Handout 11

Chapters 2.2, 2.3, 3.1 and 3.3 in Brockwell and Davis.

Important Technical issues:

This handout is devoted to addressing important technical issues such as: stationarity, identifiability, non-redundancy of time series models etc. There are many (and not all) stationary series which have two types of alternate representations - causal and invertible. Apart from theoretical considerations, these representations are also useful in practice. Invertible representations allow for straightforward forecasting formulas. If a time series has a causal representation, then it is auotmatically stationary. Moreover, causal representations allow simple computation of the standard errors for prediction (needed for constructing prediction intervals). For these reasons,

most of the textbooks and computer packages assume that the ARMA models being used are both causal and inneverible.

Causality.

A random sequence $\{X_t\}$ is called causal if it can be written as an $MA(\infty)$ sequence

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

where $\{\varepsilon_t\}$ is white noise with variance σ^2 and $\psi_1, \psi_2, ...$ are constants satisfying the constraint $\sum |\psi_j| < \infty$. Heuristically, causality means that X_t should not depend on the future values, because we would like to predict the future knowing the past. 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. Here are some theoretical results.

Fact 1: A causal sequence as given in (1) is stationary.
Fact 2: Any $MA(q)$ sequence is stationary (and causal).

In general, explicit expressions for these ψ weights are difficult to obtain (unless it is an $MA(q)$ model). Fortunately, packages such as R will calculate these ψ 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 has an $AR(1)$ model with $-1 < \phi < 1$, then it can be re-written as (details given on page 4)

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

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. We will only write down only 12 of them starting with $\psi_1 = 1.400$ (note that the values of ψ_j become small for large j)

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

Condition for stationarity of $AR(p)$ models

If $\{X_t\}$ follows an $AR(p)$ model, then it is stationary (and causal) if it can be written in the form (1). Conditions on the autoregressive coefficients $\phi_1, ..., \phi_p$ are needed to guarantee that the sequence can be written in the form (1). Let $\phi(z) = 1 - \phi_1 z - \cdots - \phi_p z^p$ be a polynomial in z. This polynomial has p roots which can be real or complex valued.

Condition for stationarity (and causality): Absolute values of all the roots of the polynomial $\phi(z)$ are larger than 1.

When $p = 1$, the root of the polynomial is $1/\phi_1$. The condition that the absolute value of $1/\phi_1$ is larger than 1 is equivalent to the condition $-1 < \phi_1 < 1$. In this case, the autocorrelation function is $\rho(j) = \phi_1^j$, $j = 0, 1, 2....$ Note that $\rho(j)$ declines to zero exponentially as j increases.

For $p = 2$, the condition on ϕ_1 and ϕ_2 for stationarity is a bit more complicated. The condition is: $-1 < \phi_2 < 1$ and $-1 < \phi_1/(1 - \phi_2) < 1$. This is equivalent to the pictorial condition that ϕ_1 and ϕ_2 must lie inside the triangle given in the attached graph which is borrowed from the book on time series by Box, Jenkins and Reinsel (in the graph, ρ_k and ϕ_{kk} are the autocorrelation and partial autocorrelation of lag k, respectively). When ϕ_1 and ϕ_2 are in regions 1 and 2, the roots of the polynomial $\phi(z)$ are real valued and the autocorrelations decay as mixture of exponentials. Whereas, if they are in regions 3 and 4, the roots are complex valued and the autocorrelations decay as damped exponentials.

In general, for $AR(p)$ the autocorrelations decay as mixture of exponentials or as damped exponentials. The ACF plot of recruitment data (Handout 9) display a damped exponentials decay. Such damped exponential decay means that series may display (random) cyclical behavior as is evident in the recruitment series.

Identifiabilty of $MA(q)$ models:

Consider the polynomial $\theta(z) = 1 + \theta_1 z + \cdots + \theta_q z^q$. This polynomial has q roots which can be real or complex valued.

Condition for identifiability: Absolute values of all the roots of the polynomial $\theta(z)$ are bigger than or equal to 1.

For $MA(1)$ model, this translates into the condition that $-1 \leq \theta_1 \leq 1$. For $MA(2)$ this condition is equivalent to the condition that $-\theta_1$ and $-\theta_2$ lie in the triangle as in the $AR(2)$ case, except that $-\theta_1$ and $-\theta_2$ are allowed to be on the boundaries of the triangle.

Invertibility of $ARMA(p, q)$ models.

A sequence $\{X_t\}$ is invertible if it can be written as an $AR(\infty)$ model, i.e.,

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

where $\{\varepsilon_t\}$ is white noise. Invertibility is desirable but not a necessity. Here is an obvious result.

Fact 3. Any $AR(p)$ model is invertible.

Condition for invertibility:

Absolute values of all the roots of the polynomial $\theta(z)$ are larger than 1 (in an $ARMA$ or an $MA(q)$ model).

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

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

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

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)

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

$ARMA(p, q)$ models.

a) The condition for stationarity (and causality) for $ARMA(p, q)$ model is that the AR coefficients $\phi_1, ..., \phi_p$ satisfy the condition for stationarity, i.e., the absolute values of the roots of $\phi(z)$ are all larger than 1. In such a case, the R function $ARMAtoMA$ can convert it as a model in (1) by producing the ψ weights. For instance

b) Condition for identifiability is the that the MA coefficients $\theta_1, ..., \theta_q$ satisfy the conditions for identifiability.

c) Condition for non-redundancy is that the all the roots of $\phi(z)$ are distinct from the roots of $\theta(z)$.

Usefulness of invertible representation

Usefulness of invertibilty is seen for MA or ARMA models.

If we know a sequence $\{X_t\}$ and we know the weights π_j, then it is easy to do the forecasting from the data $\{X_1, ..., X_n\}$. For notational simplicity let us assume that $\mu = 0$. Then we can write this sequence as

$$X_t = \varepsilon_t - \pi_1 X_{t-1} - \pi_2 X_{t-2} - \cdots.$$

So if we have the data $\{X_1, ..., X_n\}$, then the forecasted value of of X_{n+1} is

$$\hat{X}_{n+1} = -\pi_1 X_n - \pi_2 X_{n-1} - \pi_3 X_{n-2} - \cdots.$$

Our forecast of X_{n+2} is

$$-\pi_1 X_{n+1} - \pi_2 X_n - \pi_3 X_{n-1} - \cdots.$$

Since X_{n+1} is unknown, then we can substitute it by $\hat{X}_{n+1}$ thus leading to the forecasted value of X_{n+2} as

$$\hat{X}_{n+2} = -\pi_1 \hat{X}_{n+1} - \pi_2 X_n - \pi_3 X_{n-1} - \cdots.$$

This method can now be replicated to forecast X_{n+3}, X_{n+4} etc.

Typically μ is not equal to zero, but obtaining forecasts are not difficult with the known π values. For instance, the forecasting formula for X_{n+1} is given by

$$\hat{X}_{n+1} - \mu = -\pi_1(X_n - \mu) - \pi_2(X_{n-1} - \mu) - \pi_3(X_{n-2} - \mu) - \cdots.$$

Usefulness of causal representation

We have talked about forecasting a series, but have not addressed the issue of prediction limits (or prediction intervals). If the observed series is $\{X_1, ..., X_n\}$, then h step ahead forecast is denoted by $\hat{X}_{n+h}$. The forecast error is $X_{n+h} - \hat{X}_{n+h}$ which is not known since X_{n+h} is unknown. For all the cases we consider in this course, the mean of the forecast error is equal to zero (or close to zero when the parameters of the model being used for forecasting are estimated). The variance of the forecast error is denoted by

$$\sigma^2(h) = Var(X_{n+h} - \hat{X}_{n+h}) = E(X_{n+h} - \hat{X}_{n+h})^2.$$

If an estimate $\sigma^2(1)$ of the variance of the error for forecasting $\hat{X}_{n+1}$ is available, we can give a 95% prediction interval for X_{n+1} as $\hat{X}_{n+1} \pm 1.96\sigma(1)$. We can do the same for X_{n+2} as $\hat{X}_{n+2} \pm 1.96\sigma(2)$, where $\sigma^2(2)$ is the variance of the error for forcasting X_{n+2}.

However getting the values $\sigma^2(1)$, $\sigma^2(2)$ etc. are not easy except in the $MA(q)$ case. This is where the causal representation with ψ are useful. We want to predict X_{n+1}, X_{n+2},... along with the prediction intervals from the observed series $\{X_1, ..., X_n\}$. Suppose that we want to predict X_{n+1} given the data $\{X_1, ..., X_n\}$. Since

$$X_{n+1} - \mu = \varepsilon_{n+1} + \psi_1\varepsilon_n + \psi_2\varepsilon_{n-1} + \psi_3\varepsilon_{n-2} + \cdots.$$

Then the best predictor of X_{n+1} is

$$\hat{X}_{n+1} - \mu = \psi_1\varepsilon_n + \psi_2\varepsilon_{n-1} + \psi_3\varepsilon_{n-2} + \cdots. \tag{2}$$

Hence the variance of the forecast error is

$$\sigma^2(1) = E(X_{n+1} - \hat{X}_{n+1})^2 = E\varepsilon_{n+1}^2 = \sigma^2.$$

Incidentally, the π representation will yield

$$\hat{X}_{n+1} - \mu = -\pi_1(X_n - \mu) - \pi_2(X_{n-1} - \mu) - \cdots, \tag{3}$$

and these two expressions (2) and (3) yield the same value of the forecast.

Note that X_{n+2} has the representation

$$X_{n+2} - \mu = \varepsilon_{n+2} + \psi_1\varepsilon_{n+1} + \psi_2\varepsilon_n + \psi_3\varepsilon_{n-1} + \cdots,$$

and hence the best linear predictor of X_{n+2} based on the past up to time n is

$$\hat{X}_{n+2} - \mu = \psi_2 \varepsilon_n + \psi_3 \varepsilon_{n-1} + \cdots, and$$
$$X_{n+2} - \hat{X}_{n+2} = \varepsilon_{n+2} + \psi_1 \varepsilon_{n+1}.$$

So the variance of the forecast error $X_{n+2} - \hat{X}_{n+2}$ is

$$\sigma^2(2) = E(X_{n+2} - \hat{X}_{n+2})^2 = E(\varepsilon_{n+2} + \psi_1 \varepsilon_{n+1})^2 = \sigma^2 + \psi_1^2 \sigma^2 = (1 + \psi_1^2)\sigma^2.$$

A similar argument will show that the variances for prediction error for X_{n+3}, X_{n+4} are given by

$$\sigma^2(3) = E(X_{n+3} - \hat{X}_{n+3})^2 = (1 + \psi_1^2 + \psi_2^2)\sigma^2,$$
$$\sigma^2(4) = E(X_{n+4} - \hat{X}_{n+4})^2 = (1 + \psi_1^2 + \psi_2^2 + \psi_3^2)\sigma^2.$$

Example 6. Consider the $ARMA(2,2)$ model given in Example 3. Assume that $Var(\varepsilon_t) = \sigma^2 = 3$. We have

$$\psi_1 = 1.400, \psi_2 = 1.050, \psi_3 = 0.630, \psi_4 = 0.347.$$

So we can obtain the variances of the prediction errors

$$\begin{aligned}
\sigma^2(1) &= \sigma^2 = 3, \\
\sigma^2(2) &= (1 + \psi_1^2)\sigma^2 = (1 + (1.400)^2)(3) = 8.880, \\
\sigma^2(3) &= (1 + \psi_1^2 + \psi_2^2)\sigma^2 = (1 + (1.400)^2 + (1.050)^2)(3) = 12.188, \\
\sigma^2(4) &= (1 + \psi_1^2 + \psi_2^2 + \psi_3^2)\sigma^2 = (1 + (1.400)^2 + (1.050)^2 + (0.630)^2)(3) = 13.378.
\end{aligned}$$

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 argment 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

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

Appendix

Fact 1. We will need to show that

i) $E(X_t)$ is the same for all t, ii) given a positive integer h, $Cov(X_t, X_{t+h})$ is the same for all t.

Since ε_t has zero mean, we have $E(X_t - \mu) = 0$ and hence $E(X_t) = \mu$, for all t.

We can write X_t as

$$X_t = \sum_{j=0}^{\infty}\psi_j\varepsilon_{t-j} = \sum_{j=0}^{h-1}\psi_j\varepsilon_{t-j} + \sum_{j=h}^{\infty}\psi_j\varepsilon_{t-j} = \sum_{j=0}^{h-1}\psi_j\varepsilon_{t-j} + \sum_{j=0}^{\infty}\psi_{j+h}\varepsilon_{t+h-j}.$$

Note that $\sum_{j=0}^{h-1}\psi_j\varepsilon_{t-j}$ is independent of $\sum_{j=0}^{\infty}\psi_j\varepsilon_{t+h-j}$. Hence

$$\begin{aligned}
Cov(X_t, X_{t+h}) &= Cov(\sum_{j=0}^{\infty}\psi_j\varepsilon_{t-j}, \sum_{j=0}^{\infty}\psi_j\varepsilon_{t+h-j}) = Cov(\sum_{j=0}^{h-1}\psi_j\varepsilon_{t-j} + \sum_{j=0}^{\infty}\psi_{j+h}\varepsilon_{t+h-j}, \sum_{j=0}^{\infty}\psi_j\varepsilon_{t+h-j}) \\
&= Cov(\sum_{j=0}^{h-1}\psi_j\varepsilon_{t-j}, \sum_{j=0}^{\infty}\psi_j\varepsilon_{t+h-j}) + Cov(\sum_{j=0}^{\infty}\psi_{j+h}\varepsilon_{t+h-j}, \sum_{j=0}^{\infty}\psi_j\varepsilon_{t+h-j}) \\
&= 0 + \sum_{j=0}^{\infty}\psi_{j+h}\psi_j\sigma^2 = \sum_{j=0}^{\infty}\psi_{j+h}\psi_j\sigma^2, \text{ for all } t.
\end{aligned}$$

Since (i) and (ii) hold, the sequence $\{X_t\}$ is stationary.

Fact 2. If $\{X_t\}$ is MA(q), then

$$X_t - \mu = \varepsilon_t + \theta_1\varepsilon_{t-1} + \cdots + \theta_q\varepsilon_{t-q} = \sum_{j=0}^{\infty}\psi_j\varepsilon_{t-j},$$

with $\psi_0 = 1, \psi_1 = \theta_1, ..., \psi_q = \theta_q\,, 0 = \psi_{q+1} = \psi_{q+2} = \cdots$.

Fact 3: Note that the $AR(p)$ model

$$X_t - \mu = \phi_1(X_{t-1} - \mu) + \cdots + \phi_p(X_{t-p} - \mu) + \varepsilon_t$$

can be rewritten as

$$X_t - \mu - \phi_1(X_{t-1} - \mu) - \cdots - \phi_p(X_{t-p} - \mu) = \varepsilon_t, i.e.$$
$$\sum_{j=0}^{\infty} \pi_j(X_{t-j} - \mu) = \varepsilon_t,$$

with $\pi_0 = 1, \pi_1 = -\phi_1, ..., \pi_p = -\phi_p, 0 = \pi_{p+1} = \pi_{p+2} = \cdots$.

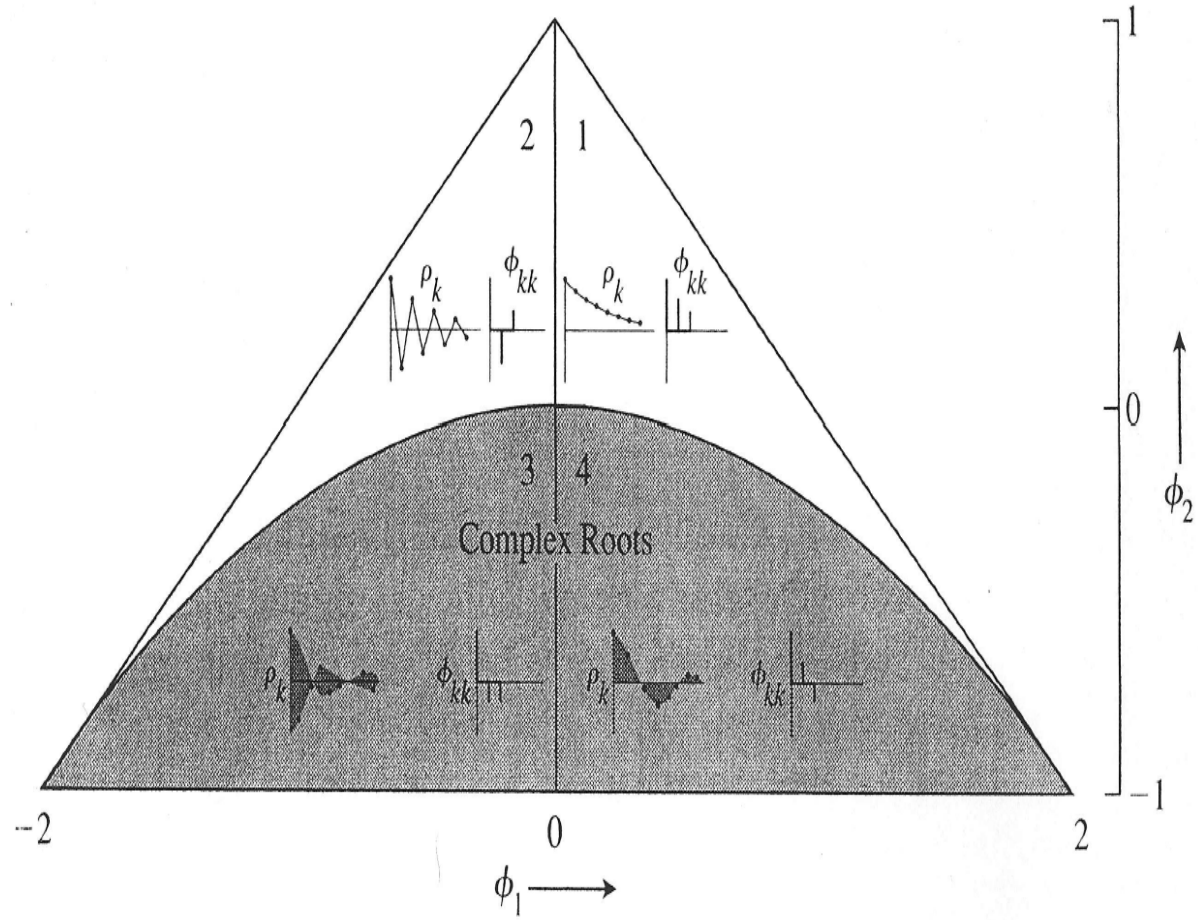

Figure 3.2 Typical autocorrelation and partial autocorrelation functions ρ_k and ϕ_{kk} for various stationary AR(2) models. (From [183].)