Time Series Lecture Notes for Stat 5362

J. D. Tubbs
Department of Statistical Sciences

Fall Semester 2005

Contents

Chapter 1

Introduction

These lecture notes are for a course using Shumway and Stoffer's book entitled, "Time Series Analysis and its Application". Some useful outside references are;

- Web resources for Shumway and Stoffer's book at

  `http://anson.ucdavis.edu/~shumway/tsa.html`

- Web resources for additional time series data (found at STATLIB at CMU) at

  `http://lib.stat.cmu.edu/datasets`

- My time series web page at

  `http://www3.baylor.edu/~Jack_Tubbs/Courses/Timeseries/`

- Data sets from a book entitled Practical Time Series by Gareth Janacek

  `http://www.uea.ac.uk/~gj/book/data/datalist.html`

- R software at

  `http://www.baylor.edu/statistics/index.php?id=10380`

1.1 Stochastic Processes

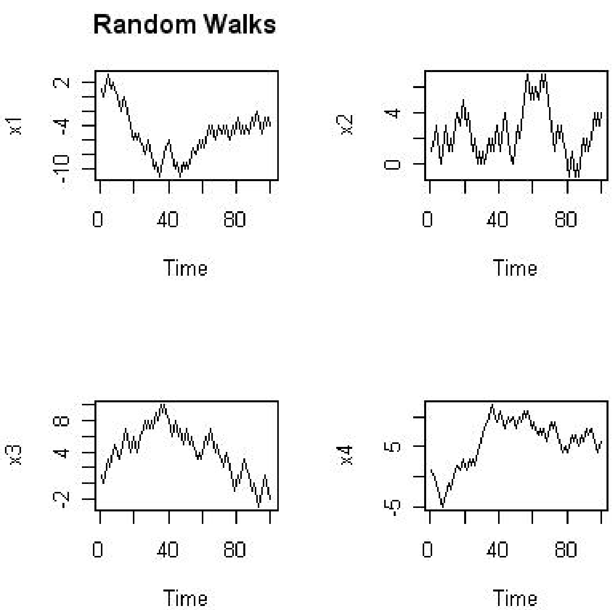
Definition: 1.1 Stochastic Processes *A stochastic process is a family of random variables* $\{X_t, t \in T\}$ *defined on the probability space* $(\Omega, \Im, P)$. T *can be defined as* $\{0 \pm 1, \pm 2, \ldots\}$ *or* $\{0, 1, 2, \ldots\}$ *in which case the stochastic process is said to be discrete with regular spacing. When* T *is given by* $[0, \infty)$ *or* $(-\infty, \infty)$*, the stochastic process is said to continuous.*

Definition: 1.2 Realizations of a Stochastic Process *The functions* $\{X_{.}(\omega), \omega \in \Omega\}$ *on* T *are known as the realizations or sample paths of the process* $\{X_t, t \in T\}$*.*

Example: Binary Process

Let $\{X_t, t = 1, 2, 3, \ldots\}$ be a sequence of random variables where $\Pr[X_t = 1] = \Pr[X_t = -1] = 1/2$. Then the stochastic process given by $\{S_T = \sum_{t=1}^{T} X_t\}$ is said to be a random walk. The following graph contains four realizations of a random walk. The R-code to generate this graph is given by

```
random.walk<-function{
 par(mfrow=c(2,2))
 x1<- cumsum(2*rbinom(100,1,.5) - 1)
 x2<- cumsum(2*rbinom(100,1,.5) - 1)
 x3<- cumsum(2*rbinom(100,1,.5) - 1)
 x4<- cumsum(2*rbinom(100,1,.5) - 1)
 ts.plot(x1);title('Random Walks')
 ts.plot(x2)
 ts.plot(x3)
 ts.plot(x4)
}
```



1.2 Stationarity and Strict Stationarity

Suppose that one has a stochastic process $\{X_t, t \in T\}$ and let $X_{t_1}, X_{t_2}, \ldots, X_{t_n}$ a finite subset of random variables at times $\{t_1, t_2, \ldots, t_n\}$ then the stochastic process or time series is said **Strictly Stationary of order n** if and only if

$$F_{X_{t_1},X_{t_2},\ldots,X_{t_n}}(x_{t_1}, x_{t_2}, \ldots, x_{t_n}) = F_{X_{t_1+k},X_{t_2+k},\ldots,X_{t_n+k}}(x_{t_1}, x_{t_2}, \ldots, x_{t_n})$$

for any choice of $t_1, t_2, \ldots, t_n$ and k. where

$$F_{X_{t_1},X_{t_2},\ldots,X_{t_n}}(x_{t_1}, x_{t_2}, \ldots, x_{t_n}) = \Pr[\{\omega \in \Omega : X_{t_1} \le x_{t_1}, \ldots X_{t_n} \le x_{t_n}\}].$$

Definition: 1.3 Strictly Stationary *The stochastic process $\{X_t, t \in T\}$is said to be strictly stationary if it is strictly stationary of order n for any positive value n.*

Definition: 1.4 Autocovariance Function *If the process $\{X_t, t \in T\}$ is such that $Var(X_t) = E[X_t - E[X_t]]^2 < \infty$ for every t, then the autocovariance function*

$$\gamma_X(t_1, t_2) = Cov(X_{t_1}, X_{t_2}) = E[(X_{t_1} - E[X_{t_1}])(X_{t_2} - E[X_{t_2}])],$$

for any t_1, t_2.

Definition: 1.5 Stationary of order 2 *The process is said to be stationary of order 2 if and only if,*

- $E[\mid X_t \mid^2] < \infty$ *for every* $t \in T$.
- $E[X_t] = \mu$ *for every* $t \in T$.
- $\gamma_X(t_1, t_2) = \gamma_X(t_1 + k, t_2 + k)$ *for every* t_1, t_2 *and* k.

Note: whenever a process is stationary of order 2 then it follows that

- $\gamma_k = \gamma_X(k) = \gamma_X(0, k) = \gamma_X(k, 0) = \gamma(t_1, t_2) = \gamma_X(-k) = \gamma_{-k}$ where $\mid t_1 - t_2 \mid = k$.
- $Var(X_t) = \gamma_X(0) = \gamma_0$
- $\rho_k = \rho_X(k) = Corr(X_{t_1}, X_{t_2}) = \gamma_X(k)/\gamma_X(0) = \rho_X(-k) = \rho_{-k}$. The function $\{\rho_k\}$ is called the **autocorrelation function** for the process $\{X_t\}$.

If $\{X_t\}$ is strictly stationary then it follows that it is stationary provided that $E[\mid X_t \mid^2] < \infty$. The converse is not true.

1.2.1 Properties of Stationary Time Series

If $\{X_t\}$ is stationary of order 2, then it can be shown that;

1. $\gamma_0 = Var(X_t)$ and $\rho_0 = 1$.
2. $\mid \gamma_k \mid \leq \gamma_0$; $\mid \rho_k \mid \leq 1$.
3. $\gamma_k = \gamma_{-k}$ and $\rho_k = \rho_{-k}$.
4. $\sum_{i=1}^{n} \sum_{j=1}^{n} \alpha_i \alpha_j \gamma_{|t_i - t_j|} \geq 0$ and $\sum_{i=1}^{n} \sum_{j=1}^{n} \alpha_i \alpha_j \rho_{|t_i - t_j|} \geq 0$, for any choice of $\alpha_i \neq 0$. That is, the autocovariance and autocorrelation functions, $\{\gamma_k\}, \{\rho_k\}$ are said to be positive semidefinite.

1.2.2 Partial Autocorrelation Function

Let $\{X_t\}$ be a stationary of order 2, then the partial autocorrelation is defined as,

$$P_k = \phi_{kk} = Corr(X_t, X_{t+k} \mid X_{t+1}, X_{t+2}, \ldots, X_{t+k-1}).$$

It can be shown that

$$P_k = \frac{\rho_k - \alpha_1 \rho_{k-1} - \alpha_2 \rho_{k-2} - \ldots - \alpha_{k-1} \rho_1}{1 - \alpha_1 \rho_1 - \alpha_2 \rho_2 - \ldots - \alpha_{k-1} \rho_{k-1}}$$

where

$$\hat{X}_{t+k} = E[X_{t+k} \mid X_{t+1}, X_{t+2}, \ldots, X_{t+k-1}] = \alpha_1 X_{t+k-1} + \alpha_2 X_{t+k-2} + \ldots + \alpha_{k-1} X_{t+1}.$$

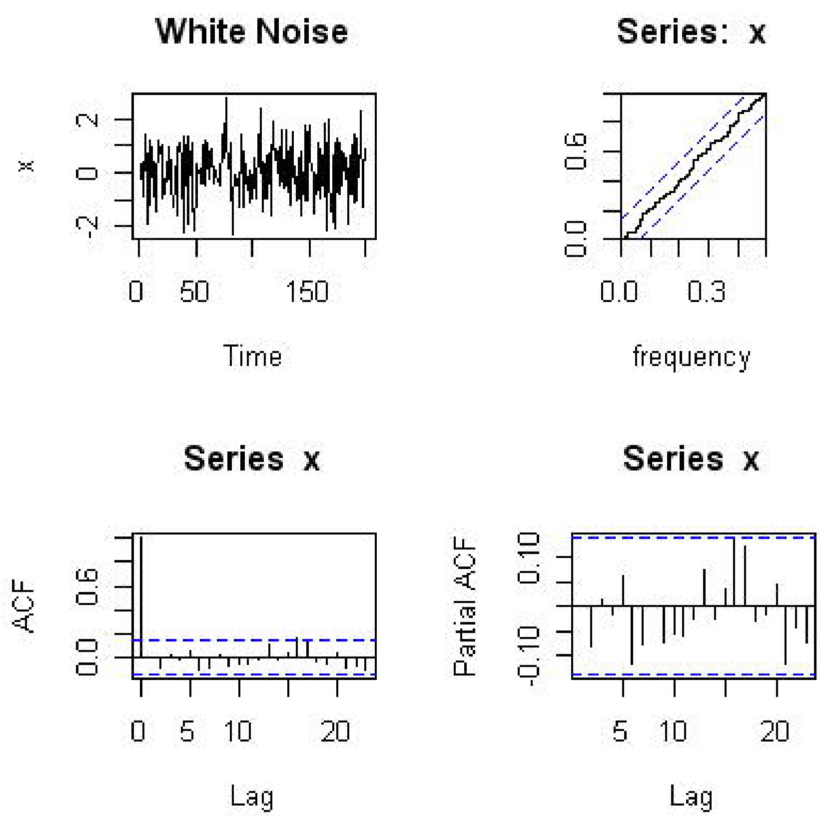
Examples

Definition: 1.6 White Noise *A process $\{X_t\}$ is said to be a white noise process if the random variables X_t are uncorrelated with*

1. $E[X_t] = 0$
2. $Var(X_t) = E[X_t^2] = \sigma_x^2$.

It follows that $\gamma_0 = \sigma_x^2$ and $\gamma_k = 0$ for $k \neq 0$, from which one has $\rho_0 = 1$ and $\rho_k = 0$ for $k \neq 0$ and $\phi_{00} = 1$ and $\phi_{kk} = 0$ for $k \neq 0$.
A graph of white noise is below. This was generated using the R-code

```
x<-rnorm(200)
timeplot(x,c(2,2),'White Noise')
```



Definition: 1.7 Markov Process *A process $\{X_t\}$ is said to be a Markov process if the following holds:*

1. $X_t = \mu$
2. $\dot{X}_t = X_t - \mu$
3. $\dot{X}_t = \phi \dot{X}_{t-1} + a_t$
4. a_t *is white noise*

This process is also called an autoregressive process of order one, AR(1).

The computation of the autocovariance and autocorrelation functions are;

$$\gamma_0 = E[\dot{X}_t^2] = E[(\phi\dot{X}_{t-1} + a_t)^2] = \phi^2\gamma_0 + \sigma_a^2.$$

or

$$Var(X_t) = \gamma_0 = \sigma_a^2/(1-\phi^2).$$

Note: ϕ must be less than one for the variance of X_t to be defined and make any sense. Now consider the first autocovariance given by

$$\begin{aligned}
\gamma_1 &= E[\dot{X}_{t+1}\dot{X}_t] \\
&= E[(\phi\dot{X}_t + a_{t+1})(\phi\dot{X}_{t-1} + a_t)] \\
&= \phi^2 E[\dot{X}_t\dot{X}_{t-1}] + \phi E[\dot{X}_t a_t] + E[a_{t+1}a_t] + \phi E[a_{t+1}\dot{X}_{t-1}] \\
&= \phi^2\gamma_1 + \phi\sigma_a^2.
\end{aligned}$$

or

$$\gamma_1 = \phi\sigma_a^2/(1-\phi^2) = \phi\gamma_0.$$

This process can be extended for any value of $k \geq 1$ where

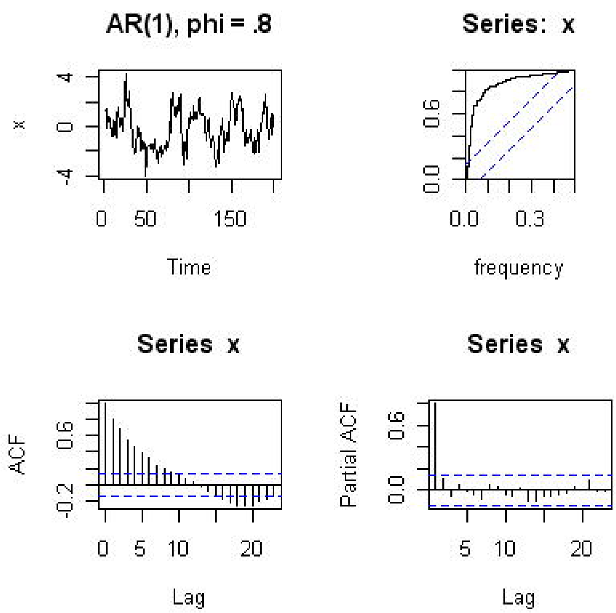
$$\gamma_k = \phi\gamma_{k-1} = \phi^k\gamma_0.$$

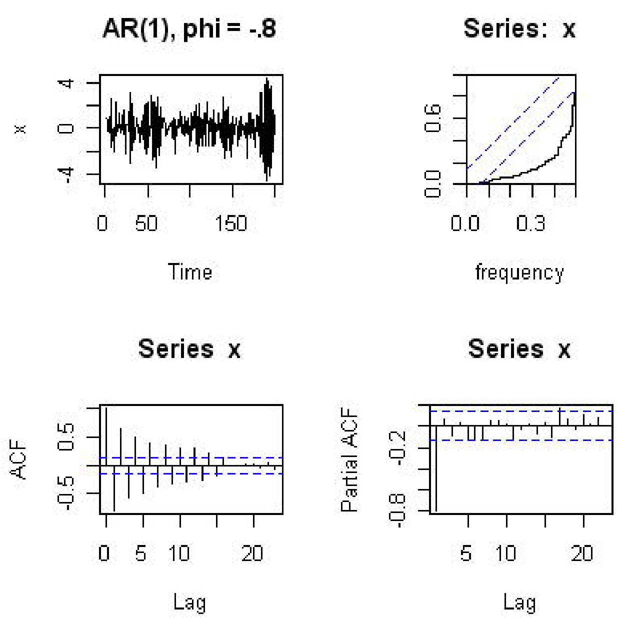
From which we have

$$\rho_k = \phi\rho_{k-1} = \phi^k.$$

It also easily follows that the partial autocorrelation function is $\phi_{11} = \rho_1 = \phi$ and $\phi_{kk} = 0$ for $k \neq 1$. The following graphs are for two AR(1) processes. Each was created in R using the following code:

```
 timeplot<-function(x, span, tname)
 {
    par(mfrow=c(2,2))
    ts.plot(x);title(tname)
    cpgram(x)
    acf(x)
    acf(x, type='partial')
 }
x<-arima.sim(list(order=c(1,0,0), ar=.8), n=200)
timeplot(x,c(2,2),'AR(1), phi = .8')
x<-arima.sim(list(order=c(1,0,0), ar=-.8), n=200)
timeplot(x,c(2,2),'AR(1), phi = -.8')
```





Definition: 1.8 Moving Average of order one – MA(1) *A process $\{X_t\}$ is said to be a Moving Average of order one if the following holds:*

1. $X_t = \mu$
2. $\dot{X}_t = X_t - \mu$
3. $\dot{X}_t = a_t - \theta a_{t-1}$

4. a_t is white noise

From the property of the white noise series $\{a_t\}$, it follows that $\{X_t\}$ is second order stationary for any value of θ and that $E[X_t] = \mu$,

$$Var(X_t) = E[\dot{X}_t^2] = (1 + \theta^2)\sigma_a^2.$$

Furthermore,

$$\gamma_1 = E[\dot{X}_{t+1}\dot{X}_t] = -\theta\sigma_a^2,$$

and

$$\gamma_k = E[\dot{X}_{t+k}\dot{X}_t] = 0$$

for $k > 1$. From which one has,

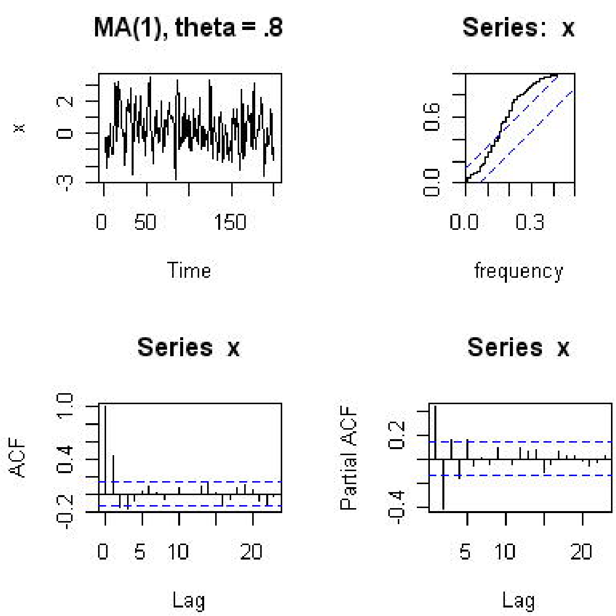
$$\rho_1 = -\theta/(1 + \theta^2)$$

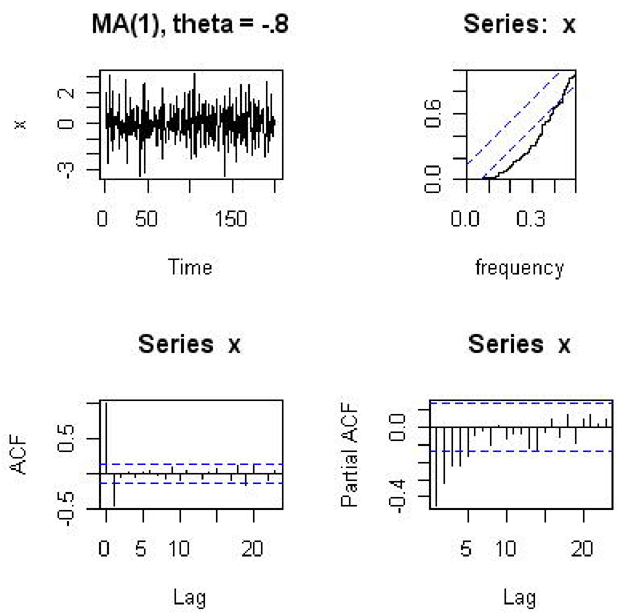
and

$$\rho_k = 0$$

for $k > 1$.
The following graphs are for two MA(1) processes. Each was created in R using the following code: Note: in R the parameter ma = -θ as I have defined above.

```
x<-arima.sim(list(order=c(0,0,1), ma=.8), n=200)
timeplot(x,c(2,2),'MA(1), theta = .8')
x<-arima.sim(list(order=c(0,0,1), ma=-.8), n=200)
timeplot(x,c(2,2),'MA(1), theta = -.8')
```





1.3 Estimation of the Mean, Autocovariance and Autocorrelation

Suppose that one observes a realization of a stationary time series $\{X_t\}$ with mean μ, autocovariance γ_k, and autocorrelation; ρ_k with $x_1, x_2, \ldots, x_n$.

1.3.1 Sample Mean

The estimate of μ is given by,

$$\bar{x} = \frac{1}{n}\sum_{t=1}^{n} x_t.$$

Using the usual properties of the expectation it follows that

$$E[\bar{x}] = \frac{1}{n}\sum_{t=1}^{n} E[x_t] = \frac{n}{n}\mu = \mu.$$

The variance of $\bar{x}$ is given by,

$$\begin{aligned}
Var(\bar{x}) &= \frac{1}{n^2}\sum_{t=1}^{n}\sum_{s=1}^{n} cov(x_t, x_s) \\
&= \frac{\gamma_0}{n}\sum_{t=1}^{n}\sum_{s=1}^{n} \rho_{t-s} \\
&= \frac{\gamma_0}{n}\sum_{k=-(n-1)}^{n-1} (1 - \frac{\mid k \mid}{n})\rho_k
\end{aligned}$$

If one can show that $\lim_{n\to\infty}[\sum_{k=-(n-1)}^{n-1}(1 - \frac{|k|}{n})\rho_k] = 0$ then $\bar{x}$ is said to be a consistent estimate for μ. A time series that has this property is said to be **Ergodic**.

1.3.2 Sample Autocovariance Function

The sample estimate for $\gamma_k = cov(X_t, X_{t+k})$ is given be either

$$\hat{\gamma}_k = \frac{1}{n}\sum_{t=1}^{n-k}(x_t - \bar{x})(x_{t+k} - \bar{x})$$

or

$$\hat{\hat{\gamma}}_k = \frac{1}{(n-k)}\sum_{t=1}^{n-k}(x_t - \bar{x})(x_{t+k} - \bar{x}).$$

It can be shown that both estimates are biased and

$$\begin{aligned} E[\hat{\gamma}_k] &\simeq (1 - k/n)[\gamma_k - Var(\bar{x})] \\ E[\hat{\hat{\gamma}}_k] &\simeq \gamma_k - Var(\bar{x}) \end{aligned}$$

Whenever the process is ergodic, it follows that both estimates are asymptotically unbiased. Usually, $\hat{\hat{\gamma}}_k$ has a smaller bias than $\hat{\gamma}_k$, however, when comparing their mean square error, $\hat{\gamma}_k$ is often smaller than that for $\hat{\hat{\gamma}}_k$, hence many prefer the estimator $\hat{\gamma}_k$.
Bartlett (1946) showed the following when he assume that the time series was Gaussian,

$$Var(\hat{\gamma}_k) \simeq \frac{1}{n}\sum_{j=-\infty}^{\infty}(\gamma_j^2 + \gamma_{j+k}\gamma_{j-k})$$

and

$$Var(\hat{\hat{\gamma}}_k) \simeq \frac{1}{n-k}\sum_{j=-\infty}^{\infty}(\gamma_j^2 + \gamma_{j+k}\gamma_{j-k}).$$

1.3.3 Sample Autocorrelation Function

The estimate for ρ_h is given by $\hat{\rho}_k = \frac{\hat{\gamma}_k}{\hat{\gamma}_0}$. From Bartlett one has the following:

$$cov(\hat{\rho}_k, \hat{\rho}_{k+j}) \simeq \frac{1}{n}\sum_{i=-\infty}^{\infty}(\rho_i\rho_{i+j} + \rho_{i-k}\rho_{i+j+k} - 2\rho_k\rho_i\rho_{i-k-j} - 2\rho_{k+j}\rho_i\rho_{i-k} + 2\rho_k\rho_{j+k}\rho_i^2).$$

When n is large, Bartlett shown that

$$Var(\hat{\rho}_k) \simeq \frac{1}{n}\sum_{i=-\infty}^{\infty}(\rho_i^2 + \rho_{i-k}\rho_{i+k} - 4\rho_k\rho_i\rho_{i-k} + 2\rho_k^2\rho_i^2).$$

For processes in which $\rho_k = 0$ for $k > m$ Bartlett showed that

$$Var(\hat{\rho}_k) \simeq \frac{1}{n}(1 + 2\rho_1^2 + 2\rho_2^2 + \ldots 2\rho_m^2)$$

and

$$\sqrt{Var(\hat{\rho}_k)} = s_{\hat{\rho}_k} \simeq \sqrt{\frac{1}{n}(1 + 2\rho_1^2 + 2\rho_2^2 + \ldots 2\rho_m^2)}.$$

Thus, when one has a white noise series it follows that

$$s_{\hat{\rho}_k} \simeq \sqrt{\frac{1}{n}}.$$

1.3.4 Sample Partial Autocorrelation Function

Durbin (1960) derived the following:

$$\hat{\phi}_{k+1,k+1} = \frac{\hat{\rho}_{k+1} - \sum_{j=1}^{k} \hat{\phi}_{kj}\hat{\rho}_{k+1-j}}{1 - \sum_{j=1}^{k} \hat{\phi}_{kj}\hat{\rho}_{j}}$$

and

$$\hat{\phi}_{k+1,j} = \hat{\phi}_{kj} - \hat{\phi}_{k+1,k+1}\hat{\phi}_{k,k+i-j}, \quad j = 1, 2, \ldots, k.$$

If one assumes that the process is white noise then

$$Var(\hat{\phi}_{kk}) \simeq \frac{1}{n}$$

and

$$s_{\hat{\phi}_{kk}} \simeq \frac{1}{\sqrt{n}}.$$

Chapter 2

Stationary Times Series Models

Before considering the models discussed in this chapter, we need to add some notation and discuss the notation of difference equations.

2.1 Added Notation and Difference Equations

2.1.1 Backshift Operator and Differencing

In order to read the time series literature you are going to need to understand the **backward operator, B** given by

$$B^d a_t = a_{t-d}.$$

Differencing is defined as,

$$\nabla x_t = (1 - B)x_t = x_t - x_{t-1}.$$

Example 1

The process given by

$$X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2},$$

can be written as,

$$X_t = \phi_1 B X_t + \phi_2 B^2 X_t + a_t - \theta_1 B a_t - \theta_2 B^2 a_t$$

or

$$(1 - \phi_1 B - \phi_2 B^2)X_t = (1 - \theta_1 B - \theta_2 B^2)a_t,$$

or more generally as,

$$\Phi_2(B)X_t = \Theta_2(B)a_t,$$

where $\Phi_2(B) = (1 - \phi_1 B - \phi_2 B^2)$ and $\Theta_2(B) = (1 - \theta_1 B - \theta_2 B^2)$.

Example 2

$$\nabla^2 x_t = (1 - B)^2 x_t = (1 - 2B + B^2)x_t = x_t - 2x_{t-1} + x_{t-2}.$$

The above differencing operation is used in the Box and Jenkins models for creating stationary models from non stationary models. Another use of differencing is modelling long memory processes which are commonly found when modelling environmental time series. For $-.5 < d < .5$, defined

$$\nabla^d x_t = (1-B)^d x_t = \sum_{j=1}^{\infty} \pi_j B^j$$

where $\pi_0 = 1$ and

$$\pi_{j+1} = \frac{(j-d)\pi_j}{(j+1)}.$$

2.1.2 Linear Difference Equation

The material found in this section is similar to that found in Differential Equations. This material is needed because we have assumed that the time series are discrete time stochastic processes. The n^{th} order Linear Difference equation is given by,

$$C_n(B)Z_t = e_t,$$

where $C_n(B) = (c_0 + c_1B + c_2B^2 + \ldots + c_nB^n)$. The above difference equation is said to homogenous if $e_t = 0$. The following lemmas are presented without proofs.

Lemma: 2.1 *If $Z_t^{(1)}$ and $Z_t^{(2)}$ are solution to the n^{th} order linear difference equation $C_n(B)Z_t = e_t$, then $b_1Z_t^{(1)} + b_2Z_t^{(2)}$ is also a solution for arbitrary choice of b_1 and b_2.*

Lemma: 2.2 *If $Z_t^{(P)}$ is a solution to the n^{th} order linear difference equation $C_n(B)Z_t = e_t$ (called a particular solution) and $Z_t^{(H)}$ is a solution to the n^{th} order homogenous linear difference equation $C_n(B)Z_t = 0$, then $Z_t^{(H)} + Z_t^{(P)}$ is also a solution to the n^{th} order linear difference equation $C_n(B)Z_t = e_t$.*

Lemma: 2.3 *Let $Z_t = bt^j$ and m be a nonnegative integer where b is any constant and j is a fixed nonnegative integer less than m, then $(1-B)^m Z_t = 0$*

Lemma: 2.4 *Let $(1-RB)^m = 0$ where $R \neq 0$ and m is a nonnegative integer, and $Z_t = R^t t^j$ where j is a nonnegative integer less than m. Then $(1-RB)^m Z_t = 0$.*

Theorem 2.1 *Let $C_n(B)Z_t = 0$ be a given homogenous linear difference equation where $C_n(B) = 1 + c_1B + c_2B^2 + \ldots + c_nB^n)$. If $C(B) = \prod_{i=1}^{N}(1-R_iB)^{m_i}$ where $\sum_{i=1}^{N} m_i = n$ and $B_i = R_i^{-1}$ are the roots of multiplicity m_i of $C(B) = 0$, then $Z_t = \sum_{i=1}^{N} R_i^t \sum_{j=0}^{m_i-1} b_{ij}t^j$. If $m_i = 1$ for all i then $Z_t = \sum_{i=1}^{n} b_iR_i^t$.*

2.2 General Stationary Time Series

The general linear process is given by,

$$\dot{X}_t = \Psi(B)a_t$$

where $\Psi(B) = \sum_{i=1}^{\infty} \psi_i B^i$, a_t is white noise with variance σ_a^2, $\psi_0 = 1$, and $\dot{X}_t = X_t - \mu$. The above process is second order stationary if and only if $\sum_{i=1}^{\infty} \psi_i^2 < \infty$. Note: Shumway and Stoffer use the term causal instead of stationary. If the process is stationary then it follows that $\lim_{k\to\infty} \mid \psi_k \mid = 0$. It can be shown that,

$$Var(X_t) = \sigma_a^2 (\sum_{i=1}^{\infty} \psi_i^2)$$

and

$$E(a_t \dot{X}_{t-j}) = \sigma_a^2$$

when $j = 0$ and equals 0 when $j > 0$. Hence, it follows that,

$$\gamma_k = \sigma_a^2 (\sum_{i=1}^{\infty} \psi_i \psi_{i+k})$$

and

$$\rho_k = \frac{\sum_{i=1}^{\infty} \psi_i \psi_{i+k}}{\sum_{i=1}^{\infty} \psi_i^2}.$$

The above process is sometimes referred to as an infinite order moving average, MA(∞), as Wold's representation. There is another expression for the process which is called an infinite order autoregressive process, AR(∞), given by

$$\pi(B)\dot{X}_t = (\sum_{j=0}^{\infty} \pi_j B^j)\dot{X}_t = a_t$$

or

$$\dot{X}_t = -(\sum_{j=1}^{\infty} \pi_j B^j)\dot{X}_t + a_t$$

where $\pi_0 = 1$. This process is said to be invertible if and only if $\sum_{j=1}^{\infty} \mid \pi_j \mid < \infty \quad (\mid \pi_k \mid \to 0 \text{ as } k \to \infty)$. Using the above two representations one has,

$$\dot{X}_t = \Psi(B) a_t,$$

and

$$\pi(B)\dot{X}_t = a_t,$$

or

$$\dot{X}_t = \Psi(B)\pi(B)\dot{X}_t.$$

Which implies for a process to be both stationary and invertible that $\Psi(B)\pi(B) = I$. This means that the roots to the characteristic equations given by

$$\Psi(B) = 0 \quad and \quad \pi(B) = 0$$

should lie outside the unit ball defined by $\mid B \mid = 1$.

Definition: 2.1 Autocovariance Generating Function *The autocovariance generating function is given by*

$$\gamma(B) = \sum_{k=-\infty}^{\infty} \gamma_k B^k,$$

where $Var(X_t) = \gamma_0$ *and* $\gamma_k = cov(X_t, X_{t+k}) = cov(X_t, X_{t-k})$.

Using the stationary of X_t and $\gamma_k = \sigma_a^2 \sum_{i=0}^{\infty} \psi_i \psi_{i+k}$, one has

$$\begin{aligned}
\gamma(B) &= \sigma_a^2 \sum_{k=-\infty}^{\infty} \sum_{i=0}^{\infty} \psi_i \psi_{i+k} B^k \\
&= \sigma_a^2 \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} \psi_i \psi_j B^{j-i} \\
&= \sigma_a^2 \sum_{j=0}^{\infty} \psi_j B^j \sum_{i=0}^{\infty} \psi_i B^{-i} \\
&= \sigma_a^2 \Psi(B)\Psi(B^{-1}).
\end{aligned}$$

2.2.1 Power Spectrum for Stationary Time Series

Let $\{X_t\}$ be a second order stationary process ($\sum_k^\infty \mid \gamma_k \mid < \infty$) then its Fourier transform given by,

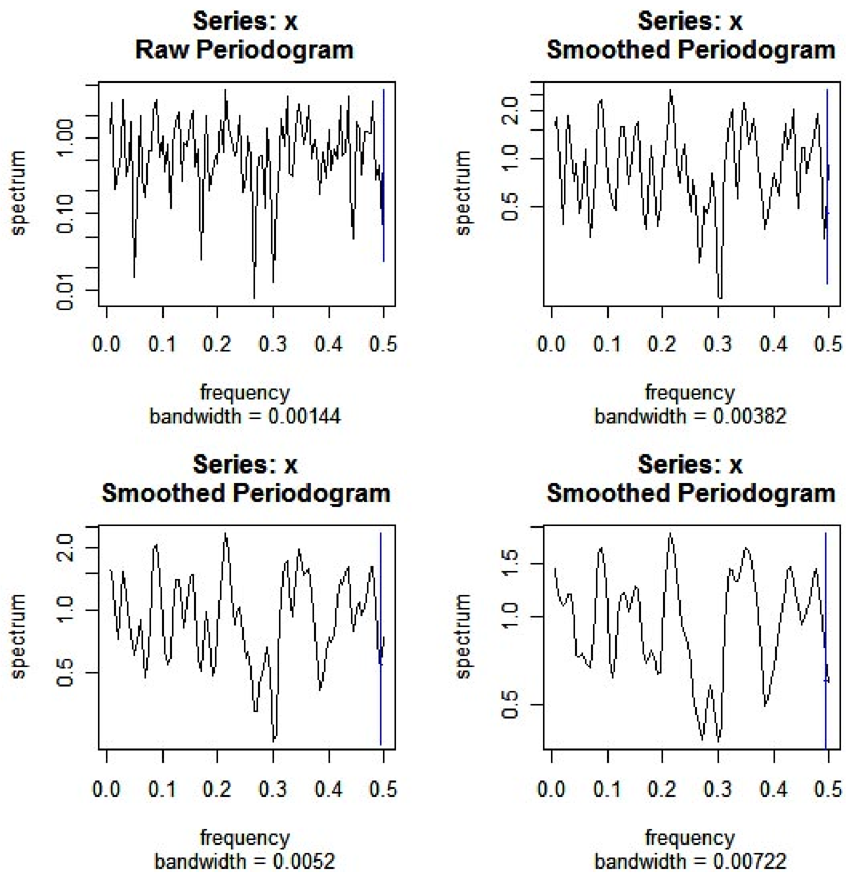
$$f(\omega) = \frac{1}{2\pi}\sum_{k=-\infty}^{\infty}\gamma_k e^{-i\omega k} = \frac{1}{2\pi}\gamma_0 + \frac{1}{\pi}\sum_{k=1}^{\infty}\gamma_k cos(\omega k),$$

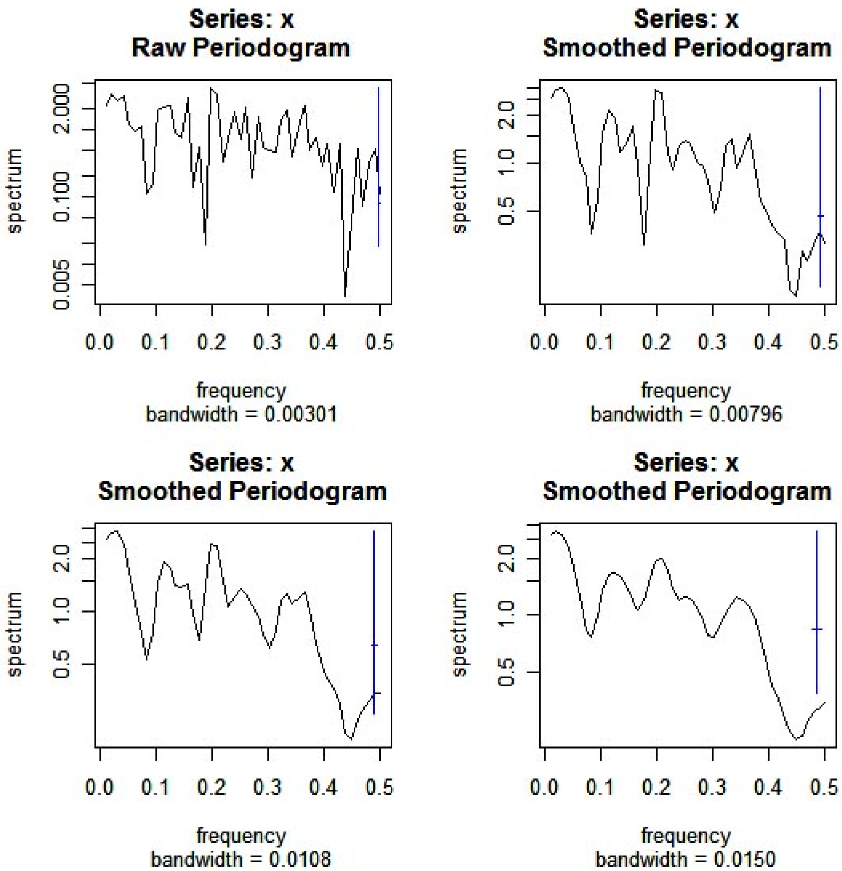
for $-\pi \leq \omega \leq \pi$. Using the first equality with the autocovariance generating function one has,

$$f(\omega) = \frac{1}{2\pi}\gamma(e^{-i\omega k}) = \frac{\sigma_a^2}{2\pi}\Psi(e^{-i\omega})\Psi(e^{i\omega}) = \frac{\sigma_a^2}{2\pi} \mid \Psi(e^{-i\omega}) \mid^2 .$$

Some examples of spectrum for stationary time series are given below. The first is for a white noise series and the second for an unknown series type. I have included the R program with four different levels of "smoothed" spectrum.

```
  specplot <- function(x,ntitle)
{
   par(mfrow=c(2,2))
   spectrum(x);title(ntitle)
   spectrum(x, spans=3)
   spectrum(x, spans=c(3,3))
   spectrum(x, spans=c(3,5))
 }
  wn<-rnorm(200)
  specplot(wn,'')
  specplot(x,'')
```


Series: x
Raw Periodogram
spectrum
0.01
0.10
1.00
0.0
0.1
0.2
0.3
0.4
0.5
frequency
bandwidth = 0.00144
Series: x
Smoothed Periodogram
spectrum
0.5
1.0
2.0
0.0
0.1
0.2
0.3
0.4
0.5
frequency
bandwidth = 0.00382
Series: x
Smoothed Periodogram
spectrum
0.5
1.0
2.0
0.0
0.1
0.2
0.3
0.4
0.5
frequency
bandwidth = 0.0052
Series: x
Smoothed Periodogram
spectrum
0.5
1.0
1.5
0.0
0.1
0.2
0.3
0.4
0.5
frequency
bandwidth = 0.00722



2.3 Autoregressive Processes

The autoregressive process of order p is given by,

$$\Phi_p(B)\dot{X}_t = a_t$$

where $\Phi_p(B) = (1-\phi_1 B-\phi_2 B^2-\ldots-\phi_p B^p)$, a_t is white noise with variance σ_a^2 and $\dot{X}_t = X_t-\mu$. Rewriting this model as,

$$\dot{X}_t = \phi_1\dot{X}_{t-1} + \phi_2\dot{X}_{t-2} + \ldots + \phi_p\dot{X}_{t-p} + a_t.$$

One can easily show that

$$\rho_k = \phi_1\rho_{k-1} + \phi_2\rho_{k-2} + \ldots + \phi_p\rho_{k-p}$$

or

$$\Phi_p(B)\rho_k = 0.$$

This last equation is called the characteristic equation. Note: it is also a p^{th} order homogenous difference equation. Suppose that one expands this equation by letting $k = p$, then one has,

$$\begin{aligned}
\rho_1 &= \phi_1 + \phi_2\rho_1 + \ldots + \phi_p\rho_{p-1} \\
\rho_2 &= \phi_1\rho_1 + \phi_2 + \ldots + \phi_p\rho_{p-2}
\end{aligned}$$

$$\vdots$$
$$\rho_p = \phi_1\rho_{p-1} + \phi_2\rho_{p-2} + \ldots + \phi_p$$

This can be written as,

$$\vec{\rho} = \vec{P}\vec{\phi}.$$

where $\vec{\rho}$ is

$$\begin{pmatrix} \rho_1 \\ \rho_2 \\ \vdots \\ \rho_p \end{pmatrix}$$

and $\vec{\phi}$ is

$$\begin{pmatrix} \phi_1 \\ \phi_2 \\ \vdots \\ \phi_p \end{pmatrix}$$

and $\vec{P}$ is

$$\begin{pmatrix} 1 & \rho_1 & \cdots & \rho_{p-1} \\ \rho_1 & 1 & \cdots & \rho_{p-2} \\ \vdots & \cdots & \cdots & \vdots \\ \rho_{p-1} & \rho_{p-2} & \cdots & 1 \end{pmatrix}$$

From the above one has,

$$\vec{\phi} = \vec{P}^{-1}\vec{\rho}$$

whenever $\vec{P}$ is nonsingular. This equation is called the Yule - Walker equation that is used for finding initial estimates of the autoregressive parameters, $\vec{\phi}$.

Partial Autocorrelations

Using the idea given by Shumway and Stoffer concerning partial correlations it follows from the model that X_t is dependent upon $X_{t-1}, X_{t-2}, \ldots, X_{t-p}$ and that once these p variables are partialed out the remaining variables are uncorrelated with X_t, hence only p partial autocorrelations are potentially nonzero.

2.3.1 AR(1) or Markov Process

Using the notation defined above the Markov or AR(1) process is,

$$\Phi_1(B)\dot{X}_t = (1 - \phi B)\dot{X}_t = a_t.$$

This process will be stationary if and only if the roots of $\Phi_1(B) = 0$ lie outside the unit ball which in this case means that $B^{-1} = \phi$ must satisfy $\mid \phi \mid < 1$. In which case one has

$$\Psi(B) = (1 - \phi B)^{-1} = \sum_{j=0}^{\infty} \phi^j B^j,$$

or $\psi_j = \phi^j$ where $\psi_0 = \phi^0 = 1$.

Autocovariance and Autocorrelation

The computation of the autocovariance and autocorrelation functions are;

$$\gamma_0 = E[\dot{X}_t^2] = E[(\phi\dot{X}_{t-1} + a_t)^2] = \phi^2\gamma_0 + \sigma_a^2.$$

or

$$Var(X_t) = \gamma_0 = \sigma_a^2/(1-\phi^2).$$

Note: ϕ must be less than one for the variance of X_t to be defined and make any sense. Now consider the first autocovariance given by

$$\begin{aligned}
\gamma_1 &= E[\dot{X}_{t+1}\dot{X}_t] \\
&= E[(\phi\dot{X}_t + a_{t+1})(\phi\dot{X}_{t-1} + a_t)] \\
&= \phi^2 E[\dot{X}_t\dot{X}_{t-1}] + \phi E[\dot{X}_t a_t] + E[a_{t+1}a_t] + \phi E[a_{t+1}\dot{X}_{t-1}] \\
&= \phi^2\gamma_1 + \phi\sigma_a^2.
\end{aligned}$$

or

$$\gamma_1 = \phi\sigma_a^2/(1-\phi^2) = \phi\gamma_0.$$

This process can be extended for any value of $k \geq 1$ where

$$\gamma_k = \phi\gamma_{k-1} = \phi^k\gamma_0.$$

From which we have

$$\rho_k = \phi\rho_{k-1} = \phi^k.$$

It also easily follows that the partial autocorrelation function is $\phi_{11} = \rho_1 = \phi$ and $\phi_{kk} = 0$ for $k \neq 1$.

Spectrum

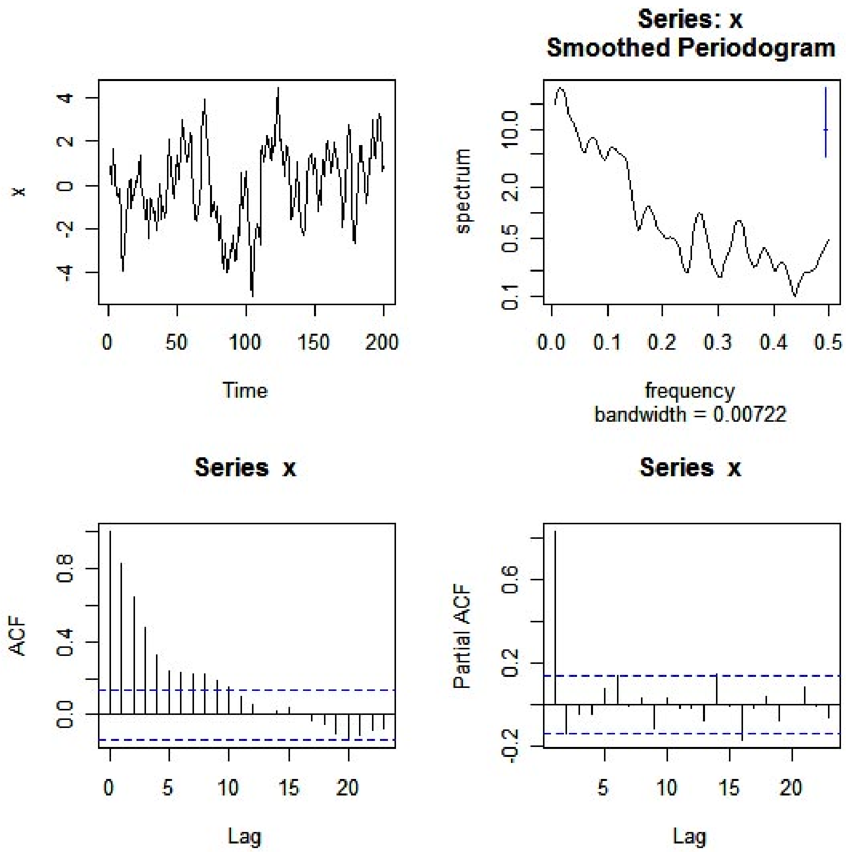
The spectrum is given by,

$$f(\omega) = \frac{2\sigma_a^2}{\mid 1 - \phi_1 e^{-i\omega} \mid^2} = \frac{2\sigma_a^2}{1 + \phi^2 - 2\phi_1\cos(\omega)}.$$

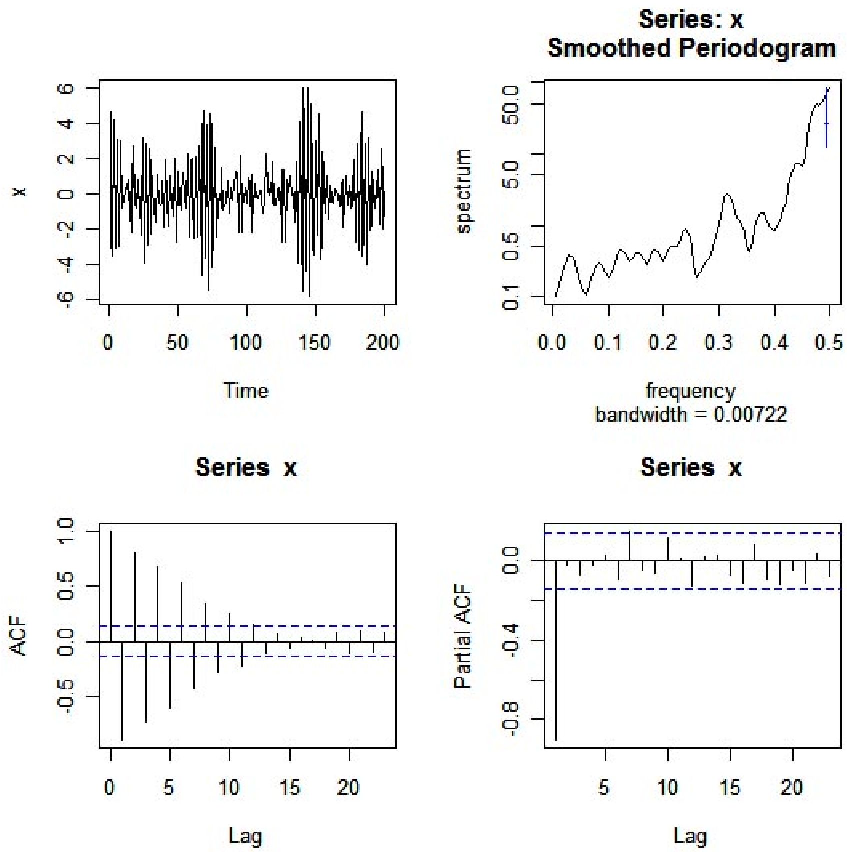
AR(1) Examples

The following graphs are for three AR(1) processes. Each was created in R using the following code:

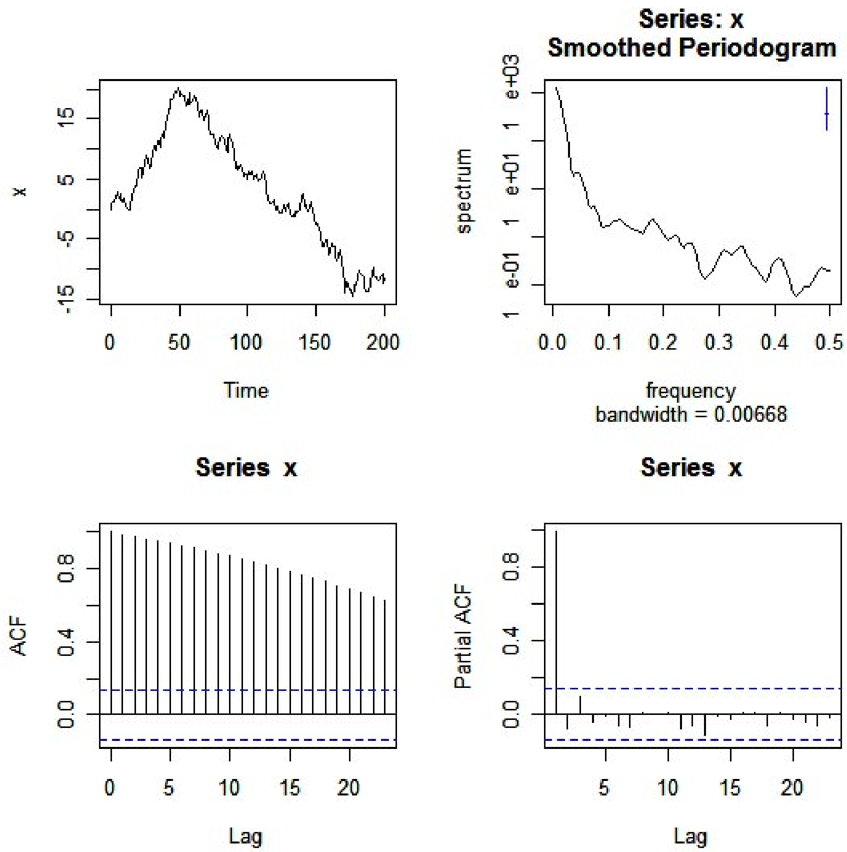
```
x<-arima.sim(list(order=c(1,0,0), ar=.9), n=200)
tsplot(x,c(3,3))
x<-arima.sim(list(order=c(1,0,0), ar=-.9), n=200)
tsplot(x,c(3,3))
z<-arima.sim(list(order=c(0,1,0)), n=200)
tsplot(z,c(3,3))
```



Series: x
Smoothed Periodogram
x
Time
spectrum
frequency
bandwidth = 0.00722
Series x
ACF
Lag
Series x
Partial ACF
Lag



The following plot was produced when $\phi = 1$. This process is not stationary and it often called a random walk.



2.3.2 AR(2)

This model is discussed as it is useful in indicating some properties for autoregressive processes that are not evident with the Markov process. The AR(2) is given by,

$$\Phi_2(B)\dot{X}_t = a_t$$

where $\Phi_2(B) = (1 - \phi_1 B - \phi_2 B^2)$. Using the relationship $\Phi_2(B)\rho_k = 0$ one has

$$\rho_k = \phi_1 \rho_{k-1} + \phi_2 \rho_{k-2}.$$

From which one can derive,

$$\rho_1 = \frac{\phi_1}{(1 - \phi_2)}$$

$$\rho_2 = \frac{\phi_1^2}{(1 - \phi_2)} = \frac{\phi_1^2 + \phi_2 - \phi_1}{1 - \phi_2}.$$

From these equation one can see that $\phi_2 < 1$. Using the Yule - Walker equations one can derive

$$\phi_1 = \frac{\rho_1(1 - \rho2)}{(1 - \rho_1^2)}$$

and

$$\phi_2 = \frac{\rho_2^2 - \rho_1}{(1-\rho_1^2)}.$$

In considering the characteristic equation $\phi_2(B) = 0$ one observes that the roots are given by,

$$\frac{-\phi_1 \pm \sqrt{\phi_1^2 + 4\phi_2}}{2\phi_2}$$

from which it follows that these conditions must hold in order for the AR(2) process to be stationary,

$$\begin{aligned} \phi_2 + \phi_1 &< 1, \\ \phi_2 - \phi_1 &< 1, \\ \mid \phi_2 \mid &< 1. \end{aligned}$$

Using the properties of the second order linear difference equations for $\Phi_2(B)\rho_k = 0$, it follows that

$$\rho_k = b_1(\frac{\phi_1 + \sqrt{\phi_1^2 + 4\phi_2}}{2\phi_2})^k + b_2(\frac{\phi_1 - \sqrt{\phi_1^2 + 4\phi_2}}{2\phi_2})^k,$$

where constant b_1 and b_2 satisfy the initial conditions. This leads to the discussion given by our authors at the top of page 105 in the text.

Variance

The variance of X_t is given by,

$$Var(X_t) = \frac{\sigma_a^2}{1 - \rho_1\phi_1 - \rho_2\phi_2} = \frac{\sigma_a^2(1-\phi_2)}{(1+\phi_2)((1-\phi_2)^2 - \phi_1^2)}.$$

Spectrum

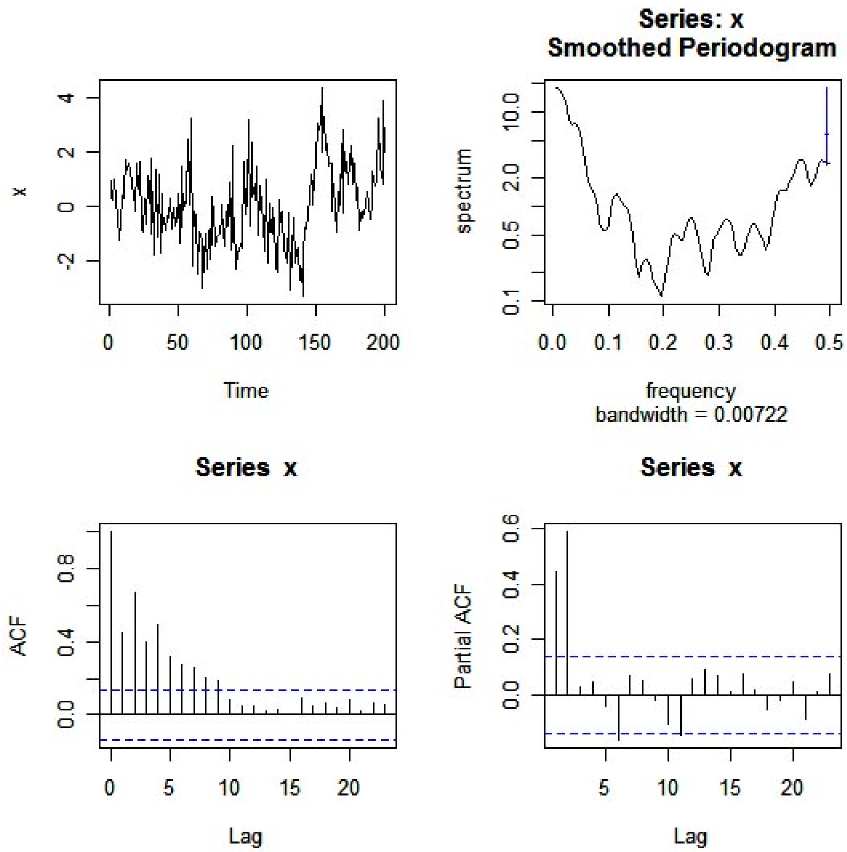
The Spectrum is given by,

$$\begin{aligned} f(\omega) &= \frac{\sigma_a^2}{\mid 1 - \phi_1 e^{-i\omega} - \phi_2 e^{-2i\omega} \mid^2} \\ &= \frac{\sigma_a^2}{1 + \phi_1^2 + \phi_2^2 - 2\phi_1(1-\phi_2)\cos(\omega) - 2\phi_2\cos(2\omega)} \end{aligned}$$

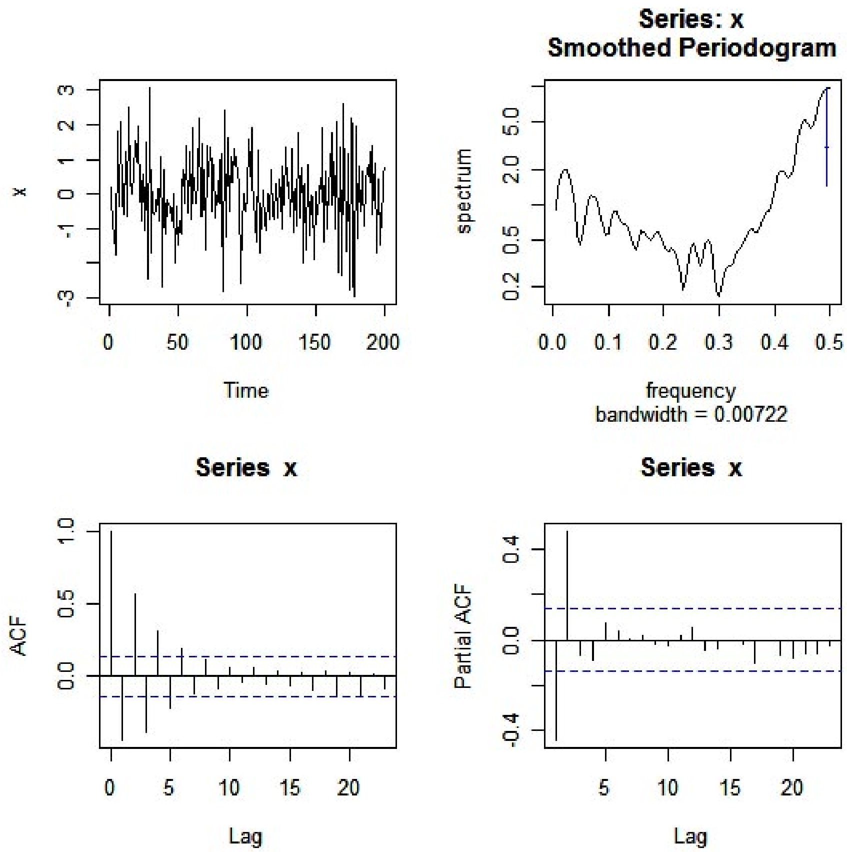
AR(2) Examples

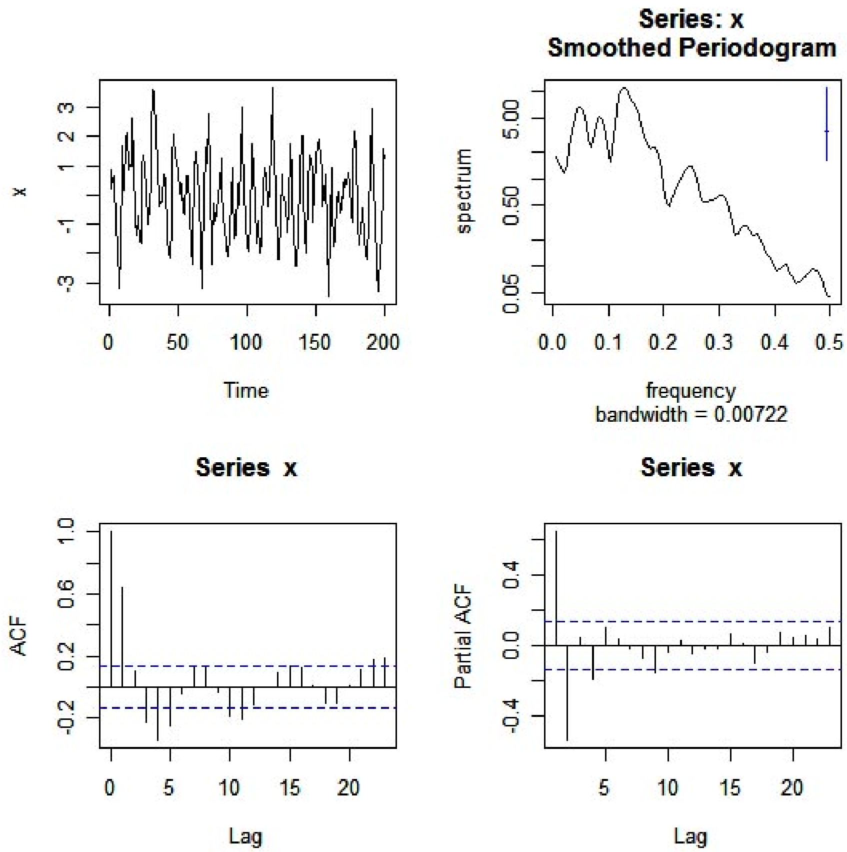
Some examples of AR(2) processes are given below. The R code is,

```
x<-arima.sim(list(order=c(2,0,0), ar=c(.25,.5)), n=200)
tsplot(x,span=c(3,5))
y<-arima.sim(list(order=c(2,0,0), ar=c(-.25,.5)), n=200)
tsplot(y,span=c(3,5))
z<-arima.sim(list(order=c(2,0,0), ar=c(1,-.5)), n=200)
tsplot(z,span=c(3,5))
```



Series: x
Smoothed Periodogram
x
Time
spectrum
frequency
bandwidth = 0.00722
Series x
ACF
Lag
Series x
Partial ACF
Lag


x
3
2
1
0
-1
-3
0
50
100
150
200
Time
Series: x
Smoothed Periodogram
spectrum
5.0
2.0
0.5
0.2
0.0
0.1
0.2
0.3
0.4
0.5
frequency
bandwidth = 0.00722
Series x
ACF
1.0
0.5
0.0
0
5
10
15
20
Lag
Series x
Partial ACF
0.4
0.0
-0.4
5
10
15
20
Lag



2.4 Moving Average Processes

X_t is said to be a moving average of order q if,

$$X_t = \mu - \theta_1 a_{t-1} - \theta_2 a_{t-2} - \ldots - \theta_q a_{t-q} + a_t = \mu - (\sum_{j=0}^{q} \theta_j B^j) a_t,$$

where $\theta_0 = -1$ and $E[X_t] = \mu$. This model can be written as,

$$X_t = \mu + \Theta_q(B) a_t,$$

where $\Theta_q(B) = -(\sum_{j=0}^{q} \theta_j B^j)$. By comparing this model with

$$X_t = \Psi(B) a_t,$$

it follows that $\psi_j = -\theta_j$ for $j = 1, 2, \ldots, q$, and $\psi_j = 0$ for $j > q$. In which case, $\sum_{j=0}^{\infty} \psi_j^2 < \infty$ for finite values of θ_j, $j = 1, 2, \ldots, q$, hence the moving average of order q is always stationary. Using the properties of the general linear process, i.e.,

$$Var(X_t) = \sigma_a^2 (\sum_{i=1}^{\infty} \psi_i^2)$$

and

$$E(a_t \dot{X}_{t-j}) = \sigma_a^2$$

when $j = 0$ and equals 0 when $j > 0$. And

$$\gamma_k = \sigma_a^2 (\sum_{i=1}^{\infty} \psi_i \psi_{i+k})$$

$$\rho_k = \frac{\sum_{i=1}^{\infty} \psi_i \psi_{i+k}}{\sum_{i=1}^{\infty} \psi_i^2}.$$

Variance, Autocovariance and Autocorrelation for MA(q)

It follows that,

$$Var(X_t) = \sigma_a^2 [1 + (\sum_{i=1}^{q} \theta_i^2)]$$

and

$$\gamma_k = \sigma_a^2 (-\theta_k + \sum_{i=1}^{q-k} \theta_i \theta_{i+k})$$

for $k = 1, 2, \ldots, q$ with $\gamma_k = 0$ for $k > q$, and

$$\rho_k = \frac{(-\theta_k + \sum_{i=1}^{q} \theta_i \theta_{i+k})}{(1 + \sum_{i=1}^{q} \theta_i^2)},$$

for $k = 1, 2, \ldots, q$ with $\rho_k = 0$ for $k > q$.

Partial Autocorrelation for MA(q)

Suppose that the moving average is invertible, that is, the roots of $\Theta_q(B) = 0$ lie outside the unit ball, then $\Theta_q(B)^{-1} = \Phi_p(B)$ where $p = \infty$. Hence, an invertible moving average is the same as an infinite order autoregressive process. Hence, the partial autocorrelations for the moving average are not necessarily zero for $k > M$, for some finite value M.

Spectrum for MA(q)

Likewise the spectrum for the MA(q) is,

$$f(\omega) = \frac{\sigma_a^2}{2\pi} \mid \Theta_q(e^{-i\omega}) \mid^2 .$$

2.4.1 MA(1)

Definition: 2.2 *The Moving Average of order one, MA(1) satisfies the following: :*

1. $X_t = \mu$
2. $\dot{X}_t = X_t - \mu$
3. $\dot{X}_t = a_t - \theta a_{t-1}$
4. a_t *is white noise*

or

$$X_t = \mu + \Theta_1(B)a_t$$

where $\Theta_1(B) = (1 - \theta B)$.
From the property of the white noise series $\{a_t\}$, it follows that $\{X_t\}$ is second order stationary for any value of θ and that $E[X_t] = \mu$,

$$Var(X_t) = E[\dot{X}_t^2] = (1 + \theta^2)\sigma_a^2.$$

Furthermore,

$$\gamma_1 = E[\dot{X}_{t+1}\dot{X}_t] = -\theta\sigma_a^2,$$

and

$$\gamma_k = E[\dot{X}_{t+k}\dot{X}_t] = 0$$

for $k > 1$. From which one has,

$$\rho_1 = -\theta/(1 + \theta^2)$$

and

$$\rho_k = 0$$

for $k > 1$.

Partial Autocorrelation

The partial autocorrelations are;

$$\phi_{kk} = \frac{-\theta^k(1 - \theta^2)}{1 - \theta^{2(k+1)}}$$

for $k \geq 1$.

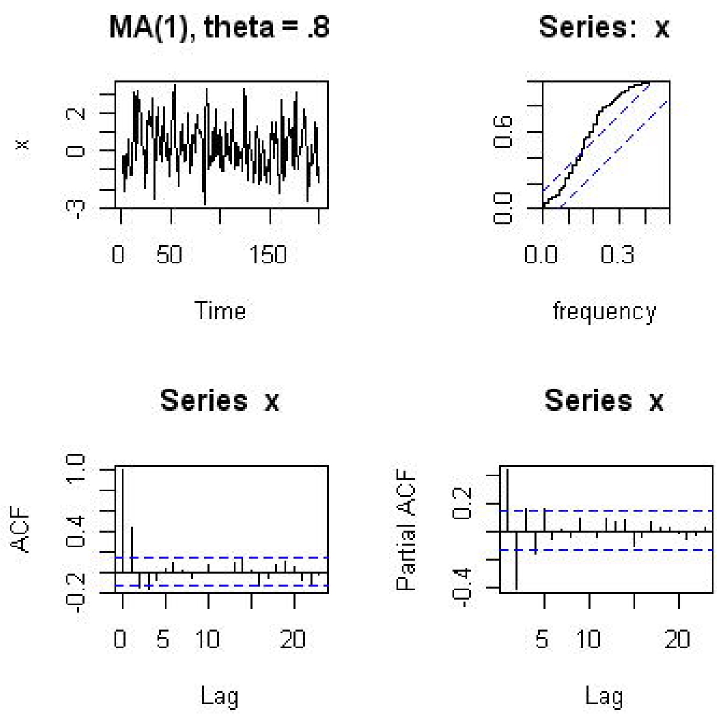
Spectrum

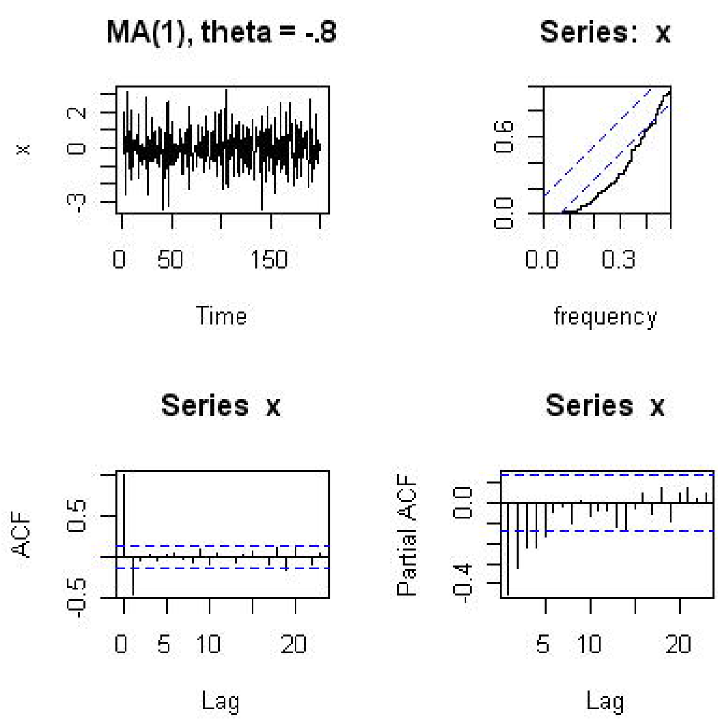
The spectrum for the MA(1) is,

$$f(\omega) = 2\sigma_a^2[1 + \theta^2 - 2\theta\cos(\omega)].$$

The following graphs are for two MA(1) processes. Each was created in R using the following code: Note: in R the parameter ma = -θ as I have defined above.

```
x<-arima.sim(list(order=c(0,0,1), ma=.8), n=200)
timeplot(x,c(2,2),'MA(1), theta = .8')
x<-arima.sim(list(order=c(0,0,1), ma=-.8), n=200)
timeplot(x,c(2,2),'MA(1), theta = -.8')
```


MA(1), theta = .8
x
Time
Series: x
frequency
Series x
ACF
Lag
Series x
Partial ACF
Lag


MA(1), theta = -.8
x
Time
Series: x
frequency
Series x
ACF
Lag
Series x
Partial ACF
Lag

2.5 Autoregressive Moving Average Processes - ARMA(p,q)

The autoregressive moving average process of order p,q, denoted by ARMA (p,q) is given by,

$$\Phi_p(B)\dot{X}_t = \Theta_q(B)a_t$$

where $\Phi_p(B) = (1 - \phi_1 B - \phi_2 B^2 - \ldots - \phi_p B^p)$,$\Theta_q(B) = (1 - \theta_1 B - \theta_2 B^2 - \ldots - \theta_q B^q)$, a_t is white noise with variance σ_a^2 and $\dot{X}_t = X_t - \mu$. Rewriting this model as,

$$\dot{X}_t = \phi_1 \dot{X}_{t-1} + \phi_2 \dot{X}_{t-2} + \ldots + \phi_p \dot{X}_{t-p} + a_t - \theta_1 a_{t-1} - \ldots - \theta_q a_{t-q}.$$

One can easily show that

$$\rho_k = \phi_1 \rho_{k-1} + \phi_2 \rho_{k-2} + \ldots + \phi_p \rho_{k-p}$$

or

$$\Phi_p(B)\rho_k = 0$$

whenever $k > q$. However, the first q autocorrelations would depend upon the values of θ_j for $j = 1, 2 \ldots, q$. Assuming that the process is stationary this model can be written as,

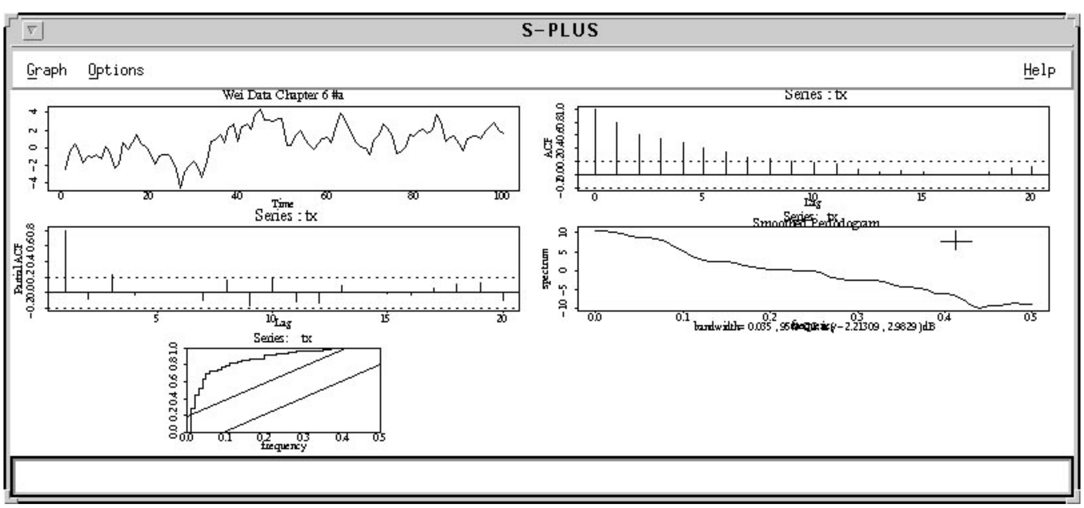
$$\begin{aligned}\dot{X}_t &= \Phi_p(B)^{-1}\Theta_q(B)a_t \\ &= \frac{(1 - \theta_1 B - \theta_2 B^2 - \ldots - \theta_q B^q)}{(1 - \phi_1 B - \phi_2 B^2 - \ldots - \phi_p B^p)}\end{aligned}$$

or $\Psi(B) = \Phi_p(B)^{-1}\Theta_q(B)$. This equation is often called a rational function which needs to be irreducible (that the two equations do not have any common roots).

Spectrum

The spectrum for the ARMA(p,q) is given by,

$$f(\omega) = 2\sigma_a^2 \frac{|\ \Theta_q(e^{-i\omega})\ |^2}{|\ \Phi_p(e^{-i\omega})\ |^2}.$$



Non Uniqueness of ARMA Processes

This is best illustrated with examples.

Example 1

Consider the AR(1) where $\phi = .8$ given by

$$\dot{X}_t = .8\dot{X}_{t-1} + a_t,$$

or

$$\begin{aligned}
(1 - .8B)\dot{X}_t &= a_t \\
\dot{X}_t &= (1 - .8B)^{-1} a_t \\
\dot{X}_t &= (1 - .8B - .8^2B^2 - .8^3B^3 - \ldots)a_t \\
&\simeq (1 - .8B - .64B^2 - .51B^3)a_t
\end{aligned}$$

Hence, the AR(1) can be similar to a MA(3).

Example 2

Suppose one has the following ARMA(2,1) model,

$$(1 - 1.3B + .4B^2)\dot{X}_t = (1 - .49B)a_t.$$

Note $(1 - 1.3B + .4B^2) = (1 - .8B)(1 - .5B)$ in which case the model can be written as an AR(1),

$$(1 - .8B)\dot{X}_t \simeq a_t.$$

In both these models one should invoke the principle of being *parsimonious*.

Chapter 3

Forecasting Stationary ARMA(p,q) Processes

The objective of forecasting is to estimate a future value, say, X_{t+k} when one has observed a realization of the time series $\{X_t\}$ given by, $x_1, x_2, \ldots, x_t$.

3.1 Minimum Mean Square Error Forecasts

Suppose that one has an ARMA(p,q) model given by

$$\Phi_p(B)\dot{X}_t = \Theta_q(B)a_t$$

and

$$\dot{X}_t = \Psi(B)a_t,$$

where $\Psi(B) = \Phi_p(B)^{-1}\Theta_q(B) = (1 + \sum_{j=1}^{\infty} \psi_j B^j)$. Letting $t = n + k$, one has

$$\dot{X}_{n+k} = \sum_{j=0}^{\infty} \psi_j a_{n+k-j},$$

since $\psi_0 = 1$.

Suppose that one wish to determine $\hat{\dot{X}}_{n+k}$ as a linear function of the observed values up to time n as,

$$\hat{\dot{X}}_{n+k} = \psi_k^* a_n + \psi_{k+1}^* a_{n-1} + \psi_{k+2}^* a_{n-2} + \ldots$$

where ψ_j^* are the values which minimize the following:

$$E[(\dot{X}_{n+k} - \hat{\dot{X}}_{n+k})^2] = \sigma_a^2 \sum_{j=0}^{k-1} \psi_j^2 + \sigma_a^2 \sum_{j=0}^{\infty} (\psi_{j+k} - \psi_{j+k}^*)^2.$$

This expression is minimized when $\psi_{j+k}^* - \psi_{j+k}$. Hence,

$$\hat{\dot{X}}_{n+k} = \psi_k a_n + \psi_{k+1} a_{n-1} + \psi_{k+2} a_{n-2} + \ldots$$

which is the conditional expectation of

$$\hat{\dot{X}}_{n+k} = E[\dot{X}_{t+k} \mid \dot{X}_n, \dot{X}_{n-1}, \ldots, \dot{X}_1].$$

The lead k error in the forecast is,

$$e_n(k) = \dot{X}_{t+k} - \hat{\dot{X}}_{n+k} = \sum_{j=0}^{k-1} \psi_j a_{n+k-j}.$$

It easily follows that $E[e_n(k)] = 0$, and

$$Var(\hat{\dot{X}}_{n+k}) = \sigma_a^2 \sum_{j=0}^{k-1} \psi_j^2.$$

Assuming a normal distribution, a $(1-\alpha) \times 100\%$ confidence interval for $\dot{X}_{t+k}$ is,

$$\hat{\dot{X}}_{n+k} \pm z_{\alpha/2}\sigma_a \sqrt{\sum_{j=0}^{k-1} \psi_j^2}.$$

Shumway and Stoffer approach this problem from a differently and derive equations that are more suitable for computational purposes.

Examples

The following examples are used to illustrate the above concepts.

Example 1 – AR(1)

Consider the stationary AR(1) given by,

$$\Phi_1(B)\dot{X}_t = (1 - \phi B)\dot{X}_t = a_t,$$

which can be written as,

$$\dot{X}_t = (1 + \sum_{j=1} \psi_j B^j) a_t,$$

where $\psi_j = \phi^j$. Hence,

$$\hat{\dot{X}}_{n+k} = \sum_{j=0}^{\infty} \phi^{k+j} a_{n-j},$$

and

$$e_n(k) = \sum_{j=0}^{k-1} \phi^j a_{n+k-j}$$

and

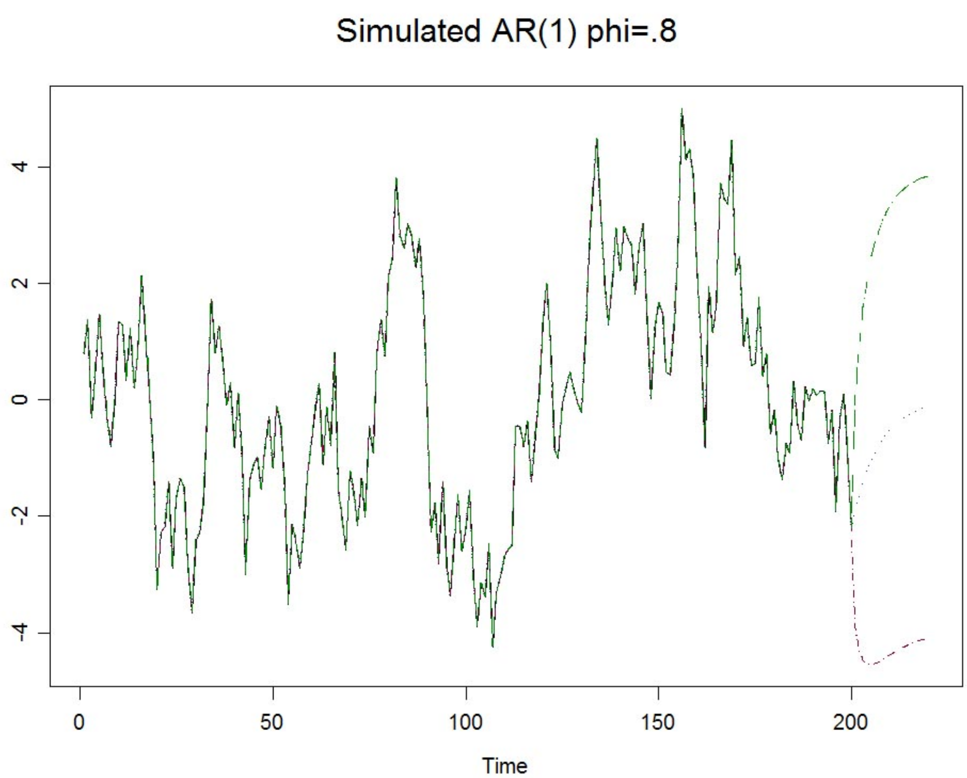
$$Var(\hat{\dot{X}}_{n+k}) = \sigma_a^2 \sum_{j=0}^{k-1} \phi^{2j} = \sigma_a^2 [\frac{1-\phi^{2k}}{1-\phi^2}],$$

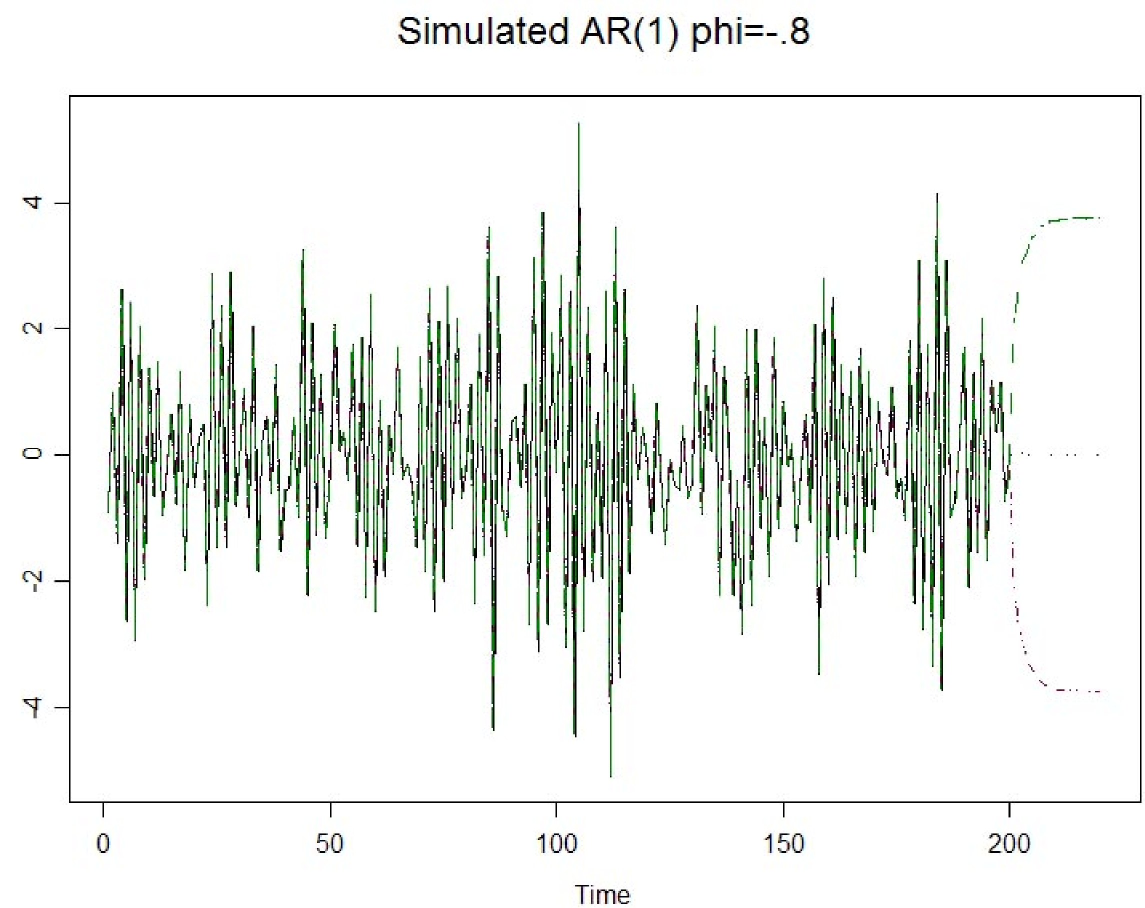
since $\sum_{j=0}^{k-1} a^j = \frac{1-a^k}{1-a}$ when $\mid a \mid < 1$. It should be mentioned that the forecast function is going to behave as the autocorrelation function, which in this case is like $\rho_k = \phi^k$.

I have plotted an example AR(1) forecast with lead = 20 and n = 200 of two simulated AR(1) processes using S-plus (difficulty with using R). The code is

```
forecast.ex.ar1<-function(ar.parm,nobs)
 {
     x<- arima.sim(list(order=c(1,0,0), ar=ar.parm),n=nobs)
     x<- x - mean(x)
     x.arima<-arima.mle(x, model=list(order=c(1,0,0)),n.cond=3)
     x.arima
     x.fore<-arima.forecast(x, n=20, model=x.arima$model)
     x.fore$mean<-x.fore$mean + mean(x)
     x.fmean<-c(x,x.fore$mean)
     x.low<-c(x,x.fore$mean-2*x.fore$std.err)
     x.high<-c(x,x.fore$mean+2*x.fore$std.err)
     ts.plot(x,x.fmean,x.low,x.high)
 }
 forecast.ex.ar1(.8,200)
 title('Simulated AR(1) phi=.8')
 forecast.ex.ar1(-.8,200)
 title('Simulated AR(1) phi=-.8')
```

The resulting graphs are:





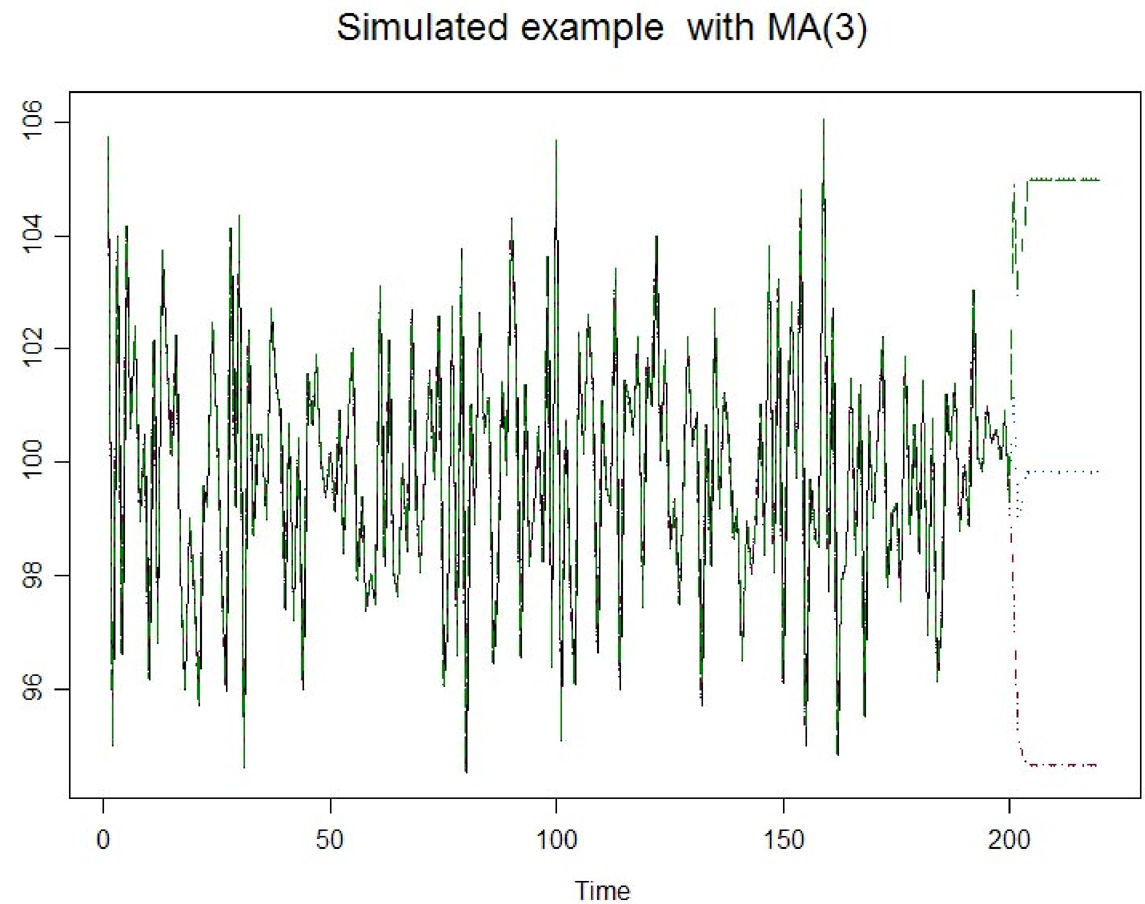
Example – MA(q)

The MA(q) is very straight forward where $\psi_j = -\theta_j$ for $j \leq q$ and $\psi_j = 0$ for $j > q$.
I have plotted an example MA(3) forecast with lead = 20 and n = 200 of a simulated MA(3) processes using S-plus. The code is

```
forecast.ex.ma3<-function(ma.parm,nobs)
 {
     x<- arima.sim(list(order=c(0,0,3), ma=ma.parm),n=nobs) + 100
     xt<- x - mean(x)
     x.arima<-arima.mle(xt, model=list(order=c(0,0,3)),n.cond=3)
     x.arima
     x.fore<-arima.forecast(xt, n=20, model=x.arima$model)
     x.fore$mean<-x.fore$mean + mean(x)
     x.fmean<-c(x,x.fore$mean)
     x.low<-c(x,x.fore$mean-2*x.fore$std.err)
     x.high<-c(x,x.fore$mean+2*x.fore$std.err)
     ts.plot(x,x.fmean,x.low,x.high)
```

```
}
forecast.ex.ma3(c(1.2,-.4,1.8),200)
title('Simulated MA(3)')
```

The resulting graph is:



Chapter 4

Estimation of Parameters

In this chapter it is assumed that one observes a realization of the stationary time series $\{X_t\} \sim ARMA(p,q)$ given by $x_1, x_2, \ldots, x_n$. Furthermore, assume that both p and q are known. That is, the time series has been identified (non trivial assumption). This means that there are $p+q+2$ parameters that need to be estimated. They are $\phi_1, \phi_2, \ldots, \phi_p$, $\theta_1, \theta_2, \ldots, \theta_q$, σ_a^2 and $E[X_t] = \mu$.

4.1 Moment Estimators – Yule-Walker Equations for AR(p) Processes

The estimation of μ with $\bar{x}$ is straight forward. Define $\dot{x}_j = x_j - \bar{x}$, for $j = 1, 2, \ldots, n$. The AR(p) is,

$$\Phi_p(B)\dot{X}_t = a_t$$

where $\Phi_p(B) = (1 - \phi_1 B - \phi_2 B^2 - \ldots - \phi_p B^p)$, or

$$\dot{X}_t = \phi_1 \dot{X}_{t-1} + \phi_2 \dot{X}_{t-2} + \ldots + \phi_p \dot{X}_{t-p} + a_t.$$

Recall that one could easily show that, One can easily show that

$$\rho_k = \phi_1 \rho_{k-1} + \phi_2 \rho_{k-2} + \ldots + \phi_p \rho_{k-p}$$

or

$$\Phi_p(B)\rho_k = 0.$$

From which one can derive the Yule-Walker equations given by,

$$\vec{\rho} = \vec{P}\vec{\phi}.$$

Since $\vec{P}$ is positive definite one has,

$$\vec{\phi} = \vec{P}^{-1}\vec{\rho}.$$

The method of moments estimates are found by substituting in $\vec{P}$ and $\vec{\rho}$ with $\hat{\rho}_k \simeq \rho_k$. This leads to,

$$\hat{\phi} = \hat{P}^{-1}\hat{\rho}$$

and

$$\hat{\sigma}_a^2 = \hat{\gamma}_0[1 - \hat{\rho}'\hat{P}^{-1}\hat{\rho}].$$

Shumway and Stoffer indicate that the Yule-Walker estimates are asymptotically optimal when the process is a AR(p) (this is not the case when one has a MA term in the model. That is,

Property 4.1 Large Sample Results for Yule-Walker Estimators for an AR(p) Process *are as follows:*

$$\sqrt{n}(\hat{\vec{\phi}} - \vec{\phi}) \xrightarrow[d]{} N(0, \sigma_a^2 \vec{P}^{-1}), \quad \hat{\sigma}_a^2 \xrightarrow[p]{} \sigma_a^2.$$

Property 4.2 Large Sample Results for PACF for an AR(p) Process *are as follows:*

$$\sqrt{n} \quad \hat{\phi}_{kk} \xrightarrow[d]{} N(0,1),$$

for $k > p$.

As example2.25 on page 127 illustrates, the moment estimates using the Yule-Walker equations with a MA model leads to suboptimal estimates when compare with the maximum likelihood estimates.

4.2 Maximum Likelihood and Least Squares Estimates

4.2.1 AR(1) Process

Suppose that,

$$\dot{X}_t = \phi \dot{X}_{t-1} + a_t,$$

where $a_t \sim WN(0, \sigma_a^2)$ and $\mid \phi \mid < 1$. The Likelihood equation is given by,

$$L(\mu, \sigma_a^2, \phi) = f_{\mu, \sigma_a^2, \phi}(x_1, x_2, \ldots, x_n),$$

or in the case of the AR(1) one has,

$$L(\mu, \sigma_a^2, \phi) = f(x_1) f(x_2 \mid x_1) f(x_3 \mid x_3) \ldots f(x_n \mid x_{n-1}),$$

where

$$f(x_t \mid x_{t-1}) = f_w((x_t - \mu) - \phi(x_{t-1} - \mu)) \sim N(0, \sigma_a^2).$$

In which case,

$$L(\mu, \sigma_a^2, \phi) = \prod_{t=2}^{n} f_w((x_t - \mu) - \phi(x_{t-1} - \mu)),$$

or

$$L(\mu, \sigma_a^2, \phi) = (2\pi\sigma_a^2)^{-n/2}(1 - \phi^2)^{1/2} exp[\frac{-S(\mu, \phi)}{2\sigma_a^2}],$$

where

$$S(\mu, \phi) = (1 - \phi^2)(x_1 - \mu)^2 + \sum_{t=2}^{n} [(x_t - \mu) - \phi(x_{t-1} - \mu)]^2.$$

$S(\mu, \phi)$ in this expression is called the **unconditional sum of squares**. The **unconditional least squares** involves minimizing this expression. The **conditional sum of squares** and **conditional least squares** are found when minimizing the following term,

$$S(\mu, \phi) = \sum_{t=2}^{n} [(x_t - \mu) - \phi(x_{t-1} - \mu)]^2.$$

This would be the appropriate expression if one assumed the value for x_1 was given. Hence, the statistics are conditioned upon knowing or specifying x_1. The conditional least squares involves minimizing,

$$S(\mu, \phi) = \sum_{t=2}^{n} [(x_t - (\alpha + \phi x_{t-1}]^2,$$

where $\alpha = \mu(1-\phi)$. This equation is the same as the linear regression equation between (x_i, y_i) when $y_i = x_t$ and $x_i = x_{t-1}$. The least squares solution is, $\hat{\alpha} = \bar{y} - \hat{\phi}\bar{x}_1$ where $\bar{x}_1 = \frac{1}{(n-1)} \sum_{t=1}^{n-1} x_t$ and $\bar{y} = \frac{1}{(n-1)} \sum_{t=2}^{n} x_t$, and

$$\hat{\mu} = \frac{\bar{y} - \hat{\phi}\bar{x}_1}{1 - \hat{\phi}},$$

and

$$\hat{\phi} = \frac{\sum_{t=2}^{n}(x_t - \bar{y})(x_{t-1} - \bar{x}_1)}{\sum_{t=2}^{n}(x_{t-1} - \bar{x}_1)^2}.$$

4.3 ARMA(p,q)

The estimation procedure is the general ARMA(p,q) is more complicated but follows a similar methodology. The ARMA(p,q) model is,

$$\Phi_p(B)\dot{X}_t = \Theta_q(B)a_t$$

where $\Phi_p(B) = (1 - \phi_1 B - \phi_2 B^2 - \ldots - \phi_p B^p)$,$\Theta_q(B) = (1 - \theta_1 B - \theta_2 B^2 - \ldots - \theta_q B^q)$, a_t is white noise with variance σ_a^2 and $\dot{X}_t = X_t - \mu$. Rewriting this model as,

$$a_t = \dot{X}_t + \phi_1 \dot{X}_{t-1} + \phi_2 \dot{X}_{t-2} + \ldots + \phi_p \dot{X}_{t-p} + \theta_1 a_{t-1} + \ldots + \theta_q a_{t-q}.$$

The joint density function for $\vec{a} = (a_1, a_2, \ldots, a_n)$ is,

$$f(\vec{a} \mid \vec{\phi}, \vec{\theta}, \sigma_a^2, \mu) = (2\pi\sigma_a^2)^{-n/2} \exp\left[\frac{-1}{2\sigma_a^2} \sum_{t=1}^{n} a_t^2\right].$$

Note, this likelihood function is dependent upon unobserved $x_* = (\dot{x}_{1-p}, \ldots, \dot{x}_0)$ and $a_* = (\dot{a}_{1-q}, \ldots, \dot{a}_0)$, hence the conditional log-likelihood is,

$$l_*(\vec{\phi}, \vec{\theta}, \sigma_a^2, \mu \mid x_*, a_*, x_1, \ldots, x_n) = \frac{-n}{2} \ln 2\pi\sigma_a^2 - \frac{S_*(\vec{\phi}, \vec{\theta}, \mu)}{2\sigma_a^2}$$

where

$$S_*(\vec{\phi}, \vec{\theta}, \mu) = \sum_{t=1}^{n} a_t^2(\vec{\phi}, \vec{\theta}, \mu \mid x_*, a_*, x_1, \ldots, x_n).$$

The unconditional estimates are derived by setting $a_* = 0$.

4.4 SAS PROC ARIMA Estimation Details

The ARIMA procedure primarily uses the computational methods outlined by Box and Jenkins. Marquardt's method is used for the nonlinear least-squares iterations. Numerical approximations of the derivatives of the sum-of-squares function are taken using a fixed delta (controlled by the DELTA= option).
The methods do not always converge successfully for a given set of data, particularly if the starting values for the parameters are not close to the least-squares estimates.

Back-forecasting

The unconditional sum of squares is computed exactly; thus, back-forecasting is not performed. Early versions of SAS/ETS software used the back-forecasting approximation and allowed a positive value of the BACKLIM= option to control the extent of the back-forecasting. In the current version, requesting a positive number of back-forecasting steps with the BACKLIM= option has no effect.

Preliminary Estimation

If an autoregressive or moving-average operator is specified with no missing lags, preliminary estimates of the parameters are computed using the autocorrelations computed in the IDENTIFY stage. Otherwise, the preliminary estimates are arbitrarily set to values that produce stable polynomials. When preliminary estimation is not performed by PROC ARIMA, then initial values of the coefficients for any given autoregressive or moving average factor are set to 0.1 if the degree of the polynomial associated with the factor is 9 or less. Otherwise, the coefficients are determined by expanding the polynomial (1 - 0.1B) to an appropriate power using a recursive algorithm.
These preliminary estimates are the starting values in an iterative algorithm to compute estimates of the parameters.

Estimation Methods

Maximum Likelihood

The METHOD= ML option produces maximum likelihood estimates. The likelihood function is maximized via nonlinear least squares using Marquardt's method. Maximum likelihood estimates are more expensive to compute than the conditional least-squares estimates, however, they may be preferable in some cases (Ansley and Newbold 1980; Davidson 1981).
The maximum likelihood estimates are computed as follows. Let the univariate ARMA model be

$$\Phi_p(B)\dot{X}_t = \Theta_q(B)a_t$$

where a_t is an independent sequence of normally distributed innovations with mean 0 and variance σ^2. The log likelihood function can be written as follows:

$$\frac{-1}{2\sigma^2}\vec{x}'\Omega^{-1}\vec{x} - \frac{1}{2}\ln(\mid \Omega \mid) - \frac{n}{2}\ln(\sigma^2).$$

In this equation, n is the number of observations, $\sigma^2\Omega$ is the variance of $\vec{x}$ as a function of the $\vec{\phi}$ and $\vec{\theta}$ parameters, and $\mid \cdot \mid$ denotes the determinant. The vector $\vec{x}$ is the time series X_t minus the structural part of the model μ.
The maximum likelihood estimate (MLE) of σ^2 is

$$s^2 = \frac{1}{n}\vec{x}'\Omega^{-1}\vec{x}.$$

Note that the default estimator of the variance divides by n-r, where r is the number of parameters in the model, instead of by n. Specifying the NODF option causes a divisor of n to be used.
The log likelihood concentrated with respect to σ^2 can be taken up to additive constants as

$$\frac{-n}{2}\ln\vec{x}'\Omega^{-1}\vec{x} - \frac{1}{2}\ln(\mid \Omega \mid).$$

Let H be the lower triangular matrix with positive elements on the diagonal such that $HH' = \Omega$. Let $\vec{e}$ be the vector $H^{-1}\vec{x}$. The concentrated log likelihood with respect to can now be written as

$$\frac{-n}{2}\ln\vec{e}'\vec{e} - \frac{1}{2}\ln(\mid H \mid)$$

or

$$\frac{-n}{2}\ln(\mid H \mid^{1/n} \vec{e}'\vec{e} \mid H \mid^{1/n}).$$

The MLE is produced by using a Marquardt algorithm to minimize the following sum of squares:

$$\mid H \mid^{1/n} \vec{e}'\vec{e} \mid H \mid^{1/n}.$$

The subsequent analysis of the residuals is done using $\vec{e}$ as the vector of residuals.

Unconditional Least Squares

The METHOD=ULS option produces unconditional least-squares estimates. The ULS method is also referred to as the *exact least-squares* (ELS) method. For METHOD=ULS, the estimates minimize

$$\sum_{t=1}^{n} a_t^2 = \sum_{t=1}^{n}(x_t - C_t V_t^{-1} \tilde{x}')^2$$

where C_t is the covariance matrix of x_t and $\tilde{x}' = (x_1, \ldots, x_{t-1})$, and V_t is the variance matrix of $\tilde{x}'$. In fact, $\sum_{t=1}^{n} a_t^2$ is the same as $\vec{x}'\Omega^{-1}\vec{x}$ and, hence, $\vec{e}'\vec{e}$. Therefore, the unconditional least-squares estimates are obtained by minimizing the sum of squared residuals rather than using the log likelihood as the criterion function.

Conditional Least Squares

The METHOD=CLS option produces conditional least-squares estimates. The CLS estimates are conditional on the assumption that the past unobserved errors are equal to 0. The series x_t can be represented in terms of the previous observations, as follows:

$$x_t = a_t + \sum_{i=1}^{\infty} \pi_i x_{t-i}.$$

The π weights are computed from the ratio of the ϕ and θ polynomials, as follows:

$$\frac{\Phi(B)}{\Theta(B)} = 1 - \sum_{i=1}^{\infty} \pi_i B^i.$$

The CLS method produces estimates minimizing

$$\sum_{t=1}^{n} a_t^2 = \sum_{t=1}^{n}(x_t - \sum_{i=1}^{\infty} \hat{\pi}_i x_{t-i})^2$$

where the unobserved past values of x_t are set to 0 and $\hat{\pi}$ are computed from the estimates of ϕ and θ at each iteration.
For METHOD=ULS and METHOD=ML, initial estimates are computed using the METHOD=CLS algorithm.

Start-up for Transfer Functions

When computing the noise series for transfer function and intervention models, the start-up for the transferred variable is done assuming that past values of the input series are equal to the first value of the series. The estimates are then obtained by applying least squares or maximum likelihood to the noise series. Thus, for transfer function models, the ML option does not generate the full (multivariate ARMA) maximum likelihood estimates, but it uses only the univariate likelihood function applied to the noise series. Because PROC ARIMA uses all of the available data for the input series to generate the noise series, other start-up options for the transferred series can be implemented by prefixing an observation to the beginning of the real data. For example, if you fit a transfer function model to the variable Y with the single input X, then you can employ a start-up using 0 for the past values by prefixing to the actual data an observation with a missing value for Y and a value of 0 for X.

Information Criteria

PROC ARIMA computes and prints two information criteria, Akaike's information criterion (AIC) (Akaike 1974; Harvey 1981) and Schwarz's Bayesian criterion (SBC) (Schwarz 1978). The AIC and SBC are used to compare competing models fit to the same series. The model with the smaller information criteria is said to fit the data better. The AIC is computed as

$$-2\ln(L) + 2k$$

where L is the likelihood function and k is the number of free parameters. The SBC is computed as

$$-2\ln(L) + \ln(n)k$$

where n is the number of residuals that can be computed for the time series. Sometimes Schwarz's Bayesian criterion is called the Bayesian Information criterion (BIC).

If METHOD=CLS is used to do the estimation, an approximation value of L is used, where L is based on the conditional sum of squares instead of the exact sum of squares, and a Jacobian factor is left out.

Tests of Residuals

A table of test statistics for the hypothesis that the model residuals are white noise is printed as part of the ESTIMATE statement output. The chi-square statistics used in the test for lack of fit are computed using the Ljung-Box formula

$$\chi^2_m = n(n+2)\sum_{k=1}^{m}\frac{r_k^2}{(n-k)}$$

where

$$r_k = \frac{\sum_{t=1}^{n-k} a_t a_{t+k}}{\sum_{t=1}^{n} a_t^2},$$

and a_t is the residual series.

This formula has been suggested by Ljung and Box (1978) as yielding a better fit to the asymptotic chi-square distribution than the Box-Pierce Q statistic. Some simulation studies of the finite sample properties of this statistic are given by Davies, Triggs, and Newbold (1977) and by Ljung and Box (1978).

Each chi-square statistic is computed for all lags up to the indicated lag value and is not independent of the preceding chi-square values. The null hypotheses tested is that the current set of autocorrelations is white noise.

t-values

The t values reported in the table of parameter estimates are approximations whose accuracy depends on the validity of the model, the nature of the model, and the length of the observed series. When the length of the observed series is short and the number of estimated parameters is large with respect to the series length, the t approximation is usually poor. Probability values corresponding to a t distribution should be interpreted carefully as they may be misleading.

Cautions During Estimation

The ARIMA procedure uses a general nonlinear least-squares estimation method that can yield problematic results if your data do not fit the model. Output should be examined carefully. The GRID option can be used to ensure the validity and quality of the results. Problems you may encounter include the following:

- Preliminary moving-average estimates may not converge. Should this occur, preliminary estimates are derived as described previously in“Preliminary Estimation.” You can supply your own preliminary estimates with the ESTIMATE statement options.
- The estimates can lead to an unstable time series process, which can cause extreme forecast values or overflows in the forecast.
- The Jacobian matrix of partial derivatives may be singular; usually, this happens because not all the parameters are identifiable. Removing some of the parameters or using a longer time series may help.
- The iterative process may not converge. PROC ARIMA's estimation method stops after n iterations, where n is the value of the MAXITER= option. If an iteration does not improve the SSE, the Marquardt parameter is increased by a factor of ten until parameters that have a smaller SSE are obtained or until the limit value of the Marquardt parameter is exceeded.
- For METHOD=CLS, the estimates may converge but not to least-squares estimates. The estimates may converge to a local minimum, the numerical calculations may be distorted by data whose sum-of-squares surface is not smooth, or the minimum may lie outside the region of invertibility or stationarity.
- If the data are differenced and a moving-average model is fit, the parameter estimates may try to converge exactly on the invertibility boundary. In this case, the standard error estimates that are based on derivatives may be inaccurate.

Chapter 5

Model Identification

Model identification of p and q in ARMA(p,q) models is as much art form as science. The basic guiding tools are the ACF and PACF. More recently other tools have been developed and incorporated into the software packages, such as SAS. This chapter considers some of these additional methods.
The following is from the SAS User's Guide concerning one of these commonly used methods.

5.1 The Inverse Autocorrelation Function

The sample inverse autocorrelation function (SIACF) plays much the same role in ARIMA modeling as the sample partial autocorrelation function (SPACF) but generally indicates subset and seasonal autoregressive models better than the SPACF.
Additionally, the SIACF may be useful for detecting over-differencing. If the data come from a nonstationary or nearly nonstationary model, the SIACF has the characteristics of a noninvertible moving average. Likewise, if the data come from a model with a noninvertible moving average, then the SIACF has nonstationary characteristics and therefore decays slowly. In particular, if the data have been over-differenced, the SIACF looks like a SACF from a nonstationary process.
The inverse autocorrelation function is not often discussed in textbooks, so a brief description is given here. More complete discussions can be found in Cleveland (1972), Chatfield (1980), and Priestly (1981).
Let X_t be generated by the ARMA(p,q) process

$$\Phi_p(B)\dot{X}_t = \Theta_q(B)a_t$$

where a_t is a white noise sequence. If $\Theta(B)$ is invertible (that is, if considered as a polynomial in B has no roots less than or equal to 1 in magnitude), then the model

$$\Theta_q(B)\dot{X}_t = \Phi_p(B)a_t$$

is also a valid ARMA(q,p) model. This model is sometimes referred to as the dual model. The autocorrelation function (ACF) of this dual model is called the inverse autocorrelation function (IACF) of the original model. Notice that if the original model is a pure autoregressive model, then the IACF is an ACF corresponding to a pure moving-average model. Thus, it cuts off sharply when the lag is greater than p; this behavior is similar to the behavior of the partial autocorrelation function (PACF).
The sample inverse autocorrelation function (SIACF) is estimated in the ARIMA procedure by the following steps. A high-order autoregressive model is fit to the data by means of the Yule-Walker equations. The order of the autoregressive model used to calculate the SIACF is the minimum of the NLAG= value and one-half the number of observations after differencing. The SIACF is then calculated as the autocorrelation

function that corresponds to this autoregressive operator when treated as a moving-average operator. That is, the autoregressive coefficients are convolved with themselves and treated as autocovariances.
Under certain conditions, the sampling distribution of the SIACF can be approximated by the sampling distribution of the SACF of the dual model (Bhansali 1980). In the plots generated by ARIMA, the confidence limit marks $(\cdot)$ are located at $\pm 2/\sqrt{n}$. These limits bound an approximate 95% confidence interval for the hypothesis that the data are from a white noise process.

5.1.1 The Inverse Autocorrelation Method

The inverse autocorrelation function (IACF) is the ACF associated with the reciprocal of the spectral density of a time series. Consider a time series with spectral density $S(\lambda)$. Define the inverse spectral density by

$$Si(\lambda) = \frac{1}{4\pi^2 S(\lambda)}, \quad \pi \leq \lambda \leq \pi.$$

If $S(\lambda) \neq 0$ for $\pi \leq \lambda \leq \pi$, then $Si(\lambda)$ has the Fourier expansion

$$Si(\lambda) = \frac{1}{2\pi} \sum_{j=-\infty}^{\infty} \gamma i(j) \exp(-i\lambda j),$$

where

$$\sum_{j=-\infty}^{\infty} \mid \gamma i(j) \mid < \infty.$$

Applying the inverse Fourier transform to $Si(\lambda)$ yields

$$\gamma i(k) = \int_{-\infty}^{\infty} Si(\lambda) \exp(i\lambda k) d\lambda, \quad k = 0, \pm 1, \pm 2, \ldots,$$

which is the inverse autocovariance at lag k. The inverse autocorrelation is then derived as,

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

5.1.2 Additional Identification Methods

SAS PROC ARIMA has three other methods for aiding in model identification. There are the ESAFC, MIMIC, and SCAN methods. The following example is given using the d data set from Wei (AR(1)).

```
                          The ARIMA Procedure

                         Name of Variable = z

                  Mean of Working Series     -0.1076
                  Standard Deviation        1.622778
                  Number of Observations         100


                 Autocorrelation Check for White Noise


 To      Chi-          Pr >
Lag    Square   DF   ChiSq
```

---------------Autocorrelations---------------

6	159.62	6	<.0001	0.759	0.603	0.488	0.396	0.337	0.282
12	168.61	12	<.0001	0.197	0.151	0.056	-0.020	-0.022	-0.122
18	182.47	18	<.0001	-0.187	-0.187	-0.141	-0.087	-0.114	-0.076
24	184.54	24	<.0001	-0.036	-0.053	-0.079	-0.045	-0.048	-0.036

Squared Canonical Correlation Estimates

Lags	MA 0	MA 1	MA 2	MA 3	MA 4	MA 5
AR 0	0.5769	0.3666	0.2402	0.1583	0.1151	0.0806
AR 1	0.0032	0.0006	<.0001	0.0016	0.0002	0.0085
AR 2	0.0008	<.0001	0.0005	0.0018	0.0023	0.0080
AR 3	<.0001	0.0006	0.0007	0.0016	0.0001	0.0086
AR 4	0.0018	0.0021	0.0018	0.0009	0.0017	0.0063
AR 5	<.0001	0.0024	<.0001	0.0019	0.0010	0.0018

SCAN Chi-Square[1] Probability Values

Lags	MA 0	MA 1	MA 2	MA 3	MA 4	MA 5
AR 0	<.0001	<.0001	0.0035	0.0307	0.0806	0.1604
AR 1	0.5718	0.8084	0.9700	0.7015	0.8803	0.3746
AR 2	0.7788	0.9576	0.8433	0.7227	0.6416	0.3938
AR 3	0.9385	0.8292	0.8279	0.7137	0.9274	0.4237
AR 4	0.6809	0.6862	0.6944	0.8168	0.7087	0.5296
AR 5	0.9511	0.6425	0.9697	0.6936	0.7921	0.7341

The ARIMA Procedure

Extended Sample Autocorrelation Function

Lags	MA 0	MA 1	MA 2	MA 3	MA 4	MA 5
AR 0	0.7591	0.6025	0.4876	0.3958	0.3371	0.2819
AR 1	-0.0742	-0.0307	-0.0048	-0.0474	0.0179	0.1095
AR 2	-0.3891	-0.0195	-0.0007	-0.0521	0.0254	0.1052
AR 3	-0.2503	-0.0524	0.0180	-0.0311	0.0231	0.0621
AR 4	-0.1640	-0.4020	-0.1463	-0.0338	0.0212	0.0526
AR 5	0.1444	0.3725	-0.0329	0.0538	0.0156	0.0151

ESACF Probability Values

Lags	MA 0	MA 1	MA 2	MA 3	MA 4	MA 5

AR 0	<.0001	<.0001	0.0041	0.0307	0.0784	0.1532
AR 1	0.4602	0.7618	0.9626	0.6431	0.8608	0.2827
AR 2	0.0001	0.8538	0.9946	0.6264	0.8139	0.3450
AR 3	0.0137	0.6271	0.8766	0.7781	0.8389	0.6058
AR 4	0.1082	<.0001	0.1580	0.7653	0.8513	0.6982
AR 5	0.1594	0.0003	0.7796	0.6841	0.8909	0.9026

Minimum Information Criterion

Lags	MA 0	MA 1	MA 2	MA 3	MA 4	MA 5
AR 0	0.964847	0.763448	0.6265	0.538289	0.502819	0.462929
AR 1	0.139087	0.181698	0.219499	0.265226	0.310305	0.338222
AR 2	0.173816	0.219446	0.265498	0.311179	0.356346	0.373383
AR 3	0.219123	0.265131	0.309312	0.353533	0.393666	0.419412
AR 4	0.265171	0.31114	0.354327	0.394095	0.439269	0.444095
AR 5	0.291709	0.337715	0.375794	0.421356	0.443363	0.486417

Error series model: AR(5)
Minimum Table Value: BIC(1,0) = 0.139087

ARMA(p+d,q) Tentative Order Selection Tests

---------SCAN--------			--------ESACF--------		
p+d	q	BIC	p+d	q	BIC
1	0	0.139087	1	0	0.139087
0	4	0.502819	3	1	0.265131
			5	2	0.375794
			0	4	0.502819

(5% Significance Level)

Chapter 6

Nonstationary Processes

In this chapter, two fixable models are considered. There are models for which the methodology described in the usual Time Series literature are not appropriate. The fixable models involve the non stationarity of the mean. These are **Deterministic Trend Models** and **Stochastic Trend Models**.

6.1 Deterministic Trend Models

This model is given by

$$X_t = \sum_{j=0}^{k} \alpha_j t^j + a_t$$

where

$$E[X_t] = \sum_{j=0}^{k} \alpha_j t^j$$

a k^{th} order polynomial. Consider the special case when $k = 1$, the linear trend model given by

$$X_t = \mu_t + a_t = \alpha_0 + \alpha_1 t + a_t.$$

Suppose one takes a first order difference given by $\nabla = (1 - B)$ or

$$\nabla X_t = X_t - X_{t-1} = \alpha_0 + \alpha_1 t + a_t - \alpha_0 - \alpha_1(t-1) + a_{t-1} = \alpha_1 + w_t,$$

where $w_t = a_t - a_{t-1} \sim WN(0, 2\sigma_a^2)$. This method can be extended to the k^{th} order linear trend model by defining the k^{th} linear difference given by $\nabla^k = (1 - B)^k$ from which one has

$$\nabla^k X_t = constant + \nabla^k a_t,$$

and $w_t = \nabla^k a_t \sim WN(0, \sigma_w^2)$.

6.2 Stochastic Trend Models

Consider the model given by

$$X_t = X_{t-1} + a_t.$$

As you recall this is called the random walk model or the AR(1) with $\phi = 1$. Since $\rho_k = \phi^k$ in the AR(1), one notes that all the correlations are equal to one. In fact this model is not stationary since the solution to the equation $(1 - B)X_t = 0$ is one and $(1 - B)^{-1}$ does not converge when $| B |= 1$. But notice that

$$\nabla X_t = (1 - B)X_t = X_t - X_{t-1} = a_t$$

which is stationary and is in fact white noise. The general stochastic nonstationary trend model is given by

$$\Upsilon(B)(X_t + C) = a_t,$$

where $\Upsilon(B) = \Phi(B)(1 - B)^d$ for some $d > 0$ and $\Phi(B)$ is a stationary autoregressive of order p operator (i.e., all p roots lie inside the unit ball). If one considered the characteristic equation given by,

$$\Upsilon(B) = 0$$

then p roots would lie inside the unit ball and d roots would lie on the unit ball. Note: if any roots lie outside the unit ball then this is an example of a non fixable nonstationary process. It should be mentioned that the differencing operation assumes that the d roots are all real with B=1 as the solution when in fact the roots could be real at B=-1 or not real but rather complex roots (pairs) at $a \pm bi$, with $a^2 + b^2 = 1$. The above model is called the ARIMA(p,d,q) when

$$\Upsilon(B)(X_t + C) = \Theta(B)a_t,$$

or

$$\Phi(B)(1 - B)^d(X_t + C) = \Theta(B)a_t.$$

This implies that one can determine the value d such that $Z_t = (1 - B)^d(X_t + C)$ is a ARMA(p,q) process. The following example will illustrate this concept

```
diff_example_1<-function(order)
  {
     t<-1:200
     temp<- arima.sim(list(order = c(1,1,0), ar = 0.7), n = 199)  #ARIMA(1,1,0)
     x<- 50*temp + 100 + 2*t            # time series plus linear trend
     x1<-diff.ts(x,lag=1,differences=1)
     x2<-diff.ts(x,lag=1,differences=2)
     if(order == 0) tsplot2(x)
     else
       if(order == 1) tsplot2(x1)
       else tsplot2(x1)
  }

diff_example_2<-function(order)
  {
     t<-1:200
     temp<- arima.sim(list(order = c(1,1,0), ar = 0.7), n = 199)  #ARIMA(1,1,0)
     x<- 50*temp + 100 - .002*(t^2)            # time series plus quadratic trend
     x1<-diff.ts(x,lag=1,differences=1)
     x2<-diff.ts(x,lag=1,differences=2)
     if(order == 0) tsplot2(x)
     else
       if(order == 1) tsplot2(x1)
```

```
      else tsplot2(x1)
  }

 diff_example_3<-function(order)
  {
    t<-1:200
    temp<- arima.sim(list(order = c(0,0,0)), n = 199)
    x<- 50*temp + 100      # time series white noise
    x1<-diff.ts(x,lag=1,differences=1)
    x2<-diff.ts(x,lag=1,differences=2)
    if(order == 0) tsplot2(x)
    else
      if(order == 1) tsplot2(x1)
      else tsplot2(x1)
  }
 #example number 1
 diff_example_1(0)
 diff_example_1(1)
 diff_example_1(2)
```

I have not included the output graphs.
A commonly used model of this type is the ARIMA(0,1,1) or the **integrated moving average IMA(1,1) of order one** given by

$$(1-B)X_t = (1-\theta B)a_t.$$

Suppose that $\theta < 1$ then $(1-\theta B)^{-1} = \sum_{j=0}^{\infty} \theta^j B^j)$ in which case

$$\begin{aligned}
\frac{(1-B)}{(1-\theta B)} &= (1-B)(\sum_{j=0}^{\infty} \theta^j B^j) \\
&= (1+\theta B+\theta^2 B^2+\ldots)(-B-\theta B^2-\theta^2 B^3-\ldots) \\
&= 1-(1-\theta)B-(1-\theta)\theta B^2-(1-\theta)\theta^2 B^3-\ldots \\
&= 1-\alpha B-\alpha(1-\alpha)B^2-\alpha(1-\alpha)^2 B^3-\ldots
\end{aligned}$$

or X_t has an AR(∞) representation given by

$$X_t = \alpha \sum_{j=1}^{\infty} (1-\alpha)^{j-1} B^j X_t + a_t,$$

where $\alpha = (1-\theta)$. The one step ahead forecast for this model is,

$$\begin{aligned}
\hat{x}_{t+1} &= \alpha \sum_{j=1}^{\infty} (1-\alpha)^{j-1} B^j X_{t+1} \\
&= \alpha x_t + (1-\alpha)\alpha \sum_{j=1}^{\infty} (1-\alpha)^{j-1} B^j X_t \\
&= \alpha x_t + (1-\alpha)\hat{x}_t.
\end{aligned}$$

The new forecast is a linear combination of the new observation and the old forecast. This method of forecasting is called the **Exponential Weighted Moving Average (EWMA)**. The value α hence θ is called the smoothing constant.

The above models have been correctable non stationary in the mean models. It is possible that the time series is not stationary in the second moment or variance. The next section discusses this problem and mentions some solutions.

6.3 Variance Stabilizing Transformations

Suppose that

$$Var(X_t) = cf(\mu_t),$$

that is, the variance of the process changes as a function of the level of the mean. The objective is to find a transformation, T, such that $T(X_t)$ has constant variance. Consider the first order Taylor series for $T(\cdot)$ given by,

$$T(X_t) \simeq T(\mu_t) + T'(\mu_t)(X_t - \mu_t),$$

where $T'(\mu_t)$ is the first derivative of $T(X_t)$ and $\mu_t = E[X_t]$. Considering the variance of $T(X_t)$ we have

$$Var(T(X_t)) \simeq [T'(\mu_t)]^2 Var(X_t) = c[T'(\mu_t)]^2 f(\mu_t).$$

Thus, in order for the variance of $T(X_t)$ to be constant it follows that,

$$T'(\mu_t) = \frac{1}{\sqrt{f(\mu_t)}},$$

or

$$T(\mu_t) = \int \frac{1}{\sqrt{f(\mu_t)}} d\mu_t.$$

Special cases include;

- $Var(T(X_t)) = c^2\mu_t^2$ then

  $$T(\mu_t) = \int \frac{1}{\mu_t} d\mu_t = ln(\mu_t).$$

  The logarithmic transformation should be used.

- $Var(T(X_t)) = c^2\mu_t$ then

  $$T(\mu_t) = \int \frac{1}{\sqrt{\mu_t}} d\mu_t = 2\sqrt{\mu_t}.$$

  The square root transformation should be used.

- $Var(T(X_t)) = c^2\mu_t^4$ then

  $$T(\mu_t) = \int \frac{1}{\sqrt{\mu_t^4}} d\mu_t = -\frac{1}{\mu_t}.$$

  The reciprocal transformation should be used.

A wider class of transformation was proposed by Box and Cox (1964) and are called the **Box-Cox Transformation** given by

$$T(X_t) = X_t^{(\lambda)} = \frac{X_t^\lambda - 1}{\lambda}, \lambda \neq 0,$$

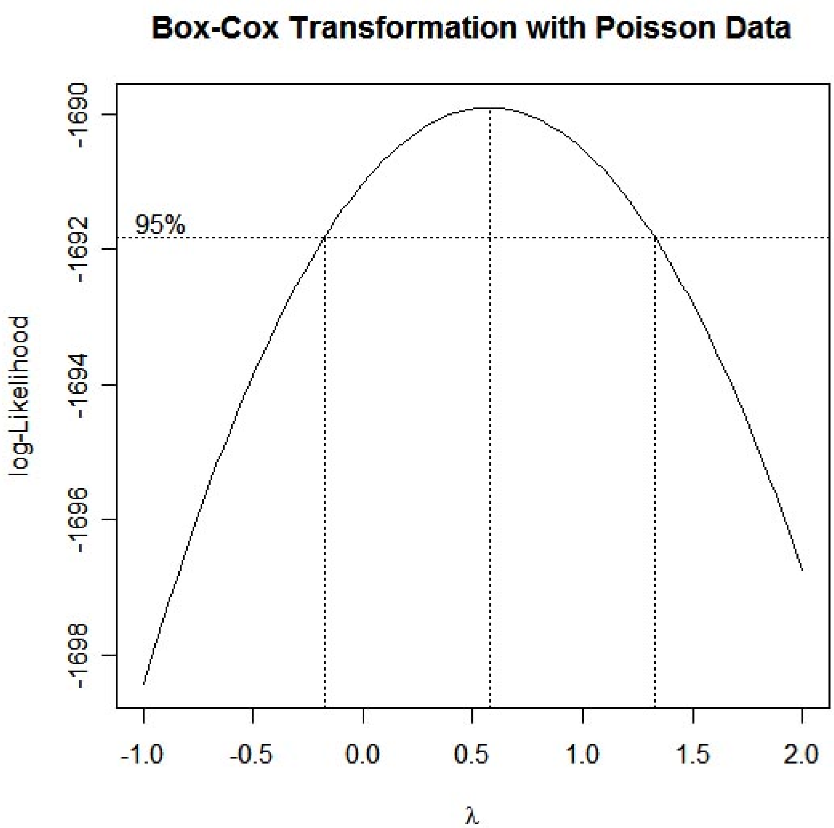
and $T(X_t) = X_t^{(\lambda)} = ln(X_t)$ when $\lambda = 0$.

Example of Box-Cox Transformation

In this example I generated Poisson data as an AR(1) process using R. The code is

```
temp<- arima.sim(list(order = c(1,0,0),ar=.8),  n=199,
+  rand.gen = function(n, ...) rpois(199,10))
t<-1:370
boxcox(temp ~ t, lambda = seq(-1, 2, 1/10))
```

The output is,



6.4 Unit Root Tests

Since, the location of the roots in an autoregressive process is important for either having stationary or for creating stationary processes the associated statistical inference for this problem has received considerable attention in the time series literature. This section considers some of the approaches taken. I have used the SAS User‘s Guide for the following discussion. The manual considers this problem of determining stationarity. For a more thorough discussion of this problem see Harris and Sollis, “Applied Time Series Modelling and Forecasting”, chapter 3.

6.4.1 Stationarity Tests

When a time series has a unit root, the series is nonstationary and the ordinary least squares (OLS) estimator is not normally distributed. Dickey (1976) and Dickey and Fuller (1979) studied the limiting distribution of the OLS estimator of autoregressive models for time series with a simple unit root. Dickey, Hasza and Fuller (1984) obtained the limiting distribution for time series with seasonal unit roots. Hamilton (1994) discusses the various types of unit root testing.

- For a description of Dickey-Fuller tests, refer to ,"PROBDF Function for Dickey-Fuller Tests" in Chapter 4.
- Refer to Chapter 4, "SAS Macros and Functions," for a description of the augmented Dickey-Fuller tests.
- Refer to Chapter 8, "The AUTOREG Procedure," for a description of Phillips-Perron tests.
- The random walk with drift test recommends whether or not an integrated times series has a drift term. Hamilton (1994) discusses this test.

6.4.2 Dickey-Fuller Tests

Theoretical Background

When a time series has a unit root, the series is nonstationary and the ordinary least squares (OLS) estimator is not normally distributed. Dickey (1976) and Dickey and Fuller (1979) studied the limiting distribution of the OLS estimator of autoregressive models for time series with a simple unit root. Dickey, Hasza and Fuller (1984) obtained the limiting distribution for time series with seasonal unit roots. Consider the $(p+1)^{th}$ order autoregressive time series

$$X_t = \sum_{j=1}^{p+1} \phi_j X_{t-j} + a_t,$$

and its characteristic equation

$$\Phi_{p+1}(B) = 0.$$

If all the characteristic roots are less than 1 in absolute value, X_t is stationary. X_t is nonstationary if there is a unit root. If there is a unit root, the sum of the autoregressive parameters is 1, and, hence, you can test for a unit root by testing whether the sum of the autoregressive parameters is 1 or not. For convenience, the model is parameterized as

$$\nabla X_t = \delta X_{t-1} + \sum_{j=1}^{p} \theta_j \nabla X_{t-j} + a_t,$$

where $\nabla X_t = X_t - X_{t-1}$ and

$$\delta = \sum_{j=1}^{p+1} \phi_j - 1,$$

$$\theta_j = -\phi_{j+1} - \ldots - \phi_{p+1}.$$

The estimators are obtained by regressing ∇X_t on $X_{t-1}, \nabla X_{t-1}, \nabla X_{t-2}, \ldots, \nabla X_{t-p}$. The t statistic of the ordinary least squares estimator of is the test statistic for the unit root test.
If the TREND = 1 option is used, the autoregressive model includes a mean term α_0. If TREND = 2, the model also includes a time trend term and the model is as follows:

$$\nabla X_t = \alpha_0 + \gamma t + \delta X_{t-1} + \sum_{j=1}^{p} \theta_j \nabla X_{t-j} + a_t.$$

If the series is an ARMA process, a large value of the AR= option may be desirable in order to obtain a reliable test statistic. To determine an appropriate value for the AR= option for an ARMA process, refer to Said and Dickey (1984).

Test Statistics

The Dickey-Fuller test is used to test the null hypothesis that the time series exhibits a lag d=1 (greater than one in seasonal models) unit root against the alternative of stationarity. The PROBDF function computes the probability of observing a test statistic more extreme than x under the assumption that the null hypothesis is true. You should reject the unit root hypothesis when PROBDF returns a small (significant) probability value.
There are several different versions of the Dickey-Fuller test. The PROBDF function supports six versions, as selected by the type argument. Specify the type value that corresponds to the way that you calculated the test statistic x.
The last two characters of the type value specify the kind of regression model used to compute the Dickey-Fuller test statistic. The meaning of the last two characters of the type value are as follows.

ZM zero mean or no intercept case. The test statistic x is assumed to be computed from the regression model

$$X_t = \delta X_{t-1} + a_t$$

SM single mean or intercept case. The test statistic x is assumed to be computed from the regression model

$$X_t = \alpha_0 + \delta X_{t-1} + a_t$$

TR intercept and deterministic time trend case. The test statistic x is assumed to be computed from the regression model

$$X_t = \alpha_0 + \gamma t + \delta X_{t-1} + a_t$$

The first character of the type value specifies whether the regression test statistic or the studentized test statistic is used. Let $\hat{\delta}$ be the estimated regression coefficient for the $d^{th} = 1^{st}$ lag of the series, and let $se_{\hat{\delta}}$ be the standard error of $\hat{\delta}$. The meaning of the first character of the type value is as follows.

R the regression coefficient-based test statistic. The test statistic is

$$x = n(\hat{\delta} - 1).$$

S the studentized test statistic. The test statistic is

$$x = \frac{(\hat{\delta} - 1)}{se_{\hat{\delta}}}.$$

Refer to Dickey and Fuller (1979) and Dickey, Hasza, and Fuller (1984) for more information about the Dickey-Fuller test null distribution. The preceding formulas are for the basic Dickey-Fuller test. The PROBDF function can also be used for the augmented Dickey-Fuller test, in which the error term et is modeled as an autoregressive process; however, the test statistic is computed somewhat differently for the augmented Dickey-Fuller test. Refer to Dickey, Hasza, and Fuller (1984) and Hamilton (1994) for information about seasonal and nonseasonal augmented Dickey-Fuller tests.
The PROBDF function is calculated from approximating functions fit to empirical quantiles produced by Monte Carlo simulation employing 10^8 replications for each simulation. Separate simulations were performed for selected values of n and for d=1,2,4,6,12.

The maximum error of the PROBDF function is $\pm 10^{-3}$ approximately for d in the set (1,2,4,6,12) and may be slightly larger for other d values. (Because the number of simulation replications used to produce the PROBDF function is much greater than the 60,000 replications used by Dickey and Fuller (1979) and Dickey, Hasza, and Fuller (1984), the PROBDF function can be expected to produce results that are substantially more accurate than the critical values reported in those papers.)

6.4.3 Phillips-Perron Tests

Unit Root and Cointegration Testing

Consider the random walk process

$$X_t = X_{t-1} + a_t$$

where the disturbances might be serially correlated with possible heteroscedasticity. Phillips and Perron (1988) proposed the unit root test of the OLS regression model.

$$X_t = \alpha X_{t-1} + a_t$$

Let $s^2 = (T-k)^{-1} \sum_{j=1}^{T} \hat{a}_t^2$ and let $\hat{\sigma}_\alpha^2$ be the variance estimate of the OLS estimator $\hat{\alpha}$, where $\hat{a}_t$ is the OLS residual. You can estimate the asymptotic variance of $T^{-1} \sum_{j=1}^{T} \hat{a}_t^2$ using the truncation lag ℓ.

$$\hat{\lambda} = \sum_{j=0}^{\ell} \kappa_j [1 - j/(\ell+1)] \hat{\gamma}_j,$$

where $\kappa_0 = 1$, $\kappa_j = 2$ for $j > 0$, and $\hat{\gamma}_j = T^{-1} \sum_{t=j+1}^{T} \hat{a}_t \hat{a}_{t-j}$. Then the Phillips-Perron $Z(\hat{\alpha})$ test (zero mean case) is written

$$Z(\hat{\alpha}) = T(\hat{\alpha} - 1) - \frac{T^2 \hat{\sigma}^2 (\hat{\lambda} - \hat{\gamma}_0)}{2s^2},$$

and has the following limiting distribution:

$$\frac{1/2[B(1)^2 - 1]}{\int_0^1 [B(x)]^2 dx}$$

where $B(\cdot)$ is a standard Brownian motion. Note that the realization $Z(x)$ from the stochastic process $B(\cdot)$ is distributed as $N(0, x)$ and thus $B(1)^2 \sim \chi_1^2$.
Therefore, you can observe that $\Pr[\hat{\alpha} < 1] \approx 0.68$ as $T \to \infty$, which shows that the limiting distribution is skewed to the left.
Let $t_{\hat{\alpha}}$ be the t-test statistic for $\hat{\alpha}$. The Phillips-Perron test is written

$$Z(\hat{\alpha}) = t_{\hat{\alpha}} \sqrt{\frac{\hat{\gamma}_0}{\hat{\lambda}}} - \frac{T\hat{\sigma}(\hat{\lambda} - \hat{\gamma}_0)}{2s\sqrt{\hat{\lambda}}}$$

and its limiting distribution is derived as

$$\frac{1/2[B(1)^2 - 1]}{[\int_0^1 [B(x)]^2 dx]^{1/2}}.$$

6.5 Long Memory Time Series

The Box-Jenkins approach to non stationary process is to take differences (assuming that one has a near unit root), yet there are data sets for which the ACF does not decay exponentially (as in short memory ARMA(p,q) processes) but for which the differencing operator would be too dramatic.
Rather than use the differencing approach given by $\nabla^d = (1-B)^d$ for $d = 1, 2, \ldots, D$ the new approach is to define

$$\nabla^d = (1-B)^d = \sum_{j=0}^{\infty} \pi_j B^j,$$

where

$$\pi_j = \frac{\Gamma(j-d)}{\Gamma(j+1)\Gamma(-d)},$$

with $\Gamma(x+1) = x\ \Gamma(x)$ being the Gamma function and $-.5 < d < .5$. This, implies that

$$\sum_{j=1}^{\infty} \pi_j\ x_{t-j} = a_t$$

which is an AR(∞) model. Alternatively, the model can be express as a MA(∞) where

$$x_t = \sum_{j=1}^{\infty} \psi_j\ a_{t-j}$$

where

$$\psi_j = \frac{\Gamma(j+d)}{\Gamma(j+1)\Gamma(d)},$$

and $\sum_{j=1}^{\infty} \mid \psi_j \mid^2 < \infty$. The ACF is

$$\rho_k = \frac{\Gamma(k+d)\Gamma(1-d)}{\Gamma(k-d+1)\Gamma(d)},$$

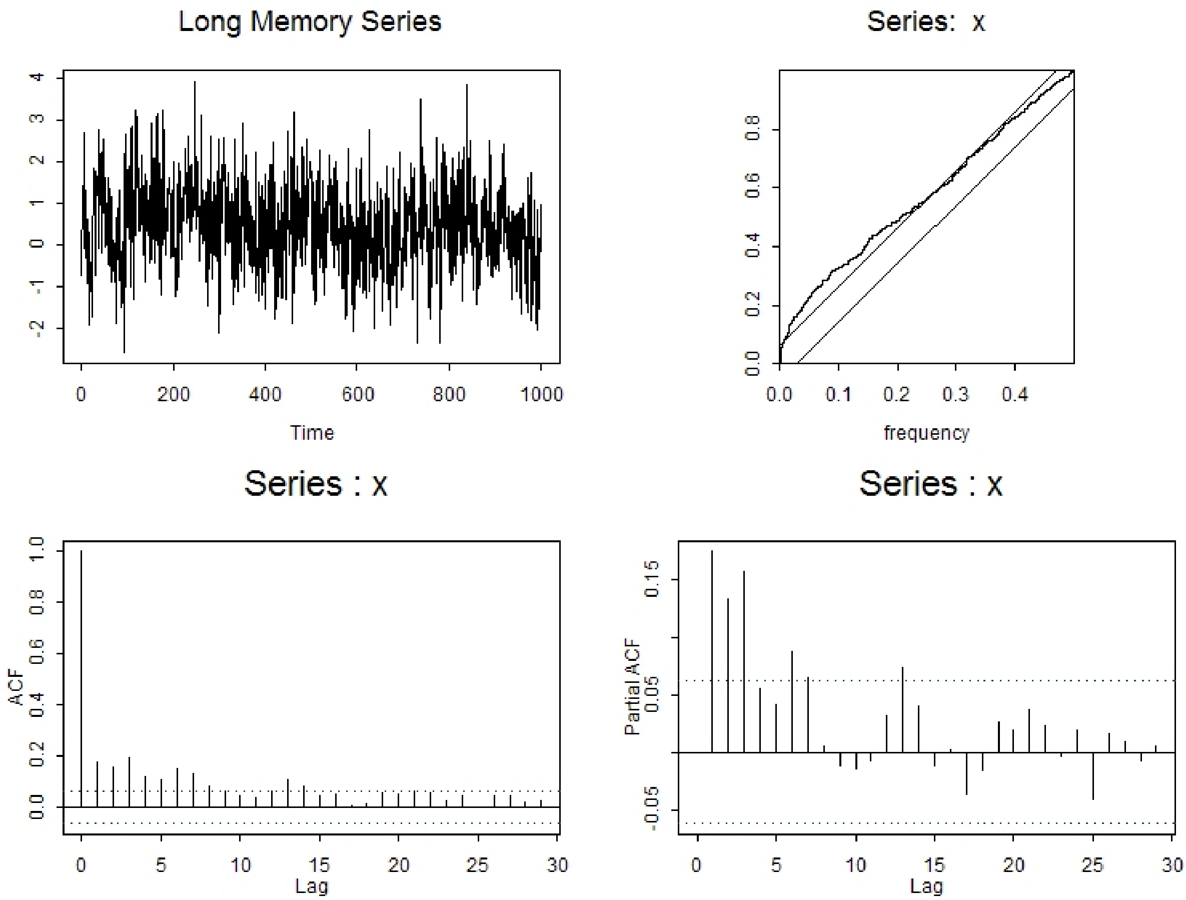
for $k = 1, 2, \ldots$
Shumway and Stoffer discuss the estimation of the parameter d on pages 168-171. R does not support any code for estimating the fractional differencing parameter d in the above model. The needed code can be found in Splus. Consider the following example using Splus

Simulated Long Memory Example

The Splus code is

```
 x<-arima.fracdiff.sim(model = list(d = .3, ar = .2, ma = .4), n = 1000)
 timeplot(x, c(2,2), 'Long Memory Series')
 arima.fracdiff(x, model = list(ar = NA, ma = NA))
  $model:
$model$ar: [1] 0.04868728

$model$ma: [1] 0.2153242

$model$d: [1] 0.2726246
```

```
$var.coef:
                d           ar1          ma1
  d  0.00003992280  0.00009521861 0.0001358843
 ar1 0.00009521861  0.03377665321 0.0327834580
 ma1 0.00013588434  0.03278345796 0.0328317538

$loglik: [1] -1423.468

$h: [1] 0.00001500921

$d.tol: [1] 0.0001220703

$M: [1] 100

$hess:
             d       ar1       ma1
  d  -26463.397 -1028.272  1136.287
 ar1  -1028.272 -1000.182  1002.967
 ma1   1136.287  1002.967 -1036.653

$call: arima.fracdiff(x = x, model = list(ar = NA, ma = NA))
```


Long Memory Series
Time
Series: x
frequency
Series : x
ACF
Lag
Series : x
Partial ACF
Lag

Chapter 7

Diagnostic Checking

The Box-Jenkins approach to modelling time series data is an iterative approach whereby one obtains an initial model identification using the ACF and PACF (and additional tools) to determine initial values for (p,d,q) in the ARIMA(p,d,q) model given by

$$\Phi_p(B)\nabla^d \dot{X}_t = \Theta_q(B)a_t.$$

For specified values of p,d,q one then obtains estimates for $\{\phi_1, \ldots, \phi_p\}$, $\{\theta_1, \ldots, \theta_q\}$ and σ_a^2. This is called the estimation stage. From here we have,

$$\hat{a}_t = \hat{\Theta}_q(B)^{-1}\hat{\Phi}_p(B)Z_t,$$

where $Z_t = \nabla^d \dot{X}_t$. If the model is correctly specified then

$$\hat{a}_t \sim N(0, \hat{\sigma}_a^2).$$

At this stage one can examine the estimated residuals $\hat{a}_t$ as a white noise series. This is done by considering the ACF for the residuals, given by $\{r_k = r_a(k) = corr(\hat{a}_t, \hat{a}_{t+k}\}$. The SAS USER's Guide has

Tests of Residuals

A table of test statistics for the hypothesis that the model residuals are white noise is printed as part of the ESTIMATE statement output. The chi-square statistics used in the test for lack of fit are computed using the Ljung-Box formula

$$\chi_m^2 = n(n+2)\sum_{k=1}^{m} \frac{r_k^2}{(n-k)}$$

where

$$r_k = \frac{\sum_{t=1}^{n-k} \hat{a}_t \hat{a}_{t+k}}{\sum_{t=1}^{n} \hat{a}_t^2},$$

and $\hat{a}_t$ is the residual series.
This formula has been suggested by Ljung and Box (1978) as yielding a better fit to the asymptotic chi-square distribution than the Box-Pierce Q statistic. Some simulation studies of the finite sample properties of this statistic are given by Davies, Triggs, and Newbold (1977) and by Ljung and Box (1978).
Each chi-square statistic is computed for all lags up to the indicated lag value and is not independent of the preceding chi-square values. The null hypotheses tested is that the current set of autocorrelations is white noise.

Often one might have acceptable results for the residual test when using several models. The question is which model should one use. In time series analysis the answer to this question is no clear cut for it is possible to have several competing models that provide good fits. In this case one relies upon additional criteria, such as the magnitude of the estimated standard error for the residuals, and the model selection criteria such as Akaike's AIC and BIC criteria.

Information Criteria

PROC ARIMA computes and prints two information criteria, Akaike's information criterion (AIC) (Akaike 1974; Harvey 1981) and Schwarz's Bayesian criterion (SBC) (Schwarz 1978). The AIC and SBC are used to compare competing models fit to the same series. The model with the smaller information criteria is said to fit the data better. The AIC is computed as

$$-2\ln(L) + 2k$$

where L is the likelihood function and k is the number of free parameters. The SBC is computed as

$$-2\ln(L) + \ln(n)k$$

where n is the number of residuals that can be computed for the time series. Sometimes Schwarz's Bayesian criterion is called the Bayesian Information criterion (BIC).

7.0.1 Example

I have included the following simple example using a data from Wei's text.

```
options linesize=75 center;
data a; N=_N_; input  z @@; datalines;
0.315 -0.458 -0.488 -0.170 0.565
-0.344 -1.176 -1.054 -0.826 0.710
-0.341 -1.809 -1.242 -0.667 -0.999
 2.812 1.286 -1.084 -1.505 -2.556
-0.144 -1.749 -3.032 -2.958 -2.827
 -3.392 -2.431 -2.757 -2.822
-3.314 -2.738 -1.979 -1.671 -2.977
 -0.709 0.718 0.736 0.879 1.642
2.180 1.963 0.716 0.769
0.973 0.334 1.309 0.878 0.062 0.169 0.677
1.851 0.242 0.828 -0.317 -1.042 -2.0930 0.6530 0.2610 2.0200
2.1360 1.6350 -0.1410 -1.7470 -2.0047 -0.7520 -0.2110 -1.0620
-1.5650 0.2320 0.0150 -0.9350 -0.3380 0.8530 0.8880 3.0690 3.3640
3.8540 4.4190 2.1450
2.2910 1.7530 1.0580 1.0480 0.2000
1.4240 0.5900 0.3560 0.4760 0.6840 -2.2600 -0.5690 -1.0140 -0.2070
0.6380 -0.6640 -0.4690 -0.2150 -0.2960 -1.5610 0.2460 ;
 symbol1 v=star i=join color=red;
  proc means;
   proc gplot; plot z*N=1;run;
proc arima;
   i var=z;
   e p=1 plot;
```

```
run;
```

The resultant output is

```
                          The MEANS Procedure

Variable     N          Mean        Std Dev     Min    Max

 N          100     50.5000000     29.0114920     1.00  100.00
 z          100     -0.1075970      1.6309537   -3.392  4.4190000


                          The ARIMA Procedure

                         Name of Variable = z

                    Mean of Working Series     -0.1076
                    Standard Deviation        1.622778
                    Number of Observations         100

                            Autocorrelations

Lag  Covariance  Correlation  -1 9 8 7 6 5 4 3 2 1 0 1 2 3 4 5 6 7
8 9 1

  0    2.633410      1.00000  |                    |********************|
  1    1.999020      0.75910  |                .   |***************     |
  2    1.586720      0.60253  |              .     |************        |
  3    1.284025      0.48759  |             .      |**********          |
  4    1.042326      0.39581  |             .      |********            |
  5    0.887709      0.33709  |            .       |*******.            |
  6    0.742284      0.28187  |            .       |****** .            |
  7    0.519781      0.19738  |            .       |****   .            |
  8    0.396955      0.15074  |            .       |***    .            |
  9    0.148691      0.05646  |            .       |*      .            |
 10   -0.053658      -.02038  |            .       |       .            |
 11   -0.057641      -.02189  |            .       |       .            |
 12   -0.322329      -.12240  |            .     **|       .            |
 13   -0.491971      -.18682  |            .   ****|       .            |
 14   -0.492720      -.18710  |            .   ****|       .            |
 15   -0.371030      -.14089  |            .    ***|       .            |
 16   -0.229414      -.08712  |            .     **|       .            |
 17   -0.299960      -.11391  |            .     **|       .            |
 18   -0.200568      -.07616  |            .     **|       .            |
 19   -0.094490      -.03588  |            .      *|       .            |
 20   -0.138715      -.05268  |            .      *|       .            |
 21   -0.208441      -.07915  |            .     **|       .            |
 22   -0.117934      -.04478  |            .      *|       .            |
 23   -0.126105      -.04789  |            .      *|       .            |
 24   -0.095753      -.03636  |            .      *|       .            |
```

"." marks two standard errors

Inverse Autocorrelations

```
Lag     Correlation    -1 9 8 7 6 5 4 3 2 1 0 1 2 3 4 5 6 7 8 9 1

  1        -0.44056    |           *********|   .                |
  2        -0.10290    |                . **|   .                |
  3         0.06032    |                .   |*  .                |
  4         0.03878    |                .   |*  .                |
  5         0.04793    |                .   |*  .                |
  6        -0.21385    |                ****|   .                |
  7         0.17056    |                .   |***.                |
  8        -0.07763    |                . **|   .                |
  9        -0.05048    |                .  *|   .                |
 10         0.19670    |                .   |****                |
 11        -0.21193    |                ****|   .                |
 12         0.05037    |                .   |*  .                |
 13         0.04739    |                .   |*  .                |
 14         0.09095    |                .   |** .                |
 15        -0.03787    |                .  *|   .                |
 16        -0.18170    |                ****|   .                |
 17         0.22496    |                .   |****                |
 18        -0.05616    |                .  *|   .                |
 19        -0.07469    |                .  *|   .                |
 20        -0.01235    |                .   |   .                |
 21         0.08001    |                .   |** .                |
 22        -0.02976    |                .  *|   .                |
 23        -0.01838    |                .   |   .                |
 24         0.01961    |                .   |   .                |
```

Partial Autocorrelations

```
Lag     Correlation    -1 9 8 7 6 5 4 3 2 1 0 1 2 3 4 5 6 7 8 9 1

  1         0.75910    |                .   |***************     |
  2         0.06207    |                .   |*  .                |
  3         0.02720    |                .   |*  .                |
  4         0.00735    |                .   |   .                |
  5         0.03922    |                .   |*  .                |
  6        -0.00276    |                .   |   .                |
  7        -0.08915    |                . **|   .                |
  8         0.02000    |                .   |   .                |
  9        -0.13863    |                .***|   .                |
 10        -0.05966    |                .  *|   .                |
 11         0.09343    |                .   |** .                |
 12        -0.22718    |               *****|   .                |
 13        -0.04493    |                .  *|   .                |
 14         0.07201    |                .   |*  .                |
```

```
    15        0.13167    |                .   |***.              |
    16        0.05415    |                .   |*  .              |
    17       -0.15207    |                .***|   .              |
    18        0.17190    |                .   |***.              |
    19        0.00873    |                .   |   .              |
    20       -0.10225    |                . **|   .              |
    21       -0.07527    |                . **|   .              |
    22        0.04245    |                .   |*  .              |
    23       -0.02037    |                .   |   .              |
    24       -0.03769    |                .  *|   .              |
```

Autocorrelation Check for White Noise

To Lag	Chi-Square	DF	Pr > ChiSq	---------------Autocorrelations---------------					
6	159.62	6	<.0001	0.759	0.603	0.488	0.396	0.337	0.282
12	168.61	12	<.0001	0.197	0.151	0.056	-0.020	-0.022	-0.122
18	182.47	18	<.0001	-0.187	-0.187	-0.141	-0.087	-0.114	-0.076
24	184.54	24	<.0001	-0.036	-0.053	-0.079	-0.045	-0.048	-0.036

Conditional Least Squares Estimation

Parameter	Estimate	Standard Error	t Value	Approx Pr > |t|	Lag
MU	-0.05014	0.41198	-0.12	0.9034	0
AR1,1	0.75981	0.06574	11.56	<.0001	1

Constant Estimate	-0.01204
Variance Estimate	1.137768
Std Error Estimate	1.066662
AIC	298.6742
SBC	303.8846
Number of Residuals	100

* AIC and SBC do not include log determinant.

Correlations of Parameter Estimates

Parameter	MU	AR1,1
MU	1.000	0.035
AR1,1	0.035	1.000

```
                                The ARIMA Procedure

                        Autocorrelation Check of Residuals

 To      Chi-         Pr >
Lag    Square   DF   ChiSq
---------------Autocorrelations---------------

  6      1.29    5  0.9358  -0.044   0.004   0.025  -0.005   0.040   0.089
 12      8.87   11  0.6334  -0.039   0.107  -0.026  -0.135   0.168  -0.081
 18     17.17   17  0.4428  -0.140  -0.109  -0.036   0.136  -0.132  -0.014
 24     20.73   23  0.5976   0.097   0.009  -0.118   0.058  -0.029  -0.005


                         Autocorrelation Plot of Residuals

Lag  Covariance  Correlation  -1 9 8 7 6 5 4 3 2 1 0 1 2 3 4 5 6 7
8 9 1

  0    1.137768      1.00000  |                    |********************|
  1   -0.049826      -.04379  |                .  *|   .                |
  2   0.0049934      0.00439  |                .   |   .                |
  3    0.027977      0.02459  |                .   |   .                |
  4  -0.0058909      -.00518  |                .   |   .                |
  5    0.045867      0.04031  |                .   |*  .                |
  6    0.100744      0.08855  |                .   |** .                |
  7   -0.044413      -.03904  |                .  *|   .                |
  8    0.122258      0.10745  |                .   |** .                |
  9   -0.029035      -.02552  |                .  *|   .                |
 10   -0.153061      -.13453  |                .***|   .                |
 11    0.191651      0.16844  |                .   |***.                |
 12   -0.092164      -.08100  |                . **|   .                |
 13   -0.159737      -.14039  |                .***|   .                |
 14   -0.124206      -.10917  |                . **|   .                |
 15   -0.040603      -.03569  |                .  *|   .                |
 16    0.155170      0.13638  |                .   |***.                |
 17   -0.150014      -.13185  |                .***|   .                |
 18   -0.016151      -.01420  |               .    |    .               |
 19    0.110586      0.09720  |               .    |**  .               |
 20    0.010383      0.00913  |               .    |    .               |
 21   -0.134062      -.11783  |               .  **|    .               |
 22    0.066500      0.05845  |               .    |*   .               |
 23   -0.033436      -.02939  |               .   *|    .               |
 24  -0.0056655      -.00498  |               .    |    .               |

                             Inverse Autocorrelations

     Lag    Correlation    -1 9 8 7 6 5 4 3 2 1 0 1 2 3 4 5 6 7 8 9 1
```

```
  1       -0.00576    |                .   |   .                |
  2       -0.07564    |                . **|   .                |
  3        0.04313    |                .   |*  .                |
  4        0.02753    |                .   |*  .                |
  5       -0.04778    |                .  *|   .                |
  6       -0.21278    |                ****|   .                |
  7        0.04422    |                .   |*  .                |
  8       -0.08875    |                . **|   .                |
  9       -0.02593    |                .  *|   .                |
 10        0.13883    |                .   |***.                |
 11       -0.10619    |                . **|   .                |
 12        0.06661    |                .   |*  .                |
 13        0.13578    |                .   |***.                |
 14        0.15921    |                .   |***.                |
 15       -0.02087    |                .   |   .                |
 16       -0.13017    |                .***|   .                |
 17        0.14700    |                .   |***.                |
 18       -0.04821    |                .  *|   .                |
 19       -0.10488    |                . **|   .                |
 20       -0.03133    |                .  *|   .                |
 21        0.06140    |                .   |*  .                |
 22       -0.00642    |                .   |   .                |
 23       -0.00717    |                .   |   .                |
 24        0.04735    |                .   |*  .                |


                     Partial Autocorrelations

Lag    Correlation    -1 9 8 7 6 5 4 3 2 1 0 1 2 3 4 5 6 7 8 9 1

  1       -0.04379    |                .  *|   .                |
  2        0.00248    |                .   |   .                |
  3        0.02494    |                .   |   .                |
  4       -0.00303    |                .   |   .                |
  5        0.03986    |                .   |*  .                |
  6        0.09184    |                .   |** .                |
  7       -0.03137    |                .  *|   .                |
  8        0.10300    |                .   |** .                |
  9       -0.02072    |                .   |   .                |
 10       -0.13945    |                .***|   .                |
 11        0.15195    |                .   |***.                |
 12       -0.07936    |                . **|   .                |
 13       -0.15496    |                .***|   .                |
 14       -0.14480    |                .***|   .                |
 15       -0.02337    |                .   |   .                |
 16        0.15526    |                .   |***.                |
 17       -0.16310    |                .***|   .                |
 18        0.04180    |                .   |*  .                |
```

```
19        0.11438    |                .   |** .               |
20        0.04337    |                .   |*  .               |
21       -0.07264    |                .  *|   .               |
22        0.00177    |                .   |   .               |
23        0.01071    |                .   |   .               |
24       -0.05765    |                .  *|   .               |
```

```
                     Model for variable z
                  Estimated Mean    -0.05014

                    Autoregressive Factors
                 Factor 1:  1 - 0.75981 B**(1)
```

The above data were modelled as a AR(1). Suppose one used a MA(1) instead. The results are;

```
                        The ARIMA Procedure

              Conditional Least Squares Estimation
```

Parameter	Estimate	Standard Error	t Value	Approx Pr > |t|	Lag
MU	-0.09211	0.20057	-0.46	0.6471	0
MA1,1	-0.59947	0.08202	-7.31	<.0001	1

```
                Constant Estimate      -0.09211
                Variance Estimate      1.592724
                Std Error Estimate     1.262032
                AIC                     332.312
                SBC                    337.5223
                Number of Residuals         100
            * AIC and SBC do not include log determinant.
```

```
                  Correlations of Parameter
                          Estimates
```

Parameter	MU	MA1,1
MU	1.000	-0.013
MA1,1	-0.013	1.000

```
                Autocorrelation Check of Residuals

 To      Chi-          Pr >
Lag    Square   DF   ChiSq
---------------Autocorrelations---------------
```

```
 6     60.50    5  <.0001   0.276   0.513   0.245   0.302   0.184   0.228
12     65.44   11  <.0001   0.089   0.132   0.035  -0.061   0.041  -0.110
18     73.56   17  <.0001  -0.130  -0.117  -0.128   0.016  -0.142  -0.013
24     75.36   23  <.0001  -0.035  -0.001  -0.098   0.018  -0.052   0.000
```

```
                   Autocorrelation Plot of Residuals

Lag  Covariance  Correlation  -1 9 8 7 6 5 4 3 2 1 0 1 2 3 4 5 6 7
8 9 1

  0    1.592724      1.00000  |                    |********************|
  1    0.439895      0.27619  |                .   |******              |
  2    0.817155      0.51306  |                .   |**********          |
  3    0.390476      0.24516  |               .    |*****               |
  4    0.480575      0.30173  |               .    |******              |
  5    0.292452      0.18362  |              .     |**** .              |
  6    0.362779      0.22777  |              .     |*****.              |
  7    0.141328      0.08873  |              .     |**   .              |
  8    0.209993      0.13185  |              .     |***  .              |
  9    0.056222      0.03530  |              .     |*    .              |
 10   -0.096479      -.06058  |              .    *|     .              |
 11    0.064563      0.04054  |              .     |*    .              |
 12   -0.175668      -.11029  |              .   **|     .              |
 13   -0.207346      -.13018  |              .  ***|     .              |
 14   -0.186695      -.11722  |              .   **|     .              |
 15   -0.204592      -.12845  |              .  ***|     .              |
 16    0.025239      0.01585  |              .     |     .              |
 17   -0.225775      -.14175  |              .  ***|     .              |
 18   -0.020762      -.01304  |              .     |     .              |
 19   -0.055211      -.03466  |              .    *|     .              |
 20  -0.0014871      -.00093  |              .     |     .              |
 21   -0.155945      -.09791  |              .   **|     .              |
 22    0.028984      0.01820  |              .     |     .              |
 23   -0.083444      -.05239  |              .    *|     .              |
 24  0.00042879      0.00027  |              .     |     .              |

                     "." marks two standard errors

                        Inverse Autocorrelations

     Lag    Correlation    -1 9 8 7 6 5 4 3 2 1 0 1 2 3 4 5 6 7 8 9 1

       1       -0.06909    |                .  *|   .                |
       2       -0.42482    |            ********|   .                |
       3        0.04087    |                .   |*  .                |
       4        0.12641    |                .   |***.                |
       5       -0.04720    |                .  *|   .                |
```

```
  6       -0.17667    |                ****|   .                |
  7        0.05633    |                .   |*  .                |
  8       -0.04769    |                .  *|   .                |
  9        0.01143    |                .   |   .                |
 10        0.12654    |                .   |***.                |
 11       -0.15076    |                .***|   .                |
 12       -0.03816    |                .  *|   .                |
 13        0.16279    |                .   |***.                |
 14        0.16165    |                .   |***.                |
 15       -0.11375    |                . **|   .                |
 16       -0.17028    |                .***|   .                |
 17        0.18615    |                .   |****                |
 18        0.01696    |                .   |   .                |
 19       -0.15940    |                .***|   .                |
 20       -0.02248    |                .   |   .                |
 21        0.09720    |                .   |** .                |
 22       -0.00381    |                .   |   .                |
 23       -0.02893    |                .  *|   .                |
 24        0.02567    |                .   |*  .                |


                     Partial Autocorrelations

Lag    Correlation    -1 9 8 7 6 5 4 3 2 1 0 1 2 3 4 5 6 7 8 9 1

  1        0.27619    |                .   |******              |
  2        0.47284    |                .   |*********           |
  3        0.05559    |                .   |*  .                |
  4        0.02613    |                .   |*  .                |
  5        0.00414    |                .   |   .                |
  6        0.06011    |                .   |*  .                |
  7       -0.07006    |                .  *|   .                |
  8       -0.01918    |                .   |   .                |
  9       -0.02780    |                .  *|   .                |
 10       -0.18222    |                ****|   .                |
 11        0.07824    |                .   |** .                |
 12       -0.05299    |                .  *|   .                |
 13       -0.17723    |                ****|   .                |
 14        0.00819    |                .   |   .                |
 15        0.04334    |                .   |*  .                |
 16        0.21086    |                .   |****                |
 17       -0.12085    |                . **|   .                |
 18        0.00671    |                .   |   .                |
 19        0.11800    |                .   |** .                |
 20       -0.01272    |                .   |   .                |
 21       -0.11874    |                . **|   .                |
 22       -0.02291    |                .   |   .                |
 23        0.03344    |                .   |*  .                |
 24       -0.03947    |                .  *|   .                |
```

```
Model for variable z
Estimated Mean    -0.09211

Moving Average Factors
Factor 1:  1 + 0.59947 B**(1)
```

As one would have hoped this model does not fit the data. What happens when one overfits the data. Suppose one tries an AR(2) instead of the AR(1).

```
                              The ARIMA Procedure

                     Conditional Least Squares Estimation

                             Standard                 Approx
  Parameter      Estimate       Error    t Value    Pr > |t|     Lag

  MU             -0.05329     0.43390      -0.12      0.9025       0
  AR1,1           0.71464     0.10235       6.98      <.0001       1
  AR1,2           0.05927     0.10278       0.58      0.5655       2


                        Constant Estimate      -0.01205
                        Variance Estimate       1.14557
                        Std Error Estimate     1.070313
                        AIC                    300.3321
                        SBC                    308.1476
                        Number of Residuals         100
                   * AIC and SBC do not include log determinant.


                      Correlations of Parameter Estimates

                    Parameter        MU     AR1,1     AR1,2

                    MU            1.000     0.021    -0.001
                    AR1,1         0.021     1.000    -0.765
                    AR1,2        -0.001    -0.765     1.000


                      Autocorrelation Check of Residuals

 To      Chi-         Pr >
Lag    Square   DF   ChiSq
---------------Autocorrelations---------------

  6      0.94    4  0.9189  -0.002  -0.022   0.005  -0.018   0.034   0.082
 12      8.00   10  0.6286  -0.035   0.101  -0.031  -0.130   0.164  -0.078
 18     16.65   16  0.4087  -0.154  -0.120  -0.029   0.133  -0.124  -0.015
```

```
 24     20.20   22  0.5707   0.103   0.010  -0.116   0.053  -0.026  -0.005
```

```
                   Autocorrelation Plot of Residuals

Lag  Covariance  Correlation  -1 9 8 7 6 5 4 3 2 1 0 1 2 3 4 5 6 7
8 9 1

  0    1.145570      1.00000  |                    |********************|
  1  -0.0019043      -.00166  |                .   |   .                |
  2   -0.024840      -.02168  |                .   |   .                |
  3   0.0054150      0.00473  |                .   |   .                |
  4   -0.020247      -.01767  |                .   |   .                |
  5    0.039242      0.03426  |                .   |*  .                |
  6    0.093828      0.08191  |                .   |** .                |
  7   -0.040205      -.03510  |                .  *|   .                |
  8    0.115424      0.10076  |                .   |** .                |
  9   -0.035318      -.03083  |                .  *|   .                |
 10   -0.149107      -.13016  |                .***|   .                |
 11    0.187575      0.16374  |                .   |***.                |
 12   -0.089261      -.07792  |                . **|   .                |
 13   -0.176918      -.15444  |                .***|   .                |
 14   -0.137710      -.12021  |                . **|   .                |
 15   -0.032978      -.02879  |                .  *|   .                |
 16    0.151899      0.13260  |                .   |***.                |
 17   -0.141519      -.12354  |                . **|   .                |
 18   -0.016778      -.01465  |               .    |    .               |
 19    0.118453      0.10340  |               .    |**  .               |
 20    0.011604      0.01013  |               .    |    .               |
 21   -0.132645      -.11579  |               .  **|    .               |
 22    0.060348      0.05268  |               .    |*   .               |
 23   -0.030273      -.02643  |               .   *|    .               |
 24  -0.0053889      -.00470  |               .    |    .               |
```

```
                        Inverse Autocorrelations

     Lag    Correlation    -1 9 8 7 6 5 4 3 2 1 0 1 2 3 4 5 6 7 8 9 1

       1       -0.04989    |                .  *|   .                |
       2       -0.05803    |                .  *|   .                |
       3        0.05636    |                .   |*  .                |
       4        0.03505    |                .   |*  .                |
       5       -0.02760    |                .  *|   .                |
       6       -0.20696    |                ****|   .                |
       7        0.06490    |                .   |*  .                |
       8       -0.08377    |                . **|   .                |
       9       -0.02684    |                .  *|   .                |
      10        0.14877    |                .   |***.                |
      11       -0.11536    |                . **|   .                |
```

```
 12        0.06963    |                .   |*  .                |
 13        0.12495    |                .   |** .                |
 14        0.15178    |                .   |***.                |
 15       -0.01767    |                .   |   .                |
 16       -0.13127    |                .***|   .                |
 17        0.15747    |                .   |***.                |
 18       -0.04858    |                .  *|   .                |
 19       -0.09711    |                . **|   .                |
 20       -0.02728    |                .  *|   .                |
 21        0.06296    |                .   |*  .                |
 22       -0.00888    |                .   |   .                |
 23       -0.00793    |                .   |   .                |
 24        0.04441    |                .   |*  .                |


                     Partial Autocorrelations

Lag    Correlation    -1 9 8 7 6 5 4 3 2 1 0 1 2 3 4 5 6 7 8 9 1

  1       -0.00166    |                .   |   .                |
  2       -0.02169    |                .   |   .                |
  3        0.00466    |                .   |   .                |
  4       -0.01814    |                .   |   .                |
  5        0.03443    |                .   |*  .                |
  6        0.08135    |                .   |** .                |
  7       -0.03332    |                .  *|   .                |
  8        0.10445    |                .   |** .                |
  9       -0.03267    |                .  *|   .                |
 10       -0.12581    |                .***|   .                |
 11        0.16164    |                .   |***.                |
 12       -0.09438    |                . **|   .                |
 13       -0.15779    |                .***|   .                |
 14       -0.14254    |                .***|   .                |
 15       -0.01192    |                .   |   .                |
 16        0.13848    |                .   |***.                |
 17       -0.17583    |                ****|   .                |
 18        0.05432    |                .   |*  .                |
 19        0.11223    |                .   |** .                |
 20        0.03298    |                .   |*  .                |
 21       -0.07248    |                .  *|   .                |
 22        0.00846    |                .   |   .                |
 23        0.00671    |                .   |   .                |
 24       -0.05402    |                .  *|   .                |


                      Model for variable z
                   Estimated Mean    -0.05329

                      Autoregressive Factors
```

```
Factor 1:  1 - 0.71464 B**(1) - 0.05927 B**(2)
```

Chapter 8

Transfer Function Models

8.1 General Concepts

Assume that x_t and y_t are stationary time series. The general single-input, single-output linear system is given by,

$$y_t = \nu(B)x_t + n_t$$

where $\nu(B) = \sum_{j=-\infty}^{\infty} \nu_j B^j$ which is referred to as the transfer function, and n_t is the noise series that is independent of the input series x_t. When one assumes that both x_t and n_t are ARMA models then the above equation is sometimes called ARMAX model.
The ARMAX models need conditions similar to stationarity in the ARMA model which is given by,

$$y_t = \nu(B)x_t + n_t$$

where $\nu(B) = \sum_{j=-\infty}^{\infty} \nu_j B^j$,$\sum_{j=0}^{\infty} \mid \nu_j \mid < \infty$ and x_t and n_t are independent of one another. In this course we will assume that

$$\nu(B) = \frac{\omega_s(B)B^b}{\delta_r(B)}$$

where $\omega_s(B) = (\omega_0 - \omega_1 B - \omega_2 B^2 - \cdots - \omega_s B^s)$, $\delta_r(B) = (1 - \delta_1 B - \cdots - \delta_r B^r)$, and b is the lag or delay parameter. This model would be called the (b, r, s) model. It follows that

$$\delta_r(B)\nu(B) = \omega_s(B)B^b,$$

or

$$(1 - \delta_1 B - \cdots - \delta_r B^r)(\nu_0 + \nu_1 B + \nu_2 B^2 + \cdots) = (\omega_0 B^b - \omega_1 B^{b+1} - \cdots - \omega_s B^{b+s}).$$

From which one has,

$$\begin{aligned}
&\nu_j = 0 && j < b,\\
&\nu_j = \textstyle\sum_{i=1}^{r} \delta_i \nu_{j-i} + \omega_0 && j = b,\\
&\nu_j = \textstyle\sum_{i=1}^{r} \delta_i \nu_{j-i} + \omega_{j-b} && j = b+1, b+2 \cdots b+s,\\
&\nu_j = \textstyle\sum_{i=1}^{r} \delta_i \nu_{j-i} && j > b+s.
\end{aligned}$$

From the above one has that the parameter b is the location of the first nonzero impulse and the r impulse response weights $\nu_{b+s}, \nu_{b+s-1}, \cdots, \nu_{b+s-r+1}$ serve as starting values for the homogenous difference equation

$$\delta_r(B)\nu_j = 0, \quad j > b + s.$$

8.2 The Cross-Correlation Function (CCF)

Define the cross autocovariance function as

$$\gamma_{xy}(k) = cov(x_t, y_{t+k}) = E[(x_t - \mu_x)(y_{t+k} - \mu_y)],$$

for $k = 0, \pm 1, \pm 2, \ldots$. From which one has the CCF as,

$$\rho_{xy}(k) = \frac{\gamma_{xy}(k)}{\sigma_x \sigma_y},$$

where σ_x, and σ_y are the standard deviations of x_t and y_t, respectively.
It can be shown that $\rho_{xy}(k) \neq \rho_{xy}(-k)$ but rather that

$$\rho_{xy}(k) = \rho_{yx}(-k)$$

and

$$\rho_{xy}(-k) = \rho_{yx}(k).$$

8.2.1 Relationship Between the CCF and the Transfer Function

Recall that

$$y_t = \nu(B)x_t + n_t$$

where $\nu(B) = \sum_{j=-\infty}^{\infty} \nu_j B^j$ and n_t is the noise series that is independent of the input series x_t. From which one has

$$y_{t+k} = \nu_0 x_{t+k} + \nu_1 x_{t+k-1} + \nu_2 x_{t+k-2} + n_{t+k}.$$

From which one has,

$$\gamma_{xy}(k) = \nu_0 \gamma_{xx}(k) + \nu_1 \gamma_{xx}(k-1) + \nu_2 \gamma_{xx}(k-2) + \ldots,$$

or

$$\rho_{xy}(k) = \frac{\sigma_x}{\sigma_y}[\nu_0 \rho_x(k) + \nu_1 \rho_x(k-1) + \nu_2 \rho_x(k-2) + \ldots].$$

From this equation, there is no way that one can understanding what $\rho_{xy}(k)$ should look like as it depends upon $\rho_x(k)$. Suppose that x_t is white noise then

$$\rho_{xx}(k) = \frac{\sigma_x}{\sigma_y}\nu_k$$

or

$$\nu_k = \frac{\sigma_y}{\sigma_x}\rho_{xy}(k).$$

Which implies that the CCF has the same shape as the impulse response function from which one has

$$\hat{\nu}_k = \frac{\hat{\sigma}_y}{\hat{\sigma}_x}\hat{\rho}_{xy}(k).$$

8.3 Estimation of the CCF

One can estimate,

$$\rho_{xy}(k) = \frac{\gamma_{xy}(k)}{\sigma_x \sigma_y},$$

with

$$\hat{\rho}_{xy}(k) = \frac{\hat{\gamma}_{xy}(k)}{s_x s_y},$$

where

$$\begin{aligned}\hat{\gamma}_{xy}(k) &= \frac{1}{n}\sum_{t=1}^{n-k}(x_t - \bar{x})(y_{t+k} - \bar{y}), \quad k \geq 0 \\ &= \frac{1}{n}\sum_{t=1-k}^{n}(x_t - \bar{x})(y_{t+k} - \bar{y}), \quad k < 0\end{aligned}$$

Where $s_x = \sqrt{\hat{\gamma}_{xx}(0)}$ and $s_y = \sqrt{\hat{\gamma}_{yy}(0)}$. As in the univariate case the CCF are not independent and the variance of the estimators depend upon the quantities that are being estimated. Through the use of some simplifying assumptions one has,

$$Var[\hat{\rho}_{xy}(k)] \simeq (n-k)^{-1}.$$

8.4 Identification of Transfer Function Models

Assume that both x_t and y_t are stationary.

1. Prewhiten the input series x_t. That is, suppose that x_t is a $ARMA(p_x, q_x)$, that is,

$$\Phi_{p_x}(B)x_t = \Theta_{q_x}(B)\alpha_t,$$

   or

$$\alpha_t = [\frac{\Phi_{p_x}(B)}{\Theta_{q_x}(B)}]x_t.$$

2. Define,

$$\beta_t = [\frac{\Phi_{p_x}(B)}{\Theta_{q_x}(B)}]y_t.$$

3. Calculate the estimated CCF between series α_t and β_t in order to estimate the impulse response $\hat{\nu}_k$,

$$\hat{\nu}_k = \frac{\hat{\sigma}_\beta}{\hat{\sigma}_\alpha}\hat{\rho}_{\alpha\beta}(k).$$

4. From $\hat{\nu}_k$ identify b, r, s and compute

$$\hat{\nu}(B) = \frac{\hat{\omega}_s(B)B^b}{\hat{\delta}_r(B)}.$$

5. Calculate $\hat{n}_t$ by

$$\hat{n}_t = y_t - \hat{\nu}(B)x_t = y_t - [\frac{\hat{\omega}_s(B)B^b}{\hat{\delta}_r(B)}]x_t.$$

   And identify $\hat{n}_t$ as a $ARMA(p_n, q_n)$ with white noise a_t.

6. The final model becomes,

$$y_t = [\frac{\hat{\omega}_s(B)}{\hat{\delta}_r(B)}]x_{t-b} + [\frac{\Phi_{p_n}(B)}{\Theta_{q_n}(B)}]a_t.$$

8.4.1 Housing Starts-Sales Example

This example is given in Wei's text in problem 13.5 on page 330 concerning monthly housing sales X_t, and housing starts Y_t, between January 1965 and December 1975. Note: it should be noted that the results given below and the models selection are the result of several runs and possible models. The SAS code is

```
options linesize=75 pagesize=60 center; data a; N=_N_;

infile 'C:\Documents and Settings\Jack_Tubbs\My
Documents\Baylor_stat_classes\Time_series\Timeseries\Wei\Wei_House/weihouse-starts.dat';

input  starts @@; data b; N=_N_;

infile 'C:\Documents and Settings\Jack_Tubbs\My
Documents\Baylor_stat_classes\Time_series\Timeseries\Wei\Wei_House/weihouse-sales.dat';

input sales @@; data c; merge a b; by N;

*proc means;
 symbol1 c=red i=join v=star;
 symbol2 c=green i=join v=plus;
 symbol3 c=blue i=join v=none;

 proc gplot; plot starts*N=1 sales*N=2 /overlay;

 proc arima data=c; /*  Look at the input series ads  */
   i var=sales(1,12) ;
 /*  Fit a model to the input series ads  */
   e   q=(1,12)  noint;
 /*  Crosscorrelation of prewhitened series  */
   i var=starts(1,12) crosscor=(sales(1,12)) ;
 /*  Fit the model with b=1 r=0 s=0  */
   e    p=1 q=(12) input=(1$  sales) noint;
   f    lead=24 back=12  id=N out=out1;
   f    lead=12 id=N out=out2;

 proc gplot data=out1; where n > 100;
  plot starts*n=1 forecast*n=2/overlay;
 proc gplot data=out2; where n > 100;
  plot starts*n=1 forecast*n=2 l95*n=3 u95*n=3/overlay;
 run;
```

With SAS output,

```
                              The SAS System
                             The ARIMA Procedure

                          Name of Variable = sales

          Period(s) of Differencing                        1,12
          Mean of Working Series                       0.092437
```

```
        Standard Deviation                        4.784414
        Number of Observations                         119
        Observation(s) eliminated by differencing       13


                           Autocorrelations

Lag  Covariance  Correlation  -1 9 8 7 6 5 4 3 2 1 0 1 2 3 4 5 6 7 8 9 1

  0   22.890615     1.00000   |                    |********************|
  1   -3.179791     -.13891   |                .***|   .                |
  2    0.738434     0.03226   |                .   |*  .                |
  3   -1.836809     -.08024   |                . **|   .                |
  4    3.635120     0.15880   |                .   |***.                |
  5    0.874085     0.03819   |                .   |*  .                |
  6    2.005501     0.08761   |                .   |** .                |
  7   -2.606604     -.11387   |                . **|   .                |
  8   -1.200984     -.05247   |                .  *|   .                |
  9   -0.684426     -.02990   |                .  *|   .                |
 10    1.489643     0.06508   |                .   |*  .                |
 11   -0.468412     -.02046   |                .   |   .                |
 12   -8.250773     -.36044   |             *******|   .                |
 13   -0.643473     -.02811   |                .  *|   .                |
 14    0.045531     0.00199   |                .   |   .                |
 15    2.452704     0.10715   |                .   |** .                |
 16   -0.498360     -.02177   |                .   |   .                |
 17   -0.318854     -.01393   |                .   |   .                |
 18   -1.331848     -.05818   |                .  *|   .                |
 19    3.972155     0.17353   |                .   |***.                |
 20   -3.360945     -.14683   |               . ***|    .               |
 21    3.254759     0.14219   |               .    |*** .               |
 22   -1.806466     -.07892   |               .  **|    .               |
 23    2.505940     0.10947   |               .    |**  .               |
 24   -2.881463     -.12588   |               . ***|    .               |

                    "." marks two standard errors


                       Inverse Autocorrelations

    Lag    Correlation    -1 9 8 7 6 5 4 3 2 1 0 1 2 3 4 5 6 7 8 9 1

      1        0.24035    |                .   |*****               |
      2        0.01881    |                .   |   .                |
      3       -0.05824    |                .  *|   .                |
      4       -0.12575    |                .***|   .                |
      5       -0.18538    |                ****|   .                |
      6       -0.22956    |               *****|   .                |
      7       -0.10702    |                . **|   .                |
```

```
  8        0.01024    |                .   |   .                |
  9        0.00926    |                .   |   .                |
 10        0.02351    |                .   |   .                |
 11        0.17920    |                .   |****                |
 12        0.46261    |                .   |*********           |
 13        0.14259    |                .   |***.                |
 14       -0.02482    |                .   |   .                |
 15       -0.07805    |                . **|   .                |
 16       -0.10466    |                . **|   .                |
 17       -0.06208    |                .  *|   .                |
 18       -0.09859    |                . **|   .                |
 19       -0.06848    |                .  *|   .                |
 20        0.08914    |                .   |** .                |
 21       -0.00687    |                .   |   .                |
 22        0.00938    |                .   |   .                |
 23        0.02688    |                .   |*  .                |
 24        0.16577    |                .   |***.                |
```

Partial Autocorrelations

```
Lag    Correlation    -1 9 8 7 6 5 4 3 2 1 0 1 2 3 4 5 6 7 8 9 1

  1       -0.13891    |                .***|   .                |
  2        0.01322    |                .   |   .                |
  3       -0.07545    |                . **|   .                |
  4        0.14084    |                .   |***.                |
  5        0.08339    |                .   |** .                |
  6        0.09893    |                .   |** .                |
  7       -0.07455    |                .  *|   .                |
  8       -0.10209    |                . **|   .                |
  9       -0.06388    |                .  *|   .                |
 10        0.01441    |                .   |   .                |
 11        0.00455    |                .   |   .                |
 12       -0.36419    |             *******|   .                |
 13       -0.11104    |                . **|   .                |
 14       -0.01310    |                .   |   .                |
 15        0.08887    |                .   |** .                |
 16        0.11609    |                .   |** .                |
 17        0.07874    |                .   |** .                |
 18        0.03701    |                .   |*  .                |
 19        0.12650    |                .   |***.                |
 20       -0.21991    |                ****|   .                |
 21        0.00892    |                .   |   .                |
 22       -0.02016    |                .   |   .                |
 23        0.05017    |                .   |*  .                |
 24       -0.25323    |               *****|   .                |
```

The ARIMA Procedure

Autocorrelation Check for White Noise

To Lag	Chi-Square	DF	Pr > ChiSq	---------------Autocorrelations---------------					
6	7.60	6	0.2687	-0.139	0.032	-0.080	0.159	0.038	0.088
12	27.84	12	0.0058	-0.114	-0.052	-0.030	0.065	-0.020	-0.360
18	30.12	18	0.0363	-0.028	0.002	0.107	-0.022	-0.014	-0.058
24	45.68	24	0.0048	0.174	-0.147	0.142	-0.079	0.109	-0.126

Conditional Least Squares Estimation

Parameter	Estimate	Standard Error	t Value	Approx Pr > |t|	Lag
MA1,1	0.17392	0.06525	2.67	0.0088	1
MA1,2	0.73015	0.07022	10.40	<.0001	12

Variance Estimate	15.88194
Std Error Estimate	3.985215
AIC	668.7471
SBC	674.3053
Number of Residuals	119

* AIC and SBC do not include log determinant.

Correlations of Parameter Estimates

Parameter	MA1,1	MA1,2
MA1,1	1.000	-0.061
MA1,2	-0.061	1.000

Autocorrelation Check of Residuals

To Lag	Chi-Square	DF	Pr > ChiSq	---------------Autocorrelations---------------					
6	2.84	4	0.5855	-0.056	0.042	0.000	0.047	-0.034	0.119
12	7.03	10	0.7225	-0.056	-0.134	0.020	0.022	-0.095	0.032
18	10.53	16	0.8377	-0.012	0.023	0.103	-0.054	-0.101	-0.024
24	23.21	22	0.3900	0.127	-0.183	0.120	-0.087	0.056	-0.109

```
                   Model for variable sales

               Period(s) of Differencing    1,12

                  No mean term in this model.

                    Moving Average Factors

       Factor 1:  1 - 0.17392 B**(1) - 0.73015 B**(12)


                    Name of Variable = starts

    Period(s) of Differencing                           1,12
    Mean of Working Series                          0.168908
    Standard Deviation                              10.41034
    Number of Observations                               119
    Observation(s) eliminated by differencing             13


                         Autocorrelations

Lag  Covariance  Correlation  -1 9 8 7 6 5 4 3 2 1 0 1 2 3 4 5 6 7 8 9 1

  0     108.375      1.00000  |                    |********************|
  1  -50.520017      -.46616  |           *********|   .                |
  2   21.244070      0.19602  |                .   |****                |
  3   -8.797187      -.08117  |               .  **|    .               |
  4    4.160342      0.03839  |               .    |*   .               |
  5    7.731932      0.07134  |               .    |*   .               |
  6    0.691688      0.00638  |               .    |    .               |
  7   -1.919851      -.01771  |               .    |    .               |
  8  -11.782844      -.10872  |               .  **|    .               |
  9   22.983795      0.21208  |               .    |****.               |
 10  -15.109063      -.13941  |               . ***|    .               |
 11   21.784301      0.20101  |               .    |****.               |
 12  -45.409081      -.41900  |            ********|    .               |
 13   22.657154      0.20906  |               .    |****.               |
 14   -1.823162      -.01682  |               .    |    .               |
 15    5.935948      0.05477  |               .    |*   .               |
 16  -13.926780      -.12851  |               . ***|    .               |
 17    4.280473      0.03950  |              .     |*    .              |
 18   -0.164335      -.00152  |              .     |     .              |
 19   -7.550818      -.06967  |              .    *|     .              |
 20   16.852092      0.15550  |              .     |***  .              |
 21  -16.550459      -.15271  |              .  ***|     .              |
 22   -6.236998      -.05755  |              .    *|     .              |
 23   17.300168      0.15963  |              .     |***  .              |
```

```
24  -10.839925      -.10002  |               .   **|     .               |

                 "." marks two standard errors


                    Inverse Autocorrelations

   Lag     Correlation   -1 9 8 7 6 5 4 3 2 1 0 1 2 3 4 5 6 7 8 9 1

     1         0.22668    |                .   |*****               |
     2        -0.17961    |                ****|   .                |
     3        -0.10772    |                . **|   .                |
     4         0.07292    |                .   |*  .                |
     5        -0.07988    |                . **|   .                |
     6        -0.08346    |                . **|   .                |
     7         0.01952    |                .   |   .                |
     8         0.05102    |                .   |*  .                |
     9        -0.17310    |                .***|   .                |
    10        -0.09932    |                . **|   .                |
    11         0.13146    |                .   |***.                |
    12         0.41948    |                .   |********            |
    13        -0.01659    |                .   |   .                |
    14        -0.12347    |                . **|   .                |
    15        -0.02268    |                .   |   .                |
    16         0.07985    |                .   |** .                |
    17        -0.03990    |                .  *|   .                |
    18         0.00949    |                .   |   .                |
    19         0.04275    |                .   |*  .                |
    20        -0.03047    |                .  *|   .                |
    21        -0.03985    |                .  *|   .                |
    22         0.02147    |                .   |   .                |
    23         0.01414    |                .   |   .                |
    24         0.11771    |                .   |** .                |


                    Partial Autocorrelations

   Lag     Correlation   -1 9 8 7 6 5 4 3 2 1 0 1 2 3 4 5 6 7 8 9 1

     1        -0.46616    |           *********|   .                |
     2        -0.02719    |                .  *|   .                |
     3         0.00002    |                .   |   .                |
     4         0.00621    |                .   |   .                |
     5         0.11487    |                .   |** .                |
     6         0.10844    |                .   |** .                |
     7         0.01402    |                .   |   .                |
     8        -0.16066    |                .***|   .                |
     9         0.12543    |                .   |***.                |
    10         0.02300    |                .   |   .                |
```

```
     11       0.14978    |                .   |***.               |
     12      -0.35562    |             *******|   .               |
     13      -0.16657    |                .***|   .               |
     14       0.08700    |                .   |** .               |
     15       0.13697    |                .   |***.               |
     16      -0.13877    |                .***|   .               |
     17      -0.00647    |                .   |   .               |
     18       0.08468    |                .   |** .               |
     19      -0.10688    |                . **|   .               |
     20      -0.04544    |                .  *|   .               |
     21       0.14190    |                .   |***.               |
     22      -0.14711    |                .***|   .               |
     23       0.12914    |                .   |***.               |
     24      -0.19030    |                ****|   .               |


                    Autocorrelation Check for White Noise

 To      Chi-         Pr >
Lag    Square   DF   ChiSq
---------------Autocorrelations---------------

  6     32.90    6  <.0001  -0.466   0.196  -0.081   0.038   0.071   0.006
 12     71.94   12  <.0001  -0.018  -0.109   0.212  -0.139   0.201  -0.419
 18     80.86   18  <.0001   0.209  -0.017   0.055  -0.129   0.039  -0.002
 24     94.33   24  <.0001  -0.070   0.155  -0.153  -0.058   0.160  -0.100

                    Variable sales has been differenced.


                       Correlation of starts and sales

           Period(s) of Differencing                         1,12
           Number of Observations                             119
           Observation(s) eliminated by differencing           13
           Variance of transformed series starts         62.98702
           Variance of transformed series sales           15.5198

                     Both series have been prewhitened.


                              Crosscorrelations

 Lag  Covariance  Correlation  -1 9 8 7 6 5 4 3 2 1 0 1 2 3 4 5 6 7 8 9 1

 -24   -1.666084      -.05329  |                .  *|   .               |
 -23   -0.051880      -.00166  |                .   |   .               |
 -22    2.568955      0.08217  |                .   |** .               |
 -21    0.361254      0.01155  |                .   |   .               |
```

```
-20   1.986899    0.06355  |                .   |*  .                |
-19  -4.211229    -.13469  |                .***|   .                |
-18   0.682247    0.02182  |                .   |   .                |
-17  -3.702894    -.11843  |                . **|   .                |
-16   0.372988    0.01193  |                .   |   .                |
-15  -4.122029    -.13184  |                .***|   .                |
-14   1.618335    0.05176  |                .   |*  .                |
-13   0.029303    0.00094  |                .   |   .                |
-12   0.143274    0.00458  |                .   |   .                |
-11   0.081786    0.00262  |                .   |   .                |
-10   2.775699    0.08878  |                .   |** .                |
 -9  -1.149301    -.03676  |                .  *|   .                |
 -8   1.098890    0.03515  |                .   |*  .                |
 -7   4.096438    0.13102  |                .   |***.                |
 -6  -7.013329    -.22431  |                ****|   .                |
 -5   3.005521    0.09613  |                .   |** .                |
 -4  -1.402161    -.04485  |                .  *|   .                |
 -3   1.869780    0.05980  |                .   |*  .                |
 -2  -0.050546    -.00162  |                .   |   .                |
 -1   3.270866    0.10462  |                .   |** .                |
  0   2.320755    0.07423  |                .   |*  .                |
  1   8.930287    0.28563  |                .   |******              |
  2   3.762984    0.12035  |                .   |** .                |
  3   3.615705    0.11564  |                .   |** .                |
  4   1.353609    0.04329  |                .   |*  .                |
  5   3.431290    0.10975  |                .   |** .                |
  6  -1.708064    -.05463  |                .  *|   .                |
  7  -0.078543    -.00251  |                .   |   .                |
  8  -1.386109    -.04433  |                .  *|   .                |
  9   0.376658    0.01205  |                .   |   .                |
 10   0.577312    0.01846  |                .   |   .                |
 11  -0.679947    -.02175  |                .   |   .                |
 12  -2.037166    -.06516  |                .  *|   .                |
 13  -0.118029    -.00378  |                .   |   .                |
 14  -0.663166    -.02121  |                .   |   .                |
 15   1.702981    0.05447  |                .   |*  .                |
 16   3.844330    0.12296  |                .   |** .                |
 17  -3.921551    -.12543  |                .***|   .                |
 18   5.035030    0.16104  |                .   |***.                |
 19  -3.901387    -.12478  |                . **|   .                |
 20  -1.420319    -.04543  |                .  *|   .                |
 21  -0.109009    -.00349  |                .   |   .                |
 22  -0.522109    -.01670  |                .   |   .                |
 23  -2.854428    -.09130  |                . **|   .                |
 24   3.282835    0.10500  |                .   |** .                |

                    "." marks two standard errors
```

```
                   Crosscorrelation Check Between Series

 To      Chi-         Pr >
Lag    Square   DF   ChiSq
---------------Crosscorrelations--------------

  5     15.34    6  0.0178   0.074   0.286   0.120   0.116   0.043   0.110
 11     16.04   12  0.1894  -0.055  -0.003  -0.044   0.012   0.018  -0.022
 17     20.62   18  0.2988  -0.065  -0.004  -0.021   0.054   0.123  -0.125
 23     26.84   24  0.3122   0.161  -0.125  -0.045  -0.003  -0.017  -0.091

       Both variables have been prewhitened by the following filter:

                            The ARIMA Procedure

                          Moving Average Factors

              Factor 1:  1 - 0.17392 B**(1) - 0.73015 B**(12)


                   Conditional Least Squares Estimation

                         Standard             Approx
Parameter    Estimate       Error  t Value  Pr > |t|   Lag
Variable  Shift

MA1,1         0.75291     0.07134    10.55    <.0001    12  starts
0 AR1,1        -0.47666     0.08295    -5.75    <.0001     1
starts        0 NUM1          0.70797     0.15878     4.46
<.0001     0  sales         1


                      Variance Estimate      49.95394
                      Std Error Estimate      7.06781
                      AIC                    799.3407
                      SBC                    807.6527
                      Number of Residuals         118
               * AIC and SBC do not include log determinant.


                    Correlations of Parameter Estimates

              Variable             starts    starts     sales
              Parameter             MA1,1     AR1,1      NUM1

              starts      MA1,1     1.000     0.134     0.035
              starts      AR1,1     0.134     1.000    -0.064
              sales        NUM1     0.035    -0.064     1.000
```

Autocorrelation Check of Residuals

To Lag	Chi-Square	DF	Pr > ChiSq	---------------	Autocorrelations	---------------			
6	6.91	4	0.1407	-0.106	-0.170	0.065	-0.065	0.083	0.026
12	11.79	10	0.2995	-0.074	-0.049	0.151	-0.070	0.031	0.034
18	14.24	16	0.5806	0.080	-0.002	0.022	-0.095	-0.045	-0.006
24	27.97	22	0.1767	-0.110	0.163	-0.119	-0.179	0.096	-0.010

Crosscorrelation Check of Residuals with Input sales

To Lag	Chi-Square	DF	Pr > ChiSq	---------------	Crosscorrelations	--------------			
5	4.65	5	0.4606	-0.040	0.104	0.133	0.040	0.086	-0.025
11	6.30	11	0.8525	-0.046	-0.014	0.090	0.034	-0.015	-0.049
17	13.02	17	0.7345	-0.061	-0.045	0.029	0.085	-0.126	0.167
23	18.14	23	0.7501	-0.031	-0.125	0.069	-0.058	-0.095	0.099

Model for variable starts

Period(s) of Differencing 1,12

No mean term in this model.

Autoregressive Factors

Factor 1: 1 + 0.47666 B**(1)

Moving Average Factors

Factor 1: 1 - 0.75291 B**(12)

Input Number 1

Input Variable	sales
Shift	1
Period(s) of Differencing	1,12
Overall Regression Factor	0.707966

Forecasts for variable starts

Obs	Forecast	Std Error	95% Confidence Limits		Actual	Residual
121	36.1656	7.0678	22.3129	50.0182	39.8000	3.6344
122	40.8728	8.4614	24.2887	57.4569	39.9000	-0.9728
123	63.5372	10.2551	43.4375	83.6369	62.5000	-1.0372
124	85.6855	11.4545	63.2352	108.1359	77.8000	-7.8855
125	88.0824	12.6763	63.2374	112.9274	92.8000	4.7176
126	84.9659	13.7293	58.0568	111.8749	90.3000	5.3341
127	78.4513	14.7341	49.5731	107.3296	92.8000	14.3487
128	76.6163	15.6624	45.9185	107.3141	90.7000	14.0837
129	67.4369	16.5442	35.0108	99.8630	84.5000	17.0631
130	67.8331	17.3789	33.7711	101.8951	93.8000	25.9669
131	53.9664	18.1764	18.3414	89.5914	71.6000	17.6336
132	38.0898	18.9398	0.9685	75.2111	55.7000	17.6102
133	33.2416	20.1700	-6.2908	72.7741	.	.
134	38.4099	21.1868	-3.1155	79.9353	.	.
135	60.8545	22.1996	17.3441	104.3650	.	.
136	83.1076	23.1191	37.7949	128.4203	.	.
137	85.4546	24.0258	38.3648	132.5443	.	.
138	82.3618	24.8892	33.5800	131.1437	.	.
139	75.8359	25.7283	25.4094	126.2625	.	.
140	74.0063	26.5387	21.9913	126.0213	.	.
141	64.8244	27.3262	11.2660	118.3827	.	.
142	65.2218	28.0911	10.1644	120.2793	.	.
143	51.3545	28.8359	-5.1628	107.8718	.	.
144	35.4782	29.5618	-22.4620	93.4183	.	.

Forecasts for variable starts

Obs	Forecast	Std Error	95% Confidence Limits	
133	52.6699	7.0678	38.8172	66.5225
134	56.0797	8.4614	39.4956	72.6638
135	78.9303	10.2551	58.8306	99.0300
136	99.4243	11.4545	76.9740	121.8747
137	104.8308	12.6763	79.9858	129.6758
138	100.9362	13.7293	74.0272	127.8452
139	96.3223	14.7341	67.4440	125.2005
140	94.5733	15.6624	63.8755	125.2712
141	86.1479	16.5442	53.7218	118.5741
142	88.6408	17.3789	54.5789	122.7028
143	72.9036	18.1764	37.2786	108.5286
144	56.9247	18.9398	19.8035	94.0460

The summary models is;

```
Model for variable starts
```

```
Period(s) of Differencing    1,12
   No mean term in this model.

      Autoregressive Factors
  Factor 1:  1 + 0.47666 B**(1)


       Moving Average Factors
  Factor 1:  1 - 0.75291 B**(12)
```

Which becomes the following

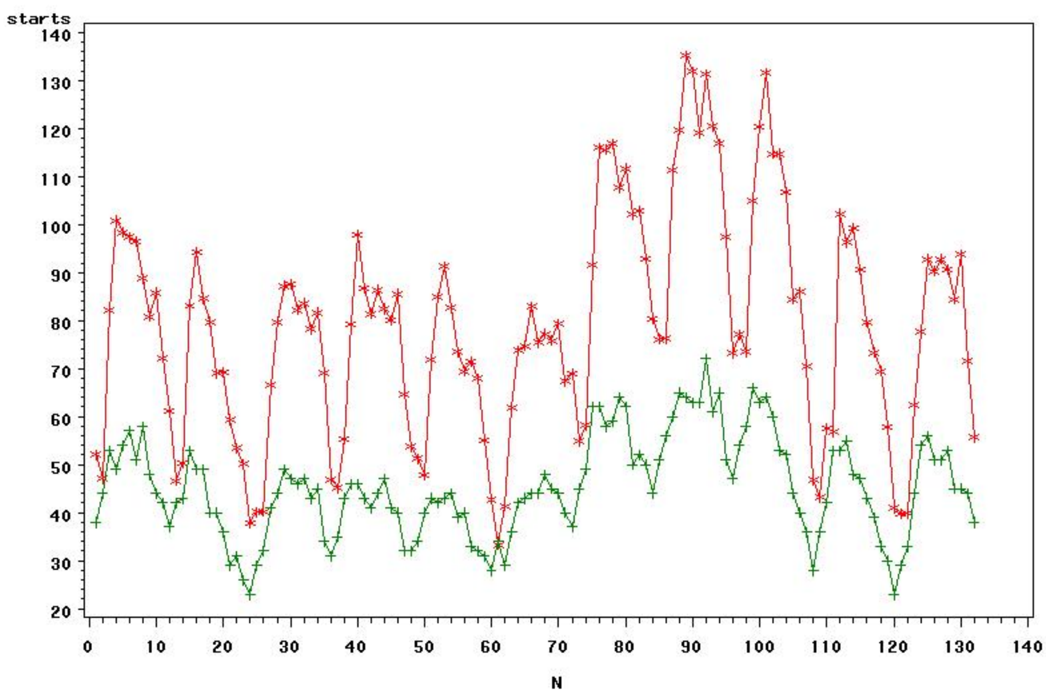
$$start_t = \frac{1}{(1-.48B)} sales_{t-1} + a_t - .75a_{t-12}$$

or

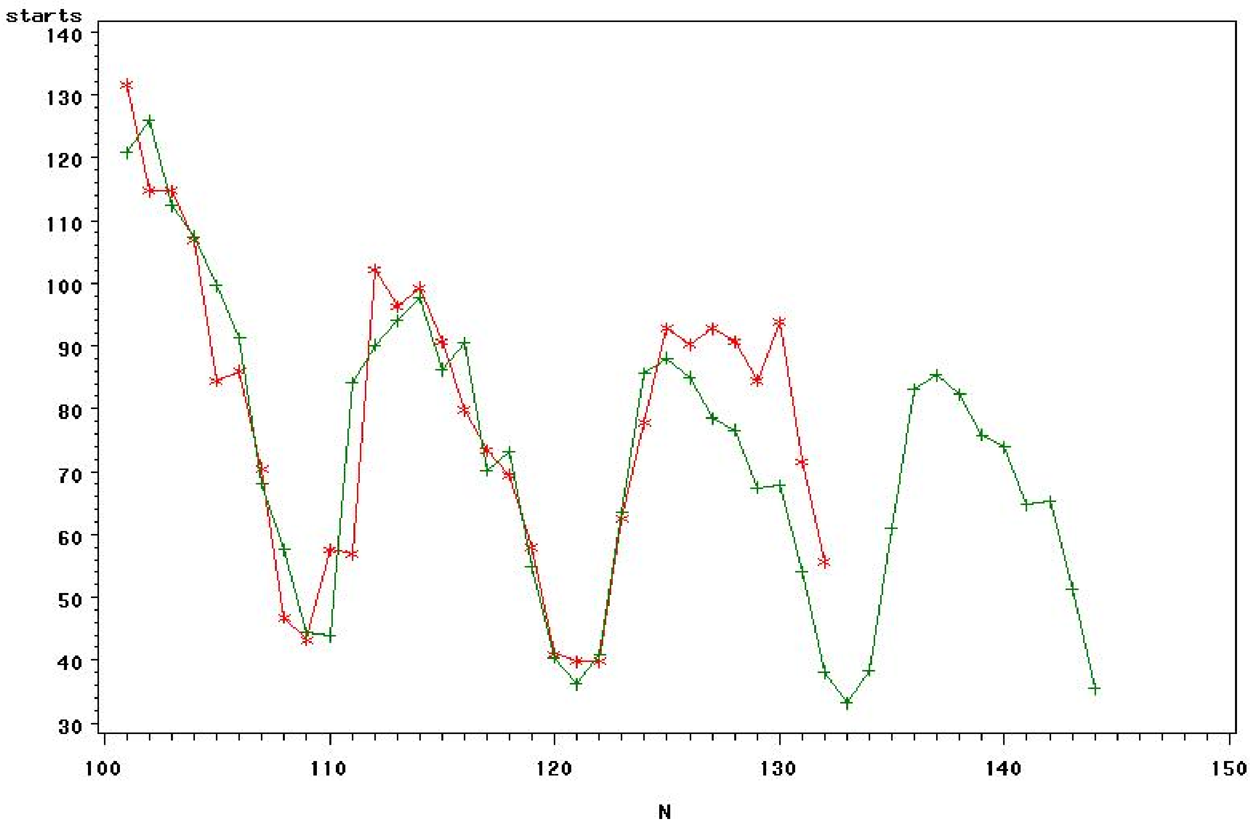
$$start_t = .48start_{t-1} + sales_{t-1} + a_t - .75a_{t-12}$$

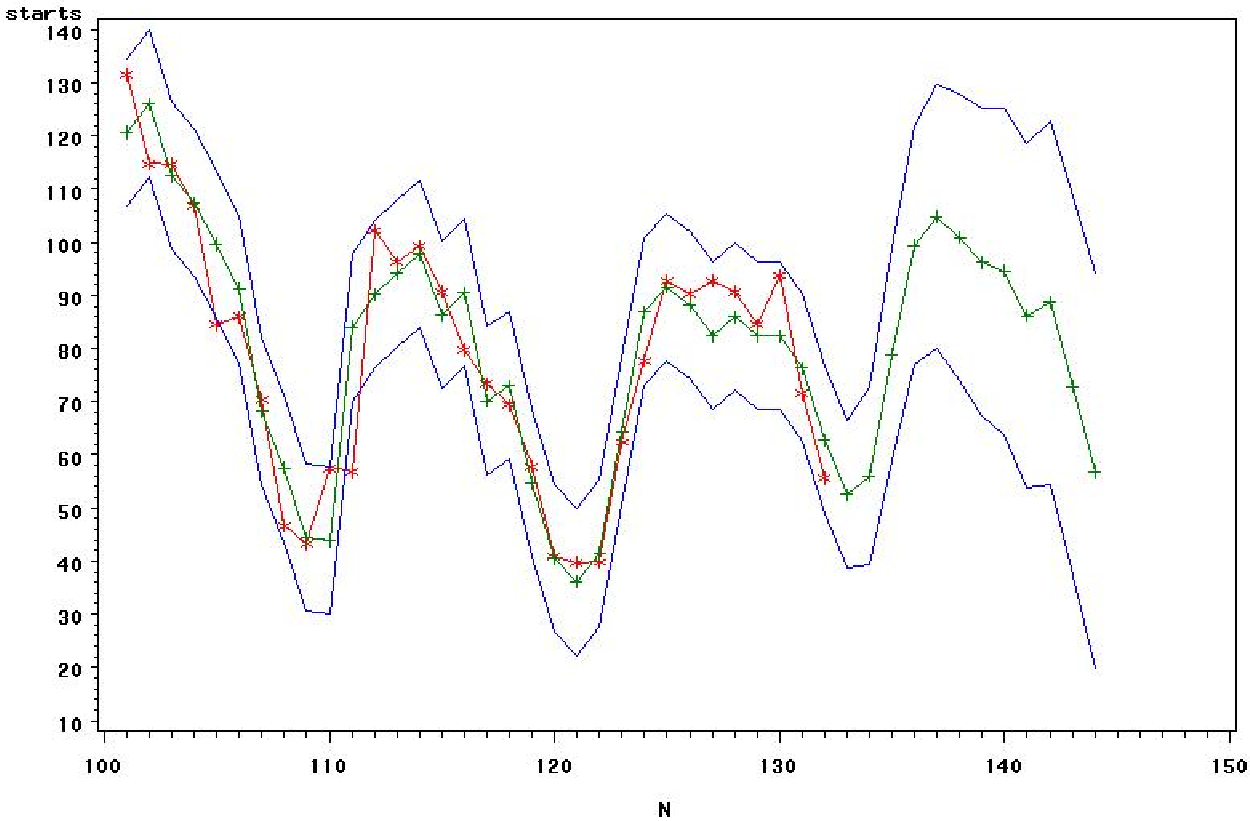
where the original prewhitening function is,

$$(1-B)(1-B^{12})sales_t = (1-.17B-.73B^{12})\alpha_t.$$



Some of the graphs are


starts
140
130
120
110
100
90
80
70
60
50
40
30
100
110
120
130
140
150
N


starts
140
130
120
110
100
90
80
70
60
50
40
30
20
10
100
110
120
130
140
150
N

Chapter 9

Autoregressive Error and GARCH Models

9.1 Regression with Autocorrelated Errors

Ordinary regression analysis is based on several statistical assumptions. One key assumption is that the errors are independent of each other. However, with time series data, the ordinary regression residuals usually are correlated over time. It is not desirable to use ordinary regression analysis for time series data since the assumptions on which the classical linear regression model is based will usually be violated.

Violation of the independent errors assumption has three important consequences for ordinary regression. First, statistical tests of the significance of the parameters and the confidence limits for the predicted values are not correct. Second, the estimates of the regression coefficients are not as efficient as they would be if the autocorrelation were taken into account. Third, since the ordinary regression residuals are not independent, they contain information that can be used to improve the prediction of future values.

The AUTOREG procedure solves this problem by augmenting the regression model with an autoregressive model for the random error, thereby accounting for the autocorrelation of the errors. Instead of the usual regression model, the following autoregressive error model is used:

$$\begin{aligned} y_t &= x_t'\,\beta + \nu_t \\ \nu_t &= \varphi_1 \nu_{t-1} + \varphi_2 \nu_{t-2} + \ldots + \varphi_p \nu_{t-p} = \epsilon_t \\ \epsilon_t &\sim WN(0, \sigma^2) \end{aligned}$$

The notation $\epsilon_t \sim WN(0, \sigma^2)$ indicates that each ϵ_t is normally and independently distributed with mean 0 and variance σ^2.

By simultaneously estimating the regression coefficients β and the autoregressive error model parameters φ_i, the AUTOREG procedure corrects the regression estimates for autocorrelation. Thus, this kind of regression analysis is often called *autoregressive error correction or serial correlation correction.*

9.1.1 Generalized Durbin-Watson Tests

Consider the following linear regression model:

$$Y = X\beta + \nu$$

where X is an $N \times k$ data matrix, β is a $k \times 1$ coefficient vector, ν and is a $N \times 1$ disturbance vector. The error term ν is assumed to be generated by the j^{th} order autoregressive process $\nu_t = \varphi_j \nu_{t-j}$ where $\mid \varphi_j \mid < 1$, ϵ_t is a sequence of independent normal error terms with mean 0 and variance σ^2. Usually, the Durbin-Watson statistic is used to test the null hypothesis $H_0 : \varphi_1 = 0$ against $H_1 : -\varphi_1 > 0$. Vinod (1973) generalized the Durbin-Watson statistic:

$$d_j = \frac{\sum_{t=j+1}^{N} (\hat{\nu}_t - \hat{\nu}_{t-1})^2}{\sum_{t=1}^{N} \hat{\nu}_t^2}$$

where $\hat{\nu}_t$ are OLS residuals. Using the matrix notation,

$$d_j = \frac{\nu' M A_j' A_j M \nu}{\nu' M \nu}$$

where $M = I_N - X(X'X)^{-1}X'$ and A_j is a $(N-j) \times N$ matrix:

$$\begin{pmatrix} -1 & 0 & \cdots & 0 & 1 & 0 & \cdots & 0 \\ 0 & -1 & 0 & \cdots & 0 & 1 & 0 & \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & \cdots & 0 & -1 & 0 & 0 & \cdots & 1 \end{pmatrix}$$

and there are $j-1$ zeros between -1 and 1 in each row of matrix A_j.

The QR factorization of the design matrix X yields a $N \times N$ orthogonal matrix Q

$$X = QR$$

where R is a $N \times k$ upper triangular matrix. There exists a $N \times (N-k)$ submatrix of Q such that $Q_1 Q_1' = M$ and $Q_1' Q_1 = I_{N-k}$. Consequently, the generalized Durbin-Watson statistic is stated as a ratio of two quadratic forms:

$$d_j = \frac{\sum_{l=1}^{N} \lambda_{jl} \xi_l^2}{\sum_{l=1}^{N} \xi_l^2}$$

where λ_{jl} are upper n eigenvalues of $MA_j'A_jM$ and ξ_t is a standard normal variate, and $n = min(N-k, N-j)$. These eigenvalues are obtained by a singular value decomposition of $Q_1'A_j'$ (Golub and Loan 1989; Savin and White 1978).

The marginal probability (or p-value) for d_j given c_0 is

$$\Pr[\frac{\sum_{l=1}^{N} \lambda_{jl} \xi_l^2}{\sum_{l=1}^{N} \xi_l^2} < c_0] = \Pr[q_j < 0]$$

where

$$q_j = \sum_{l=1}^{N} (\lambda_{jl} - c_o) \xi_l^2.$$

When the null hypothesis $H_0 : \varphi_j = 0$ holds, the quadratic form q_j has the characteristic function

$$\phi_j(t) = \Pi_{l=1}^{n} (1 - 2(\lambda_{jl} - c_0)it)^{-1/2}$$

The distribution function is uniquely determined by this characteristic function:

$$F(x) = \frac{1}{2} + \frac{1}{2\pi} \int_0^\infty \frac{e^{itx}\phi_j(-t) - e^{-itx}\phi_j(t)}{it} dt$$

For example, to test ($-\varphi_4 > 0$) given $\varphi_1 = \varphi_2 = \varphi_3 = 0$ against $H_1 : -\varphi_4 > 0$, the marginal probability (p-value) can be used:

$$F(x) = \frac{1}{2} + \frac{1}{2\pi} \int_0^{\infty} \frac{\phi_4(-t) - \phi_4(t)}{it} dt$$

where

$$\phi_4(t) = \Pi_{l=1}^{n}(1 - 2(\lambda_{4l} - \hat{d}_4)it)^{-1/2}$$

and $\hat{d}_4$ is the calculated value of the fourth-order Durbin-Watson statistic.

In the Durbin-Watson test, the marginal probability indicates positive autocorrelation ($-\varphi_j > 0$) if it is less than the level of significance (α), while you can conclude that a negative autocorrelation ($-\varphi_j < 0$) exists if the marginal probability based on the computed Durbin-Watson statistic is greater than $1-\alpha$. Wallis (1972) presented tables for bounds tests of fourth-order autocorrelation and Vinod (1973) has given tables for a five percent significance level for orders two to four. Using the AUTOREG procedure, you can calculate the exact p-values for the general order of Durbin-Watson test statistics. Tests for the absence of autocorrelation of order p can be performed sequentially; at the j^{th} step, test ($-\varphi_j > 0$) given $\varphi_1 = \varphi_2 = \ldots = \varphi_{j-1} = 0$ against $H_1 : -\varphi_j > 0$. However, the size of the sequential test is not known.

The Durbin-Watson statistic is computed from the OLS residuals, while that of the autoregressive error model uses residuals that are the difference between the predicted values and the actual values. When you use the Durbin-Watson test from the residuals of the autoregressive error model, you must be aware that this test is only an approximation. See "Regression with Autoregressive Errors" earlier in this chapter. If there are missing values, the Durbin-Watson statistic is computed using all the nonmissing values and ignoring the gaps caused by missing residuals. This does not affect the significance level of the resulting test, although the power of the test against certain alternatives may be adversely affected. Savin and White (1978) have examined the use of the Durbin-Watson statistic with missing values.

9.2 GARCH Models

Let's consider the simplest case given ,

$$y_t = x_t' \beta + \varepsilon_t,$$

where ε_t are uncorrelated errors but have a variance that changes over time. Assume that the error term can be expressed as,

$$\varepsilon_t = \sigma_t \epsilon_t,$$

where ϵ_t are i.i.d. random variables with mean zero and variance 1. Furthermore ϵ_t is independent of past ε_{t-j}, for all $j > 0$ and

$$\sigma_t^2 = \theta_0 + \theta_1 \varepsilon_{t-1}^2,$$

this model is called a $ARCH(1)$ model. The more general model is given below.

Consider the series y_t, which follows the GARCH process. The conditional distribution of the series Y for time t is written

$$y_t \mid \Psi_{t-1} \sim N(0, h_t)$$

where Ψ_{t-1} denotes all available information at time t-1. The conditional variance h_t is

$$h_t = \omega + \sum_{i=1}^{q} \alpha_i y_{t-i}^2 + \sum_{j=1}^{p} \gamma_j h_{t-j}$$

where $p \geq 0, q > 0, \omega > 0, \alpha_i \geq 0, \gamma_j \geq 0$.

The $GARCH(p,q)$ model reduces to the $ARCH(q)$ process when $p = 0$. At least one of the ARCH parameters must be nonzero ($q > 0$). The GARCH regression model can be written

$$\begin{aligned}
y_t &= x_t'\beta + \epsilon_t \\
\epsilon_t &= \sqrt{h_t}e_t \\
h_t &= \omega + \sum_{i=1}^{q} \alpha_i \epsilon_{t-i}^2 + \sum_{j=1}^{p} \gamma_j h_{t-j} \\
e_t &\sim WN(0,1)
\end{aligned}$$

In addition, you can consider the model with disturbances following an autoregressive process and with the GARCH errors. The $AR(m) - GARCH(p,q)$ regression model is denoted

$$\begin{aligned}
y_t &= x_t'\beta + \nu_t \\
\nu_t &= \epsilon_t - \varphi_1 \nu_{t-1} - \varphi_2 \nu_{t-2} - \ldots - \varphi_m \nu_{t-m} \\
\epsilon_t &= \sqrt{h_t}e_t \\
h_t &= \omega + \sum_{i=1}^{q} \alpha_i \epsilon_{t-i}^2 + \sum_{j=1}^{p} \gamma_j h_{t-j} \\
e_t &\sim WN(0,1)
\end{aligned}$$