1 Moving average models

Definition: $(X_t)_{t\in\mathbb{Z}}$ is a moving average of order q (MA(q)) if

$$X_t = \sum_{k=1}^{q} \theta_k \varepsilon_{t-k} + \varepsilon_t,\ t \in \mathbb{Z},$$
$$(\varepsilon_t)_{t\in\mathbb{Z}} \text{ a sequence of i.i.d. variables, } \mathbb{E}[\varepsilon_t] = 0.$$

Because an MA(q) is of the form $X_t = fct.(\varepsilon_t, \varepsilon_{t-1}, \ldots, \varepsilon_{t-q})$, the process is always stationary and causal.

We can represent an MA(q) with the backshift operator as follows.

$$X_t = (\Theta(B)\varepsilon)_t,\ t \in \mathbb{Z},$$
$$\Theta(z) = 1 + \sum_{k=1}^{q} \theta_k z^k \ (z \in \mathbb{C}).$$

Analogously to AR(p) models, we can invert $\Theta(\cdot)$ if its roots are outside the unit circle in the plane of complex numbers. That is, we have the following result.

Theorem 1. *Consider an MA(q) process and assume that* $\Theta(z) \neq 0$ *for* $|z| \leq 1$ *and* $\mathbb{E}|\varepsilon_t| < \infty$*. Then,*

$$\varepsilon_t = \sum_{j=0}^{\infty} \gamma_j X_{t-j},\ \gamma_0 = 1,\ t \in \mathbb{Z},$$

$$\Gamma(z) = \Theta^{-1}(z) = 1/\Theta(z) = \sum_{j=0}^{\infty} \gamma_j z^j,\ \gamma_0 = 1.$$

Sketch of a proof: Analogously to the sufficient conditions for stationarity and causality of an AR(p), we can invert

$$(\Theta^{-1}(B)X)_t = \varepsilon_t,\ t \in \mathbb{Z}.$$

Thereby, we use that $\Theta(z) \neq 0$ for $|z| \leq 1$. □

Implication: we can model an infinite conditional dependence with 1 or a few parameters. For example, in an AR(p) model, we have that

$$\mathbb{E}[X_t | X_{t-1}, X_{t-2}, \ldots] = \mathbb{E}[X_t | X_{t-1}, \ldots, X_{t-p}] = \sum_{j=1}^{p} \phi_j X_{t-j}.$$

But with an MA(q) model,

$$\mathbb{E}[X_t | X_{t-1}, X_{t-2}, \ldots]$$

depends on the infinite past. As a concrete example, consider an MA(1) model

$$X_t = \theta \varepsilon_{t-1} + \varepsilon_t.$$

Then,

$$\Theta(z) = 1 + \theta z, \quad \Gamma(z) = 1/\Theta(z) = 1 + \sum_{j=1}^{\infty} (-\theta)^j z^j.$$

For $|\theta| < 1$, $\Gamma(z)$ is well-defined for $|z| \leq 1$ and thus, for $|\theta| < 1$ and $\mathbb{E}|\varepsilon_t| < \infty$, we can represent

$$X_t = \sum_{j=1}^{\infty} (-\theta)^j X_{t-j} + \varepsilon_t, \ t \in \mathbb{Z},$$

which is an AR(∞) process, i.e. a non-Markovian process whose conditional distribution depends on an infinite past.

We say that an MA(q) is invertible if it can be represented as an AR(∞) model.

2 Moving average autoregressive models

A combination of AR(p) and MA(q) provides a flexible modeling framework.

Definition: $(X_t)_{t\in\mathbb{Z}}$ is a moving average autoregressive of orders p and q (ARMA(p, q)) if

$$X_t = \sum_{j=1}^{p} \phi_j X_{t-j} + \sum_{k=1}^{q} \theta_k \varepsilon_{t-k} + \varepsilon_t, \ t \in \mathbb{Z}.$$

With the backshift operator, the model can be represented as

$$\begin{aligned} &(\Phi(B)X)_t = (\Theta(B)\varepsilon)_t, \ t \in \mathbb{Z}, \\ &\Phi(z) = 1 - \sum_{j=1}^{p} \phi_j z^j, \quad \Theta(z) = 1 + \sum_{k=1}^{q} \theta_k z^k, \quad z \in \mathbb{C}. \end{aligned}$$

The model can be over-parameterized if we do not restrict $\Phi(\cdot)$ and $\Theta(\cdot)$. For example, consider the (seemingly) ARMA(1, 1) equation

$$\begin{aligned} &X_t = 0.8X_{t-1} - 0.8\varepsilon_{t-1} + \varepsilon_t, \\ &\text{i.e. } (\Phi(B)X)_t = (\Theta(B)\varepsilon)_t, \quad \Phi(z) = \Theta(z) = 1 - 0.8z. \end{aligned}$$

We note that the i.i.d. sequence $(\varepsilon)_{t\in\mathbb{Z}}$ satisfies the equation above (just use $X_t = \varepsilon_t$) and hence, the equation above is satisfied by an i.i.d. sequence (which we usually do not represent as an ARMA(1,1) process). The problem occurs because $\Phi(\cdot)$ and $\Theta(\cdot)$ have common roots (i.e. $z_0 = 1/0.8$) and hence, we can factor out some terms on both sides of $(\Phi(B)X)_t = (\Theta(B)\varepsilon)_t$. The problem disappears and ARMA(p, q) models become identifiable if we assume that the set of roots of $\Phi(\cdot)$ and the set of roots of $\Theta(\cdot)$ have no common element, i.e. the polynomials $\Phi(\cdot)$ and $\Theta\cdot)$ have no common factors.

Using the analogous arguments as before, we can invert $\Phi(\cdot)$ and/or $\Theta(\cdot)$ if the corresponding roots are outside the unit circle. We then obtain the following result.

Theorem 2. *Consider an ARMA(p, q) with $\Phi(z) \neq 0$ $(|z| \leq 1)$, $\Theta(z) \neq 0$ $(|z| \leq 1)$ and assume that the roots of $\Phi(\cdot)$ and $\Theta(\cdot)$ are distinct. Then, the MA(∞) representation*

$$X_t = \sum_{j=1}^{\infty} \psi_j \varepsilon_{t-j} + \varepsilon_t, \ t \in \mathbb{Z},$$
$$\Psi(z) = \frac{\Theta(z)}{\Phi(z)} = \sum_{j=0}^{\infty} \psi_j z^j, \ \psi_0 = 1 \ (|z| \leq 1),$$

holds, and the AR(∞) representation

$$\varepsilon_t = \sum_{j=0}^{\infty} \gamma_j X_{t-j}, \ \gamma_0 = 1, \ t \in \mathbb{Z},$$
$$i.e. \ X_t = \sum_{j=1}^{\infty} -\gamma_j X_{t-j} + \varepsilon_t, \ t \in \mathbb{Z},$$
$$\Gamma(z) = \frac{\Phi(z)}{\Theta(z)} = \sum_{j=0}^{\infty} \gamma_j z^j \ (|z| \leq 1)$$

holds as well.

Note that the condition $\Phi(z) \neq \ (|z| \leq 1)$ implies stationarity and causality of the ARMA(p, q) process since we can represent it as $X_t = \sum_{j=1}^{\infty} \psi_j \varepsilon_{t-j} + \varepsilon_t$ which is a function of infinitely many $\varepsilon_t, \varepsilon_{t-1}, \dots$

3 Autocorrelation function and Partial autocorrelation function

The autocorrelation function (ACF) of a weakly stationary process is defined as

$$\rho(k) = \frac{R(k)}{R(0)}.$$

Definition: The partial autocorrelation function (PACF) of a weakly stationary process is defined as

$$\alpha(k) = \text{Parcorr}(X_0, X_k | X_1, \dots, X_{k-1}), \quad k \in \mathbb{N} \ (k \geq 1).$$

The partial autocorrelation is defined as

$$\alpha(k) \frac{\text{Cov}(X_0 - \hat{X}_{0|1,\dots,k-1}, X_k - \hat{X}_{k|1,\dots,k-1})}{\sqrt{\text{Var}(X_0 - \hat{X}_{0|1,\dots,k-1}) \text{Var}(X_k - \hat{X}_{k|1,\dots,k-1})}}, \tag{1}$$

where $\hat{X}_{t|1,\dots,k-1}$ is the best linear prediction of X_t based on $X_1, \dots, X_{k-1}$. Note that all the quantities involved in (1) involve only first and second moments and hence, weak stationarity is sufficient to define $\alpha(\cdot)$ as a function of the lag k only.

3.1 Qualitative behavior of ACF and PACF

For an MA(q) model, it is easy to see that $\rho(k) = 0$ for all $k \geq q+1$ (because X_t ad X_{t+k} are independent for $k \geq q+1$).

For an AR(p) model, we note that for $k \geq p+1$

$$\hat{X}_{k|1,\dots k-1} = \sum_{j=1}^{p} \phi_j X_{k-j},$$

and hence

$$X_k - \hat{X}_{k|1,\dots k-1} = \varepsilon_k.$$

Therefore, the numerator in (1) equals zero if $k \geq p+1$ and if the AR(p) model is causal (since then ε_k is independent from $\{X_t; t \leq k-1\}$). In summary, for a stationary and causal AR(p),

$$\alpha(k) = 0 \text{ for } k \geq p+1.$$

The following "duality" scheme holds in general.

model	ACF	PACF
AR(p)	$\rho(k)$ decays exp. fast as $k \to \infty$	$\alpha(k) = 0$ for $k \geq p+1$
MA(q)	$\rho(k) = 0$ for $k \geq q+1$	$\alpha(k)$ decays exp. fast as $k \to \infty$
ARMA(p, q)	$\rho(k)$ decays exp. fast as $k \to \infty$	$\alpha(k)$ decays exp. fast as $k \to \infty$