9.2 The ARMA models

Contents

(All variables are real and one-dimensional unless otherwise specified.)

1. Overview

The two basic members of the ARMA family are the **AR model** and the **MA model**. Let ε_t be white noise.

- **The AR(*p*) model.** The pth-order **autoregressive** model is defined as

$$\begin{aligned} X_t &= \varphi_0 + \varphi_1 X_{t-1} + \cdots + \varphi_p X_{t-p} + \varepsilon_t \\ &= \varphi_0 + \sum_{i=1}^{p} \varphi_i X_{t-i} + \varepsilon_t \end{aligned}$$

  for some constants $\{\varphi\}$. If $\varphi_0 \neq 0$, X is not stationary until t is large enough.
- **The MA(*q*) model.** The qth-order **moving-average** model is defined as

$$\begin{aligned} X_t &= \theta_0 + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \cdots + \theta_q \varepsilon_{t-q} \\ &= \theta_0 + \varepsilon_t + \sum_{j=1}^{q} \theta_j \varepsilon_{t-j} \end{aligned}$$

  for some constants $\{\theta\}$, and θ_0 will become the mean of X.

We can create various sophisticated models by combining the two.

- **The ARMA *(p, q)* model*.*** It is simply the sum of AR(*p*) and MA(*q*).

$$X_t = \varphi_0 + \sum_{i=1}^{p} \varphi_i X_{t-i} + \varepsilon_t + \sum_{j=1}^{q} \theta_j \varepsilon_{t-j}$$

- **The ARIMA(*p*, *d*, *q*) model.** "I" stands for "integrated". If X is an ARIMA(*p*, *d*, *q*) process,

$$
\begin{aligned}
\underbrace{(1-\mathrm{L})^d}_{\sum_{k=0}^{d}\binom{d}{k}(-1)^k\,\mathrm{L}^k} \quad X_t &= (1-\mathrm{L})^{d-1}\,(X_t - X_{t-1}) \\
&= (1-\mathrm{L})^{d-2}\underbrace{(X_t - 2X_{t-1} + X_{t-2})}_{(X_t - X_{t-1})-(X_{t-1}-X_{t-2})} \\
&= \cdots \\
&= \sum_{k=0}^{d}\binom{d}{k}(-1)^k X_{t-k}
\end{aligned}
$$

  is an ARMA(*p*, *q*) process. It could be useful when there are complicated seasonal trends.
- **The VARMA(*p*, *q*) model.** "V" stands for "vectorial". It describes the mutual influence of n time series with vectors and matrices.

$$
\underbrace{\begin{bmatrix} X_t^{(1)} \\ \vdots \\ X_t^{(n)} \end{bmatrix}}_{\mathbf{X}_t} = \underbrace{\begin{bmatrix} \varphi_0^{(1)} \\ \vdots \\ \varphi_0^{(n)} \end{bmatrix}}_{\boldsymbol{\varphi}_0} + \sum_{i=1}^{p} \underbrace{\begin{bmatrix} \varphi_i^{(11)} & \cdots & \varphi_i^{(1n)} \\ \vdots & \ddots & \vdots \\ \varphi_i^{(n1)} & \cdots & \varphi_i^{(nn)} \end{bmatrix}}_{\boldsymbol{\Phi}_i} \underbrace{\begin{bmatrix} X_{t-i}^{(1)} \\ \vdots \\ X_{t-i}^{(n)} \end{bmatrix}}_{\mathbf{X}_{t-i}} + \underbrace{\begin{bmatrix} \varepsilon_t^{(1)} \\ \vdots \\ \varepsilon_t^{(n)} \end{bmatrix}}_{\boldsymbol{\varepsilon}_t} + \sum_{j=1}^{q} \begin{bmatrix} \theta_j^{(11)} & \cdots \\ \vdots & \\ \theta_j^{(n1)} & \cdots \end{bmatrix}
$$

$$
\Rightarrow \mathbf{X}_t = \boldsymbol{\varphi}_0 + \sum_{i=1}^{p} \boldsymbol{\Phi}_i \mathbf{X}_{t-i} + \boldsymbol{\varepsilon}_t + \sum_{j=1}^{q} \boldsymbol{\Theta}_j \boldsymbol{\varepsilon}_{t-j}
$$

2. Autoregressive model

The autoregressive model suggests that every data point in a time series depends on its previous values linearly. ("Auto" means "self".) It is also said to be an **infinite-impulse-response** system because its current value will affect its future values forever (despite an exponentially weak influence).

The pth-order autoregressive model, i.e. AR(*p*), is defined as

$$
X_t = \varphi_0 + \sum_{i=1}^{p} \varphi_i X_{t-i} + \varepsilon_t
$$

for some constants $\{\varphi\}$ and discrete time $t \geq 1$ as well as a fixed and given initial point X_0. The **white noise** ε is a set of random variables drawn independently from the same distribution with **zero mean** and **finite variance** σ_ε^2. (Remember the distribution may not be normal; if it happens to be normal, the noise may be called Gaussian white noise.)

Using Gauss's **ordinary least squares** (OLS) method, we can get $\boldsymbol{\varphi} = \begin{pmatrix}\varphi_0 & \varphi_1 & \dots & \varphi_p\end{pmatrix}^\top$ after transforming $X \leftarrow X - \bar{X}$ and then solving $\mathbf{y} = \mathbf{X}\boldsymbol{\varphi}$ with

$$\mathbf{X} = \begin{pmatrix} 1 & X_1 & X_2 & \cdots & X_p \\ 1 & X_2 & X_3 & \cdots & X_{p+1} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & X_{t-p} & X_{t-p+1} & \cdots & X_{t-1} \end{pmatrix}$$

and $\mathbf{y} = \begin{pmatrix}X_{p+1} & X_{p+2} & \dots & X_t\end{pmatrix}^\top$. The general solution is

$$\boldsymbol{\varphi} = \left(\mathbf{X}^\top \mathbf{X}\right)^{-1} \mathbf{X}^\top \mathbf{y}\,.$$

2.1 The AR(1) model

Despite its simplicity, the first-order autoregressive model

$$X_t = c + mX_{t-1} + \varepsilon_t$$

already shows nontrivial properties.

Mean. We can calculate its mean μ analytically.

$$\begin{aligned} E(X_t) &= E(c + mX_{t-1} + \varepsilon_t) \\ &= c + mE(X_{t-1}) \end{aligned}$$

Since μ does not change with time (by definition), we may rewrite the equation as

$$\begin{aligned} \mu &= c + m\mu \\ \Rightarrow \mu &= \frac{c}{1-m}\,. \end{aligned}$$

At $m = 1$, μ diverges, so the AR(1) model has a mean if and only if $m \neq 1$.

Variance. We can similarly calculate its variance σ^2. Let $\mathbf{var}$ and $\mathbf{cov}$ be the operators of variance and covariance.

$$\begin{aligned} \mathbf{var}(X_t) &= \mathbf{var}(c + mX_{t-1} + \varepsilon_t) \\ &= \mathbf{var}(mX_{t-1} + \varepsilon_t) \\ &= m^2\mathbf{var}(X_{t-1}) + \sigma_\varepsilon^2 + 2m\,\mathbf{cov}(X_{t-1}, \varepsilon_t) \end{aligned}$$

The third term involves the covariance between X and the noise.

$$
\begin{aligned}
\operatorname{cov}(X_{t-1}, \varepsilon_t) &= E(X_{t-1}\varepsilon_t) - E(X_{t-1})E(\varepsilon_t) \\
&= E(X_{t-1}\varepsilon_t) \\
&= E(c\varepsilon_t + mX_{t-2}\varepsilon_t + \varepsilon_t\varepsilon_{t-1}) \\
&= 0 + m\underbrace{E(X_{t-2}\varepsilon_t)}_{\operatorname{cov}(X_{t-2},\varepsilon_t)} + 0 \\
&= m^2 \operatorname{cov}(X_{t-3}, \varepsilon_t) \\
&= \cdots \\
&= m^{(t-2)} \operatorname{cov}(X_0, \varepsilon_t) \\
&= 0
\end{aligned}
$$

In the fourth line, the last term is $E(\varepsilon_t\varepsilon_{t-1}) = 0$ because the white noise at each time step is **independent**. As a result, the covariance on a day recursively depends on its value on the previous day, and the recursion stops at $\operatorname{cov}(X_0, \varepsilon_t) = 0$ because X_0 is just a constant.

Finally, because σ^2 does not change with time, we conclude that

$$
\begin{aligned}
\sigma^2 &= m^2\sigma^2 + \sigma_\varepsilon^2 + 0 \\
\Rightarrow \sigma^2 &= \frac{\sigma_\varepsilon^2}{1 - m^2} \,.
\end{aligned}
$$

In addition to the divergence at $|m| = 1$, σ^2 absurdly becomes negative if $|m| > 1$. Therefore, the AR(1) model has a variance if and only if $|m| < 1$.

General solution. Through iteration, we get

$$
\begin{aligned}
X_t &= c + m\left(c + mX_{t-2} + \varepsilon_{t-1}\right) + \varepsilon_t \\
&= c\left(1 + m\right) + m^2 X_{t-2} + \left(m\varepsilon_{t-1} + \varepsilon_t\right) \\
&= \cdots \\
&= c\sum_{a=0}^{t-1} m^a + m^t X_0 + \sum_{b=0}^{t-1} m^b \varepsilon_{t-b} \,.
\end{aligned}
$$

The first two terms represent a trend.

- If $|m| \geq 1$, the second term diverges and goes to $\pm\infty$ rapidly.
- If $|m| = 1$, X is a **random walk**, and the divergence of the first term makes it not stationary either.
- If $|m| < 1$, the first term becomes a constant and the second term vanishes after some time.

$$
\lim_{t\to\infty} X_t = \frac{c}{1 - m} + \sum_{b=0}^{\infty} m^b \varepsilon_{t-b}
$$

Autocorrelation. The model's autocorrelation $B(n)$, where n is the lag, can thus be derived.

$$
\begin{aligned}
B(n) &= E[(X_{t+n} - \mu)(X_t - \mu)] \\
&= E\left[\left(\sum_{a=0}^{\infty} m^a \varepsilon_{t-a+n}\right)\left(\sum_{b=0}^{\infty} m^b \varepsilon_{t-b}\right)\right]
\end{aligned}
$$

All terms but the ones with $a = b + n$ vanish in the product of the sums because of the independence of the white noise, which implies $E\left(\varepsilon_{t-a+n}\varepsilon_{t-b}\right) = \begin{cases} \sigma_\varepsilon^2 & (a = b + n) \\ 0 & (a \neq b + n) \end{cases}$.

$$
\therefore\ B(n) = \sum_{b=0}^{\infty} m^{(b+n)+b}\sigma_\varepsilon^2 = \sigma_\varepsilon^2 m^n \sum_{b=0}^{\infty} \left(m^2\right)^b = \frac{\sigma_\varepsilon^2}{1 - m^2} m^n
$$

As $B(n) \sim m^n$ for some $|m| < 1$, it shows an **exponential decay**.

Fitting. We can estimate m and c with a much simpler version of the OLS method.

$$
\hat{m} = \frac{\sum_{i=1}^{t-1}\left(X_i - \bar{X}\right)\left(X_{i+1} - \bar{X}\right)}{\sum_{i=1}^{t-1}\left(X_i - \bar{X}\right)^2} \quad \text{and} \quad \hat{c} = \bar{X} - \hat{m}\bar{X}
$$

Finally, you must check if the measured noise $\hat{\varepsilon}_t = X_t - \hat{m}X_{t-1} - \hat{c}$ violates the assumptions of white noise: namely, you need to use

- a *t*-test to check whether its mean $\hat{\mu}_\varepsilon \approx 0$ and
- a Ljung-Box test to check whether it has independent data points and thus an autocorrelation that vanishes like $\hat{B}_\varepsilon(n) \approx \begin{cases} \sigma_{\hat{\varepsilon}}^2 & (n = 0) \\ 0 & (n \neq 0) \end{cases}$.

Large deviance from the assumptions distorts the quality of inference.

3. Moving-average model

The *q*th-order moving-average model, i.e. MA(*q*), is defined as

$$
X_t = \theta_0 + \varepsilon_t + \sum_{j=1}^{q} \theta_j \varepsilon_{t-j}
$$

for some constants $\{\theta\}$ and, again, discrete time $t \geq 1$, while it simply starts from $X_0 = \theta_0$. Although its current value also affects its future values, the MA model is fundamentally different from an AR model because it is a **finite-impulse-response** system, meaning that X_t and $X_{t+\Delta t}$ are not directly related any more once $\Delta t > q$.

On one hand, it is easier to analyze the MA model than the AR model because it is not recursively defined. On the other hand, it is harder to fit the MA model than the AR model because we cannot estimate $\{\theta\}$ without simultaneously estimating the noise ε.

3.1 The MA(1) model

The first-order moving-average model is simply defined as

$$X_t = c + \varepsilon_t + m\varepsilon_{t-1} \,,$$

which is already the solution to the process. (Nothing in terms of X on the right-hand side.)

Mean. Its mean is

$$\mu = E(c + \varepsilon_t + m\varepsilon_{t-j}) = c + 0 + 0 = c \,.$$

Variance. Its variance is

$$\sigma^2 = \mathrm{var}(c + \varepsilon_t + m\varepsilon_{t-1}) = 0 + \sigma_\varepsilon^2 + m^2\sigma_\varepsilon^2 = \left(1 + m^2\right)\sigma_\varepsilon^2 \,.$$

Autocorrelation. Its autocorrelation is

$$\begin{aligned} B(n) &= E[(X_{t+n} - \mu)\,(X_t - \mu)] \\ &= E[(\varepsilon_{t+n} + m\varepsilon_{t-1+n})\,(\varepsilon_t + m\varepsilon_{t-1})] \\ &= E(\varepsilon_{t+n}\varepsilon_t) + mE(\varepsilon_{t+n}\varepsilon_{t-1}) + mE(\varepsilon_{t-1+n}\varepsilon_t) + m^2E(\varepsilon_{t-1+n}\varepsilon_{t-1}) \,. \end{aligned}$$

The four terms cannot survive at the same time because of the independence of the noise.

- At $n = 0$, only the first and the last terms survive, trivially giving $B(n) = \left(1 + m^2\right)\sigma_\epsilon^2 \equiv \mathrm{var}(X_t)$.
- At $n = 1$, only the third term survives, giving $B(n) = m\sigma_\varepsilon^2$.
- At $n \geq 2$, no terms can survive at all.

Therefore,

$$B(n) \sim \begin{cases} (1 + m)^2 & (n = 0) \\ m & (n = 1) \\ 0 & (n \geq 2) \end{cases}$$

and **vanishes** once $n > q = 1$.

Unlike the AR(1) model, we can see that the mean, variance, and autocorrelation of the MA(1) model do not impose any requirements on its parameters, so it is **always stationary**.

Fitting. While $\hat{c} = \bar{X}$, we have to estimate m via an iterative approach after transforming $X \leftarrow X - \hat{c}$.

1. Let $\hat{\varepsilon}^{(i)}$ be the ith guess of the noise. Its initial point is fixed to be zero in all guesses.
2. Let $\hat{m}^{(0)} = 0$ be the initial guess of m.
3. For $j \geq 1$,
    1. compute the evolution of $\hat{\varepsilon}^{(j-1)}$ using $\hat{\varepsilon}_t^{(j-1)} = X_t - \hat{m}^{(j-1)}\hat{\varepsilon}_{t-1}^{(j-1)}$, then
    2. obtain a new guess $\hat{m}^{(j)}$ by fitting $X_t = \hat{\varepsilon}_t^{(j-1)} + \hat{m}^{(j)}\hat{\varepsilon}_{t-1}^{(j-1)}$.
4. Repeat until $\hat{m}$ converges.
5. Check whether the final noise $\hat{\varepsilon}$ is indeed white noise.

4. Box-Jenkins method

Suppose you want to fit an ARMA(*p*, *q*) model to some data. How can you determine its orders, though?

You can guess them primitively using the **Box-Jenkins method**.

If the data follows...	its **autocorrelation**...	its **partial autocorrelation**...
ARMA(*p*, 0), i.e. AR(*p*),	decays exponentially.	vanishes once the lag $n > p$.
ARMA(0, *q*), i.e. MA(*q*),	vanishes once the lag $n > q$.	decays exponentially.
ARMA(*p*, *q*),	starts to decay after some lags.	

Partial autocorrelation. Let $\rho(n) = \dfrac{B(n)}{B(0)}$ be the normalized autocorrelation of X. Its partial autocorrelation with lag n is defined as

$$\phi_{n,n} = \frac{\rho(n) - \sum_{k=1}^{n} \phi_{n-1,k}\rho(n-k)}{1 - \sum_{k=1}^{n} \phi_{n-1,k}\rho(k)}$$

for $\phi_{n,k} = \phi_{n-1,k} - \phi_{n,n}\phi_{n-1,n-k}$ and $\phi_{1,1} \equiv \rho(n)$. It measures the correlation between X_t and X_{t+n} after removing the influence from $\{X_{t+1}, X_{t+2}, \ldots, X_{t+n-1}\}$.

After all, the Box-Jenkins method is a primitive way for you to determine the orders. If you want more accurate results, you have to spend (a lot of) time sweeping through various combinations of (*p*, *q*) and computing the **Akaike information criterion** (AIC) of the resultant ARMA models—the one that yields the most negative AIC is considered to be the best.