BOX-JENKINS MODEL NOTATION

The Box-Jenkins ARMA(p,q) model is denoted by the equation

$$y_t = \phi_0 + \phi_1 y_{t-1} + \cdots + \phi_p y_{t-p} + a_t - \theta_1 a_{t-1} - \cdots - \theta_q a_{t-q} . \qquad (1)$$

The autoregressive (AR) part of the model is $\phi_1 y_{t-1} + \cdots + \phi_p y_{t-p}$ while the moving average (MA) part of the model is $-\theta_1 a_{t-1} - \cdots - \theta_q a_{t-q}$. The model's intercept is ϕ_0 while a_t is the "white noise" error. The parameters (coefficients) $\phi_0, \phi_1, \cdots, \phi_p, \theta_1, \cdots, \theta_p$ of the model are determined from the data by the method of moments, least squares, the method of maximum likelihood, or some other method that is consistent. The white noise errors terms a_t are assumed to have the following properties:

1. $E(a_t) = 0$, $\forall t$ (zero mean assumption)
2. $E(a_t^2) = \sigma_a^2$, $\forall t$ (constant variance assumption)
3. $E(a_s a_t) = 0$, $\forall s \neq t$ (independence of errors assumption)
4. a_t are normally distributed

Assumption 4 is not always needed for deriving certain results with respect to the Box-Jenkins model but we will assume it here.

Equation (1) is referred to as the **intercept-form** of the Box-Jenkins ARMA(p,q) model. An algebraically equivalent form of the model (often reported by computer programs like SAS) is the so-called **deviation-from-the mean form** of the Box-Jenkins model:

$$y_t - \mu = \phi_1 (y_{t-1} - \mu) + \cdots + \phi_p (y_{t-p} - \mu) + a_t - \theta_1 a_{t-1} - \cdots - \theta_q a_{t-q} \quad . \qquad (2)$$

In this form, the mean of y, denoted by μ, is related to the intercept ϕ_0 of equation (1) by the formula $\mu = [\phi_0 / (1 - \phi_1 - \cdots - \phi_p)]$.

Probably the most compact way to write the Box-Jenkins ARMA(p,q) model is by using "backshift" polynomials $\phi(B) = 1 - \phi_1 B - \phi_2 B^2 - \cdots - \phi_p B^p$ and $\theta(B) = 1 - \theta_1 B - \theta_2 B^2 - \cdots - \theta_q B^q$. $\phi(B)$ is called the **autoregressive backshift polynomial** and $\theta(B)$ is called the **moving average backshift polynomial**. The **backshift operators** $B, B^2, B^3, \cdots$ simply "shift back" in time observations y_t so that $B^s y_t \equiv y_{t-s}$, for example. Using these polynomials we can write equation (2) compactly as

$$\phi(B)(y_t - \mu) = \theta(B)a_t \quad . \tag{2'}$$

For example, the ARMA(1,0) model (in brief AR(1)) can be written as

$$(1 - \phi_1 B)(y_t - \mu) = a_t \; .$$

Using the distributive law and the properties of the backshift operator we have

$$y_t - \mu - \phi_1 B y_t + \phi_1 B \mu = a_t$$

and

$$y_t - \mu - \phi_1 y_{t-1} + \phi_1 \mu = a_t$$

and finally

$$y_t - \mu = \phi_1 (y_{t-1} - \mu) + a_t \quad .$$

This last equation is the deviation-from-the-mean form of the AR(1) Box-Jenkins model. Similarly the ARMA(0,1) model (in short MA(1)) can be written as

$$y_t - \mu = a_t - \theta_1 a_{t-1}$$

and in polynomial form

$$y_t - \mu = \theta(B) a_t$$

where the moving average polynomial is defined by $\theta(B) = 1 - \theta_1 B$. The ARMA(1,1) "mixed" Box-Jenkins model can be written as

$$y_t - \mu = \phi_1 (y_t - \mu) + a_t - \theta_1 a_{t-1}$$

or more compactly using backshift polynomials as in

$$\phi(B)(y_t - \mu) = \theta(B) a_t$$

where $\phi(B) = 1 - \phi_1 B$ and $\theta(B) = 1 - \theta_1 B$.

For the Box-Jenkins model to be "estimable" the so-called **stationarity** and **invertibility** conditions must hold. If we replace the back shift operators $B, B^2, \cdots, B^p$ in the autoregressive polynomial and the moving average polynomial with corresponding powers of z, $z, z^2, \cdots, z^p$, and set these polynomials to zero we have what are called the **autoregressive polynomial** of the ARMA(p,q) Box-Jenkins model, namely,

$$\phi(z) = 1 - \phi_1 z - \phi_2 z^2 - \cdots - \phi_p z^p = 0 \quad (3)$$

and the **moving average polynomial** of the ARMA(p,q) model, namely

$$\theta(z) = 1 - \theta_1 z - \theta_2 z^2 - \cdots - \theta_p z^p = 0 \ . \quad (4)$$

Treating the parameters $\phi_1, \phi_2, \cdots, \phi_p$ and $\theta_1, \theta_2, \cdots, \theta_p$ as known in these homogenous equations (eventually, to be practical, we will have to use estimates of these parameters) let $z_1^{AR}, z_2^{AR}, \cdots, z_p^{AR}$ denote the p roots (zeroes) of the autoregressive polynomial (3) and $z_1^{MA}, z_2^{MA}, \cdots, z_q^{MA}$ denote the q roots of the moving average polynomial (4). For the Box-Jenkins model (1) to be **stationary** it must be the case that **all** of the roots of the autoregressive polynomial (3) must be **greater than one in magnitude** (or, if complex, have modulus greater than one). For the Box-Jenkins model (1) to be **invertible** it must be the case that all of the roots of the moving average polynomial **must be greater than one in magnitude** (or, if complex, have modulus greater than one).

STATIONARITY

Consider the AR(1) model $y_t - \mu = \phi_1(y_{t-1} - \mu) + a_t$. For this model the autoregressive polynomial equation is $1 - \phi_1 z_1 = 0$ and therefore $z_1^{AR} = 1/\phi_1$ is the root of the autoregressive polynomial. Thus, for the AR(1) model to be stationary it is required that $|1/\phi_1| > 1$ and therefore that $|\phi_1| < 1$. Similarly, for an MA(1) model $y_t - \mu = a_t - \theta_1 a_{t-1}$ to be invertible it is required that $z_1^{MA} = |1/\theta_1| > 1$ and therefore that $|\theta_1| < 1$. For the stationarity and invertibility conditions for other popular Box-Jenkins models like the AR(2), MA(2), and ARMA(1,1) models, see my ADF and PACF table in the document ACF_PACF.doc. By definition, **all** AR(p) models are invertible while **all** MA(q) models are stationary.

Now consider the practical implications of stationarity and invertiblility in Box-Jenkins models. When a Box-Jenkins model is stationary its observations y_t satisfy the following three properties:

1. $E(y_t) = \mu \quad \forall t$ (i.e. the mean of y_t is constant for all time periods)
2. Var(y_t) = $\sigma_y^2 \quad \forall t$ (i.e. the variance of y_t is constant for all time periods)
3. Cov(y_t, y_{t-j}) = γ_j (i.e. the covariance between y_t and y_{t-j} is constant for all time periods and fixed j, j = 1, 2, $\cdots$)

These three conditions give rise to what is called **weak stationarity** (or just stationarity for short). The practical implication of stationarity is that only **one realization** of the time series y_t is needed for us to be able to consistently estimate the mean μ, the variance σ_y^2, the covariance γ_j, and the autocorrelation ρ_j with the sample statistics $\bar{y}$, s_y^2, c_j, and r_j. These statistics are defined as

$$\bar{y} = \frac{\sum_{t=1}^{T} y_t}{T} \quad r_j$$ where T denotes the total number of observations available on y_t (sample mean)

$$s_y^2 = \frac{\sum_{t=1}^{T}(y_t - \bar{y})^2}{T} \quad \text{(sample variance)}$$

$$c_j = \frac{\sum_{t=j+1}^{T}(y_t - \bar{y})(y_{t-j} - \bar{y})}{T} \quad \text{(sample covariance)}$$

$$r_j = \frac{c_j}{s_y^2} = \frac{\sum_{t=j+1}^{T}(y_t - \bar{y})(y_{t-j} - \bar{y})}{\sum_{t=1}^{T}(y_t - \bar{y})^2}, \; j = 1, 2, \cdots.$$

(sample autocorrelation)

On the other hand, we can see that if our time series y_t has, for example, an ever increasing mean (trend), then the sample mean $\bar{y}$ would not be appropriate for characterizing the trend in the data. The sample statistic s_y^2 would likely not be appropriate for series data that exhibits, for example, an increasing or decreasing volatility overtime. Furthermore, data that exhibits a changing autocorrelation structure overtime would be incorrectly characterized by the sample statistics c_j and r_j. For some examples of various forms of nonstationarity in time series data you should run the SAS program MCARLO.sas that is available on the website for this course. I will go over some of these graphs in class and what data transformations are necessary to change a nonstationary time series into a stationary time series so that the transformed series can be properly modeled using the Box-Jenkins model. For example, when a time series has a linear trend, a typical transformation to use to make the data y_t stationary is to take the first difference of the data, $\Delta y_t \equiv y_t - y_{t-1} = y_t^*$, where Δ represents the first difference operator and y_t^* represents the transformed data.

Since many economic and business time series have trend in them the Box-Jenkins model is generalized to the case where Δy_t is modeled as an ARMA(p,q) process. Such a model can be written in intercept form as

$$\Delta y_t = \phi_0 + \phi_1 \Delta y_{t-1} + \cdots + \phi_p \Delta y_{t-p} + a_t - \theta_1 a_{t-1} - \cdots - \theta_q a_{t-q} \quad (5)$$

or in deviation-from-the-mean form as

$$\Delta y_t - \mu_{\Delta y} = \phi_1(\Delta y_{t-1} - \mu_{\Delta y}) + \cdots + \phi_p(\Delta y_{t-p} - \mu_{\Delta y}) + a_t - \theta_1 a_{t-1} - \cdots - \theta_q a_{t-q} \quad (5')$$

where $\mu_{\Delta y}$ denotes the mean of the Δy_t series. In the case of taking the first difference of the data to make it stationary (and thus amenable to Box-Jenkins analysis) the models (5) or (5') or denoted by ARIMA(p,1,q) where the middle number represents the number of times the data has to be differenced in order to make the data stationary. In general, if the data has to be differenced d consuecutive times to render the data stationary, as in $\Delta^d y_t$, then the d-differenced Box-Jenkins ARMA(p,q) model is denoted by ARIMA(p,d,q) and is written in intercept form as

$$\Delta^d y_t = \phi_0 + \phi_1 \Delta^d y_{t-1} + \cdots + \phi_p \Delta^d y_{t-p} + a_t - \theta_1 a_{t-1} - \cdots - \theta_q a_{t-q} \quad . \quad (6)$$

It cannot be overemphasized how important the proper choice of the order of differencing (d) is in properly modeling economic and business time series by means of the Box-Jenkins model. If an inappropriate choice of d is made before proceeding to build a Box-Jenkins model for the data, poor forecasting models will result. More will be discussed on this topic when we address the issue of "unit root" testing later in the course.

INVERTIBILITY

The implication of the invertibility condition can best be appreciated by considering the MA(1) model with, for simplicity, a zero mean for y_t assumed ($\phi_0 = 0$). We can rewrite the MA(1) model by iterative substitution.

$$\begin{aligned} y_t &= a_t - \theta_1 a_{t-1} \\ &= a_t - \theta_1 (y_{t-1} + \theta_1 a_{t-2}) \\ &= a_t - \theta_1 y_{t-1} - \theta_1^2 a_{t-2} \\ &= a_t - \theta_1 y_{t-1} - \theta_1^2 y_{t-2} - \cdots - \theta_1^t y_1 - \theta_1^{t+1} a_0 \quad . \end{aligned} \quad (7)$$

From equation (7) we can see that an MA(1) model can be rewritten as an infinite order (as $t \to \infty$) AR model (AR(∞)). But for (7) to be meaningful, we should have $|\theta_1| < 1$ so that current values of y (y_t) become less and less dependent on distant past values of y as time proceeds. If say, $\theta_1 > 1$, a current y_t will be infinitely dependent on the past observations as $t \to \infty$. This last circumstance doesn't seem appropriate to real world applications. Therefore, in the MA(1) case, imposing the invertibility condition eliminates models that put infinite weight on distant past values of a time series in determining current values of the time series.

Another rationale for imposing the invertibility conditions on MA models is to ensure **identification** of MA models. Consider the following two MA(1) models:

$$y_t = a_t - \theta_1 a_{t-1} \tag{8}$$

and

$$y_t = a_t - \theta_1^* a_{t-1} \text{ where } \theta_1^* = 1/\theta_1 . \tag{9}$$

The first MA(1) model has the same autocorrelation function as the second MA(1) model! The autocorrelation function of the first model is $\rho_1 = -\theta_1/(1+\theta_1^2)$ and 0 for $j \geq 2$ while the autocorrelation function of the second model is $\rho_1 = -\theta_1^*/(1+\theta_1^{*2}) = -(1/\theta_1)/(1+(1/\theta_1)^2) = -\theta_1/(1+\theta_1^2)$ and 0 for $j \geq 2$. Since the autocorrelation functions of the two models are exactly the same, the autocorrelation function cannot be used to distinguish between the two parametrizations. Therefore, a second reason for imposing the invertibility conditions for MA models is to ensure their uniqueness (as it relates to the autocorrelation function) for given values of the MA coefficients.

When building Box-Jenkins models it should be recognized that such models are **identified up to a common factor**. That is, if a **common** root exists between the autoregressive and moving average polynomials, then a given ARMA(p,q) model can be reduced to a ARMA(p-1,q-1) model. Consider the following ARMA(1,1) model

$$(1-\phi_1 B)(y_t - \mu) = (1-\theta_1 B)a_t \ . \tag{10}$$

Obviously, if $\phi_1 = \theta_1$, then the ARMA(1,1) model of (10) can be reduced to the white noise model (ARMA(0,0)),

$$y_t - \mu = a_t \tag{11}$$

by canceling out the common factors that exist across the autoregressive and moving average backshift polynomials. The existence of this common factor is the result of the autoregressive and moving average polynomials ((3) and (4)) have the common roots $z_1^{AR} = 1/\phi_1 = z_1^{MA} = 1/\theta_1$ because $\phi_1 = \theta_1$. Thus, in the presence of these common factors (common roots), the data will not allow us to distinguish between equation (10) and (11) and we say that ARMA models are identified up to a common factor.

The lesson to be drawn here is that when building a Box-Jenkins model for a given time series we should be careful in comparing two competing model, one of which has one more autoregressive parameter and one more moving average parameter than the other competing model, that is when we are comparing a ARMA(p-1, q-1) model with a ARMA(p,q) model. If the bigger model has an (estimated) autoregressive polynomial with a root that is “almost” equal to a root of the (estimated) moving average polynomial then we might suspect that there is in fact a common root in the population model and thus the simpler model is to be preferred. So when you have a satisfactory (in the sense of goodness-of-fit statistics and white noise residuals) ARMA(p-1,q-1) model but a “somewhat” better ARMA(p,q) model you should inspect the empirical roots of the autoregressive and moving average polynomials of the ARMA(p,q) model to see if there is an “almost” common root across the polynomials. If there is, then you are probably better served (in terms of forecasting accuracy) in going with the simpler ARMA(p-1,q-1) model.