CHAPTER 5. Box Jenkins ARIMA Modeling.
Copyright by Prof. H. D. Vinod, Fordham University. Jan. 1993. All rights reserved.

The general expression for ARIMA models should include seasonal factors as well, denoted by ARIMA(p,d,q)(P,D,Q)s is

\[(1 - \phi_1 L - \phi_2 L^2 - \cdots - \phi_p L^p) (1 - \Phi_1 L^s - \Phi_2 L^{2s} - \cdots - \Phi_P L^{Ps}) y_t = \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \cdots + \theta_q \varepsilon_{t-q} + \varepsilon_t\]  

where \(y_t\) is the original data, with \(t=\text{time}\), the lag (backshift) operator \(L\) is defined by: \(L \hat{y}_t = y_{t-1}\), \(LL \hat{y}_t = y_{t-2}\) as before. The difference operator \(\nabla\) is subject to the identity: \(\nabla y_t = y_t - y_{t-1}\), and is usually used to induce stationarity into a time series.

- \(\phi\) mean of the data.
- \(p\) order of the nonseasonal autoregressive term
- \(q\) order of the nonseasonal moving average term
- \(P\) order of the seasonal autoregressive term
- \(Q\) order of the seasonal moving average term
- \(d\) order of the nonseasonal differencing
- \(D\) order of the seasonal differencing
- \(s\) length of seasonality
- \(\varepsilon_t\) random error or white noise term
- \((L)\) moving average operator other than the seasonal
- \((1 - \Theta_1 L^s - \Theta_2 L^{2s} - \cdots - \Theta_Q L^{Qs})\) autoregressive operator other than the seasonal
- \((\Theta_1 L^s - \Theta_2 L^{2s} - \cdots - \Theta_Q L^{Qs})\) seasonal operator.
- \((\Theta_1 L^s - \Theta_2 L^{2s} - \cdots - \Theta_Q L^{Qs})\) moving average seasonal operator.
- \((\Theta_1 L^s - \Theta_2 L^{2s} - \cdots - \Theta_Q L^{Qs})\) autoregressive seasonal operator.
- \((\Theta_1 L^s - \Theta_2 L^{2s} - \cdots - \Theta_Q L^{Qs})\) moving average seasonal operator.

The methodology of estimating these models is discussed in many books including Box and Jenkins (1970). We shall now review it briefly. The estimation usually uses Marquardt or similar nonlinear least squares algorithm. It may need backward forecasting (backcasting) to estimate the model parameters. Estimation

Estimation is generally carried out by the maximum likelihood (ML) or approximate maximum likelihood method. The exact likelihood function for a mixed ARMA(p,q) model is given by Newbold (1974). Denote the entire set of AR and MA parameters by \(\theta\), the error by \(\varepsilon_t\), and the whole time series data on \(y_t\) is denoted by \(y\). Also, denote error sum of squares by \(\text{ESS}(\theta, y) = \sum_{t=1}^{T} (y_t - \hat{y}_t)^2\) Now, the likelihood is proportional to:

\[\frac{1}{T} \exp \left( -\frac{1}{2} \text{ESS}(\theta, y) \right)\]  

There are three main methods for estimating (1). First, is ML, which maximizes (4). A practical transformation is suggested by Ansley (1979). The second method of estimating (1) is called exact least squares, which ignores the term \(\hat{y}_t\) in (4), and minimizes the ESS. Observe that the lower limit of the summation in (3) is \(-\infty\), and we would generally only have finite data. Box and Jenkins suggest back-casting (forecasting backwards in time) a number of pre sample values of \(\varepsilon_t\). The third method is even simpler and more approximate. It is called conditional least squares, and ignores the back-casting step. The word conditional refers to the fact that the conditional expectation of pre sample values is zero, and when we ignore pre sample forecasting, we are replacing them by their conditional expectations.

Diagnostic Checking
After estimation of (1) is completed, it is desirable to test the adequacy of the model. It is obvious that the residuals after the fit should not have any serial correlation. Instead of the Durbin-Watson type test, the time series literature uses tests based on autocorrelations of residuals, $\hat{z}_j$, their sums of squares, etc. A common graphical procedure plots the actual $\hat{z}_j$ against $j$ along with the 95% confidence band based on asymptotic standard errors. If the observed $\hat{z}_j$ are beyond the 95% confidence interval, it indicates that one has to modify the model.

Instead of confidence intervals on $\hat{z}_j$ one may seek to use a single statistic. Hosking (1980) reviews a number of Lagrange Multiplier tests for this purpose. Note that in time series ARMA(p,q) modeling one may specify the orders $p$ and $q$, which describes the null hypothesis. There is a variety of alternative hypotheses possible with other orders of $p$ and $q$, possible interactions, non-linearities, etc. If the statistical test requires one to specify a specific alternative, it would be difficult to use. The Lagrange Multiplier tests are designed against general, non-specific alternatives. Hence, the popular tests are of this type.

**Portmanteau Test by Box and Pierce (1970)**

The word portmanteau means a large leather suitcase. Assume that certain conditions given by Box and Pierce (1970) are satisfied: (i) that the number of lags considered $h_T$ increase as the sample size $T$ increase, (ii) that the Wold decomposition of $y_t$ has coefficients of order $O(1/T)$ for $j \neq h_T$, and (iii) that $h_T$ is of order $O(T)$. Then the statistic

$$Q = \sum_{j=1}^{h_T} \hat{z}_j^2$$

where $h_T$ is arbitrary and there may be a loss of power if $Q$ is used for MA models. For the MA(1) model, the Durbin-Watson type (von Neumann ratio based) test based on $\hat{z}_j$ may have higher power than the Q test. To improve the performance of $Q$ in small samples the following modification is suggested.

**Ljung Box Modification of the Portmanteau Test**

Ljung and Box (1978) argue that under the null hypothesis that the estimated model is adequate, the following statistic gives a closer approximation to the true critical region:

$$Q^* = \sum_{j=1}^{h_T} \frac{\hat{z}_j^2}{(T \bar{A} j)}$$

Davies et al (1977) point out that the variance of $Q^*$ statistic exceeds that of the Chi-square.

**Autocorrelation of Squared Residuals**

Engle (1982) and others. A simple portmanteau test for ARCH effects may be based on the autocorrelation of squared residuals $\hat{z}_j$, Granger and Anderson (1978b) report examples where there may be significant $\hat{z}_j$ even though the $\hat{z}_j$ are uncorrelated. The following statistic suggested by McLeod and Li (1983)

$$Q^{W*} = \sum_{j=1}^{h_W} \frac{\hat{z}_j^2}{(T \bar{A} j)}$$

Various portmanteau tests $Q$, $Q^*$ and $Q^{W*}$ considered above are found to be useful in rejecting inadequate models. They are not very robust in distinguishing between models for ultimate selection. Hosking (1980)
views the Q tests as Lagrange Multiplier tests and shows that if the alternative hypothesis is an AR(p+1) or an MA(q+1) process, the statistic is the same.

Turning Point Tests

Brockwell and Davis (1987 p.302) discuss many of these and related tests. If there is a turning point at i if 1<i<n and if y_i<y_{i-1} and at the same time y_{i+1}>y_i. If y_1, y_2, ..., y_n are random iid sequence, then the probability of a turning point at time i is 2/3. The expected number of turning points (T) is E(T)=2(n-2)/3 and variance Var(T)=16n-29)/90 hence T–E(T)/sqrt(VAR(T)) is asymptotically unit normal.

Normality Tests:

The assumption of normality is present in maximum likelihood and many related estimation methods as well as in testing. Testing for normality is usually based on Pearson's measures of skewness and kurtosis

\[ \text{Skew} = r^{-1} \cdot j(-.5)^{5/2} \] and \[ \text{Kurt} = (\cdot j(-.5)^{3/2} \]

where \( r = \frac{\text{D}(y - \bar{y})}{T} \), the j-th central moment. Bowman and Shenton (1975) show that the estimates \( \hat{\text{Skew}} \) are asymptotically normal, or \( \hat{\text{Skew}} \sim \text{AN}(0, 6/T) \) and \( \hat{\text{Kurt}} \sim \text{AN}(3, 24/T) \). Since the estimated Skew and Kurt are independent of each other, an asymptotic test statistic is \( (T/6)^{\text{Skew}} + (T/24)(\text{Kurt}-3)^2 \). If the distribution is normal (under the null), this statistic has \( \chi^2 \) as the asymptotic distribution with 2 degrees of freedom. Granger and Newbold (1977, p. 315) note that the asymptotic variance of \( \hat{\text{Skew}} \) is \( (6/T) \) and the asymptotic variance of \( \hat{\text{Kurt}} \) is \( (24/T) \) where the \( \mu_i \) may be approximated by \( \mu_0 \) and the infinite sums replaced by finite sums as an approximation. The approximation is reasonable for moderate sample sizes, but not for small samples.


Given a time series \( y_t \) with \( t \) defined on the index set \( g = \{ -2, -1, 0, 1, 2, \ldots \} \) the difference operator is defined by \( y_{t-1} - y_t \). This is called the “first” difference. Second difference is defined by \( y_{t-2} - 2y_{t-1} + y_t \), which uses the identity \( y_{t-1} - y_t = (y_{t-1} - y_t) - (y_t - 1) \) between the difference operator and the lag operator. In general, \( y_{t-p} (1 - L)^p \) will obviously involve the Binomial coefficients.

6.1 Linear Difference Equations:

Definition: It is convenient to define the \( p^{th} \) order (ordinary) linear difference equation as follows.

\[ y_t = \sum_{i=0}^{p} a_i y_{t-i} \]

where \( p < n \), where the coefficients \( a_i \) are assumed to be constant, and where \( a_i \) is called an input function or a forcing function.

The homogeneous form of the difference equation is defined by setting \( a_i = 0 \), yielding

\[ y_t = \sum_{i=0}^{p} a_i y_{t-i} \]

(2)

The solution of the homogeneous form: This is obtained by first writing (2) in terms of the so-called characteristic polynomial in \( L \), the lag operator written with \( (Lz^{-1}) \):

\[ (1 - L)^p y_t = \sum_{i=0}^{p} a_i y_{t-i} \]

(3)

We define the roots -1, \( z = 1, 2, \ldots, p \) of the (autoregressive) polynomial \( g(z) \) or \( g(L) \) by writing \( (1 - L)^p y_t = \sum_{i=0}^{p} a_i y_{t-i} \), \( (Lz^{-1}) y_t = \sum_{i=0}^{p} a_i y_{t-i} \), \( (z - L) y_t = \sum_{i=0}^{p} a_i y_{t-i} \), \( (z - L^p) y_t = \sum_{i=0}^{p} a_i y_{t-i} \)

(4)

The solution will contain the \( p \) values and the starting value at \( t=0 \) to be \( y_0 \). For clarity of exposition let us first assume \( p=1 \), yielding a first order homogeneous difference equation: \( y_t = \sum_{i=0}^{1} a_i y_{t-i} \). At \( t=1 \), we have \( y_1 = \sum_{i=0}^{1} a_i y_{1-i} \). At \( t=2 \), we have \( y_2 = \sum_{i=0}^{1} a_i y_{2-i} \). Similarly at \( t=3 \) we have \( y_3 = \sum_{i=0}^{1} a_i y_{3-i} \). In general the solution of the first order homogeneous difference equation is \( y_t = \sum_{i=0}^{1} a_i y_{t-i} \). This is called a solution because it evaluates \( y_t \) from the two assumed known quantities \( y_1 \) and \( y_0 \).

First Order Homogeneous Difference Equation: Note that the polynomial (3) with \( p=1 \) for our first order equation is: \( (1 - L) y_t = \sum_{i=0}^{1} a_i y_{t-i} \). This makes it obvious that the root of the autoregressive (AR) characteristic polynomial in \( z \) \( L^{-1} \) is -1, \( L_1 \). In the solution \( y_t = \sum_{i=0}^{1} a_i y_{t-i} \), we raise the root \( L_1 \) to the power \( t \). If \( t \) is large and \( |a_1| > 1 \), it is clear that when such a \( L_1 \) is raised to a large power, \( L_1^t \) term may become a very large positive or
negative number depending on the sign of \( g_t \). Thus, if \(|g_t| > 1\), the solution \(|y_t| f_{-} t\) as \( t f_{-}\), where we have used the absolute value of \( y_t \) because its sign will depend on the sign of the starting value \( y_0 \), which may be negative.

**Exercise:** Verify that any negative starting value of \( y_0 \) will lead to oscillatory behavior with changing signs of successive \( y_t \).

By contrast, if \(|g_t| < 1\), a fraction, it will become smaller and smaller as it is raised to a higher power. Then the solution \(|y_t| f_{-} 0\) as \( t f_{-}\) if \(|g_t| < 1\). Finally, if \(|g_t| \neq 1\), raising it to any power makes no difference, and the solution is \(|y_t| y_{-} t\) if \( g_t \neq 1 \) and \(|y_t| A_{-} t\) as \( t f_{-}\) if \( g_t = 1 \). It is customary to describe the roots larger than \( 1 \) as being outside the unit circle, roots smaller than \( 1 \) as being inside the unit circle, and the roots equal to \( 1 \) as being on the unit circle. The dynamic system whose solution \(|y_t| t\) is called unstable or non-stationary, and if \(|y_t| 0\), it is called stable or stationary. When \( g_t \neq 1 \), it is called a random walk model. Since this has considerable importance in econometrics it is discussed separately. For the random walk model the solution \( y_t \) is such that the random shocks accumulate indefinitely. It can be shown that its Green's function is unity, that is the memory is long and constant, and the process never dies out. The solutions of higher order homogeneous difference equations are analogous, having somewhat complicated expressions involving all \(-i\) roots of the characteristic polynomial in \( z \) defined above in (3).

Now, let us consider the first order non-homogeneous case: \( y_t A_{-} g_t y_{t-1} \) \( \neq \) \( A_{-} \) \( A_{-} \) for all \( t \). This is called non-homogeneous because \( A_{-} 0 \). At \( t=1 \), we have \( y_1 \) \( A_{-} g_1 y_0 + A_{-} \). At \( t=2 \), we have \( y_2 \) \( A_{-} g_2 y_1 + A_{-} \) \( A_{-} \) \( g_2 y_0 + g_1 A_{-} + A_{-} \). Similarly at \( t=3 \) we have \( y_3 \) \( A_{-} g_3 y_2 + g_2 A_{-} + g_1 A_{-} + A_{-} \). In general the solution of the first order non-homogeneous difference equation is

\[
y_t \begin{cases} 
y_1 & \text{if } g_1 \neq 1, \\
y_0 + A a_j & \text{if } g_1 = 1. \end{cases}
\]

This is called a solution because it evaluates \( y_t \) from the three assumed known quantities \( g_t \), \( A_{-} \) and \( y_0 \). In deriving the above formula we have assumed that the input \( a_t \) \( A_{-} \) for all \( t \). If we let \( a_t \) be a function of time \( t \), and substitute \( a_t \) values at \( t=1,2,3,\ldots \) for \( A_{-} \), we get the following solution:

\[
y_t \begin{cases} 
y_1 & \text{if } g_1 \neq 1, \\
y_0 + A j=0 a_{i-j} & \text{if } g_1 = 1. \end{cases}
\]

The contribution of \( y_0 \) to the solution remains bounded as long as \(|g_t| < 1\).

**Second Order Non-homogeneous Difference Equation and Business Cycles:**

Samuelson (1939) applied second order difference equations to explain the business cycles as follows. The accelerator principle states that the investment (\( I \)) is dependent on the changes in income rather than the level of income: \( I_{-} u_{-} b_{-}(Y_{-1} \ A_{-} Y_{-2}) \). Consumption (\( C_{-} \)) depends on the past income by the first order nonhomogeneous equation \( C_{-} \) \( u_{-} b_{-} \) \( Y_{-1} + b_{-} \) \( t \) \( C_{-} \) \( C_{-} \). Substituting the first two equations in this identity, we have \( Y_{-} \) \( u_{-} b_{-} (Y_{-1} \ A_{-} Y_{-2}) \) \( + b_{-} Y_{-1} + b_{-} \). For certain values of \( b_{-} \) when the roots of the characteristic polynomial are imaginary, this second order equation can be shown to lead to oscillation is income.

6.2 Stochastic difference equations:

So far, we have considered the ordinary (nonstochastic) difference equations. The stochastic difference equations are characterized by the fact that the input function \( a_t \) of the difference equation is a white noise random variable with zero mean and variance \( \sigma^2 \), that is

\[
a_t \mu N(0, \sigma^2) \text{ for all } t. \quad \text{Note that } E(a_{i+j}) \mu 0, \text{ for all } j \neq 0.
\]

6.2.1 First Order AR(1) Stochastic difference equations:

Consider the first order autoregressive model denoted by AR(1) and defined by the stochastic difference equation for \( y_t \) measured from the mean, i.e., \( E y_t \mu 0 \).

\[
y_t A_{-} g_{t-1} \mu (1 A_{-} gL) y_{t-1} \mu a_t, \quad \text{where } a_t \sim N(0, \sigma^2) \text{ and } |g| < 1 \] (1)
In a higher order autoregressive model \( y \) has subscripts 1, 2 etc. Now the solution may be obtained by the brute force method of substitution similar to the one above, and it is also obtained by a formal division of both sides of the equation by the characteristic polynomial

\[
y_t \hat{y} (\phi L)^{-1} a_t \hat{y} \frac{9}{j=0} a_{t-j} \hat{y} G_j a_{t-j} \quad \text{(where } G_0=1) \tag{2}\]

and where the coefficients \( G_j \) are called Green's functions, Miller (1968). A random shock at time \( t \) gets multiplied by the weight \( G_j \) suggesting that the Green's function measures the extent to which the system remembers a random shock. Hence, \( G_j \) may be used to represent the memory of the dynamic system. It is interesting to note that Green's function can be regarded as an “orthogonal decomposition” of \( y \) first proved by the econometrician H. Wold (1938). To visualize the orthogonality imagine infinite dimensional space with axes marked \( a_t, a_{t-1}, \ldots \) along which we have the orthogonal (independent) random inputs, and \( y_t \) is a vector through the origin.

For future reference we note that the solution of the first order difference equation (1) arising from AR(1) model having one root (= 9) of the characteristic polynomial is given by \( y_t \hat{y} 9^t \).

**Variance and Autocovariance for the AR(1):**

Now we turn to the autocovariance \( \# \hat{y} E(y_{t+k}) \) of order \( k \), which is the numerator of the autocorrelation coefficient \( \beta_k \). When \( k=0 \) we simply have the variance \( \# \) defined as follows.

\[
\# \hat{y} E(y_t^2) \hat{y} E \frac{9}{j=0} a_{t-j} \hat{y} G_j a_{t-j} \delta \delta \]

\[
\hat{y} E(a_t + 9a_{t-1} + 9^2 a_{t-2}, \ldots) (a_t + 9a_{t-1} + 9^2 a_{t-2}, \ldots) d \]

\[
\# \hat{y} E(a_t)(a_{t+k} + 9a_{t+k-1} + 9^2 a_{t+k-2}, \ldots) + 9a_{t-1}(a_{t+k} + 9a_{t+k-1} + 9^2 a_{t+k-2}, \ldots) d \]

\[
\hat{y} 5^2 + 9^2 5^2 + 9^4 5^2 \ldots \]  
[Using \( E(a_t a_{t-j}) \hat{y} 0 \) when \( j \neq 0 \); and that \( E(a_t^2) \hat{y} 5^2 \) for all \( t \)]

Using the geometric series \( \frac{1}{1+2z} \) \( \hat{y} 1+z z^2 \ldots \) for \( z \) \( \hat{y} 9^2 \) we have

\[
\# \hat{y} \left( \frac{1}{1+9^2} \right) 5^2 \tag{3}\]

To evaluate \( \# \), we use (2) to write

\[
y_{t+k} \hat{y} (\phi L)^{-1} a_{t-k} \hat{y} \frac{9}{j=0} a_{t-j-k} \hat{y} G_j a_{t-j-k} \quad \text{(where } G_0=1) \tag{4}\]

From (2) and (4) we write

\[
\# \hat{y} E(y_{t+k}) \hat{y} E \frac{9}{j=0} a_{t-j} \hat{y} G_j a_{t-j} \delta \delta \]

Hence,

\[
\# \hat{y} \left( E(a_t + 9a_{t-1} + 9^2 a_{t-2}, \ldots) (a_{t-k} + 9a_{t-k-1} + 9^2 a_{t-k-2}, \ldots) d \right) \]

\[
\hat{y} E G_j (a_t) (a_{t-k} + 9a_{t-k-1} + 9^2 a_{t-k-2}, \ldots) + 9a_{t-1} (a_{t-k} + 9a_{t-k-1} + 9^2 a_{t-k-2}, \ldots) \]

\[
+ 9^2 a_{t-2} (a_{t-k} + 9a_{t-k-1} + 9^2 a_{t-k-2}, \ldots) \hat{y} \]

For \( k=1 \) we have:

\[
\# \hat{y} 0 + 95^2 + 9^3 5^2 + \ldots \hat{y} 9^2 \# \]  
using (3), and the geometric series used there.
To derive an expression for a larger $k (>1)$ it is convenient to write these results in terms of the Kronecker delta function $\delta_{ij}$ not to be confused with the Kronecker product of matrices $\tilde{A}$ defined below.

$$E(a_ia_i^\prime) \tilde{a}_{i,i} \delta_{j,j} \tilde{a}_{j,j} \delta_{k,k} \tilde{a}_{k,k} \delta_{l,l} \tilde{a}_{l,l}$$ (5)

Delta function can be used when multiplying two infinite sums above, because the nonzero terms occur only when the subscripts of $a$ are equal to each other.

$$\delta_{ij} \delta_{kl}$$

Now one infinite sum over $i$ can be eliminated by choosing the subscripts as follows. We require $t \tilde{a}_{i,i} = \tilde{A} j \tilde{A} k$, which implies that $\tilde{A} i \tilde{a}_{i,i} \tilde{A} j \tilde{A} k$, that is $i \tilde{a}_{i,i} j \tilde{A} k$, which allows us to replace $i$ by $j+k$ or $\tilde{a}_{i,i} j \tilde{A} k$ when $i \tilde{a}_{i,i} j \tilde{A} k$. Thus

$$E(y_{t+k}) \tilde{a}_{t,t} \delta_{k,k} \tilde{a}_{k,k} \delta_{l,l} \tilde{a}_{l,l}$$ (6)

using the geometric series. The autocorrelation coefficient is

$$\rho_k = \tilde{a}_{t,t} \tilde{a}_{k,k} \delta_{l,l} \tilde{a}_{l,l}$$ (7)

Exercise: Use (3) and (5) to evaluate $\rho_k$ for the AR(1) model, where $5^2(1 \tilde{A} 9^2)^{-1}$ cancels, to show that $\rho_k = \tilde{a}_{t,t} \tilde{a}_{k,k} \delta_{l,l} \tilde{a}_{l,l}$ for AR(1)

Exercise: Show that $\tilde{a}_{t,t} \tilde{a}_{k,k} \delta_{l,l} \tilde{a}_{l,l}$, the autocovariance function is also meaningful for negative values of $k$. Show that the $k$ in the last equation of (6) should be replaced by $|k|$ to accommodate the negative $k$ values.

**Autocovariance Generating Function (AGF)**

This concept requires some familiarity with complex numbers and Laurent expansions. Nerlove, et al (1979, p.38) and Sargent(1979,p.221) are some of the econometric references. Box and Jenkins(1970, p.81) is a statistical reference. Hamilton (1994, p.61) describes AGF's as a function constructed by taking $k$-th autocovariance and multiplying it by $k$-th power of some (complex) number $z$ and summing it over all possible values of $k$. AGF($z$) = $\sum_{k=-\infty}^{\infty} \tilde{a}_{t,t} \tilde{a}_{k,k} \delta_{l,l} \tilde{a}_{l,l}$

The AGF is simply a doubly infinite sum of a covariance times $z^k$, where $z$ is a complex variable and $k$ is the order of the lag. One may also think of $z$ as the familiar $L$ the lag operator. Using (6) defined for both negative and positive values of $k$ we try to get rid of the infinite sum by using a geometric progression in four steps (9a) to (9d). The following infinite sums converge on the unit circle and in an annulus region around it : $|9|<|z|<1/|9|$.

$$AGF(z) = \sum_{k=-\infty}^{\infty} \tilde{a}_{t,t} \tilde{a}_{k,k} \delta_{l,l} \tilde{a}_{l,l}$$

To verify the last equality use $D(A_{t,t} \to l) = D(A_{t,t} \to 1) + D(1 \to l)$ and $j = \tilde{A} k$. To remove the remaining infinite sum we take out the common term with $j=1$ as follows.

$$AGF(z) = \sum_{k=-\infty}^{\infty} \tilde{a}_{t,t} \tilde{a}_{k,k} \delta_{l,l} \tilde{a}_{l,l}$$ (9b)

where the bracketed first term gets a $(1 \tilde{A} 9^2)^2$ in the numerator which cancels.

For a one-sided general moving process $y_t = \sum_{j=0}^{\infty} G_j a_{t,j}$ the autocovariances
\(\#_\#_\# = E(y_t y_{t+k}) = E \sum_{0}^{\infty} G_j G_{j+k} \delta_{j} \delta_{j+k} \)

where the expectation of \(a_i a_j\) is 5 only when the subscripts match, and is zero otherwise. Thus

\[ \#_\#_\#_\# = 52^{k} \sum_{j=0}^{\infty} G_j G_{j+k} \]

may be substituted in the definition of AGF to yield

\[
AGF(z) = 5^{2} \sum_{k=-j}^{j} G_j G_{j+k} z^k = 5^{2} \sum_{k=-j}^{j} G_j G_{j+k} z^k = 5^{2} \sum_{h=0}^{\infty} G_j G_{h} z^{h-j} \tag{9c}
\]

To verify the second equality use \(G_j = 0\) when \(j < 0\). To verify the third equality use \(h = j + k, k = h - j\).

\[
AGF(z) = 5^{2} \sum_{h=0}^{\infty} G_j G_{h} z^{h-j} = 5^{2} G_z z = 5^{2} G_z G(z) \tag{9d}
\]

where the particular case (9b) reveals what is meant by the function \(G(z)\). The last equality is sometimes called canonical factorization. Thus \(z = \cos(\theta) - i \sin(\theta) = \exp(-i \theta)\) where \(i = -1\) and \(\theta\) is the radian angle that \(z\) makes with the real axis. If the AGF is evaluated at the \(z = \exp(-i \theta)\) and divided by 21, the resulting function of \(\exp(-i \theta)\) is called the population spectrum \(s(\theta) = (1/2) AGF\). If autocovariances are absolutely summable and if two process have common AGF, then they will also have identical autocovariances. This is why it is a generating function.

**Examples of AGF for MA processes:** Recall that the only nonzero autocovariances are \#_\#_\#_\#, \#_\#_\#, \#_\#. Hence the infinite sum only has 3 terms with powers of \(z\) equal to the subscripts of \#_\#_\#_\#.

\[
AGF(\text{for MA(1)}) = z + (1 + \theta) z + z = [1+z](1+z) \]

More generally, for the MA(\(\infty\) ) process, \(y_t = \sum_{t} G(L) y_{t+k} \) where \(G(L) = G_0 + G_1 L + G_2 z^2 + \ldots\) is an infinite polynomial. Now,

\[
AGF(\text{for MA(\(\infty\))}) = 52G(z)G(z) \]

This is a powerful result because ARMA processes can often be written as MA(\(\infty\)) processes by Wold’s theorem. AGF(\(\text{for AR(1)}\)) is \(5^2[[1-z][1-z]]\)

**Stationarity of the Stochastic AR(1) Difference Equation:**

Now we turn to checking the stationarity of the AR(1) model. For a stochastic process to be weakly stationary \(\hat{y}\) also called covariance stationary \(\hat{y}\) recall that there are three requirements: (i) \(E(y_t) \leq \text{ for all } t\), and is not a function of \(t\). (ii) \(E(y_t \hat{A} \hat{y}_{t+k}) \leq \text{ for all } t\), and is not a function of \(t\). (iii) \(E(\hat{y}_t \hat{A} \hat{y}_{t+k}) \leq \text{ for all } t\), and is not a function of \(t\). For the AR(1) model the variance \#_\#_\#_\# from (3) is \(5^2(1+9^t+9^{2t}+\ldots)\), where the infinite series in \(9^t\) will be finite, if and only if \(|9| < 1\). This stationarity condition will also ensure the finiteness of \#_\#_\#_\#.

**AGF and the Power Spectrum**

Some econometricians introduce the idea of a power spectrum by using the relationship between the complex number \(z\) and the familiar lag operator. Nerlove(1979, App. C) gives a thorough discussion of the related material. Here we state some facts for convenience of the reader, and not attempt a complete discussion. When we substitute \(z = \exp(2i \theta)\) we obtain half the power spectrum at frequency \(0 \leq \theta \leq \pi/2\). Thus the power spectrum of the general moving average process is

\[
p(\theta) = 25^2 G(z) G(z^{-1}) = 25^2 G(e^{2i \theta}) G(e^{-2i \theta}) = 25^2 |G(e^{2i \theta})|^2 \tag{9d}
\]

The spectral density function \(s(\theta) = (5^2/21) G(e^{i \theta}) G(e^{-i \theta})\) is a Fourier transform of the autocovariance function. It is proportional to the AGF defined on the unit circle. It is always real even though AGF involves complex numbers. The integral of sdf from -\(\pi\) to \(\pi\) is the variance of the process. In a usual density function the area under the density represents a probability. By analogy, the area under a spectral density represents the part of the total variance attributed to the range represented by the corresponding angular frequencies.

**Digression:** Autocorrelations satisfy the Homogeneous Difference Equation
We note an interesting fact that the autocorrelation coefficients \( \mathbf{Z}_k \) satisfy the homogeneous form of the difference equation (1), viz., \( y_i \) \( y_{i-1} \) if we replace \( y_i \) by \( \mathbf{Z}_k \) and define \( \mathbf{Z}_k \) \( \mathbf{Z}_{k-1} \). We have shown in (8) above that \( \mathbf{Z}_k \) \( \mathbf{Z}^k \). Hence all we have to do is to verify

\[
\mathbf{Z}_k \mathbf{Z}_{k-1} \mathbf{Z} (\mathbf{Z}^{k-1}) = \mathbf{Z}^k \quad (10)
\]

We shall see that this generalizes to higher order processes, and equations similar to (10) are called Yule-Walker equations.

### 6.2.2 Stochastic Difference Equation of the AR(2) Model:

Consider the AR(2) model

\[
y_t = A y_{t-1} + A y_{t-2} + a_t \quad (1)
\]

Its characteristic polynomial

\[
(1 - \lambda L)(1 - \lambda L) = 0
\]

where the roots satisfy two relations:

\[
\lambda_1 + \lambda_2 = 1, \quad \lambda_1 \lambda_2 = 1
\]

The roots of the quadratic are known explicitly by the formula

\[
\lambda = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}
\]

which are both real if the square root term in the above expression \( (\lambda_1^2 + 4 \lambda_2) \) \( \neq 0 \), otherwise they are complex conjugate to be discussed later. When they are real, the solution of AR(2) is obtained by dividing both sides of (1) by the characteristic polynomial (2). Then we have:

\[
y_t = \frac{a_t A \lambda_1 + B \lambda_2}{(1 - \lambda_1 L)(1 - \lambda_2 L)}
\]

where \( A \) and \( B \) are obtained below by the method of partial fractions. When \( \lambda_1 \) and \( \lambda_2 \) are real numbers, the method of partial fractions involves the following steps. To get \( A \) we solve the denominator \( (1 - \lambda_1 L) \) \( \neq 0 \), to yield \( \lambda_1 \) \( \neq 1 \). Now substitute this \( L \) in the long denominator of (5). That is,

\[
(1 - \lambda_2 L)(1 - \lambda_2 L) = (1 - \lambda_2 L)(1 - \lambda_2^{-1})(1 - \lambda_2^{-1}) = 0(0)(1 - \lambda_2^{-1}), \text{ which is taken to be } (1 - \lambda_2^{-1}) \text{ upon ignoring the zero multiplier. Similarly for } B. \text{ We have:}
\]

\[
A = \frac{1}{(1 - \lambda_1 L)} \quad \text{and } B = \frac{1}{(1 - \lambda_2 L)}
\]

To verify (5a) obtained by ignoring zeros it suffices to see that the numerators of expressions on the left hand and the right hand side of the last equal sign in (5) are identical. Upon simplification the question becomes:

\[
1 \frac{1}{(1 - \lambda_1 L)} + \frac{1}{(1 - \lambda_2 L)} = \frac{1}{(1 - \lambda_1 L)} + \frac{1}{(1 - \lambda_2 L)}
\]

In numerical problems it is easy to write the partial fractions. For example.

\[
\frac{1}{(1 - \lambda_1 L)} + \frac{1}{(1 - \lambda_2 L)}
\]

where all we have done is to substitute the solution \( x \) \( \neq 5 \) in the first denominator, except when it becomes zero; and \( x \) \( \neq 4 \) in the second denominator, except when the term becomes zero.

### Real Number Solution of the Stochastic AR(2) Difference Equation:

We use the partial fractions to write the second order problem as a sum of two AR(1) problems in (5).

\[
y_t = \sum_{j=0}^{\infty} \mathbf{G}_j a_{t-j}
\]

Hence in terms of Green’s function, \( \mathbf{G}_j \) \( A - \lambda_1^{-1} + B - \lambda_2^{-1} \), \( A \mathbf{G}_j \) \( \{1 (\lambda_1^{-1} - 1)\}^{-1} \), and \( B \mathbf{G}_j \) \( \{1 (\lambda_2^{-1} - 1)\}^{-1} \) when the roots \( \lambda_1 \) and \( \lambda_2 \) are real numbers. If the roots satisfy \( |\lambda_1| < 1 \), and \( |\lambda_2| < 1 \) the solution is stable, since the effect of past errors \( a_{t-j} \) is weighted by a smaller and smaller number as \( j \) increases. A plot of \( \mathbf{G}_j \) against \( j \) will show a sum of two declining curves, representing the dynamic memory of the system.

### Complex Number Solution of the Stochastic AR(2) Difference Equation:

\[
y_t \mathbf{A} \mathbf{d} - \mathbf{L} + \frac{1}{\mathbf{L}} \mathbf{L}^2 + \mathbf{d}_a + \mathbf{B} \mathbf{d} - \mathbf{L} + \frac{1}{\mathbf{L}} \mathbf{L}^2 + \mathbf{d}_a \mathbf{G}_j a_{t-j}
\]

Hence in terms of Green’s function, \( \mathbf{G}_j \) \( A - \lambda_1^{-1} + B - \lambda_2^{-1} \), \( A \mathbf{G}_j \) \( \{1 (\lambda_1^{-1} - 1)\}^{-1} \), and \( B \mathbf{G}_j \) \( \{1 (\lambda_2^{-1} - 1)\}^{-1} \) when the roots \( \lambda_1 \) and \( \lambda_2 \) are real numbers. If the roots satisfy \( |\lambda_1| < 1 \), and \( |\lambda_2| < 1 \) the solution is stable, since the effect of past errors \( a_{t-j} \) is weighted by a smaller and smaller number as \( j \) increases. A plot of \( \mathbf{G}_j \) against \( j \) will show a sum of two declining curves, representing the dynamic memory of the system.
When the roots given in (4) of the AR(2) polynomial in (2) are imaginary, because \((9_1^2 + 49_2) < 0\), they are complex conjugate, and the coefficients \(A\) and \(B\) in the expression for \(G_t\) are also complex conjugate. If the reader is familiar with equation (10) below, the elementary discussion between here and (10) may be skipped.

Let us review the use of the imaginary number \(\sqrt{-1} \) for convenience of students. Any complex number can be written as \(C = \text{Re} + \text{Im} \sqrt{-1}\), which is usually represented by a vector \((X, Y)\) in a two dimensional space. The horizontal coordinate \(X\) represents the real part and the vertical coordinate \(Y\) represents the imaginary part. The modulus of a complex number \(|C|\) is its length. A companion of a complex number \(C = X + \sqrt{-1}Y\) is called its complex conjugate.

**Exercise:** What is the length of the vector \(C\)? Verify that \(|C|^2 = X^2 + Y^2\) by the Pythagorean Theorem. Plot \(C\) and \(\text{Re}\) vectors for \(X = 2, Y = 1\) and verify that \(C\) is directly under \(\text{Re}\). Geometrically show the product of two vectors using the parallelogram, and note that it will land on the horizontal axis at the squared distance \(C \cdot C = (X+\sqrt{-1}Y) \cdot (X-\sqrt{-1}Y) = X^2 + Y^2\).

The complex roots, especially in their polar form are useful in Time Series analysis for representing cyclical behavior of the series. The polar coordinates have two parts: the amplitude (\(|\theta|\)) and the phase angle (\(\arg\)), and there is a simple, unique, one-to-one mapping of the coordinates \((X,Y)\) to \((|\theta|, \arg)\), when the angle is restricted to the half open interval \([0,2\pi)\) in radians or to \([0,360^\circ)\) in degrees.

\[ C \cdot C = X \cdot X + \sqrt{-1}Y \cdot \sqrt{-1}Y = X^2 + Y^2 \]

where the notation \(\arg\) represents the argument of the complex number or the phase angle. In the mapping \((X,Y) \rightarrow (|\theta|, \arg)\), we also have \(|\theta| = \sqrt{X^2 + Y^2}\) and \(\arg = \tan^{-1}(Y/X)\), respectively the real and imaginary coordinates of the complex number \(C\). Recall that the sine function is periodic in the sense that it has the same value after every interval of \(360^\circ\). In other words, \(\sin \theta = \sin (\theta + 360^\circ)\). Similarly the cosine function is periodic. Hence, it is intuitively clear that the complex numbers may be useful for representing periodic or cyclical behavior of time series. If we permit \(\theta = 2\pi = 360\) we lose uniqueness of the mapping \((X,Y) \rightarrow (|\theta|, \arg)\), because \(360^\circ\) cannot be distinguished from \(0^\circ\). Losing uniqueness is actually a good thing, because the same complex number \(C\) represents a new cycle starting at \(2\pi\) and ending at \(720\), that is in the interval \(\theta \in [0, 2\pi)\). Yet another cycle has \(\theta \in [2\pi, 4\pi)\). In general, \(\theta \in [k, k+2\pi)\) for an arbitrary integer multiple \(k\), which can be negative.

The trigonometric and exponential functions are also represented as power series:

\[
\sin \theta = \sum_{n=0}^{\infty} \frac{(-1)^n \theta^{2n+1}}{(2n+1)!}, \quad \cos \theta = \sum_{n=0}^{\infty} \frac{(-1)^n \theta^{2n}}{(2n)!}, \quad \exp \theta = \sum_{n=0}^{\infty} \frac{\theta^n}{n!}
\]

**Exercise** Expand \(\exp \theta\) by the above formula, substitute \(\theta = 2\pi\), and use the power series for sine and cosine to write \(\exp \theta = \cos \theta + \sin \theta\). Hence verify that \(C = X + \sqrt{-1}Y\) \(\exp \theta\) with \(X \cos \theta\) and \(Y \sin \theta\) as before. Similarly show that the complex conjugate \(\bar{C} = X - \sqrt{-1}Y\) \(\exp \theta\).

Now we turn to the roots of the quadratic polynomial \((1 - \lambda X + \lambda Y^2)\)

\[ X \cdot (1/2) = 9_1 \lambda L \lambda 9_2 \lambda L^2 \]

which are complex conjugate when \((9_1^2 + 49_2) < 0\) in the present case. From the second equality of (10), note that

\[ X \cdot (1/2) = 9_1 \lambda \lambda 9_2 \lambda \]

are the X and Y coordinates. In defining Y we have moved the negative sign inside the square root sign because \((9_1^2 + 49_2) > 0\), and its square root can be evaluated. Since \(-1\) \(X + Y\) and \(-1\) \(X - Y\) their sum \(-1 - 1\) \(9_1\), and their product \(-1 - 1\) \(9_2\) are still real numbers, as they should be for substitution in the expression for the polynomial. Now evaluate the absolute value of each of the two roots, and note that they are identical to each other. We have the squared amplitude \(9_1^2 \lambda \lambda 9_2^2\) \(X^2 + Y^2\). In terms of the coefficients \(9\) we have the following remarkable result:

\[ l = 9_1^2/4 \lambda (1/4)(9_1^2 + 49_2) = 9_2 \lambda - l \]

**Exercise** What is the phase angle \(\theta\) in terms of \(9_1\) and \(9_2\)? One way to find \(\theta\) might be to use the usual formulas for polar coordinates in (8), write \(\theta = \tan^{-1}(Y/X)\) and use (10a). Since the expression for Y in (10a) is
somewhat complicated, we can avoid using it by considering \( \cos \theta = \frac{\text{adjacent side}}{\text{hypotenuse}} = \frac{X}{X^2+Y^2} \). Hence from (10a) and (10b) we have:

\[
< \mathbf{u} \cos \mathbf{A} \mathbf{i} + \mathbf{g} \mathbf{A} \mathbf{i} = \mathbf{u} \cos \mathbf{A} \mathbf{i} \left( \frac{e^{i \frac{\pi}{2}} - e^{-i \frac{\pi}{2}}}{2i} \right)
\]  

(11)

**Exercise** Plot the above roots (10) in polar coordinate system. First draw the vector \( \mathbf{C}=(X,Y) \) from the origin with \( X=2 \) and \( Y=1 \). Verify that the angle \( \theta \) of the vector \( \mathbf{C} \) with the horizontal axis satisfies the above relation (11).

Now we turn to the solution of the stochastic difference equation. In terms of Green’s function,

\[
G_j \mathbf{u} = \mathbb{A} - \frac{j}{2} + \bar{\mathbb{A}} - \frac{j}{2} \cdot \exp(\mathbb{S}!) \exp(j<m \mathbb{S}>) \cdot \exp(\bar{\mathbb{A}} \mathbf{j} < \mathbb{S})
\]

We have denoted the second coefficient by \( A=\exp(\mathbb{S}) \), because it is a complex conjugate of \( A=\exp(\mathbb{S}) \), having \( \mathbb{S} \) = amplitude and \( m \) =phase angle for the polar form of the coefficients. Thus it can be verified that a plot of \( G_j \) against \( j \) is a damped cosine wave when AR(2) model has complex roots which satisfy \( | \mathbb{S} | < 1 \), and \( | \mathbb{A} | < 1 \). The damping implies eventual dying of the memory, and the cosine wave captures the cyclical behavior. In the complex plane we require the roots to be inside the unit disk.

**Stationarity conditions on the autoregressive parameters** \( \mathbb{S} \) **in the AR(2) Model:**

Stationarity conditions on the autoregressive parameters \( \mathbb{S} \) in the AR(2) Model:

The stationarity condition \( | \mathbb{S} | < 1 \), and \( | \mathbb{A} | < 1 \) and the relations (3) and (10) connecting them with \( \mathbb{S}_1 \) and \( \mathbb{S}_2 \) mean the following.

(a) \( \mathbb{A} 2 \mathbf{u} \mathbf{9} \mathbf{u} \mathbf{2} \) for \( i=1,2 \). (b) \( | \mathbb{S}_1 | < 1 \). (c) \( \mathbb{S}_1 + \mathbb{S}_2 < 1 \). (d) \( \mathbb{A} \mathbf{A} \mathbf{9} \mathbf{9} \mathbf{1} \mathbf{1} \).

**Exercise:** Should we use the absolute values \( | \mathbb{S}_1 | \) and \( | \mathbb{A} | \) in (c) and (d) in (13) above? Draw a triangular figure and indicate the stable regions satisfying the conditions (a) to (d) above.

**Autocovariances for the Stochastic AR(2)**

In the section for AR(1) model the equation ??(5) defined the \# \( \mathbf{u} \) \( E(y_i y_{i-k}) \) \# \( \mathbf{9} \# \), and ??(10) stated the Yule-Walker equations. Let us now find analogous results for the second order case. If the roots \( -\mathbb{A} \) and \( -\bar{\mathbb{A}} \) of the autoregressive polynomial are real, recall that the solution is

\[
y_i \mathbf{u} = \mathbb{G}_j \mathbb{a}_{i-j} \text{ with } \mathbb{G}_j \mathbf{u} = \mathbb{A} - \frac{j}{2} + \bar{\mathbb{A}} \mathbf{u} \left( \begin{array}{c} \mathbb{A} \left( -\frac{j}{2} \right) \end{array} \right), \text{ and } \mathbb{B} \mathbf{u} \left( \begin{array}{c} \mathbb{B} \left( -\frac{j}{2} \right) \end{array} \right)
\]

(14)

\[
\# \mathbf{u} \mathbf{E}(y_i y_{i-k}) \mathbf{u} \mathbf{E} \mathbf{1} = \mathbb{G}_j \mathbf{a}_{i-j} \mathbb{H} \mathbf{u} \left( \begin{array}{c} \mathbb{G}_j \mathbf{a}_{i-j} \mathbf{6} \mathbf{u} \left( \begin{array}{c} \mathbb{G}_j \mathbf{G}_j \mathbf{S}_i \left( -\frac{j}{2} \right) \end{array} \right) \right) \mathbf{5}^2
\]

(15)

where the Kronecker delta mentioned earlier has the defining property that: \( \mathbf{S}_{m} \cdot \mathbf{u} = 1 \) if \( k=m \), and \( \mathbf{S}_{m} \cdot \mathbf{u} = 0 \) otherwise. Hence the only nonzero terms are the ones with the two subscripts of \( \mathbf{S} \) equal to each other; that is only when:

\[
t \mathbb{A} i \mathbf{u} t \mathbb{A} j \mathbf{A} k, \text{ or } \mathbb{A} i \mathbf{u} \mathbb{A} j \mathbf{A} k, \text{ or } i \mathbf{u} j \mathbf{k}
\]

Hence one of the summations can be eliminated if \( i \) is replaced by \( j+k \). Thus

\[
\# \mathbf{u} \mathbf{5}^2 \mathbf{r} \mathbf{G}_{j+k} \mathbf{G}_j
\]

(17)

If \( k=0 \) we have

\[
\# \mathbf{u} \mathbf{5}^2 \mathbf{r} \mathbf{A} - \frac{j}{2} + \mathbf{B} - \frac{j}{2} \mathbf{2}^2
\]

(18)

for \( \mathbf{A} \) and \( \mathbf{B} \) in (14). In general,

\[
\# \mathbf{u} \mathbf{5}^2 \mathbf{r} \mathbf{(A - \frac{j}{2} + B - \frac{j}{2})} (\mathbf{A - \frac{j+k}{1} + B - \frac{j+k}{2}})
\]

(19)

Thus we note that an explicit expression for the autocovariances is available. There is an interesting alternative derivation of \# based on (1) written as \( y_i \mathbf{u} \mathbf{9}_1 y_{i-1} + \mathbf{9}_2 y_{i-2} + \mathbf{a}_i \). Now multiply both sides by \( y_{i-k} \) to yield \( y_{i-k} \)

\[
y_i \mathbf{u} \mathbf{9}_1 y_{i-k} y_{i-1} + \mathbf{9}_2 y_{i-k} y_{i-2} + \mathbf{a}_i y_{i-k} \). Now take expectation of both sides to give
\[ \hat{y}(y_{t+k}) \hat{y} E(y_{t+k}y_{t+k}+9_1 y_{t+2+k} y_{t+k}+9_2 y_{t+3+k} y_{t+k}+9_3 y_{t+4+k} y_{t+k}) \]  
where we use \( a \mathcal{N}(0,5^2) \), \( E(a_t a_m) \hat{y} s_{x_m} s_{5^2}, \) (with \( s_{x_m} \hat{y} 0 \) if \( k \neq m \) and \( s_{x_m} \hat{y} 1 \) if \( k=m \)), and symmetric autocovariances

\[ \hat{y} E(y_{t+k}) \hat{y} E(y_{t+k}y_{t+k}) \hat{y} E(y_{t+k}y_{t+k}) \hat{y} =_{k \neq m} \left\{ \begin{array}{ll} 0 & \text{if } k > l \ 5^2 & \text{if } k \leq l \end{array} \right. \]  

To evaluate the last term of (20) let us consider \( a_{t-l} \) instead of \( a_{t} \) (future reference), and evaluate expectations in terms of Green's function similar to (15).

\[ \hat{y} E(a_t y_{t+k}) \hat{y} E(a_{t-j-k} y_{t+j-k}) \hat{y} =_{0 \leq j \leq k \leq l} \]  

\[ \hat{y} E(a_{t-j} y_{t+k}) \hat{y} E(a_{t-j-k} y_{t+j-k}) \hat{y} =_{0 \leq j \leq k \leq l} \]

Exercise: Verify the last equality of (22). Note that only nonzero expectation \(( \hat{y} 5^2 \) obtains when \( t \hat{y} 1 \hat{y} t \hat{y} j+k \hat{y} \), that is when \( j \hat{y} l \hat{y} k \). Hence check the subscript of last \( G \) in (22). Next, apply (22) to verify that the last term of (20) is zero, i.e., \( E(a_t y_{t+k})=0 \). In (22) let \( l=0 \), and note that \( k \geq l \).

If \( k=0 \), we have

\[ \hat{y} \begin{array}{c} 0 \ 5^2 \end{array} \hat{y} \]

For \( k=1 \),

\[ \hat{y} \begin{array}{c} 0 \ 5^2 \end{array} \hat{y} \]

For \( k=2 \),

\[ \hat{y} \begin{array}{c} 0 \ 5^2 \end{array} \hat{y} \]

The relations (24) and (25) will be used below in deriving Yule-Walker relations of the AR(2) model. For \( k \geq 2 \),

\[ \hat{y} \begin{array}{c} 0 \ 5^2 \end{array} \hat{y} \]

Thus, for \( k \geq 2 \) the autocovariances are seen to satisfy the same difference equation (1) as \( y_t \). This suggests that one can understand the behavior of the original series \( y_t \) by studying the behavior of the autocovariances \#.

For example, the roots \(-1 \) and \(-2 \) of the AR(2) model studied above can then be directly used to solve the difference equation (26) for \# with a solution similar to (14) or (12).

\[ \begin{array}{c} 0 \ 5^2 \end{array} \]

if the roots are real. If the roots are complex conjugate as in (10), we replace \( G_j \) by

\[ \begin{array}{c} 0 \ \exp(j \theta) \end{array} \]

Yule-Walker Relations to get \( 9_1 \) from \( 3_i \):

Let us divide (24) and (25) by \# to write them in terms of correlations instead of covariances.

\[ \begin{array}{c} 0 \ 5^2 \end{array} \]

with \( 3_i=1 \). In matrix notation (29) and (30) become:

\[ \hat{y} \begin{array}{c} 0 \ 5^2 \end{array} \hat{y} \]

which can be solved by the usual methods to obtain \( 9_1 \) and \( 9_2 \hat{A} \) the autoregressive parameters \( \hat{A} \) in terms of the autocorrelation coefficients \( 3_i \) and \( 3_i \). It is somewhat instructive to use the Cramer's rule to obtain the so called Yule-Walker relations.

\[ \begin{array}{c} 0 \ \exp(j \theta) \end{array} \]

(28)
For higher order AR(p) models Yule-Walker relations use ratios of higher order determinants.

6.3 Partial Autocorrelation Function (PACF)

When we consider higher order autoregressive models AR(1), AR(2),..., AR(p) where do we stop? This is similar to the problem of deciding when to stop when adding regressors in a multiple regression, and does not have a clear cut solution. There are statistical tests and other considerations including parsimony. In the methodology popularized by Box and Jenkins (1970) partial autocorrelation functions are used to help choose the p of the AR(p). The usual partial correlation coefficient of y on x₁ holding x₂ fixed is defined by the following formula:

\[ r_{y1,2} = \frac{\sum_{i=1}^{\infty} A_{i,2} - \sum_{i=1}^{\infty} A_{i,1} \sum_{i=1}^{\infty} A_{i,2}}{\sum_{i=1}^{\infty} A_{i,1}^2} \]

\[ \sum_{i=1}^{\infty} A_{i,1} \]

(34)

where \( C_{ij} \) denotes the cofactor [determinant of the submatrix when i-th row and j-th column are stripped out, multiplied by \((-1)^{i+j}\)] of (i,j)th element of a \( 3 \times 3 \) matrix of correlation coefficients of y, x₁ and x₂ respectively. We have

\[ C \begin{pmatrix} 1 & r_{y1} & r_{y2} \\ r_{y1} & 1 & r_{y2} \\ r_{y2} & r_{y2} & 1 \end{pmatrix} \]

(35)

The interpretation of \( r_{y1,2} \) is that it measures the marginal contribution of x₁ when y is regressed on x₁ and x₂. Roughly speaking, one may include x₁ in the regression if \( y_{y1,2} \) is statistically significant. Now \( r_{y2,1} \) is obtained by switching subscripts 1 and 2 in (34) and (35); and measures the marginal contribution of x₂, helpful in deciding whether x₂ should be included. More generally, the partial correlation of y on x₁, while holding both x₂ and x₃ fixed, will be based on a \( 4 \times 4 \) matrix similar to (35) having the last column \( (r_{y3}, r_{x1,2}, r_{x1,3}) \).

In time series analysis, the partial autocorrelation coefficients of order k are denoted by \( h_{kk} \) and defined in terms of regressions as follows.

\[ y_{i} \sim A_{i} + \epsilon_{i} \]

For AR(1), we have \( h_{11} \sim y_{1} \), from (1) of that section above

(36)

\[ y_{i} \sim A_{i} + \epsilon_{i} \]

For AR(2), we have \( h_{22} \sim y_{2} \), from (1) above

(37)

\[ y_{i} \sim A_{i} + \epsilon_{i} \]

For AR(3), we have \( h_{33} \sim y_{3} \)

(38)

If the AR(2) model of (1) above is valid, and we fitted AR(3) of (38) it is obvious that \( h_{33} \sim 0 \), and similarly \( h_{kk} \sim 0 \) for k>3 if we fit an AR(k) model. This fact is used as a diagnostic in determining the order of the AR model, especially emphasized in the Box-Jenkins style of modeling. The numerical estimation of partial autocorrelation coefficients is based on the relationship between \( \hat{h}_{kk} \) and \( \hat{\theta}_{kk} \) discussed above in (33) for the AR(2) case of (37). The AR(3) model of (38) suggests the following interesting relation.

Relation Between Yule-Walker Equations and Partial Autocorrelation Coefficients:

For the AR(1) model the Yule-Walker equation is simply \( \hat{\theta}_{11} \sim \hat{\theta}_{11} \), and we have \( h_{11} \sim \hat{h}_{11} \). For the AR(2) they are given by equations (29) and (30), and the solution for \( h_{22} \) is indicated in (33) as:

\[ h_{22} = \frac{1}{\hat{\theta}_{22}} \int_{\hat{\theta}_{22}}^{\hat{\theta}_{22}} \frac{1}{\hat{\theta}_{22}} - \int_{\hat{\theta}_{22}}^{\hat{\theta}_{22}} \frac{1}{\hat{\theta}_{22}} \]

(39)

\[ \frac{1}{\hat{\theta}_{22}} \sim \hat{\theta}_{22} \sim \hat{\theta}_{22} \]

\[ \frac{1}{\hat{\theta}_{22}} \sim \hat{\theta}_{22} \sim \hat{\theta}_{22} \]

\[ \frac{1}{\hat{\theta}_{22}} \sim \hat{\theta}_{22} \sim \hat{\theta}_{22} \]
where the determinant of the 3 × 3 matrix of autocorrelations in the denominator is directly estimated. The determinant in the numerator is the same as the one in the denominator except for the last column, which is (3, 3, 3). This pattern holds for higher orders also, and the last column for \( u_{kk} \) becomes (3, 3, 3, …, 3) in the numerator. Computer programs routinely calculate the partial autocorrelation coefficients of all orders by this method without running any regressions similar to (38). The partial autocorrelation function \( u_{kk} \) has a sharp cutoff at the order \( p \) when the underlying true model is AR(p), so that all \( u_{kk} \) \( \neq 0 \) for all \( k > p \). Durbin(1960) proposed a recursive method of finding the partial autocorrelations using the Yule-Walker relations.

Significance tests for Partial Autocorrelation Coefficients

Quenouille(1949) shows that the variance of the estimate \( \hat{\theta}_{kk} \) of \( u_{kk} \) from a sample of size \( T \) is given by

\[
\text{Var}[\hat{\theta}_{kk}] \propto \frac{1}{T},
\]

whence the standard error is

\[
\text{SE}[\hat{\theta}_{kk}] \propto \frac{1}{\sqrt{T}} \quad \text{for } k \uparrow p+1,
\]

where the null hypothesis is the AR(p) model (41).

Many computer packages plot the partial autocorrelations along with a band for standard errors based on (41). These plots are an important cornerstone in Box-Jenkins style identification of ARIMA models.

6.5 General Solution to \( \text{ARMA}(n,n-1) \) Stochastic Difference Equations:

In general, consider the n-th order dynamics represented by the \( \text{ARMA}(n,n-1) \) model:

\[
C_{i=1}^n (1 - \hat{\alpha} L) y_i = b_i + C_{j=1}^{n-1} (1 - \hat{\alpha} L) a_i
\]

The solution is:

\[
y_i \propto \sum_{j=1}^p G_j \cdot a_j
\]

where

\[
G_j \propto g^\frac{1}{\hat{\alpha}} + g^\frac{2}{\hat{\alpha}} + \ldots + g^\frac{n}{\hat{\alpha}}
\]

These \( G_j \), \( i=1,2,\ldots,n \) are the n solutions of the familiar n-th order characteristic polynomial in \( z \) (inverse of the lag operator \( L \)) on the autoregressive side involving the AR parameters \( \hat{\alpha} \). For example, in the AR(1) case the polynomial is simply \( (1 - \hat{\alpha} L) \), with the root \( -\hat{\alpha} \). In the ARMA(2,1) case the polynomial is \( (1 - \hat{\alpha} L)(1 - \hat{\alpha} L^2) \) \( \propto (1 - L)(1 - L^2 \hat{\alpha}^2) \), with the two roots \( -1 \) and \( -2 \), which can be complex conjugate if the dynamics involves cyclical behavior. Furthermore, in (3) the coefficients \( g_i \) are explicitly known by the following remarkable formula, Pandit and Wu(1983, p.105).

\[
g_i \propto \frac{(-\hat{\alpha})^{n-3} A_1 \ldots A_{n-2}}{(-A_{n-1} A_{n-2}) \ldots (-A_{n-1} A_{n-2}) \ldots (-A_{n-1} A_{n-2})}
\]

In particular, when \( n=2 \), \( g_1=(-\hat{\alpha})^1/(\hat{\alpha} - 1) \) and \( g_2=(-\hat{\alpha})^1/(\hat{\alpha} - 2) \).

Box-Jenkins style \( \text{ARIMA Model Identification} \) (See Granger Watson survey)

In the initial stages of time series analysis one looks at various plots of the data and other statistics to make a guess regarding the set of models that are most likely to be effective in forecasting the series or otherwise successful for the purpose at hand. The determination of the order \( d \) of differencing is usually the first step in model identification. The \( d \geq 0 \) is often used to transform a nonstationary series into a stationary one. If the correlogram of the original series declines too slowly to zero, differencing makes it go to zero more quickly.

The correlogram is a plot of autocorrelations \( 3_k \) against \( k \).
If a mixed ARMA(p,q) process has a positive AR order, (p>0), then $\varphi_k$ should generally tend to be small eventually. However, if the $p=0$ in AR(p), we have $\varphi_k=0$ for all $k > q+1$, where $q$ is the order of the moving average part.

The partial autocorrelation coefficients $\varphi_{kk}$ are completely symmetric for determination of the MA order.

Box-Jenkins principle of parsimony states that the choice of the model should be such that $p+q$ should be as small as possible.

**Estimation of ARIMA Models**

The analytical expression for the likelihood function is given by Newbold(1974) as follows.

Ansley(1979) provides an exact maximum likelihood estimator.

Ansley and Newbold(1979 unpublished?) report that the method of conditional least squares performs poorly compared to the exact maximum likelihood.

**Diagnostic Checking and Criticism of Estimated Model**

Hosking(1980) suggests Lagrange Multiplier tests for this purpose.

Akaike Information Criterion (AIC) has been used to choose between time series models.