «τιμωρεί» τη χρησιμοποίηση μιας επιπλέον παραμέτρου στο μοντέλο. Δεδομένου ότι το BIC κριτήριο επιβάλλει μία μεγαλύτερη τιμωρία για τον αριθμό των παραμέτρων από το AIC κριτήριο, η χρήση του ελάχιστου BIC για την επιλογή μοντέλου θα οδηγήσει σε ένα μοντέλο, του οποίου ο αριθμός των παραμέτρων δεν είναι μεγαλύτερος από αυτόν, που επιλέχθηκε από το AIC.
- › Τα AIC και BIC θα πρέπει να αντιμετωπίζονται σαν επιπρόσθετες διαδικασίες, που βοηθούν στην επιλογή μοντέλου. Δεν πρέπει να χρησιμοποιούνται σαν υποκατάστατα για μία προσεκτική εξέταση των αυτοσυσχετίσεων και μερικών αυτοσυσχετίσεων δείγματος.

ARIMA Μοντέλα για Εποχικά Δεδομένα

- › Εποχικά ARIMA μοντέλα περιέχουν κανονικούς αυτοπαλινδρομικούς όρους και όρους κινούμενου μέσου, που εξηγούν τη συσχέτιση σε χαμηλά lags και εποχικούς αυτοπαλινδρομικούς όρους και όρους κινούμενου μέσου, που εξηγούν τη συσχέτιση σε εποχικά lags.
- › Για μη στάσιμες εποχικές σειρές, μία επιπλέον εποχική διαφορά απαιτείται συχνά για να καθοριστεί ολοκληρωτικά το μοντέλο.

ARIMA Μοντέλα για Εποχικά Δεδομένα

Τα εποχικά ARIMA μοντέλα συμβολίζονται με $ARIMA(p,d,q)(P,D,Q)_s$ με s την εποχικότητα, όπου:

- › p : κανονικοί αυτοπαλινδρομικοί όροι
- › d : κανονικές διαφορές
- › q : κανονικοί όροι κινούμενου μέσου
- › P : εποχικοί αυτοπαλινδρομικοί όροι
- › D : εποχικές διαφορές
- › Q : εποχικοί όροι κινούμενου μέσου

Πλεονεκτήματα ARIMA Μοντέλων

- › Παροχή ακριβών βραχυπρόθεσμων μεσοπρόθεσμων προβλέψεων.
- › Είναι αρκετά ευέλικτα και μπορούν να αναπαριστούν ένα ευρύ φάσμα των χαρακτηριστικών των χρονοσειρών, που μπορούν να συμβούν στην πράξη.
- › Οι τυπικές διαδικασίες για έλεγχο της επάρκειας του μοντέλου είναι διαθέσιμες.
- › Προβλέψεις και διαστήματα πρόβλεψης έπονται απευθείας από το προσαρμοσμένο μοντέλο.

Μειονεκτήματα ARIMA Μοντέλων

- › Απαιτείται μία σχετικά μεγάλη ποσότητα δεδομένων.
- › Δεν υπάρχουν εύκολοι τρόποι, για να αναβαθμιστούν οι παράμετροι ενός ARIMA μοντέλου, καθώς νέα δεδομένα γίνονται διαθέσιμα.
- › Η κατασκευή ενός ικανοποιητικού ARIMA μοντέλου συχνά απαιτεί μεγάλο χρόνο επένδυσης και άλλες πηγές.

Ερωτήσεις???

