

# Modele ARMA

ASC 2024

---

Piotr Żoch

- Model AR
- Model MA
- Modele ARMA
- Prognozy

ARMA



- AR: *autoregressive*
- MA: *moving average*

$$y_t = a_0 + \underbrace{\sum_{i=1}^p a_i y_{t-i}}_{AR(p)} + \underbrace{\sum_{i=0}^q \theta_i \varepsilon_{t-i}}_{MA(q)}$$

- Stacjonarny proces ARMA

$$A(L)y_t = a_0 + \theta(L)\varepsilon_t$$

można zapisać jako

$$y_t = \sum_{j=0}^{\infty} \psi_j L^j \varepsilon_t + \kappa_t$$

$$\psi_0 = 1$$

gdzie  $\kappa_t$  jest składnikiem deterministycznym.

- Mnożnik dynamiczny mierzy efekt  $\varepsilon_t$  na wartości  $\{y_{t+h}\}_{h=0}^{\infty}$ :

$$\frac{\partial y_{t+h}}{\partial \varepsilon_t} = \frac{\partial y_h}{\partial \varepsilon_0} = \psi_h$$

- Funkcja reakcji na impuls to  $\{\psi_h\}_{h=0}^{\infty}$  - mówi nam o tym, jak reagują wartości  $\{y_{t+h}\}_{h=0}^{\infty}$  na jednorazowy impuls/szok  $\varepsilon_t$ .

- Mnożnik skumulowany: efekt permanentnej zmiany  $\varepsilon_t = \varepsilon_{t+1} = \varepsilon_{t+2} = \dots$ .

$$\frac{\partial y_{t+h}}{\partial \varepsilon_t} + \frac{\partial y_{t+h}}{\partial \varepsilon_{t+1}} + \frac{\partial y_{t+h}}{\partial \varepsilon_{t+2}} + \dots + \frac{\partial y_{t+h}}{\partial \varepsilon_{t+h}} = \sum_{j=0}^h \psi_{h-j}$$

- Dla  $h \rightarrow \infty$  mamy  $\sum_{j=0}^{\infty} \psi_j = \Psi(1)$ .

Model AR

---



- Model AR( $p$ ) (rzędu  $p$ ):

$$y_t = a_0 + a_1 y_{t-1} + a_2 y_{t-2} + \dots + a_p y_{t-p} + \epsilon_t$$

gdzie  $\epsilon_t$  i.i.d. z wartością oczekiwaną 0 i wariancją  $\sigma^2$ .

- Przykład - AR(1)

$$y_t = \rho y_{t-1} + \epsilon_t$$

można zapisać jako

$$y_t = \epsilon_t + \rho \epsilon_{t-1} + \rho^2 \epsilon_{t-2} + \dots$$

jeśli  $|\rho| < 1$ .

# Model AR(1)

- Mamy

$$y_t = \epsilon_t + \rho\epsilon_{t-1} + \rho^2\epsilon_{t-2} + \dots$$

- Mnożnik dynamiczny:

$$\frac{\partial y_{t+h}}{\partial \epsilon_t} = \rho^h$$

- Mnożnik skumulowany:

$$1 + \rho + \rho^2 + \dots = \frac{1}{1 - \rho}$$

- Model AR(p) to

$$y_t = a_0 + a_1 y_{t-1} + a_2 y_{t-2} + \dots + a_p y_{t-p} + \epsilon_t$$

- Możemy zbadać stabilność analizując pierwiastki *wielomianu opóźnień* (szczegółowa dyskusja wielomianów opóźnień na kolejnych zajęciach)

$$1 - a_1 L - a_2 L^2 - \dots - a_p L^p.$$

- Stabilność gdy wszystkie pierwiastki są *poza* kołem jednostkowym.

# Model AR(p)

- Inne spojrzenie na model AR(p): niech

$$x_{0,t} = y_t$$

$$x_{1,t} = y_{t-1}$$

$$x_{2,t} = y_{t-2}$$

i tak dalej.

- Możemy zapisać AR(p) jako

$$\begin{bmatrix} x_{0,t} \\ x_{1,t} \\ \vdots \\ x_{p,t} \end{bmatrix} = \begin{bmatrix} a_0 \\ 0 \\ \vdots \\ 0 \end{bmatrix} + \begin{bmatrix} a_1 & a_2 & \dots & a_p & 0 \\ 1 & 0 & \dots & 0 & 0 \\ \vdots & \vdots & \ddots & 0 & 0 \\ 0 & 0 & \dots & 0 & 1 \end{bmatrix} \begin{bmatrix} x_{0,t-1} \\ x_{1,t-1} \\ \vdots \\ x_{p,t-1} \end{bmatrix} + \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \epsilon_t$$

# Model AR(p)

- Sprawdźmy:

$$\begin{bmatrix} x_{0,t} \\ x_{1,t} \\ \vdots \\ x_{p,t} \end{bmatrix} = \begin{bmatrix} a_0 \\ 0 \\ \vdots \\ 0 \end{bmatrix} + \begin{bmatrix} a_1 & a_2 & \dots & a_p & 0 \\ 1 & 0 & \dots & 0 & 0 \\ \vdots & \vdots & \ddots & 0 & 0 \\ 0 & 0 & \dots & 0 & 1 \end{bmatrix} \begin{bmatrix} x_{0,t-1} \\ x_{1,t-1} \\ \vdots \\ x_{p,t-1} \end{bmatrix} + \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \epsilon_t$$

- Pierwszy wiersz to

$$\underbrace{x_{0,t}}_{y_t} = a_0 + a_1 \underbrace{x_{0,t-1}}_{y_{t-1}} + a_2 \underbrace{x_{1,t-1}}_{y_{t-2}} + \dots + a_p \underbrace{x_{p-1,t-1}}_{y_{t-p}} + \epsilon_t.$$

- Drugi wiersz to

$$\underbrace{x_{1,t}}_{y_{t-1}} = \underbrace{x_{0,t-1}}_y.$$

- Kolejne wiersze są podobne do drugiego.

# Model AR(p)

- Możemy zapisać AR(p) jako

$$\mathbf{x}_t = \begin{bmatrix} a_0 \\ 0 \\ \vdots \\ 0 \end{bmatrix} + \mathbf{A}\mathbf{x}_{t-1} + \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \epsilon_t$$

- Ważne: występuje tylko pierwsze opóźnienie wektora  $\mathbf{x}_t$ !
- Korzyść: możemy badać czy AR(p) da się zapisać jako MA( $\infty$ ) analizując macierz  $\mathbf{A}$ :

$$\mathbf{x}_t = (\mathbf{I} - \mathbf{A})^{-1} \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \epsilon_t$$

## Model AR(p) - momenty

- Wartość oczekiwana:

$$E[y_t] = a_0 + a_1 E[y_{t-1}] + a_2 E[y_{t-2}] + \dots + a_p E[y_{t-p}] + E[\epsilon_t],$$

gdzie

$$E[\epsilon_t] = 0.$$

- Dla procesu stacjonarnego

$$E[y_t] = E[y_{t-1}] = \dots$$

czyli

$$E[y_t] = \frac{a_0}{1 - a_1 - a_2 - \dots}$$



- Równania Yule'a-Walkera: kowariancja  $\gamma_k \equiv \text{Cov}(y_t, y_{t-k})$  spełnia

$$\gamma_k = a_1 \gamma_{k-1} + \dots + a_p \gamma_{k-p}$$

$$\gamma_0 = a_1 \gamma_1 + \dots + a_p \gamma_p + \sigma^2$$

- Korelacja:

$$\rho_k = a_1 \rho_{k-1} + \dots + a_p \rho_{k-p}$$

- Mnożniki:

$$\psi_k = a_1 \psi_{k-1} + \dots + a_p \psi_{k-p}$$

## Model MA

---

- Model MA( $q$ ) (rzędu  $q$ )

$$y_t = \mu + \epsilon_t + \theta_1 \epsilon_{t-1} + \dots + \theta_q \epsilon_{t-q}$$

gdzie  $\epsilon_t$  i.i.d. z wartością oczekiwaną 0 i wariancją  $\sigma^2$ .

- Model średniej ruchomej (*moving average*) to średnia ważona procesu białego szumu z bieżącego oraz  $q$  poprzednich okresów.

- Trudność: jak oszacować parametry tego modelu?
  - Nie obserwujemy  $\epsilon_t, \epsilon_{t-1}, \dots$  w danych.

- Momenty modelu MA(q):

$$E[y_t] = a_0$$

$$\text{Var}[y_t] = (1 + \theta_1^2 + \dots + \theta_q^2) \sigma^2$$

$$\text{Cov}[y_t, y_{t-k}] = \begin{cases} 0 & k > q \\ (\theta_k + \theta_{k+1}\theta_1 + \dots + \theta_q\theta_{q-k}) \sigma^2 & k \leq q \end{cases}$$

- Rozważmy model MA(1)

$$\begin{aligned}y_t &= \epsilon_t + \theta_1 \epsilon_{t-1} \\ &= (1 + \theta_1 L) \epsilon_t\end{aligned}$$

- Można go zapisać jako

$$\epsilon_t = y_t - \theta_1 y_{t-1} + \theta_1^2 y_{t-2} + \dots$$

jeśli  $|\theta_1| < 1$ .

- W tym wypadku mówimy, że model MA(1) jest **odwracalny** i można przedstawić MA(1) jako AR( $\infty$ ).

## Momenty MA(1)

- Momenty

$$y_t = \epsilon_t + \theta_1 \epsilon_{t-1}$$

to

$$\gamma_0 = (1 + \theta_1^2) \sigma^2$$

$$\gamma_1 = \theta_1 \sigma^2$$

- Autokorelacja

$$\begin{aligned} \rho_1 &= \frac{\theta_1}{1 + \theta_1^2} \\ &= \frac{\frac{1}{\theta_1}}{1 + \left(\frac{1}{\theta_1}\right)^2} \end{aligned}$$

- Dwa modele MA(1)

$$y_t = \epsilon_t + \theta_1 \epsilon_{t-1}$$

$$y_t = \epsilon_t + \frac{1}{\theta_1} \epsilon_{t-1}$$

mają taką samą autokorelację (ACF)

- Jeden z nich jest odwracalny, drugi nie.
- Ten odwracalny nazywamy **postacią fundamentalną** modelu MA(1).



## Modele ARMA w praktyce

---

$$y_t = a_0 + \sum_{i=1}^p a_i y_{t-i} + \sum_{i=0}^q \theta_i \epsilon_{t-i}$$

- Jak dobrać  $(p,q)$ ?
  - Metoda Boxa-Jenkinsa (1970)
  - Kryteria informacyjne

- Idea: użyć wykresów ACF i PACF do identyfikacji rzędów  $p$  i  $q$ .

- Funkcja autokorelacji (Autocorrelation Function) to współczynnik korelacji między dwiema realizacjami

$$\rho_k = \frac{\text{Cov}(y_t, y_{t-k})}{\text{Var}(y_t)}$$

i  $\rho_k \in [-1, 1]$

- Funkcja autokorelacji cząstkowej (Partial Autocorrelation Function) to współczynnik korelacji między dwiema realizacjami oddalonymi od siebie o  $k$  okresów bez uwzględnienia wpływu wszystkich obserwacji “pomiędzy”
- Funkcja ta jest równa wyestymowanemu współczynnikowi  $\alpha_k$  w modelu autoregresyjnym  $k$ -tego rzędu

$$y_t = \mu + \alpha_1 y_{t-1} + \dots + \alpha_k y_{t-k} + \varepsilon_t$$

- Idea: użyć wykresów ACF i PACF do identyfikacji rzędów  $p$  i  $q$ .
  - Dla MA:

$$\text{Cov}[y_t, y_{t-k}] = \begin{cases} 0 & k > q \\ (\theta_k + \theta_{k+1}\theta_1 + \dots + \theta_q\theta_{q-k}) \sigma^2 & k \leq q \end{cases}$$

więc w ACF istotnych jest  $q$  pierwszych opóźnień

- Dla AR w PACF istotnych jest  $p$  pierwszych opóźnień

- Ustalenie wartości  $p$  i  $q$  na podstawie wykresu ACF i PACF.
- Estymacja parametrów modelu ARMA( $p,q$ )
- Weryfikacja modelu ARMA( $p,q$ ):
  - Dla modelu ARMA( $p^*,q^*$ ), gdzie  $p^* \geq p$  i  $q^* \geq q$ , werykuje się istotność dodatkowych opóźnień.
  - Sprawdza się, czy występuje autokorelacja reszt.



- **Kryteria informacyjne:** wybrać specyfikację o wysokiej wartości logarytmu funkcji wiarygodności  $\ell = \ln L$  i niskiej liczbie szacowanych parametrów  $k = p + q + 1$ .
- Najczęściej stosowane: Akaike'a (AIC), Schwarza (BIC) oraz Hannana-Quinna (HQIC):

$$AIC = -2\frac{\ell}{T} + 2\frac{k}{T}$$

$$BIC = -2\frac{\ell}{T} + 2\frac{k}{T} \ln T$$

$$HQIC = -2\frac{\ell}{T} + 2\frac{k}{T} \ln(\ln T)$$

- Nagroda za dobre dopasowanie do danych, kara za liczbę parametrów.

- Polega na stopniowym upraszczaniu możliwie najogólniejszego modelu początkowego
- Ograniczenia narzucane na model definiowane przez zagnieżdżone hipotezy:
  - $H_0^K$  zawiera najwięcej ograniczeń,  $H_0^1$  najmniej.
- Hipotezą alternatywną jest zawsze postać ogólna.

## Metoda od ogólnego do szczegółowego

- Przykład: ogólna postać

$$y_t = \beta_1 y_{t-1} + \beta_2 y_{t-2} + \epsilon_t$$

i szczególne postaci

$$y_t = \beta_1 y_{t-1} + 0 \cdot y_{t-2} + \epsilon_t$$

$$y_t = 0 \cdot y_{t-1} + \beta_2 \cdot y_{t-2} + \epsilon_t$$

$$y_t = 0 \cdot y_{t-1} + 0 \cdot y_{t-2} + \epsilon_t$$

- Problem: w jakiej kolejności testować?
- W przypadku szeregów czasowych zazwyczaj zagnieżdżanie pod względem kolejności opóźnień.

- Test LR (likelihood ratio):

$$LR = -2 [\ell(\theta_R) - \ell(\theta)]$$

gdzie  $L(\theta_R)$  odpowiada modelowi z restrykcjami. Można użyć tylko w przypadku modeli zagnieżdżonych.

- Przy  $H_0$  (restrykcje) statystyka testowa ma rozkład  $\chi^2$  z liczbą stopni swobody równą liczbie niezależnych liniowo restrykcji.

$$y_t = a_0 + \sum_{i=1}^p a_i y_{t-i} + \sum_{i=0}^q \theta_i \epsilon_{t-i}$$

przyjmijmy, że znamy  $y_t$  oraz reszty  $e_t$  dla  $t \leq T$ . Interesuje nas prognoza na okresy  $T+1$ ,  $T+2$  i tak dalej.

$$y_{T+1} = a_0 + \sum_{i=1}^p a_i y_{T+1-i} + \sum_{i=1}^q \theta_i e_{T+1-i} + \epsilon_{T+1}$$

czyli

$$E_T[y_{T+1}] = a_0 + \sum_{i=1}^p a_i y_{T+1-i} + \sum_{i=1}^q \theta_i e_{T+1-i}$$

- Prognozy punktowe obliczamy w sposób rekurencyjny.

$$\begin{aligned} E_T[y_{T+2}] &= a_0 + E_T \left[ \sum_{i=1}^p a_i y_{T+2-i} \right] + \sum_{i=2}^q \theta_i e_{T+2-i} \\ &\quad + \theta_1 E_T[\epsilon_{T+1}] + E_T[e_{T+2}] \\ &= a_0 + a_1 E_T[y_{T+1}] + \sum_{i=2}^p a_i y_{T+2-i} + \sum_{i=2}^q \theta_i e_{T+2-i} \end{aligned}$$

- Dla  $k > p$  i  $k > q$  mamy

$$E_T[y_{T+k}] = a_0 + \sum_{i=1}^p a_i E_T y_{T+k-i}$$

# Wariancja błędu prognozy

- Postać  $MA(\infty)$ :

$$y_t = \epsilon_t + \psi_1 \epsilon_{t-1} + \psi_2 \epsilon_{t-2} + \dots$$

- Błąd losowy prognozy:

$$y_{T+1} - E_T[y_{T+1}] = \epsilon_{T+1}$$

$$y_{T+2} - E_T[y_{T+2}] = \epsilon_{T+2} + \psi_1 \epsilon_{T+1}$$

i tak dalej:

$$y_{T+k} - E_T[y_{T+k}] = \epsilon_{T+k} + \sum_{i=1}^{k-1} \psi_i \epsilon_{T+k-i}$$

- Ponieważ  $\text{Cov}[\epsilon_t, \epsilon_s] = 0$  dla  $s \neq t$  oraz  $\text{Var}[\epsilon_t] = \sigma_\epsilon^2$  mamy

$$\sigma_k^2 = \text{Var}[y_{T+k}] = \sigma_\epsilon^2 \left( 1 + \sum_{i=1}^{k-1} \psi_i^2 \right)$$

- Nie uwzględnia to innych błędów prognozy (np. błędnej specyfikacji).
- $\sigma_k^2$  wykorzystujemy do stworzenia prognozy przedziałowej.