



ΔΗΜΟΚΡΙΤΕΙΟ
ΠΑΝΕΠΙΣΤΗΜΙΟ
ΘΡΑΚΗΣ

DEMOCRITUS
UNIVERSITY
OF THRACE

Στοχαστική Ανάλυση Χρονοσειρών

Περιεχόμενο:

Πέρα από τα ARIMA



ΔΗΜΟΚΡΙΤΕΙΟ ΠΑΝΕΠΙΣΤΗΜΙΟ
ΘΡΑΚΗΣ

DEMOCRITUS
UNIVERSITY
OF THRACE

Παλινδρόμηση με ARIMA σφάλματα (ARIMAX)



ARIMAX

Όπως έχουμε δει, ένα ARMA(p, q) μοντέλο περιγράφεται από μία εξίσωση της μορφής:

$$y_t = \varphi_1 y_{t-1} + \dots + \varphi_p y_{t-p} - \theta_1 \varepsilon_{t-1} - \dots - \theta_q \varepsilon_{t-q} + \varepsilon_t, \varepsilon_t \sim N(0, \sigma^2)$$

Ένα ARMAX μοντέλο απλά περιέχει μία ακόμα χρονοσειρά x_t

ως επεξηγηματική ανεξάρτητη μεταβλητή στο δεξί μέλος της εξίσωσης.

$$y_t = \beta x_t + \varphi_1 y_{t-1} + \dots + \varphi_p y_{t-p} - \theta_1 \varepsilon_{t-1} - \dots - \theta_q \varepsilon_{t-q} + \varepsilon_t, \varepsilon_t \sim N(0, \sigma^2)$$

Στην παραπάνω εξίσωση ο συντελεστής β δεν είναι εύκολο να ερμηνευθεί, καθώς εκφράζει την επιρροή της x ύστερα από την αφαίρεση των p αυτοπαλινδρομικών συσχετίσεων και της επιρροής των προηγούμενων q σφαλμάτων.

Ένα μοντέλο που αντιμετωπίζει αυτή την αδυναμία και μπορεί να προσαρμοστεί στη χρονοσειρά είναι το γραμμικό μοντέλο με ARMA σφάλματα w_t :

$$y_t = \beta x_t + w_t$$

$$w_t = \varphi_1 w_{t-1} + \dots + \varphi_p w_{t-p} - \theta_1 \varepsilon_{t-1} - \dots - \theta_q \varepsilon_{t-q} + \varepsilon_t, \varepsilon_t \sim N(0, \sigma^2)$$

Στη συνέχεια, θα δούμε ορισμένους τρόπους προσαρμογής τέτοιων μοντέλων στην R.



ARIMAX: Παράδειγμα

Στο dataframe uschange της βιβλιοθήκης fpp2 παρουσιάζεται (μεταξύ άλλων) η μεταβολή στα έξοδα και στο εισόδημα, για τις ΗΠΑ ανά τετράμηνο από το έτος 1970 έως το έτος 2016.

```
library(fpp2)
```

```
head(uschange)
```

```
autoplot(uschange[,c("Consumption", "Income")]) +  
  ylab("% change") + xlab("Year")
```

Μοντέλο $\text{Consumption}(t) = \beta \text{Income}(t) + \varepsilon_t$

```
Consumption = as.vector(uschange[, "Consumption"])  
Income = as.vector(uschange[, "Income"])
```

```
my.model <- Arima(Consumption, xreg=Income)  
checkresiduals(my.model)
```

Το μοντέλο που προτείνεται είναι το: $\text{Consumption}(t) = 0.5451 + 0.2806 \cdot \text{Income}(t) + \varepsilon_t$

Ωστόσο, παρατηρούμε πως τα σφάλματα είναι εξαρτημένα μεταξύ τους γεγονός που υποδεικνύει πως υπάρχει και κάποια δομή στα υπόλοιπα της y που πρέπει να εξηγηθεί.



ARIMAX: Παράδειγμα

Δραστηριότητα: Δοκιμάζουμε με ARIMA(1, 0, 1) στα σφάλματα του προηγούμενου μοντέλου

```
my.model <- Arima(Consumption, order = c(1, 0, 1), xreg=Income)
```

```
summary(my.model)
```

```
checkresiduals(my.model)
```

(α) Να κριθεί η ποιότητα της προσαρμογής. Ποιο το μοντέλο;

(β) Να γίνει πρόβλεψη των επόμενων τιμών

```
my.forecast = forecast(my.model, xreg=rep(mean(Income), 8))
```

```
summary(my.forecast)
```

```
plot(my.forecast, showgap = FALSE)
```



ARIMAX: Παράδειγμα 2

Δοκιμάζουμε να προβλέψουμε την κατανάλωση από το εισόδημα του τρέχοντος τριμήνου όπως και το εισόδημα του προηγούμενου τριμήνου.

```
Income.lag1 = c(NA, head(Income, -1))  
my.model <- Arima(Consumption, xreg = cbind(Income, Income.lag1))  
summary(my.model)  
checkresiduals(my.model)  
  
aa.model = auto.arima(Consumption, xreg = cbind(Income, Income.lag1))  
summary(aa.model)  
checkresiduals(aa.model, lag = 20)  
  
my.forecast = forecast(aa.model, xreg=cbind(rep(mean(Income),8),  
rep(mean(Income.lag1, na.rm = T),8)))  
summary(my.forecast)  
  
plot(my.forecast, showgap = FALSE)
```



ARIMAX: Παράδειγμα 3

Πρόβλεψη Μετοχής ΕΤΕ

Θα προσπαθήσουμε να προβλέψουμε την τιμή της μετοχής της εθνικής τράπεζας από το γενικό δείκτη τιμών 15 συνεδριάσεις πριν την τρέχουσα ημέρα.

Πρώτα, διαμορφώνουμε κατάλληλα τις τιμές ώστε να είναι σε αντίστοιχη κλίμακα και δημιουργούμε τις χρονοσειρές με τις υστερήσεις που θα χρειαστούμε:

```
relative.value = function(x){  
  return((x - min(x)) / (max(x) - min(x)))  
}  
my.lag = function(x, steps = 1){  
  return(c(rep(NA, steps), head(x, -steps)))  
}  
FTSE = relative.value(stocks.df$FTSE)  
ETE = relative.value(stocks.df$ETE)  
FTSE.lag15 = my.lag(FTSE, 15)
```

Προβλέπουμε την τιμή της ΕΤΕ από την τιμή του γενικού δείκτη, 15 συνεδριάσεις πριν τη μέτρηση:

```
library(forecast)  
my.model <- Arima(ETE, xreg = cbind(FTSE.lag15))  
summary(my.model)  
  
checkresiduals(my.model)
```



ARIMAX: Παράδειγμα 3

Καθώς, είναι φανερό πως τα υπόλοιπα έχουν (στοχαστική) τάση, προχωρούμε σε μία διαφοροποίηση 1ης τάξης και επαναλαμβάνουμε τη διαδικασία:

```
library(forecast)
my.model <- Arima(ETE, xreg = cbind(FTSE.lag15), order=c(0,1,0))
summary(my.model)

checkresiduals(my.model)
```

Προχωρούμε σε πρόβλεψη με την εντολή:

```
my.forecast = forecast(my.model, xreg= tail(FTSE, 15))
summary(my.forecast)
```




**Μοντέλα όπου το ARIMA μέρος και
οι μεταβλητές είναι στην ίδια εξίσωση
Dynamic Linear Models (DLM)**



DLM - Συνάρτηση μεταφοράς

Η γενική εξίσωση ενός DLM μοντέλου είναι η παρακάτω:

$$y_t = \beta x_t + \varphi_1 y_{t-1} + \dots + \varphi_p y_{t-p} - \theta_1 \varepsilon_{t-1} - \dots - \theta_q \varepsilon_{t-q} + \varepsilon_t, \varepsilon_t \sim N(0, \sigma^2)$$

Παρατηρούμε ότι αυτή μπορεί να γραφεί ως

$$\varphi(B)y_t = \beta x_t + \theta(B)\varepsilon_t, \varepsilon_t \sim N(0, \sigma^2)$$

ή ακόμα ως

$$y_t = \frac{\beta}{\varphi(B)} x_t + \frac{\theta(B)}{\varphi(B)} \varepsilon_t, \varepsilon_t \sim N(0, \sigma^2)$$

Περαιτέρω, επιτρέποντας τη συμμετοχή χρονικών υστερήσεων της x_t μπορούμε να γράψουμε μία γενική εξίσωση της μορφής:

$$y_t = \frac{\beta(B)}{\varphi(B)} x_t + \frac{\theta(B)}{\varphi(B)} \varepsilon_t, \varepsilon_t \sim N(0, \sigma^2)$$

...και ακόμα γενικότερα...

$$y_t = \frac{\beta(B)}{u(B)} x_t + \frac{\theta(B)}{\varphi(B)} \varepsilon_t, \varepsilon_t \sim N(0, \sigma^2)$$

Ένα τέτοιο μοντέλο περιγράφεται και προσδιορίζεται από το βαθμό των πολυωνύμων β , u , φ , θ .



DLM - Συνάρτηση μεταφοράς

Η συνάρτηση `arimax` έχει τη δυνατότητα προσδιορισμού των βαθμών των πολυωνύμων αυτών. Οι σειρές πρέπει να (α) στάσιμες (β) να έχουν μέση τιμή θ .

```
library(tseries)
```

```
adf.test(Consumption)
```

```
Consumption.center <- Consumption - mean(Consumption)
```

```
Income.center <- Income - mean(Income)
```

Τώρα, είμαστε σε θέση να προσαρμόσουμε ένα μοντέλο της μορφής:

$$y_t = \beta x_t + \varphi_1 y_{t-1} + \dots + \varphi_p y_{t-p} - \theta_1 \varepsilon_{t-1} - \dots - \theta_q \varepsilon_{t-q} + \varepsilon_t, \quad \varepsilon_t \sim N(\theta, \sigma^2)$$



DLM - Συνάρτηση μεταφοράς

Επιλέγουμε, για τα σφάλματα ένα μοντέλο ARMA(0, 1) και για την επεξηγηματική μεταβλητή μία απλή γραμμική σχέση. Πιο συγκεκριμένα, προσαρμόζουμε ένα μοντέλο της μορφής:

$$y_t = \beta x_t - \theta \varepsilon_{t-1} + \varepsilon_t, \varepsilon_t \sim N(0, \sigma^2)$$

Η προσαρμογή γίνεται με την συνάρτηση `arimax` βάζοντας κατά σειρά τα εξής ορίσματα:

- Όρισμα 1: Η χρονοσειρά `Consumption.center` ως στόχος προσαρμογής.
- Όρισμα 2: Προσδιορισμός ARIMA(0, 0, 1) = MA(1) για τη μοντελοποίηση των σφαλμάτων.
- Όρισμα 3: Η `Income.center` ως επεξηγηματική μεταβλητή.
- Όρισμα 4: Θέτουμε τους συντελεστές 0 και 1 στη συνάρτηση μεταφοράς (αντιστοιχεί στο γεγονός πως η `Income` θα τοποθετηθεί δεξιά σε απλή μορφή $\alpha_1 \cdot \text{Income} + \alpha_2$
- Όρισμα 5: Ορίζουμε υποχρεωτικά $\alpha_2=0$
- Όρισμα 6: Μέθοδος εκτίμησης η μέγιστη πιθανοφάνεια.

```
my.model = arimax(Consumption.center,          # Χρονοσειρά
                  order=c(0,0,1),              # ARIMA σφάλματα
                  xtransf=data.frame(Income.center), # Επεξηγηματική χρονοσειρά
                  transfer=list(c(0,1)),        # deg(u) = 0: u(B) = 1 και deg(β) = 1: β(B) = λ1 + λ2B
                  fixed=c(NA,NA,0,NA),          # λ2 = 0.
                  method='ML')
summary(my.model); checkresiduals(my.model)
```



DLM - Συνάρτηση μεταφοράς

Call:

```
arimax(x = Consumption.center, order = c(0, 0, 1), fixed = c(NA, NA, 0, NA), method = "ML", xtransf  
= data.frame(Income.center), transfer = list(c(0, 1)))
```

Coefficients:

	ma1	intercept	Income.center-MA0	Income.center-MA1
	0.2234	0.0000	0	0.1035
s.e.	0.0715	0.0554	0	0.0497

sigma^2 estimated as 0.3821: log likelihood = -174.48, aic = 354.95

Από το output συμπεραίνουμε ότι το μοντέλο που προτείνεται είναι το:

$$\text{Consumption.center}(t) = 0.1035 \cdot \text{Income.center}(t) + 0.2234 \cdot \varepsilon_{t-1} + \varepsilon_t, \varepsilon_t \sim N(0, 0.3821)$$