Lecture 4. Moving Average Models

MA model.

1. Properties

AR : present value with $x_t = x_{t-1}, x_{t-2} \cdots$

MA : " $x_t = \varepsilon_t, \varepsilon_{t-1}, \cdots$

Definition 1.1. An *moving average* (MA) model of order q is defined by

$$x_t = \mu + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \cdots + \theta_q \varepsilon_{t-q} = \mu + \theta(L)\varepsilon_t,$$

where $\theta(L) = 1 + \theta_1 L + \theta_2 L^2 + \cdots + \theta_q L^q$ and $\varepsilon_t \sim WN(0, \sigma_\varepsilon^2)$.

- The MA model is "always" stationary because it is a linear combination of white noise errors that are stationary.
- For an MA(q) model, it shows that

$k=1$

q

$$\begin{aligned} E[x_t] &= E[\mu + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \cdots + \theta_q \varepsilon_{t-q}] \\ &= \mu, \qquad q \neq 1. \end{aligned}$$

$\varepsilon_t + \theta_1\varepsilon_{t-1} + \theta_2\varepsilon_{t-2} + \cdots + \theta_{q-1}\varepsilon_{t-q+1} + \theta_q\varepsilon_{t-q}$

$\varepsilon_{t-k} + \theta_1\varepsilon_{t-k-1} + \theta_2\varepsilon_{t-k-2} + \cdots + \theta_q\varepsilon_{t-k-q}$

$$\begin{aligned} Var[x_t] &= Var[\mu + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \cdots + \theta_q \varepsilon_{t-q}] \\ &= Var[\varepsilon_t] + \theta_1^2 Var[\varepsilon_{t-1}] + \cdots + \theta_q^2 Var[\varepsilon_{t-q}] \\ &= \underline{(1 + \theta_1^2 + \cdots + \theta_q^2)\sigma_\varepsilon^2}, \end{aligned}$$

$Var(\varepsilon_{t-1}) = \sigma_\varepsilon^2$

and

$$\begin{aligned} \gamma_k &= E[(x_t - E[x_t])(x_{t-k} - E[x_{t-k}])] \\ &= E[(\underline{\varepsilon_t + \theta_1 \varepsilon_{t-1} \cdots + \theta_q \varepsilon_{t-q}})(\varepsilon_{t-k} + \theta_1 \varepsilon_{t-1-k} \cdots + \theta_q \varepsilon_{t-q-k})] \\ &= \begin{cases} (\theta_k + \theta_{k+1}\theta_1 + \theta_{k+2}\theta_2 + \cdots + \theta_q \theta_{q-k})\sigma_\varepsilon^2 & \text{for } k = 1, 2, \ldots, q \\ 0 & \text{for } k > q. \end{cases} \end{aligned}$$

Remark 1.2. For an MA(q) model, the kth order autocorrelation is given by

$$\rho_k = \begin{cases} \frac{\theta_k + \theta_{k+1}\theta_1 + \theta_{k+2}\theta_2 + \cdots + \theta_q \theta_{q-k}}{1 + \theta_1^2 + \cdots + \theta_q^2} & \text{for } k = 1, 2, \ldots, q \\ 0 & \underline{\text{for } k > q.} \end{cases}$$

Notice that the ACF cuts off at lag q. auto collection of correlation

If MA(3) $\rho_1, \rho_2, \rho_3 = ?$

$\rho_4 = 0$.

Example 1.3. Consider an MA(1) process

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

$$
\begin{aligned}
x_t &= \mu + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \cdots + \theta_q \varepsilon_{t-q} \\
&= \mu + \varepsilon_t + \theta_1 L \varepsilon_t + \cdots + \theta_q L^q \varepsilon_t \\
&= \mu + \varepsilon_t \underbrace{(1 + \theta_1 L + \cdots + \theta_q L^q)}_{\theta(L)}
\end{aligned}
$$

MA(1).

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

$$
\begin{aligned}
\gamma_k &= E[(x_t - E[x_t])(x_{t-k} - E[x_{t-k}])] \\
&= E[(\varepsilon_t + \theta_1 \varepsilon_{t-1})(\varepsilon_{t-k} + \theta_1 \varepsilon_{t-k-1})] \\
&= E[\varepsilon_t \cdot \varepsilon_{t-k} + \theta_1 \varepsilon_t \cdot \varepsilon_{t-k-1} + \theta_1 \varepsilon_{t-1} \varepsilon_{t-k} + \theta_1^2 \varepsilon_{t-1} \varepsilon_{t-k-1}]
\end{aligned}
$$

$k=1$ $\quad E[\theta_1 \cdot \varepsilon_{t-1}^2] = \theta_1 \underline{E[\varepsilon_{t-1}^2]} = \theta_1 \cdot \underline{\sigma_\varepsilon^2}$

$Var(\varepsilon_{t-1}) = E[(\varepsilon_{t-1} - \cancel{E(\varepsilon_{t-1})})^2] = E[\varepsilon_{t-1}^2]$

$k=2,\ k=3 \Rightarrow 0.$

MA(2) process.

$$x_t = \mu + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2}$$

ACF : auto correlation Function.

$E[x_t]$, $Var[x_t]$, γ_k , ρ_k

$$E[x_t] = E[\mu + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2}] = \mu$$

$$\begin{aligned} Var[x_t] &= Var[\mu + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2}] \\ &= Var[\varepsilon_t] + \theta_1^2 Var[\varepsilon_{t-1}] + \theta_2^2 Var[\varepsilon_{t-2}] \\ &= (1 + \theta_1^2 + \theta_2^2)\sigma_\varepsilon^2 \end{aligned}$$

$$\begin{aligned} \gamma_k &= Cov[x_t, x_{t-k}] \\ &= E[(x_t - \mu)(x_{t-k} - \mu)] \\ &= E[(\varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2})(\varepsilon_{t-k} + \theta_1 \varepsilon_{t-k-1} + \theta_2 \varepsilon_{t-k-2})] \\ &= E[\cancel{\varepsilon_t \cdot \varepsilon_{t-k}} + \cancel{\varepsilon_t \cdot \theta_1 \varepsilon_{t-k-1}} + \cancel{\varepsilon_t \cdot \theta_2 \varepsilon_{t-k-2}} \\ &\quad + \theta_1 \varepsilon_{t-1} \varepsilon_{t-k} + \cancel{\theta_1 \varepsilon_{t-1} \cdot \theta_1 \varepsilon_{t-k-1}} + \cancel{\theta_1 \varepsilon_{t-1} \theta_2 \varepsilon_{t-k-2}} \\ &\quad + \theta_2 \varepsilon_{t-2} \varepsilon_{t-k} + \theta_2 \varepsilon_{t-2} \theta_1 \varepsilon_{t-k-1} + \cancel{\theta_2 \varepsilon_{t-2} \theta_2 \varepsilon_{t-k-2}}] \\ &= \theta_1 E[\varepsilon_{t-1} \varepsilon_{t-k}] + \theta_2 E[\varepsilon_{t-2} \varepsilon_{t-k}] + \theta_1 \theta_2 E[\varepsilon_{t-2} \varepsilon_{t-k-1}] \end{aligned}$$

$$\gamma_k \begin{cases} (\theta_1 + \theta_1 \theta_2)\sigma_\varepsilon^2 & k = 1. \\ \theta_2 \sigma_\varepsilon^2 & k = 2 \\ 0 & k > 2 \end{cases}$$

$$\rho_k \begin{cases} \dfrac{\theta_1 + \theta_1 \theta_2}{1 + \theta_1^2 + \theta_2^2} & k = 1 \\ \dfrac{\theta_2}{1 + \theta_1^2 + \theta_2^2} & k = 2 \\ 0 & k > 2 \end{cases}$$

where $\varepsilon_t \sim WN(0, \sigma_\varepsilon^2)$. Then, the kth-order autocovariance is

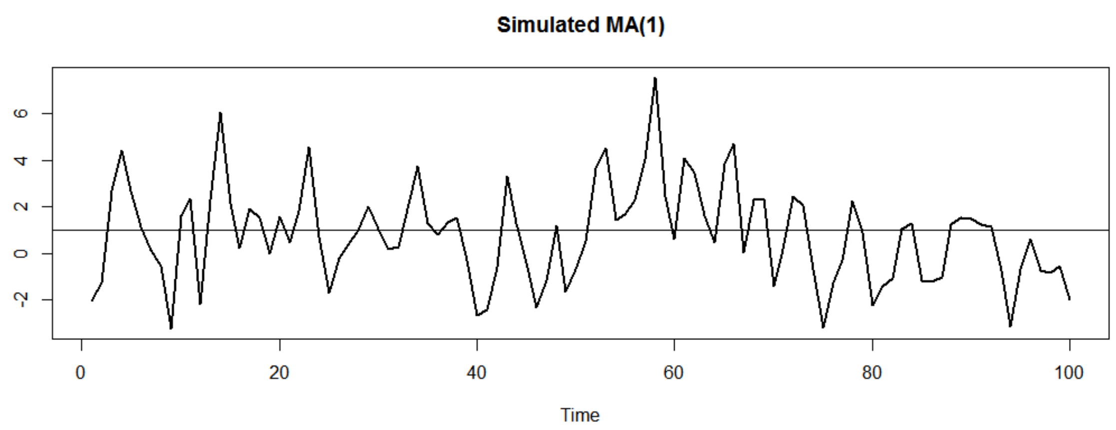
$$\begin{aligned}
\gamma_k &= E[(x_t - \mu)(x_{t-k} - \mu)] \\
&= E[(\varepsilon_t + \theta_1 \varepsilon_{t-1})(\varepsilon_{t-k} + \theta_1 \varepsilon_{t-1-k})] \\
&= E[\varepsilon_t \varepsilon_{t-k} + \theta_1 \varepsilon_t \varepsilon_{t-1-k} + \theta_1 \varepsilon_{t-1} \varepsilon_{t-k} + \theta_1^2 \varepsilon_{t-1} \varepsilon_{t-1-k}] \\
&= \begin{cases} \theta_1 \sigma_\varepsilon^2 & \text{for } k = 1 \\ 0 & \text{for } k > 1 \end{cases}
\end{aligned}$$

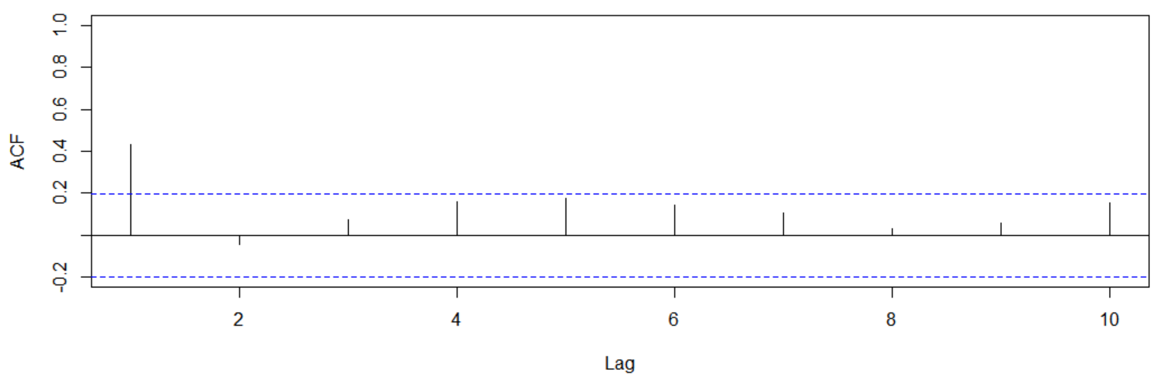
and the kth-order autocorrelation is

$$\rho_k = \begin{cases} \frac{\theta_1}{1+\theta_1^2} & \text{for } k = 1 \\ 0 & \text{for } k > 1. \end{cases}$$

Example 1.4. Simulate an MA(1) process with $\mu = 1$, $\theta_1 = 0.8$, $\sigma_\varepsilon^2 = 3$, and $T = 100$.

```
> x <- 1 + arima.sim(list(order = c(0, 0, 1), ma = 0.8), sd = sqrt(3), n = 100)
> par(mfrow = c(2, 1))
> plot(x, lwd = 2, xlab = "Time", ylab ="", main = "Simulated MA(1)")
> abline(h = 1)
> acf(x, lag = 10, main ="")
```





Theorem 1.5. *Consider the infinite series*

$$1+\phi_1 L+\phi_1^2 L^2+\cdots$$

and its nth partial sum

$$S_n = 1+\phi_1 L+\phi_1^2 L^2+\cdots+\phi_1^{n-1}L^{n-1}.$$

If $|\phi_1|<1$, *then it shows*

$$\lim_{n\to\infty} S_n = \frac{1}{1-\phi_1 L}.$$

Proof. Note

$$\begin{aligned} S_n-\phi_1 L S_n &= (1+\phi_1 L+\phi_1^2 L^2+\cdots+\phi_1^{n-1}L^{n-1})-(\phi_1 L+\phi_1^2 L^2+\phi_1^3 L^3+\cdots+\phi_1^n L^n) \\ &= 1-\phi_1^n L^n \end{aligned}$$

and hence

ϕ_1

$\|$

$$S_n = \frac{1-\phi_1^n L^n}{1-\phi_1 L}.$$

If $|\phi_1|<1$, then $|\phi_1 L|<1$ and $\phi_1^n L^n \to 0$ as $n\to\infty$, and hence

$$\lim_{n\to\infty} S_n = \frac{1}{1-\phi_1 L}.$$

□

Theorem 1.6. *The Wold theorem states that "any stationary" process* $\{x_t\}$ *can be represented as a "linear" time series in the form*

$$x_t = \mu + \sum_{i=0}^{\infty} \psi_i \varepsilon_{t-i},$$

where $\varepsilon_t \sim WN(0,\sigma_\varepsilon^2)$, $\mu = E[x_t]$, $\psi_0 = 1$, *and* $\sum_{i=0}^{\infty}\psi_i^2 < \infty$.

- Importantly, the Wold theorem implies that any stationary AR model has the corresponding MA model.

Example 1.7. Consider a stationary AR(1) model

$$x_t-\mu = \phi_1(x_{t-1}-\mu)+\varepsilon_t,$$

where $\varepsilon_t \sim WN(0,\sigma_\varepsilon^2)$ and $|\phi_1|<1$. Using a lag operator, one obtains

$$(1-\phi_1 L)(x_t-\mu) = \varepsilon_t. \tag{1.1}$$

Since the stationarity condition $|\phi_1|<1$ ensures

$$1+\phi_1 L+\phi_1^2 L^2+\cdots = \sum_{i=0}^{\infty}\phi_1^i L^i = \frac{1}{1-\phi_1 L},$$

$x_t - \mu = \phi_1 (x_{t-1} - \mu) + \varepsilon_t$ AR(1)

$|\phi_1| < 1$

$\phi_1 L (x_t - \mu)$

$\Rightarrow (x_t - \mu)(1 - \phi_1 L) = \varepsilon_t$

$$x_t = \mu + \frac{\varepsilon_t}{1 - \phi_1 L}$$

$$= \mu + \varepsilon_t \underbrace{(1 + \phi_1 L + \phi_1^2 L^2 + \phi_1^3 L^3 + \cdots)}_{= \frac{1}{1-\phi_1 L}}$$

$$= \underbrace{\mu + \varepsilon_t + \phi_1 \varepsilon_{t-1} + \phi_1^2 \varepsilon_{t-2} + \phi_1^3 \varepsilon_{t-3} + \cdots}_{MA(\infty)}$$

AR $\Rightarrow$ MA.

stationarity

$\phi(L)$

AR $\Rightarrow$ MA

$\theta(L)$

MA $\Rightarrow$ AR

invertible.

(solution of characteristic eq are grater than 1 in modulus)

AR $\xrightarrow{\text{stationary}}$ MA $\to$ $\phi(L) = 0$ all sols : greater than 1 in modulus

MA $\xrightarrow{\text{invertible}}$ AR. $\to$ $\theta(L) = 0$

stationary AR(2) model. find corresponding MA model.

$$x_t = \phi_0 + \phi_1 x_{t-1} + \phi_2 x_{t-2} + \varepsilon_t$$

$$|\phi_1| < 1 \quad |\phi_2| < 1$$

$$x_t - \mu = \phi_1 (x_{t-1} - \mu) + \phi_2 (x_{t-2} - \mu) + \varepsilon_t$$

$$x_t - \mu = \phi_1 L (x_t - \mu) + \phi_2 L^2 (x_{t-2} - \mu) + \varepsilon_t$$

$$(x_t - \mu)(1 - \phi_1 L - \phi_2 L^2) = \varepsilon_t .$$

$$x_t = \mu + \frac{\varepsilon_t}{1 - \phi_1 L - \phi_2 L^2} \qquad ??$$

one writes (1.1) as

$$x_t = \mu + \frac{\varepsilon_t}{1-\phi_1 L} = \mu + \sum_{i=0}^{\infty} \phi_1^i \varepsilon_{t-i},$$

which is the Wold representation of the AR(1) model.

Definition 1.8. An MA(q) process is *invertible* if all of the solutions of the characteristic equation

$$\theta(z) = 1 + \theta_1 z + \theta_2 z^2 + \cdots + \theta_q z^q = 0$$

are greater than 1 in modulus.

Theorem 1.9. *If an MA process is invertible, there exists a corresponding AR process.*

Example 1.10. Consider an MA(1) model of the form

$$x_t = \varepsilon_t - \theta_1 \varepsilon_{t-1}, \tag{1.2}$$

where $\{\varepsilon_t\}$ is a white noise process. The characteristic equation of the MA(1) model is

$$\theta(z) = 1 - \theta_1 z = 0$$

and, hence, the solution is $z = \theta_1^{-1}$. Rewriting (1.2) as $\varepsilon_t = x_t + \theta_1 \varepsilon_{t-1}$, one obtains

$$\begin{aligned} \varepsilon_t &= x_t + \theta_1(x_{t-1} + \theta_1 \varepsilon_{t-2}) \\ &= x_t + \theta_1 x_{t-1} + \theta_1^2 \varepsilon_{t-2} \\ &= x_t + \theta_1 x_{t-1} + \theta_1^2 x_{t-2} + \cdots \\ &= (1 + \theta_1 L + \theta_1^2 L^2 + \cdots) x_t. \end{aligned} \tag{1.3}$$

Since ε_t is stationary by definition, it must hold that $1 + \theta_1 L + \theta_1^2 L^2 + \cdots$ converges to $(1-\theta_1 L)^{-1}$ in the RHS of (1.3), which requires the condition $|\theta_1| < 1$. Therefore, if $|\theta_1| < 1$ or equivalently $|z| > 1$, the MA(1) model in (1.2) has a corresponding AR process of the form

$$\begin{aligned} x_t &= -\theta_1 \varepsilon_{t-1} + \varepsilon_t \\ &= -\theta_1(L + \theta_1 L^2 + \theta_1^2 L^3 + \cdots) x_t + \varepsilon_t \\ &= -\theta_1 x_{t-1} - \theta_1^2 x_{t-2} - \theta_1^3 x_{t-3} - \cdots + \varepsilon_t. \end{aligned}$$

2. Order Determination

For a time series x_t with ACF ρ_k, if $\rho_k \neq 0$, but $\rho_k = 0$ for $k > q$, then x_t follows an MA(q) model.

Example 2.1. Consider monthly returns of the CRSP equal-weighted index from January 1926 to December 2008.

```
> mydat <- read.table("data1.txt", header = T)
> head(mydat)
      date    ibmrtn    vwrtn    ewrtn    sprtn
```

MA(1).

$$x_t = \varepsilon_t - \theta_1 \varepsilon_{t-1}$$

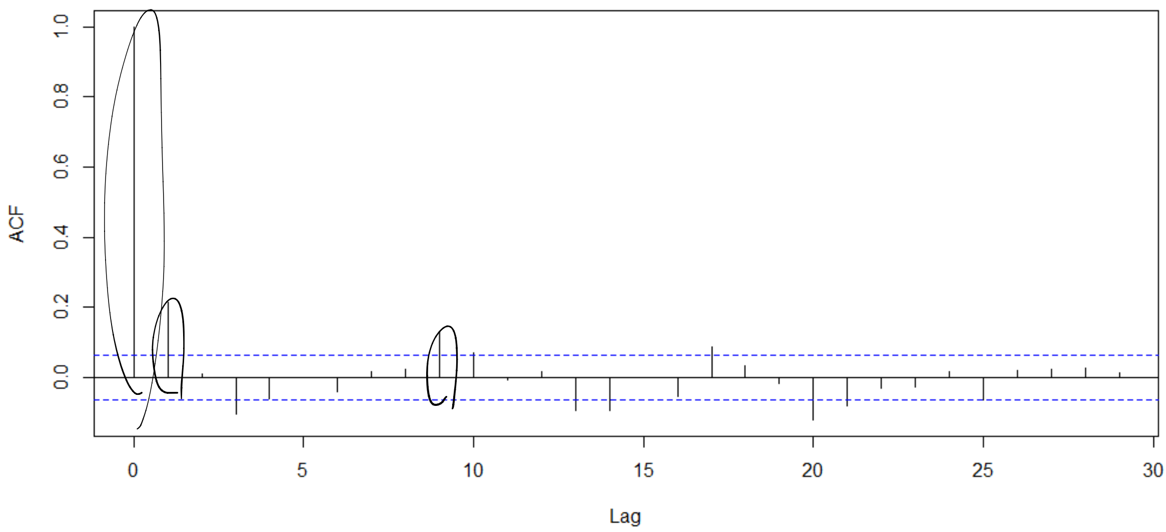
$$\theta(z) = 1 - \theta_1 z = 0.$$

$$z = \theta_1^{-1}$$

$$\begin{aligned}
\varepsilon_t &= x_t + \theta_1 \varepsilon_{t-1} \\
&= x_t + \theta_1 (x_{t-1} + \theta_1 \varepsilon_{t-2}) \\
&= x_t + \theta_1 x_{t-1} + \theta_1^2 \cdot \varepsilon_{t-2} \\
&= x_t + \theta_1 x_{t-1} + \theta_1^2 x_{t-2} + \theta_1^3 x_{t-3} + \cdots \\
&= x_t (1 + \theta_1 L + \theta_1^2 L^2 + \theta_1^3 L^3 + \cdots)
\end{aligned}$$

$$\begin{aligned}
x_t &= -\theta_1 \varepsilon_{t-1} + \varepsilon_t \\
&= -\theta_1 (x_{t-1} (1 + \theta_1 L + \theta_1^2 L^2 + \cdots)) + \varepsilon_t \\
&= -\theta_1 x_{t-1} - \theta_1^2 x_{t-2} - \theta_1^3 x_{t-3} - \cdots + \varepsilon_t
\end{aligned}$$

```
1 19260130 -0.010381 0.000724 0.023174 0.022472
...
6 19260630  0.068493 0.056888 0.050487 0.043184
> ewrtn <- mydat$ewrtn
> acf(ewrtn, main = "")
```



- The autocorrelation is significant at lags 1, 3, and 9.

3. Parameter Estimation

order q

Consider an MA(q) model of the form

$$x_t = \mu + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \cdots + \theta_q \varepsilon_{t-q},$$

$\theta_1 \sim \theta_q, \mu$

where $\varepsilon_t \sim WN(0, \sigma_\varepsilon^2)$. *Maximum likelihood estimation* is used to estimate the parameters of the MA(q) model.

- In estimating MA(q) models, one needs to decide how to treat the first q values of ε_t. In the *conditional likelihood* method, the missing lagged shocks are set to zero. In the *exact likelihood* method, the missing lagged shocks are treated as additional parameters of the model and estimated jointly with other parameters. The exact likelihood estimates are often preferred to the conditional likelihood estimates. If the sample size is large enough, both estimates are close to each other.

Remark 3.1. In `R`, the MA(q) model is treated in the form

$$x_t = \mu + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \cdots + \theta_q \varepsilon_{t-q},$$

where $\mu = E[x_t]$ is referred to as the `intercept`.

Example 3.2. Consider monthly returns of the CRSP equal-weighted index from January 1926 to December 2008.

→ MA(3) cf) AR(3) : c(3,0,0)

```
> (reg1 <- arima(ewrtn, order = c(0, 0, 3), method = "ML"))
Call:
arima(x = ewrtn, order = c(0, 0, 3), method = "ML")
Coefficients:
        ma1     ma2      ma3 intercept
     0.2114  0.0388  -0.0917    0.0122
s.e. 0.0319  0.0342   0.0328    0.0026
sigma^2 estimated as 0.005168: log likelihood = 1208.77, aic = -2407.54
```

ma parameter exact likelihood method.//

- In the exact likelihood method, the fitted model is

$$x_t = 0.0122 + \underbrace{\varepsilon_t}_{WN(0,\sigma_\varepsilon^2)} + 0.2114\varepsilon_{t-1} + 0.0388\varepsilon_{t-2} - 0.0917\varepsilon_{t-3}.$$

4. Model Checking

If the fitted model is adequate, the residual series $\{\hat{\varepsilon}_t\}$ should behave as a white noise. The Ljung-Box test is used to check the closeness of the residuals to a white noise.

Theorem 4.1. *Under the null hypothesis that ε_t is a white noise in an MA(q) model, it shows*

$$Q(m) \approx \chi^2_{(m-g)},$$

lag ma coeff

where g is the number of MA coefficients used in the model.

Example 4.2. Consider monthly returns of the CRSP equal-weighted index from January 1926 to December 2008.

```
> reg <- arima(ewrtn, order = c(0, 0, 9), method = "ML")
> tstat <- reg$coef/sqrt(diag(reg$var.coef))
> print(tstat, digits = 3)
  ma1   ma2    ma3    ma4   ma5    ma6   ma7    ma8   ma9 intercept
6.787 1.166 -3.664 -1.263 0.728 -0.951 1.323 -0.779 4.181     4.388
> (reg1 <- arima(ewrtn, order = c(0, 0, 9), fixed = c(NA, 0, NA, 0, 0, 0, 0, 0,
    NA, NA), method = "ML"))
Coefficients:
        ma1 ma2     ma3 ma4 ma5 ma6 ma7 ma8    ma9 intercept
     0.1909   0 -0.1199   0   0   0   0   0 0.1227    0.0122
s.e. 0.0293   0  0.0338   0   0   0   0   0 0.0312    0.0027
sigma^2 estimated as 0.005097: log likelihood = 1215.61, aic = -2421.22
> e1 <- reg1$residuals
> mytest <- Box.test(e1, lag = 12, type = "Ljung")
> (pvalue <- 1 - pchisq(as.numeric(mytest$statistic), 9))
[1] 0.04007017
```

< 1.64 ⇒ insignificant.

m−g

12 − 3 (1, 3, 9 th)

< 0.05

reject H₀ : MA

- The fitted model is

$$x_t = 0.0122 + \varepsilon_t + 0.1909\varepsilon_{t-1} - 0.1199\varepsilon_{t-3} + 0.1227\varepsilon_{t-9}.$$

애매하기.

The p-value is 0.04, so that the fitted model is "only" marginally adequate.

unit root AR, MA apply x.

5. Forecasting

The s-step ahead forecast of an MA model is the conditional expectation of x_{t+s} given I_t; i.e.,

$$\hat{x}_t[s] = E[x_{t+s}|I_t]$$

for $s \geq 1$.

theorical prefere MA ↕ AR

but 1-1 correspond arg

- For an MA(1) model, it shows

$$\begin{aligned}
\hat{x}_t[1] &= E[x_{t+1}|I_t] \\
&= E[\mu + \varepsilon_{t+1} + \theta_1\varepsilon_t|I_t] \\
&= \mu + \theta_1\varepsilon_t && (5.1) \\
e_t[1] &= x_{t+1} - \hat{x}_t[1] \\
&= \varepsilon_{t+1} && (5.2) \\
\hat{x}_t[s] &= E[x_{t+s}|I_t] \\
&= E[\mu + \varepsilon_{t+s} + \theta_1\varepsilon_{t+s-1}|I_t] \\
&= \mu && (5.3) \\
e_t[s] &= x_{t+s} - \hat{x}_t[s] \\
&= \varepsilon_{t+s} + \theta_1\varepsilon_{t+s-1} && (5.4)
\end{aligned}$$

for $s > 1$. Based on (5.1) and (5.3), one sees that it takes one time period for its forecast to go to the mean. From (5.2) and (5.4), it shows

$$Var[e_t[1]] = \sigma_\varepsilon^2 < Var[e_t[s]] = (1+\theta_1^2)\sigma_\varepsilon^2.$$

- For an MA(2) model, it shows

$$\begin{aligned}
\hat{x}_t[1] &= E[\mu + \varepsilon_{t+1} + \theta_1\varepsilon_t + \theta_2\varepsilon_{t-1}|I_t] \\
&= \mu + \theta_1\varepsilon_t + \theta_2\varepsilon_{t-1} \\
\hat{x}_t[2] &= E[\mu + \varepsilon_{t+2} + \theta_1\varepsilon_{t+1} + \theta_2\varepsilon_t|I_t] \\
&= \mu + \theta_2\varepsilon_t \\
\hat{x}_t[s] &= \mu
\end{aligned}$$

for $s > 2$. Thus, it takes two time periods for its forecast to go to the mean. In general, for an MA(q) model, the forecasts go to the mean after the first q periods.

MA(1).

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

$$\begin{aligned}\hat{x}_t[1] &= E[x_{t+1} \mid I_t] \\ &= E[\mu + \varepsilon_{t+1} + \theta_1 \varepsilon_t \mid I_t] \\ &= \mu + \theta_1 \varepsilon_t\end{aligned}$$

$$\begin{aligned}e_t[1] &= x_{t+1} - \hat{x}_t[1] \\ &= \varepsilon_{t+1}\end{aligned}$$

$$Var[e_t[1]] = \sigma_\varepsilon^2$$

$$\begin{aligned}\hat{x}_t[2] &= E[x_{t+2} \mid I_t] \\ &= E[\mu + \varepsilon_{t+2} + \theta_1 \varepsilon_{t+1} \mid I_t] \\ &= \mu\end{aligned}$$

$$\begin{aligned}e_t[2] &= x_{t+2} - \hat{x}_t[2] \\ &= \varepsilon_{t+2} + \theta_1 \varepsilon_{t+1}\end{aligned}$$

$$\begin{aligned}Var[e_t[2]] &= Var[\varepsilon_{t+2} + \theta_1 \varepsilon_{t+1}] \\ &= \sigma_\varepsilon^2 + \theta_1^2 \sigma_\varepsilon^2 \\ &= \sigma_\varepsilon^2 (1 + \theta_1^2).\end{aligned}$$

$$\begin{aligned}\hat{x}_t[s] &= E[x_{t+s} \mid I_t] \\ &= E[\mu + \varepsilon_{t+s} + \theta_1 \varepsilon_{t+s-1} \mid I_t] \\ &= \mu\end{aligned}$$

$$e_t[s] = \varepsilon_{t+s} + \theta_1 \varepsilon_{t+s-1}$$

$$\begin{aligned}Var[e_t[s]] &= Var[\varepsilon_{t+s} + \theta_1 \varepsilon_{t+s-1}] \\ &= \sigma_\varepsilon^2 + \theta_1^2 \sigma_\varepsilon^2 \\ &= \sigma_\varepsilon^2 (1 + \theta_1^2).\end{aligned}$$

MA(2).

$$x_t = \mu + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2}$$

$$\hat{x}_t[1] = E[x_{t+1} \mid I_t]$$
$$= E[\mu + \varepsilon_{t+1} + \theta_1 \varepsilon_t + \theta_2 \varepsilon_{t-1} \mid I_t]$$
$$= \mu + \theta_1 \varepsilon_t + \theta_2 \varepsilon_{t-1}$$
$$e_t[1] = \varepsilon_{t+1}$$
$$Var[e_t[1]] = \sigma_\varepsilon^2$$

$$\hat{x}_t[2] = E[\mu + \varepsilon_{t+2} + \theta_1 \varepsilon_{t+1} + \theta_2 \varepsilon_t \mid I_t]$$
$$= \mu + \theta_2 \varepsilon_t$$
$$e_t[2] = \varepsilon_{t+2} + \theta_1 \varepsilon_{t+1}$$
$$Var[e_t[2]] = \sigma_\varepsilon^2 (1 + \theta_1^2)$$

$$\hat{x}_t[3] = E[\mu + \varepsilon_{t+3} + \theta_1 \varepsilon_{t+2} + \theta_2 \varepsilon_{t+1} \mid I_t]$$
$$= \mu$$
$$e_t[3] = \varepsilon_{t+3} + \theta_1 \varepsilon_{t+2} + \theta_2 \varepsilon_{t+1}$$
$$Var[e_t[3]] = \sigma_\varepsilon^2 (1 + \theta_1^2 + \theta_2^2)$$

Example 5.1. Consider monthly returns of the CRSP equal-weighted index from January 1926 to December 2008.

```
> ewrtn <- ts(ewrtn, start = c(1926, 1), freq = 12)
> ewrtn1 <- window(ewrtn, end = c(2008, 2))
> reg <- arima(ewrtn1, order = c(0, 0, 9), fixed = c(NA, 0, NA, 0, 0, 0, 0, 0,
     NA, NA), method = "ML")
> x.ahead <- predict(reg, 10)
> x.ahead$pred[10]
[1] 0.01279257
> mean(ewrtn1)
[1] 0.0127984
```

- As expected, the 10-step ahead forecast is equal to the sample mean.

```
> ewrtn2 <- window(ewrtn, start = c(2007, 2))
> plot(ewrtn2, lwd = 2, xlab = "", ylab = "", ylim = c(-0.3, 0.2))
> pred <- x.ahead$pred
> lines(ts(c(ewrtn2[13], pred), start = c(2008, 2), freq = 12), lty = 4,lwd = 2)
> upper <- x.ahead$pred + 1.96*x.ahead$se
> lines(ts(c(ewrtn2[13], upper), start = c(2008, 2), freq = 12), lty = 3)
> lower <- x.ahead$pred - 1.96*x.ahead$se
> lines(ts(c(ewrtn2[13], lower), start = c(2008, 2), freq = 12), lty = 3)
> legend("bottomleft", c("Observations", "Forecasts", "95% confidence interval"),
     lty = c(1, 4, 3), lwd = c(2, 2, 1), inset = 0.01)
```