Lectures 11-12

MA and ARMA Models; Properties and Technical Details

Textbook Sections: 2.1-2.3, 3.1-3.3, 5.5

Definitions and Notation

A random sequence $\{X_t\}$ is a **linear process** if it has the expression

$$X_t = \sum_{j=-\infty}^{\infty} \psi_j \varepsilon_{t-j},$$

where $\{\varepsilon_t\}$ is $WN(0, \sigma^2)$, and the ψ_j are constants satisfying $\sum_{j=-\infty}^{\infty} |\psi_j| < \infty$.

The condition $\sum_{j=-\infty}^{\infty} |\psi_j| < \infty$ is the condition of **absolute summability**. In words, we say that the coefficients ψ_j are absolutely summable. This is a technical condition on the coefficients that ensures that the infinite sum will actually converge. Roughly speaking, this condition makes it so that after a certain point, the coefficients ψ_j decrease at a high enough rate as j increases. We won't get into the details in this class, but we will see this condition come up a few times when we deal with infinite sums.

Linear processes form a large and important class of random sequences. When fitting models to data that seems stationary, we will restrict our choice of model to a subclass of linear processes, called **autoregressive moving average, or ARMA** processes.

A random sequence $\{X_t\}$ is called $ARMA(p, q)$ if it has the expression

$$X_t - \mu = \phi_1(X_{t-1} - \mu) + \phi_2(X_{t-2} - \mu) + \cdots + \phi_p(X_{t-p} - \mu) + \theta_1\varepsilon_{t-1} + \theta_2\varepsilon_{t-2} + \cdots + \theta_q\varepsilon_{t-q} + \varepsilon_t,$$

where $\{\varepsilon_t\}$ is $WN(0, \sigma^2)$.

It is typical to condense the expression using sum notation as

$$X_t - \mu = \sum_{j=1}^{p} \phi_j(X_{t-j} - \mu) + \sum_{j=1}^{q} \theta_j\varepsilon_{t-j} + \varepsilon_t,$$

and to move the X_t terms to the left hand side as

$$(X_t - \mu) - \sum_{j=1}^{p} \phi_j(X_{t-j} - \mu) = \sum_{j=1}^{q} \theta_j\varepsilon_{t-j} + \varepsilon_t.$$

For notational simplicity, we can restrict our attention to zero-mean $ARMA(p, q)$ processes

$$X_t - \sum_{j=1}^{p} \phi_j X_{t-j} = \sum_{j=1}^{q} \theta_j \varepsilon_{t-j} + \varepsilon_t,$$

since we can always center the data before fitting a model, and fit a zero-mean model to the centered data.

The last expression (with X terms on one side, and ε terms on the other) leads to two polynomials.

- The AR polynomial is $\phi(z) = 1 - \phi_1 z - \phi_2 z^2 - \cdots - \phi_p z^p$.
- The MA polynomial is $\theta(z) = 1 + \theta_1 z + \theta_2 z^2 + \cdots + \theta_q z^q$.

It turns out that important properties of an $ARMA(p, q)$ process can be expressed in terms of the roots of the AR and MA polynomials. Below are descriptions of the properties we are concerned with, and their relation to the polynomials.

Causality

A random sequence $\{X_t\}$ is called **causal** (or possesses the property of **causality**) if it can be written as an $MA(\infty)$ sequence

$$X_t - \mu = \sum_{j=0}^{\infty} \psi_j \varepsilon_{t-j}, \text{ with } \psi_0 = 1,$$

where $\{\varepsilon_t\}$ is white noise with variance σ^2 and $\psi_1, \psi_2, ...$ are constants satisfying the constraint $\sum |\psi_j| < \infty$.

If a time series has a causal representation, then it is automatically stationary. Moreover, causal representations allow simple computation of the standard errors for prediction (needed for constructing prediction intervals). For these reasons, **most textbooks and computer packages assume that the ARMA models being used are causal.** Non-causal sequences are useless if the goal is to forecast. In this course we will consider only those time series models which are causal.

In general, explicit expressions for these ψ_j weights are difficult to obtain (unless it is an $MA(q)$ model). Fortunately, packages such as R will calculate these ψ_j coefficients for an $ARMA(p, q)$ model (R function `ARMAtoMA()`).

Example 1: If X_t has the following $MA(2)$ model

$$X_t - \mu = \varepsilon_t + 1.1\varepsilon_{t-1} - 0.3\varepsilon_{t-2},$$

then $\psi_0 = 1, \psi_1 = 1.1, \psi_2 = -0.3$, and $\psi_3, \psi_4, ...$ are all zeros.

Example 2: If X_t is $AR(1)$ with $-1 < \phi < 1$, then it can be re-written (details given in appendix) as

$$X_t - \mu = \sum_{j=0}^{\infty} \phi^j \varepsilon_{t-j}.$$

In such a case $\psi_j = \phi^j, j = 0, 1, 2, ...$ So an $AR(1)$ sequence with $-1 < \phi < 1$ is causal and hence stationary.

Example 3: Consider an $ARMA(2, 2)$ model with $\phi_1 = 0.8, \phi_2 = -0.15, \theta_1 = 0.6, \theta_2 = 0.08$. Then the R function `ARMAtoMA()` gives us the ψ_j values. Here are the first12 of them starting with $\psi_1 = 1.400$ (note that the values of ψ_j become small for large j):

$$[1.400, 1.050, 0.630, 0.347, 0.184, 0.094, 0.050, 0.024, 0.012, 0.006, 0.003, 0.002]$$

Here is how causality of an $ARMA(p, q)$ process is determined:

- All $MA(q)$ processes are causal.
- All stationary $AR(p)$ processes are causal. (Roots of the AR polynomial are larger than 1 in magnitude (i.e. are outside the unit circle)).
- An $ARMA(p, q)$ process is causal if the AR polynomial satisfies the condition of causality.

Invertibility

A random sequence $\{X_t\}$ is called **invertible** (or possesses the property of **invertibility**) if it can be written as an $AR(\infty)$ sequence

$$\sum_{j=0}^{\infty} \pi_j (X_{t-j} - \mu) = \varepsilon_t, with\ \pi_0 = 1,$$

where $\{\varepsilon_t\}$ is white noise with variance σ^2, and $\pi_1, \pi_2, ...$ are constants satisfying the constraint $\sum |\pi_j| < \infty$.

The invertible representation (above) allows for straightforward computation of forecasts. Invertibility is a desirable property when fitting a model, but is not a necessity. **Most textbooks and computer packages assume that the ARMA models being used are invertible.**

Example 4. An $MA(1)$ model with zero mean and $-1 < \theta < 1$ is invertible since it can be re-expressed as (details given in appendix)

$$X_t - \mu - \theta(X_{t-1} - \mu) + \theta^2(X_{t-2} - \mu) + \cdots = \varepsilon_t, i.e., \sum_{j=0}^{\infty} \pi_j (X_{t-j} - \mu) = \varepsilon_t,$$

where $\pi_j = (-\theta)^j$, $j = 0, 1, 2, \cdots$.

Example 5. Can we use the computer to find the π values? R **does not** have a function to convert an $ARMA$ model to an $AR(\infty)$ model. However, you can trick the computer in giving you the π values using the function `ARMAtoMA()`. Let us see how to do this for Example 3. If you pretend that it is an $ARMA$ series in $\{\varepsilon_t\}$ (it is really not) with $\{X_t\}$ as the white noise, then the AR coefficients are -0.60 and -0.08 and the MA coefficients are -0.8 and 0.15. With this trick, the π values (starting with $\pi_1 = -1.400$) computed by R are (note the rapid decay of π_j values as j increases)

$$[-1.400, 0.910, -0.434, 0.188, -0.078, 0.032, -0.013, 0.005, -0.002, 0.001, -0.0003, 0.0001]$$

Here is how invertibility of an $ARMA(p, q)$ process is determined:

- All $AR(p)$ processes are invertible.
- An $MA(q)$ process is invertible if the roots of the MA polynomial are larger than 1 in magnitude (i.e. are outside the unit circle).
- An $ARMA(p, q)$ process is invertible if the MA polynomial satisfies the condition of invertibility.

Identifiability

In general, moving average models are not unique. This means, that several different $MA(q)$ processes (with different sets or parameters $\theta_1, \cdots \theta_q, \sigma$) can have the same second-order structure.

As an example, consider an $MA(1)$ model $X_t - \mu = \varepsilon_t + \theta\varepsilon_{t-1}$, where $\{\varepsilon_t\}$ white noise with variance σ^2. Under the assumption of normality (i.e. $\{\varepsilon_t\}$ are normally distributed), any stationary series is completely characterized by the mean and autocovariances. Thus if if two sequences have the same mean and autocovariance functions, they are equally good descriptions of the data, i.e., they provide the same fit and they have the same predictive performances.

Consider the following two models

$$X_t - \mu = \varepsilon_t + \theta\varepsilon_{t-1} \quad (1)$$
$$X_t - \mu = \varepsilon'_t + (1/\theta)\varepsilon'_{t-1}, \quad (2)$$

where $\theta \neq 0$, $\{\varepsilon_t\}$ white noise with variance σ^2, and $\{\varepsilon'_t\}$ is white noise with variance $\theta^2\sigma^2$. Note that we only observe the data $\{X_t\}$, not ε_t or ε'_t. Both models have the same mean μ. All the autocovariances of lag 2 or higher are zero for both models.

For model 1,

$$\gamma(0) = (1+\theta^2)\sigma^2, \qquad \gamma(1) = \theta\sigma^2, \qquad 0 = \gamma(2) = \gamma(3) = \cdots.$$

For model 2,

$$\begin{aligned}\gamma(0) &= (1+(1/\theta)^2)(\theta^2\sigma^2) = (1+\theta^2)\sigma^2,\\ \gamma(1) &= (1/\theta)(\theta^2\sigma^2) = \theta\sigma^2,\\ 0 &= \gamma(2) = \gamma(3) = \cdots.\end{aligned}$$

So both the models have identical mean and autocovariance structures. Hence they will provide identical fits and predictions. This non-uniqueness is called lack of "identifiability". This can pose a big problem for numerical calculations. Computer packages will often fail to provide estimates of the parameters θ and σ^2 if there is more than one "correct" solution.

What do we do about it? To see this, let us examine more it closely. If the value of θ is larger than 1 in magnitude, and $Var(\varepsilon_t) = \sigma^2$, then we may as well consider model (2) for which the coefficient associated with ε_{t-1} is $1/\theta$, whose magnitude is less than 1, and $Var(\varepsilon'_t) = \theta^2\sigma^2$. So if $\theta = 2$, then we should work with model (2) in which the coefficient associated with ε'_{t-1} is $1/\theta = 1/2$ and $Var(\varepsilon'_t) = \theta^2\sigma^2 = 4\sigma^2$. Thus we can always choose a model where the coefficient associated with ε_{t-1} is no larger than 1 in magnitude, and this is what is done in practice.

How about general MA(q) models? The same issue on non-identifiability comes up. One can restrict attention to those models with appropriate conditions on the MA parameters $\theta_1, ..., \theta_q$.

A $MA(q)$ process is identifiable if the roots of the MA polynomial are equal to or larger than 1 in magnitude (i.e. lie on or outside the unit circle).

Non-Redundancy

This is an issue that arises only for $ARMA(p, q)$ models. An $ARMA(p, q)$ model is said to be **redundant** if it uses more parameters than necessary. Below is a simple example to illustrate this.

Suppose that the series $\{X_t\}$ can be described as $X_t = \varepsilon_t$, where $\{\varepsilon_t\}$ is white noise. So $\{X_t\}$ white noise. Subtract $0.5X_{t-1} = 0.5\varepsilon_{t-1}$ from this series to get

$$X_t - 0.5X_{t-1} = \varepsilon_t - 0.5\varepsilon_{t-1}, \quad \text{or} \quad X_t = 0.5X_{t-1} + \varepsilon_t - 0.5\varepsilon_{t-1}.$$

Now it seems that the series $\{X_t\}$ is $ARMA(1, 1)$, whereas in reality it is white noise.

Note that there is nothing special about using 0.5, and we can really use any $-1 < \phi < 1$, and rewrite X_t as

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

Moreover, we can complicate the model even further by adding more and more redundant terms on each side. There is an infinite number of redundant models.

In general if $\theta = -\phi$ in an $ARMA(1, 1)$ model $X_t = \phi X_{t-1} + \varepsilon_t + \theta\varepsilon_{t-1}$, then there is redundancy, since $\{X_t\}$ can be described as white noise.

Naturally, we don't want to fit a redundant model, so it is important to place constraints on the parameters in order to avoid redundancy. An $ARMA(p, q)$ model is not redundant if the AR and MA polynomials have no roots in common.

Appendix - Details for Examples 2 and 4

Example 2. Let us consider the case when $\mu = 0$. The case when $\mu \neq 0$ is similar. Note that

$$\begin{aligned} X_t &= \varepsilon_t + \phi X_{t-1} = \varepsilon_t + \phi(\varepsilon_{t-1} + \phi X_{t-2}) \\ &= \varepsilon_t + \phi\varepsilon_{t-1} + \phi^2 X_{t-2} \\ &= \varepsilon_t + \phi\varepsilon_{t-1} + \phi^2(\varepsilon_{t-2} + \phi X_{t-3}) \\ &= \varepsilon_t + \phi\varepsilon_{t-1} + \phi^2\varepsilon_{t-2} + \phi^3 X_{t-3} \end{aligned}$$

We can repeat this argument to find that, for any positive integer j, the following is true

$$X_t = \varepsilon_t + \phi\varepsilon_{t-1} + \cdots + \phi^{j-1}\varepsilon_{t-j+1} + \phi^j X_{t-j}.$$

Since ϕ^j goes to zero as j becomes large, we can see that

$$X_t = \varepsilon_t + \phi\varepsilon_{t-1} + \cdots = \sum_{j=0}^{\infty} \phi^j \varepsilon_{t-j}.$$

Example 4. We will look at the case when $\mu = 0$. The case when $\mu \neq 0$ is similar. We can write the $MA(1)$ model as

$$\begin{aligned} \varepsilon_t &= X_t - \theta\varepsilon_{t-1} = X_t - \theta(X_{t-1} - \theta\varepsilon_{t-2}) = X_t - \theta X_{t-1} + \theta^2\varepsilon_{t-2} \\ &= X_t - \theta X_{t-1} + \theta^2(X_{t-2} - \theta\varepsilon_{t-3}) \\ &= X_t - \theta X_{t-1} + \theta^2 X_{t-2} - \theta^3\varepsilon_{t-3} = X_t + (-\theta)X_{t-1} + (-\theta)^2 X_{t-2} + (-\theta)^3\varepsilon_{t-3}. \end{aligned}$$

We can repeat this argument to find that, for any positive integer j

$$\varepsilon_t = X_t + (-\theta)X_{t-1} + \cdots(-\theta)^{j-1}X_{t-j+1} + (-\theta)^j\varepsilon_{t-j}.$$

Since $(-\theta)^j$ converges to zero as j becomes large, we can find that

$$\varepsilon_t = X_t + (-\theta)X_{t-1} + \cdots = \sum_{j=0}^{\infty}(-\theta)^j X_{t-j}.$$