3
Time Series Concepts

3.1 Introduction

This chapter provides background material on time series concepts that are used throughout the book. These concepts are presented in an informal way, and extensive examples using `S-PLUS` are used to build intuition. Section 3.2 discusses time series concepts for stationary and ergodic univariate time series. Topics include testing for white noise, linear and autoregressive moving average (ARMA) process, estimation and forecasting from ARMA models, and long-run variance estimation. Section 3.3 introduces univariate nonstationary time series and defines the important concepts of $I(0)$ and $I(1)$ time series. Section 3.4 explains univariate long memory time series. Section 3.5 covers concepts for stationary and ergodic multivariate time series, introduces the class of vector autoregression models, and discusses long-run variance estimation.

Rigorous treatments of the time series concepts presented in this chapter can be found in Fuller (1996) and Hamilton (1994). Applications of these concepts to financial time series are provided by Campbell, Lo and MacKinlay (1997), Mills (1999), Gourieroux and Jasiak (2001), Tsay (2001), Alexander (2001) and Chan (2002).

3.2 Univariate Time Series

3.2.1 Stationary and Ergodic Time Series

Let $\{y_t\} = \{\ldots y_{t-1}, y_t, y_{t+1}, \ldots\}$ denote a sequence of random variables indexed by some time subscript t. Call such a sequence of random variables a *time series.*

The time series $\{y_t\}$ is *covariance stationary* if

$$\begin{aligned} E[y_t] &= \mu \text{ for all } t \\ cov(y_t, y_{t-j}) &= E[(y_t - \mu)(y_{t-j} - \mu)] = \gamma_j \text{ for all } t \text{ and any } j \end{aligned}$$

For brevity, call a covariance stationary time series simply a *stationary* time series. Stationary time series have time invariant first and second moments. The parameter γ_j is called the j^{th} order or lag j *autocovariance* of $\{y_t\}$ and a plot of γ_j against j is called the *autocovariance function.* The *autocorrelations* of $\{y_t\}$ are defined by

$$\rho_j = \frac{cov(y_t, y_{t-j})}{\sqrt{var(y_t)var(y_{t-j})}} = \frac{\gamma_j}{\gamma_0}$$

and a plot of ρ_j against j is called the *autocorrelation function* (ACF). Intuitively, a stationary time series is defined by its mean, variance and ACF. A useful result is that any function of a stationary time series is also a stationary time series. So if $\{y_t\}$ is stationary then $\{z_t\} = \{g(y_t)\}$ is stationary for any function $g(\cdot)$.

The lag j *sample autocovariance* and lag j *sample autocorrelation* are defined as

$$\hat{\gamma}_j = \frac{1}{T}\sum_{t=j+1}^{T}(y_t - \bar{y})(y_{t-j} - \bar{y}) \tag{3.1}$$

$$\hat{\rho}_j = \frac{\hat{\gamma}_j}{\hat{\gamma}_0} \tag{3.2}$$

where $\bar{y} = \frac{1}{T}\sum_{t=1}^{T} y_t$ is the sample mean. The sample ACF (SACF) is a plot of $\hat{\rho}_j$ against j.

A stationary time series $\{y_t\}$ is *ergodic* if sample moments converge in probability to population moments; i.e. if $\bar{y} \xrightarrow{p} \mu, \hat{\gamma}_j \xrightarrow{p} \gamma_j$ and $\hat{\rho}_j \xrightarrow{p} \rho_j$.

Example 1 *Gaussian white noise (GWN) processes*

Perhaps the most simple stationary time series is the *independent Gaussian white noise* process $y_t \sim iid\ N(0, \sigma^2) \equiv GWN(0, \sigma^2)$. This process has $\mu = \gamma_j = \rho_j = 0$ $(j \neq 0)$. To simulate a $GWN(0,1)$ process in `S-PLUS` use the `rnorm` function:

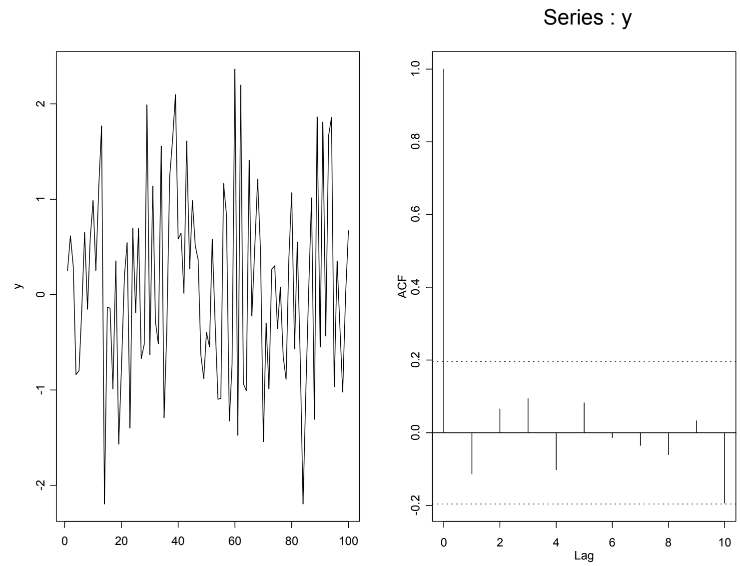

FIGURE 3.1. Simulated Gaussian white noise process and SACF.

```
> set.seed(101)
> y = rnorm(100,sd=1)
```

To compute the sample moments $\bar{y}$, $\hat{\gamma}_j$, $\hat{\rho}_j$ $(j = 1, \ldots, 10)$ and plot the data and SACF use

```
> y.bar = mean(y)
> g.hat = acf(y,lag.max=10,type="covariance",plot=F)
> r.hat = acf(y,lag.max=10,type="correlation",plot=F)
> par(mfrow=c(1,2))
> tsplot(y,ylab="y")
> acf.plot(r.hat)
```

By default, as shown in Figure 3.1, the SACF is shown with 95% confidence limits about zero. These limits are based on the result (c.f. Fuller (1996) pg. 336) that if $\{y_t\} \sim iid\ (0, \sigma^2)$ then

$$\hat{\rho}_j \overset{A}{\sim} N\left(0, \frac{1}{T}\right),\ j > 0.$$

The notation $\hat{\rho}_j \overset{A}{\sim} N\left(0, \frac{1}{T}\right)$ means that the distribution of $\hat{\rho}_j$ is approximated by normal distribution with mean 0 and variance $\frac{1}{T}$ and is based on the central limit theorem result $\sqrt{T}\hat{\rho}_j \overset{d}{\to} N(0,1)$. The 95% limits about zero are then $\pm\frac{1.96}{\sqrt{T}}$.

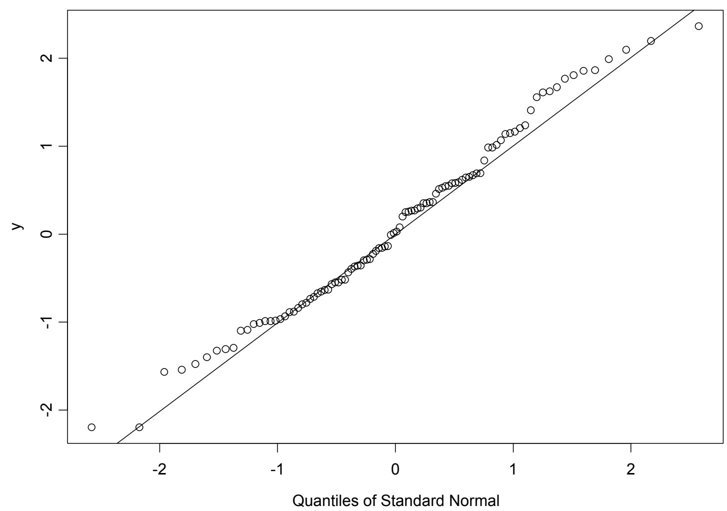

FIGURE 3.2. Normal qq-plot for simulated GWN.

Two slightly more general processes are the independent *white noise* (IWN) process, $y_t \sim IWN(0, \sigma^2)$, and the *white noise* (WN) process, $y_t \sim WN(0, \sigma^2)$. Both processes have mean zero and variance σ^2, but the IWN process has independent increments, whereas the WN process has uncorrelated increments.

Testing for Normality

In the previous example, $y_t \sim GWN(0, 1)$. There are several statistical methods that can be used to see if an *iid* process y_t is Gaussian. The most common is the normal quantile-quantile plot or *qq-plot*, a scatterplot of the standardized empirical quantiles of y_t against the quantiles of a standard normal random variable. If y_t is normally distributed, then the quantiles will lie on a 45 degree line. A normal qq-plot with 45 degree line for y_t may be computed using the `S-PLUS` functions `qqnorm` and `qqline`

```
> qqnorm(y)
> qqline(y)
```

Figure 3.2 shows the qq-plot for the simulated GWN data of the previous example. The quantiles lie roughly on a straight line. The `S+FinMetrics` function `qqPlot` may be used to create a Trellis graphics qq-plot.

The qq-plot is an informal graphical diagnostic. Two popular formal statistical tests for normality are the *Shapiro-Wilks* test and the *Jarque-*

Bera test. The Shapiro-Wilk's test is a well-known goodness of fit test for the normal distribution. It is attractive because it has a simple, graphical interpretation: one can think of it as an approximate measure of the correlation in a normal quantile-quantile plot of the data. The Jarque-Bera test is based on the result that a normally distributed random variable has skewness equal to zero and kurtosis equal to three. The Jarque-Bera test statistic is

$$JB = \frac{T}{6}\left(\widehat{skew}^2 + \frac{(\widehat{kurt}-3)^2}{4}\right) \tag{3.3}$$

where $\widehat{skew}$ denotes the sample skewness and $\widehat{kurt}$ denotes the sample kurtosis. Under the null hypothesis that the data is normally distributed

$$JB \overset{A}{\sim} \chi^2(2).$$

Example 2 *Testing for normality using the `S+FinMetrics` function `normalTest`*

The Shapiro-Wilks and Jarque-Bera statistics may be computed using the `S+FinMetrics` function `normalTest`. For the simulated GWN data of the previous example, these statistics are

```
> normalTest(y, method="sw")
Test for Normality: Shapiro-Wilks

Null Hypothesis: data is normally distributed

Test Statistics:

Test Stat 0.9703
  p.value 0.1449

Dist. under Null: normal
   Total Observ.: 100

> normalTest(y, method="jb")

Test for Normality: Jarque-Bera

Null Hypothesis: data is normally distributed

Test Statistics:

Test Stat 1.8763
  p.value 0.3914
```

```
Dist. under Null: chi-square with 2 degrees of freedom
   Total Observ.: 100
```

The null of normality is not rejected using either test.

Testing for White Noise

Consider testing the null hypothesis

$$H_0 : y_t \sim WN(0, \sigma^2)$$

against the alternative that y_t is not white noise. Under the null, all of the autocorrelations ρ_j for $j > 0$ are zero. To test this null, Box and Pierce (1970) suggested the *Q-statistic*

$$Q(k) = T \sum_{j=1}^{k} \hat{\rho}_j^2 \tag{3.4}$$

where $\hat{\rho}_j$ is given by (3.2). Under the null, $Q(k)$ is asymptotically distributed $\chi^2(k)$. In a finite sample, the Q-statistic (3.4) may not be well approximated by the $\chi^2(k)$. Ljung and Box (1978) suggested the *modified Q-statistic*

$$MQ(k) = T(T+2) \sum_{j=1}^{k} \frac{\hat{\rho}_j^2}{T-j} \tag{3.5}$$

which is better approximated by the $\chi^2(k)$ in finite samples.

Example 3 *Daily returns on Microsoft*

Consider the time series behavior of daily continuously compounded returns on Microsoft for 2000. The following `S-PLUS` commands create the data and produce some diagnostic plots:

```
> r.msft = getReturns(DowJones30[,"MSFT"],type="continuous")
> r.msft@title = "Daily returns on Microsoft"
> sample.2000 = (positions(r.msft) > timeDate("12/31/1999")
+ & positions(r.msft) < timeDate("1/1/2001"))
> par(mfrow=c(2,2))
> plot(r.msft[sample.2000],ylab="r.msft")
> r.acf = acf(r.msft[sample.2000])
> hist(seriesData(r.msft))
> qqnorm(seriesData(r.msft))
```

The daily returns on Microsoft resemble a white noise process. The qq-plot, however, suggests that the tails of the return distribution are fatter than the normal distribution. Notice that since the `hist` and `qqnorm` functions do not have methods for "`timeSeries`" objects the extractor function `seriesData` is required to extract the data frame from the data slot of `r.msft`.

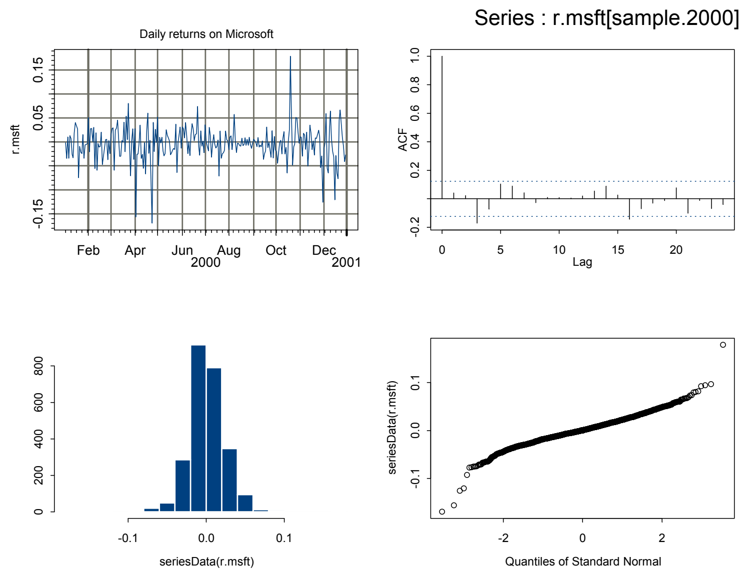

FIGURE 3.3. Daily returns on Microsoft with diagnostic plots.

The S+FinMetrics functions histPlot and qqPlot will produce a histogram and qq-plot for a "timeSeries" object using Trellis graphics. For example,

```
> histPlot(r.msft,strip.text="MSFT monthly return")
> qqPlot(r.msft,strip.text="MSFT monthly return")
```

However, Trellis plots cannot be displayed in a multipanel plot created using par.

The S+FinMetrics function autocorTest may be used to compute the Q-statistic and modified Q-statistic to test the null that the returns on Microsoft follow a white noise process:

```
> autocorTest(r.msft, lag.n=10, method="lb")

Test for Autocorrelation: Ljung-Box
Null Hypothesis: no autocorrelation

Test Statistics:

Test Stat 11.7746
  p.value  0.3004

Dist. under Null: chi-square with 10 degrees of freedom
```

```
Total Observ.: 2527
```

The argument `lag.n=10` specifies that $k = 10$ autocorrelations are used in computing the statistic, and `method="lb"` specifies that the modified Box-Pierce statistic (3.5) be computed. To compute the simple Box-Pierce statistic, specify `method="bp"`. The results indicate that the white noise null cannot be rejected.

3.2.2 Linear Processes and ARMA Models

Wold's decomposition theorem (c.f. Fuller (1996) pg. 96) states that any covariance stationary time series $\{y_t\}$ has a *linear process* or infinite order moving average representation of the form

$$\begin{aligned} y_t &= \mu + \sum_{k=0}^{\infty} \psi_k \varepsilon_{t-k} \\ \psi_0 &= 1, \ \sum_{k=0}^{\infty} \psi_k^2 < \infty \\ \varepsilon_t &\sim WN(0, \sigma^2) \end{aligned} \tag{3.6}$$

In the Wold form, it can be shown that

$$\begin{aligned} E[y_t] &= \mu \\ \gamma_0 &= var(y_t) = \sigma^2 \sum_{k=0}^{\infty} \psi_k^2 \\ \gamma_j &= cov(y_t, y_{t-j}) = \sigma^2 \sum_{k=0}^{\infty} \psi_k \psi_{k+j} \\ \rho_j &= \frac{\sum_{k=0}^{\infty} \psi_k \psi_{k+j}}{\sum_{k=0}^{\infty} \psi_k^2} \end{aligned}$$

Hence, the pattern of autocorrelations in any stationary and ergodic time series $\{y_t\}$ is determined by the moving average weights $\{\psi_j\}$ in its Wold representation. To ensure convergence of the linear process representation to a stationary and ergodic process with nice properties, it is necessary to further restrict the behavior of the moving average weights $\{\psi_j\}$. A standard assumption used in the econometrics literature (c.f. Hamilton (1994) pg. 504) is *1-summability*

$$\sum_{j=0}^{\infty} j|\psi_j| = 1 + 2|\psi_2| + 3|\psi_3| + \cdots < \infty.$$

The moving average weights in the Wold form are also called *impulse responses* since

$$\frac{\partial y_{t+s}}{\partial \varepsilon_t} = \psi_s, s = 1, 2, \ldots$$

For a stationary and ergodic time series $\lim_{s\to\infty}\psi_s = 0$ and the *long-run cumulative impulse response* $\sum_{s=0}^{\infty}\psi_s < \infty$. A plot of ψ_s against s is called the *impulse response function* (IRF).

The general Wold form of a stationary and ergodic time series is handy for theoretical analysis but is not practically useful for estimation purposes. A very rich and practically useful class of stationary and ergodic processes is the *autoregressive-moving average* (ARMA) class of models made popular by Box and Jenkins (1976). ARMA(p,q) models take the form of a *pth* order stochastic difference equation

$$\begin{aligned} y_t - \mu &= \phi_1(y_{t-1}-\mu) + \cdots + \phi_p(y_{t-p}-\mu) \\ &\quad + \varepsilon_t + \theta_1\varepsilon_{t-1} + \cdots + \theta_q\varepsilon_{t-q} \\ \varepsilon_t &\sim WN(0,\sigma^2) \end{aligned} \tag{3.7}$$

ARMA(p,q) models may be thought of as parsimonious approximations to the general Wold form of a stationary and ergodic time series. More information on the properties of ARMA(p,q) process and the procedures for estimating and forecasting these processes using `S-PLUS` are in the *S-PLUS Guide to Statistics Vol. II*, chapter 27, Venables and Ripley (1999) chapter 13, and Meeker (2001)[1].

Lag Operator Notation

The presentation of time series models is simplified using *lag operator* notation. The lag operator L is defined such that for any time series $\{y_t\}$, $Ly_t = y_{t-1}$. The lag operator has the following properties: $L^2y_t = L\cdot Ly_t = y_{t-2}$, $L^0 = 1$ and $L^{-1}y_t = y_{t+1}$. The operator $\Delta = 1 - L$ creates the first difference of a time series: $\Delta y_t = (1-L)y_t = y_t - y_{t-1}$. The ARMA$(p,q)$ model (3.7) may be compactly expressed using lag polynomials. Define $\phi(L) = 1 - \phi_1 L - \cdots - \phi_p L^p$ and $\theta(L) = 1 + \theta_1 L + \cdots + \theta_q L^q$. Then (3.7) may be expressed as

$$\phi(L)(y_t - \mu) = \theta(L)\varepsilon_t$$

Similarly, the Wold representation in lag operator notation is

$$\begin{aligned} y_t &= \mu + \psi(L)\varepsilon_t \\ \psi(L) &= \sum_{k=0}^{\infty}\psi_k L^k,\ \psi_0 = 1 \end{aligned}$$

and the long-run cumulative impulse response is $\psi(1)$ (i.e. evaluate $\psi(L)$ at $L=1$). With ARMA(p,q) models the Wold polynomial $\psi(L)$ is approx-

[1] William Meeker also has a library of time series functions for the analysis of ARMA models available for download at
`http://www.public.iastate.edu/~stat451/splusts/splusts.html`.

imated by the ratio of the AR and MA polynomials

$$\psi(L) = \frac{\theta(L)}{\phi(L)}$$

3.2.3 Autoregressive Models

AR(1) Model

A commonly used stationary and ergodic time series in financial modeling is the AR(1) process

$$y_t - \mu = \phi(y_{t-1} - \mu) + \varepsilon_t,\ t = 1, \ldots, T$$

where $\varepsilon_t \sim WN(0, \sigma^2)$ and $|\phi| < 1$. The above representation is called the *mean-adjusted form*. The *characteristic equation* for the AR(1) is

$$\phi(z) = 1 - \phi z = 0 \tag{3.8}$$

so that the root is $z = \frac{1}{\phi}$. Stationarity is satisfied provided the absolute value of the root of the characteristic equation (3.8) is greater than one: $|\frac{1}{\phi}| > 1$ or $|\phi| < 1$. In this case, it is easy to show that $E[y_t] = \mu$, $\gamma_0 = \frac{\sigma^2}{1-\phi^2}$, $\psi_j = \rho_j = \phi^j$ and the Wold representation is

$$y_t = \mu + \sum_{j=0}^{\infty} \rho^j \varepsilon_{t-j}.$$

Notice that for the AR(1) the ACF and IRF are identical. This is not true in general. The long-run cumulative impulse response is $\psi(1) = \frac{1}{1-\phi}$.

The AR(1) model may be re-written in *components form* as

$$\begin{aligned} y_t &= \mu + u_t \\ u_t &= \phi u_{t-1} + \varepsilon_t \end{aligned}$$

or in *autoregression form* as

$$\begin{aligned} y_t &= c + \phi y_{t-1} + \varepsilon_t \\ c &= \mu(1 - \phi) \end{aligned}$$

An AR(1) with $\mu = 1$, $\phi = 0.75$, $\sigma^2 = 1$ and $T = 100$ is easily simulated in S-PLUS using the components form:

```
> set.seed(101)
> e = rnorm(100,sd=1)
> e.start = rnorm(25,sd=1)
> y.ar1 = 1 + arima.sim(model=list(ar=0.75), n=100,
```

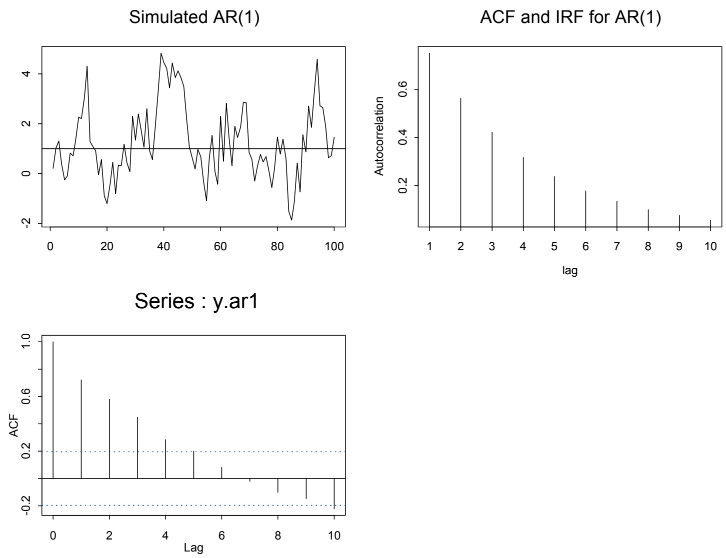

FIGURE 3.4. Simulated AR(1), ACF, IRF and SACF.

```
+ innov=e, start.innov=e.start)
> mean(y.ar1)
[1] 1.271
> var(y.ar1)
[1] 2.201
```

The ACF and IRF may be computed as

```
> gamma.j = rep(0.75,10)^seq(10)
```

The simulated data, ACF and SACF are illustrated in Figure 3.4 using

```
> par(mfrow=c(2,2))
> tsplot(y.ar1,main="Simulated AR(1)")
> abline(h=1)
> tsplot(gamma.j, type="h", main="ACF and IRF for AR(1)",
+ ylab="Autocorrelation", xlab="lag")
> tmp = acf(y.ar1, lag.max=10)
```

Notice that $\{y_t\}$ exhibits *mean-reverting* behavior. That is, $\{y_t\}$ fluctuates about the mean value $\mu = 1$. The ACF and IRF decay at a geometric rate. The decay rate of the IRF is sometimes reported as a *half-life* – the lag j^{half} at which the IRF reaches $\frac{1}{2}$. For the AR(1) with positive ϕ, it can be shown that $j^{half} = \ln(0.5)/\ln(\phi)$. For $\phi = 0.75$, the half-life is

```
> log(0.5)/log(0.75)
```

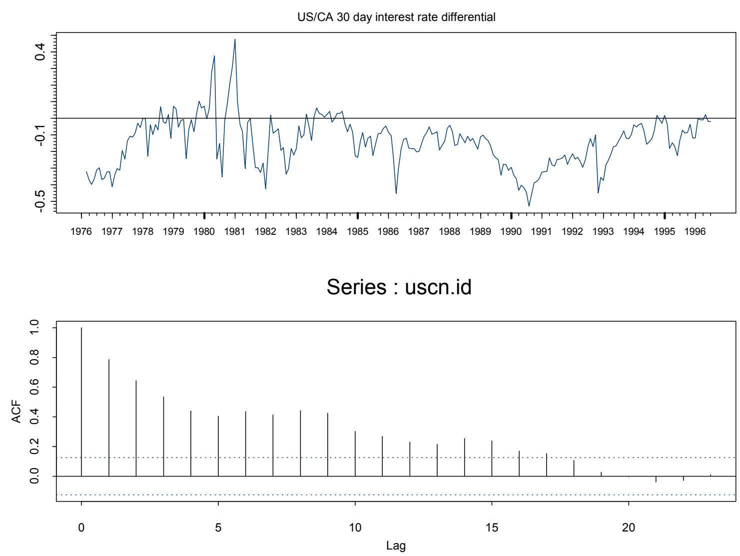

FIGURE 3.5. US/CA 30 day interest rate differential and SACF.

```
[1] 2.409
```

Many economic and financial time series are well characterized by an AR(1) process. Leading examples in finance are valuation ratios (dividend-price ratio, price-earning ratio etc), real exchange rates, interest rates, and interest rate differentials (spreads). To illustrate, consider the 30-day US/CA interest rate differential[2] constructed from the `S+FinMetrics` "`timeSeries`" object `lexrates.dat`:

```
> uscn.id = 100*(lexrates.dat[,"USCNF"]-
+ lexrates.dat[,"USCNS"])
> colIds(uscn.id) = "USCNID"
> uscn.id@title = "US/CA 30 day interest rate differential"
> par(mfrow=c(2,1))
> plot(uscn.id,reference.grid=F)
> abline(h=0)
> tmp = acf(uscn.id)
```

The interest rate differential is clearly persistent: autocorrelations are significant at the 5% level up to 15 months.

[2]By covered interest rate parity, the nominal interest rate differential between risk free bonds from two countries is equal to the difference between the nominal forward and spot exchange rates.

AR(p) Models

The AR(p) model in mean-adjusted form is

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

or, in lag operator notation,

$$\phi(L)(y_t - \mu) = \varepsilon_t$$

where $\phi(L) = 1 - \phi_1 L - \cdots - \phi_p L^p$. The autoregressive form is

$$\phi(L)y_t = c + \varepsilon_t.$$

It can be shown that the AR(p) is stationary and ergodic provided the roots of the *characteristic equation*

$$\phi(z) = 1 - \phi_1 z - \phi_2 z^2 - \cdots - \phi_p z^p = 0 \tag{3.9}$$

lie outside the complex unit circle (have modulus greater than one). A necessary condition for stationarity that is useful in practice is that $|\phi_1 + \cdots + \phi_p| < 1$. If (3.9) has complex roots then y_t will exhibit sinusoidal behavior. In the stationary AR(p), the constant in the autoregressive form is equal to $\mu(1 - \phi_1 - \cdots - \phi_p)$.

The moments of the AR(p) process satisfy the *Yule-Walker equations*

$$\begin{aligned} \gamma_0 &= \phi_1\gamma_1 + \phi_2\gamma_2 + \cdots + \phi_p\gamma_p + \sigma^2 \\ \gamma_j &= \phi_1\gamma_{j-1} + \phi_2\gamma_{j-2} + \cdots + \phi_p\gamma_{j-p} \end{aligned} \tag{3.10}$$

A simple recursive algorithm for finding the Wold representation is based on matching coefficients in $\phi(L)$ and $\psi(L)$ such that $\phi(L)\psi(L) = 1$. For example, in the AR(2) model

$$(1 - \phi_1 L - \phi_2 L^2)(1 + \psi_1 L + \psi_2 L^2 + \cdots) = 1$$

implies

$$\begin{aligned} \psi_1 &= 1 \\ \psi_2 &= \phi_1\psi_1 + \phi_2 \\ \psi_3 &= \phi_1\psi_2 + \phi_2\psi_1 \\ &\vdots \\ \psi_j &= \phi_1\psi_{j-1} + \phi_2\psi_{j-2} \end{aligned}$$

Partial Autocorrelation Function

The *partial autocorrelation function* (PACF) is a useful tool to help identify AR(p) models. The PACF is based on estimating the sequence of AR

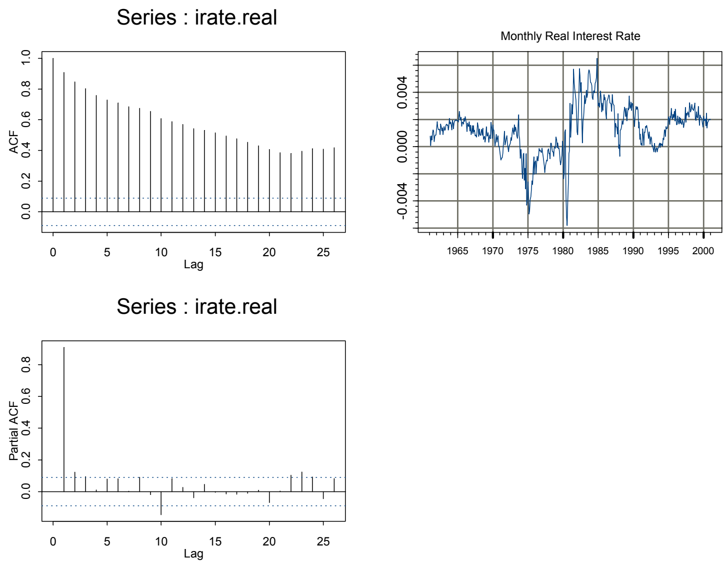

FIGURE 3.6. Monthly U.S. real interest rate, SACF and SPACF.

models

$$\begin{aligned} z_t &= \phi_{11} z_{t-1} + \varepsilon_{1t} \\ z_t &= \phi_{21} z_{t-1} + \phi_{22} z_{t-2} + \varepsilon_{2t} \\ &\vdots \\ z_t &= \phi_{p1} z_{t-1} + \phi_{p2} z_{t-2} + \cdots + \phi_{pp} z_{t-p} + \varepsilon_{pt} \end{aligned}$$

where $z_t = y_t - \mu$ is the demeaned data. The coefficients ϕ_{jj} for $j = 1, \ldots, p$ (i.e., the last coefficients in each AR(p) model) are called the partial autocorrelation coefficients. In an AR(1) model the first partial autocorrelation coefficient ϕ_{11} is non-zero, and the remaining partial autocorrelation coefficients ϕ_{jj} for $j > 1$ are equal to zero. Similarly, in an AR(2), the first and second partial autocorrelation coefficients ϕ_{11} and ϕ_{22} are non-zero and the rest are zero for $j > 2$. For an AR(p) all of the first p partial autocorrelation coefficients are non-zero, and the rest are zero for $j > p$. The sample partial autocorrelation coefficients up to lag p are essentially obtained by estimating the above sequence of p AR models by least squares and retaining the estimated coefficients $\hat{\phi}_{jj}$.

Example 4 *Monthly real interest rates*

The "timeSeries" object varex.ts in the S+FinMetrics module contains monthly data on real stock returns, real interest rates, inflation and real output growth.

```
> colIds(varex.ts)
[1] "MARKET.REAL" "RF.REAL"     "INF"          "IPG"
```

Figure 3.6 shows the real interest rate, RF.REAL, over the period January 1961 through December 2000 produced with the S-PLUS commands

```
> smpl = (positions(varex.ts) > timeDate("12/31/1960"))
> irate.real = varex.ts[smpl,"RF.REAL"]
> par(mfrow=c(2,2))
> acf.plot(acf(irate.real, plot=F))
> plot(irate.real, main="Monthly Real Interest Rate")
> tmp = acf(irate.real, type="partial")
```

The SACF and SPACF indicate that the real interest rate might be modeled as an AR(2) or AR(3) process.

3.2.4 Moving Average Models

MA(1) Model

The MA(1) model has the form

$$y_t = \mu + \varepsilon_t + \theta\varepsilon_{t-1},\ \varepsilon_t \sim WN(0, \sigma^2)$$

For any finite θ the MA(1) is stationary and ergodic. The moments are $E[y_t] = \mu$, $\gamma_0 = \sigma^2(1+\theta^2)$, $\gamma_1 = \sigma^2\theta$, $\gamma_j = 0$ for $j > 1$ and $\rho_1 = \theta/(1+\theta^2)$. Hence, the ACF of an MA(1) process cuts off at lag one, and the maximum value of this correlation is ± 0.5.

There is an identification problem with the MA(1) model since $\theta = 1/\theta$ produce the same value of ρ_1. The MA(1) is called *invertible* if $|\theta| < 1$ and is called *non-invertible* if $|\theta| \geq 1$. In the invertible MA(1), the error term ε_t has an infinite order AR representation of the form

$$\varepsilon_t = \sum_{j=0}^{\infty} \theta^{*j}(y_{t-j} - \mu)$$

where $\theta^* = -\theta$ so that ε_t may be thought of as a prediction error based on past values of y_t. A consequence of the above result is that the PACF for an invertible MA(1) process decays towards zero at an exponential rate.

Example 5 *Signal plus noise model*

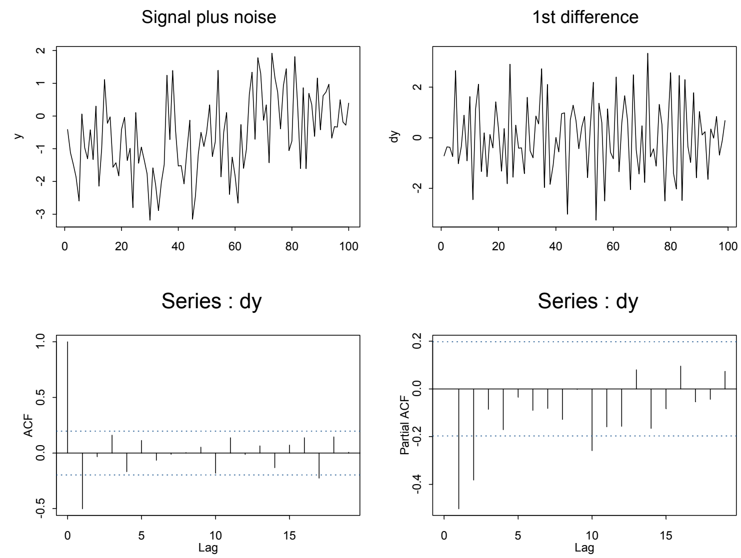

FIGURE 3.7. Simulated data, SACF and SPACF from signal plus noise model.

MA(1) models often arise through data transformations like aggregation and differencing[3]. For example, consider the signal plus noise model

$$\begin{aligned} y_t &= z_t + \varepsilon_t,\ \varepsilon_t \sim WN(0, \sigma_\varepsilon^2) \\ z_t &= z_{t-1} + \eta_t,\ \eta_t \sim WN(0, \sigma_\eta^2) \end{aligned}$$

where ε_t and η_t are independent. For example, z_t could represent the fundamental value of an asset price and ε_t could represent an *iid* deviation about the fundamental price. A stationary representation requires differencing y_t:

$$\Delta y_t = \eta_t + \varepsilon_t - \varepsilon_{t-1}$$

It can be shown, e.g. Harvey (1993), that Δy_t is an MA(1) process with $\theta = \frac{-(q+2)+\sqrt{q^2+4q}}{2}$ where $q = \frac{\sigma_\eta^2}{\sigma_\varepsilon^2}$ is the signal-to-noise ratio and $\rho_1 = \frac{-1}{q+2} < 0$.

Simulated data with $\sigma_\varepsilon^2 = 1$ and $\sigma_\eta^2 = (0.5)^2$ created with the `S-PLUS` commands

```
> set.seed(112)
> eps = rnorm(100,sd=1)
> eta = rnorm(100,sd=0.5)
```

[3] MA(1) type models for asset returns often occur as the result of no-trading effects or bid-ask bounce effects. See Campbell, Lo and MacKinlay (1997) chapter 3 for details.

```
> z = cumsum(eta)
> y = z + eps
> dy = diff(y)
> par(mfrow=c(2,2))
> tsplot(y, main="Signal plus noise",ylab="y")
> tsplot(dy, main="1st difference",ylab="dy")
> tmp = acf(dy)
> tmp = acf(dy,type="partial")
```

are illustrated in Figure 3.7. The signal-to-noise ratio $q = 0.25$ implies a first lag autocorrelation of $\rho_1 = -0.444$. This negative correlation is clearly reflected in the SACF.

MA(q) Model

The MA(q) model has the form

$$y_t = \mu + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \cdots + \theta_q \varepsilon_{t-q}, \text{ where } \varepsilon_t \sim WN(0, \sigma^2)$$

The MA(q) model is stationary and ergodic provided $\theta_1, \ldots, \theta_q$ are finite. It is *invertible* if all of the roots of the MA characteristic polynomial

$$\theta(z) = 1 + \theta_1 z + \cdots \theta_q z^q = 0 \tag{3.11}$$

lie outside the complex unit circle. The moments of the MA(q) are

$$\begin{aligned}
E[y_t] &= \mu \\
\gamma_0 &= \sigma^2(1 + \theta_1^2 + \cdots + \theta_q^2) \\
\gamma_j &= \begin{cases} (\theta_j + \theta_{j+1}\theta_1 + \theta_{j+2}\theta_2 + \cdots + \theta_q \theta_{q-j})\,\sigma^2 \text{ for } j = 1, 2, \ldots, q \\ 0 \text{ for } j > q \end{cases}
\end{aligned}$$

Hence, the ACF of an MA(q) is non-zero up to lag q and is zero afterwards. As with the MA(1), the PACF for an invertible MA(q) will show exponential decay and possibly pseudo cyclical behavior if the roots of (3.11) are complex.

Example 6 *Overlapping returns and MA(q) models*

MA(q) models often arise in finance through data aggregation transformations. For example, let $R_t = \ln(P_t/P_{t-1})$ denote the monthly continuously compounded return on an asset with price P_t. Define the annual return at time t using monthly returns as $R_t(12) = \ln(P_t/P_{t-12}) = \sum_{j=0}^{11} R_{t-j}$. Suppose $R_t \sim WN(\mu, \sigma^2)$ and consider a sample of monthly returns of size T, $\{R_1, R_2, \ldots, R_T\}$. A sample of annual returns may be created using *overlapping* or *non-overlapping* returns. Let $\{R_{12}(12), R_{13}(12), \ldots, R_T(12)\}$ denote a sample of $T^* = T - 11$ monthly overlapping annual returns and $\{R_{12}(12), R_{24}(12), \ldots, R_T(12)\}$ denote a sample of $T/12$ non-overlapping annual returns. Researchers often use overlapping returns in analysis due to the apparent larger sample size. One must be careful using overlapping returns because the monthly annual return sequence $\{R_t(12)\}$ is not a white noise process even if the monthly return sequence $\{R_t\}$ is. To see this, straightforward calculations give

$$
\begin{aligned}
E[R_t(12)] &= 12\mu \\
\gamma_0 &= var(R_t(12)) = 12\sigma^2 \\
\gamma_j &= cov(R_t(12), R_{t-j}(12)) = (12-j)\sigma^2 \text{ for } j < 12 \\
\gamma_j &= 0 \text{ for } j \geq 12
\end{aligned}
$$

Since $\gamma_j = 0$ for $j \geq 12$ notice that $\{R_t(12)\}$ behaves like an MA(11) process

$$
\begin{aligned}
R_t(12) &= 12\mu + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \cdots + \theta_{11}\varepsilon_{t-11} \\
\varepsilon_t &\sim WN(0, \sigma^2)
\end{aligned}
$$

To illustrate, consider creating annual overlapping continuously compounded returns on the S&P 500 index over the period February 1990 through January 2001. The `S+FinMetrics` "`timeSeries`" `singleIndex.dat` contains the S&P 500 price data and the continuously compounded monthly returns are computed using the `S+FinMetrics` function `getReturns`

```
> sp500.mret = getReturns(singleIndex.dat[,"SP500"],
+ type="continuous")
> sp500.mret@title = "Monthly returns on S&P 500 Index"
```

The monthly overlapping annual returns are easily computed using the `S-PLUS` function `aggregateSeries`

```
> sp500.aret = aggregateSeries(sp500.mret,moving=12,FUN=sum)
> sp500.aret@title = "Monthly Annual returns on S&P 500 Index"
```

The optional argument `moving=12` specifies that the `sum` function is to be applied to moving blocks of size 12. The data together with the SACF and SPACF of the monthly annual returns are displayed in Figure 3.8.

The SACF has non-zero values up to lag 11. Interestingly, the SPACF is very small at all lags except the first.

3.2.5 ARMA(p,q) Models

The general ARMA(p, q) model in mean-adjusted form is given by (3.7). The regression formulation is

$$
y_t = c + \phi_1 y_{t-1} + \cdots + \phi_p y_{t-p} + \varepsilon_t + \theta\varepsilon_{t-1} + \cdots + \theta\varepsilon_{t-q} \tag{3.12}
$$

It is stationary and ergodic if the roots of the characteristic equation $\phi(z) = 0$ lie outside the complex unit circle, and it is invertible if the roots of the

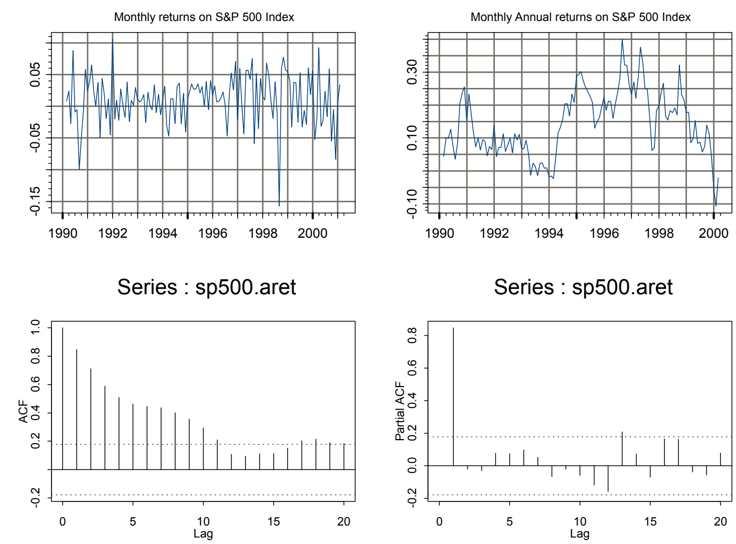

FIGURE 3.8. Monthly non-overlapping and overlapping annual returns on the S&P 500 index.

MA characteristic polynomial $\theta(z) = 0$ lie outside the unit circle. It is assumed that the polynomials $\phi(z) = 0$ and $\theta(z) = 0$ do not have canceling or common factors. A stationary and ergodic ARMA(p, q) process has a mean equal to

$$\mu = \frac{c}{1 - \phi_1 - \cdots - \phi_p} \tag{3.13}$$

and its autocovariances, autocorrelations and impulse response weights satisfy the recursive relationships

$$\begin{aligned}
\gamma_j &= \phi_1 \gamma_{j-1} + \phi_2 \gamma_{j-2} + \cdots + \phi_p \gamma_{j-p} \\
\rho_j &= \phi_1 \rho_{j-1} + \phi_2 \rho_{j-2} + \cdots + \phi_p \rho_{j-p} \\
\psi_j &= \phi_1 \psi_{j-1} + \phi_2 \psi_{j-2} + \cdots + \phi_p \psi_{j-p}
\end{aligned}$$

The general form of the ACF for an ARMA(p, q) process is complicated. See Hamilton (1994) chapter five for details. In general, for an ARMA(p, q) process, the ACF behaves like the ACF for an AR(p) process for $p > q$, and the PACF behaves like the PACF for an MA(q) process for $q > p$. Hence, both the ACF and PACF eventually show exponential decay.

ARMA(p, q) models often arise from certain aggregation transformations of simple time series models. An important result due to Granger and Morris (1976) is that if y_{1t} is an ARMA(p_1, q_1) process and y_{2t} is an ARMA(p_2, q_2) process, which may be contemporaneously correlated with y_{1t}, then $y_{1t} + y_{2t}$ is an ARMA(p, q) process with $p = p_1 + p_2$ and $q = \max(p_1 + q_2, q_1 + p_2)$. For example, if y_{1t} is an AR(1) process and y_2 is a AR(1) process, then $y_1 + y_2$ is an ARMA(2,1) process.

High order ARMA(p, q) processes are difficult to identify and estimate in practice and are rarely used in the analysis of financial data. Low order ARMA(p, q) models with p and q less than three are generally sufficient for the analysis of financial data.

ARIMA(p, d, q) Models

The specification of the ARMA(p, q) model (3.7) assumes that y_t is stationary and ergodic. If y_t is a trending variable like an asset price or a macroeconomic aggregate like real GDP, then y_t must be transformed to stationary form by eliminating the trend. Box and Jenkins (1976) advocate removal of trends by differencing. Let $\Delta = 1-L$ denote the *difference operator*. If there is a linear trend in y_t then the first difference $\Delta y_t = y_t - y_{t-1}$ will not have a trend. If there is a quadratic trend in y_t, then Δy_t will contain a linear trend but the second difference $\Delta^2 y_t = (1-2L+L^2)y_t = y_t - 2y_{t-1} + y_{t-2}$ will not have a trend. The class of ARMA(p, q) models where the trends have been transformed by differencing d times is denoted ARIMA(p, d, q)[4].

3.2.6 Estimation of ARMA Models and Forecasting

ARMA(p, q) models are generally estimated using the technique of maximum likelihood, which is usually accomplished by putting the ARMA(p, q) in state-space form from which the prediction error decomposition of the log-likelihood function may be constructed. Details of this process are given in Harvey (1993). An often ignored aspect of the maximum likelihood estimation of ARMA(p, q) models is the treatment of initial values. These initial values are the first p values of y_t and q values of ε_t in (3.7). The *exact likelihood* utilizes the stationary distribution of the initial values in the construction of the likelihood. The *conditional likelihood* treats the p initial values of y_t as fixed and often sets the q initial values of ε_t to zero. The exact maximum likelihood estimates (MLEs) maximize the exact log-likelihood, and the conditional MLEs maximize the conditional log-likelihood. The exact and conditional MLEs are asymptotically equivalent but can differ substantially in small samples, especially for models that are close to being nonstationary or noninvertible.[5]

---

[4] More general ARIMA(p, d, q) models allowing for seasonality are discussed in chapter 27 of the *S-PLUS Guide to Statistics, Vol. II.*

[5] As pointed out by Venables and Ripley (1999) page 415, the maximum likelihood estimates computed using the `S-PLUS` function `arima.mle` are conditional MLEs. Exact MLEs may be easily computed using the `S+FinMetrics` state space modeling functions.

For pure AR models, the conditional MLEs are equivalent to the least squares estimates from the model

$$y_t = c + \phi_1 y_{t-1} + \cdots + \phi_p y_{t-p} + \varepsilon_t \tag{3.14}$$

Notice, however, that c in (3.14) is not an estimate of $E[y_t] = \mu$. The least squares estimate of μ is given by plugging in the least squares estimates of $c, \phi_1, \ldots, \phi_p$ into (3.13).

Model Selection Criteria

Before an ARMA(p, q) may be estimated for a time series y_t, the AR and MA orders p and q must be determined by visually inspecting the SACF and SPACF for y_t. Alternatively, statistical *model selection criteria* may be used. The idea is to fit all ARMA(p, q) models with orders $p \leq p_{\max}$ and $q \leq q_{\max}$ and choose the values of p and q which minimizes some model selection criteria. Model selection criteria for ARMA(p, q) models have the form

$$MSC(p, q) = \ln(\tilde{\sigma}^2(p, q)) + c_T \cdot \varphi(p, q)$$

where $\tilde{\sigma}^2(p, q)$ is the MLE of $var(\varepsilon_t) = \sigma^2$ *without a degrees of freedom correction* from the ARMA(p, q) model, c_T is a sequence indexed by the sample size T, and $\varphi(p, q)$ is a penalty function which penalizes large ARMA(p, q) models. The two most common information criteria are the Akaike (AIC) and Schwarz-Bayesian (BIC):

$$\begin{aligned} AIC(p, q) &= \ln(\tilde{\sigma}^2(p, q)) + \frac{2}{T}(p + q) \\ BIC(p, q) &= \ln(\tilde{\sigma}^2(p, q)) + \frac{\ln T}{T}(p + q) \end{aligned}$$

The AIC criterion asymptotically overestimates the order with positive probability, whereas the BIC estimate the order consistently under fairly general conditions if the true orders p and q are less than or equal to $p_{\max}$ and $q_{\max}$. However, in finite samples the BIC generally shares no particular advantage over the AIC.

Forecasting Algorithm

Forecasts from an ARIMA(p, d, q) model are straightforward. The model is put in state space form, and optimal h-step ahead forecasts along with forecast standard errors (not adjusted for parameter uncertainty) are produced using the Kalman filter algorithm. Details of the method are given in Harvey (1993).

Estimation and Forecasting ARIMA(p, d, q) Models Using the S-PLUS Function arima.mle

Conditional MLEs may be computed using the S-PLUS function `arima.mle`. The form of the ARIMA(p, d, q) assumed by `arima.mle` is

$$\begin{aligned} y_t &= \phi_1 y_{t-1} + \cdots + \phi_p y_{t-p} \\ &\quad + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \cdots - \theta_q \varepsilon_{t-q} \\ &\quad + \boldsymbol{\beta}' \mathbf{x}_t \end{aligned}$$

where $\mathbf{x}_t$ represents additional explanatory variables. It is assumed that y_t has been differenced d times to remove any trends and that the unconditional mean μ has been subtracted out so that y_t is demeaned. Notice that `arima.mle` assumes that the signs on the MA coefficients θ_j are the opposite to those in (3.7).

The arguments expected by `arima.mle` are

```
> args(arima.mle)
function(x, model = NULL, n.cond = 0, xreg = NULL, ...)
```

where `x` is a univariate "`timeSeries`" or vector, `model` is a list object describing the specification of the ARMA model, `n.cond` sets the number of initial observations on which to condition in the formation of the log-likelihood, and `xreg` is a "`timeSeries`", vector or matrix of additional explanatory variables. By default, `arima.mle` assumes that the ARIMA(p, d, q) model is stationary and in mean-adjusted form with an estimate of μ subtracted from the observed data y_t. To estimate the regression form (3.12) of the ARIMA(p, q) model, simply set `xreg=1`. ARIMA(p, d, q) models are specified using list variables the form

```
> mod.list = list(order=c(1,0,1))
> mod.list = list(order=c(1,0,1),ar=0.75,ma=0)
> mod.list = list(ar=c(0.75,-0.25),ma=c(0,0))
```

The first list simply specifies an ARMA(1,0,1)/ARMA(1,1) model. The second list specifies an ARIMA(1,0,1) as well as starting values for the AR and MA parameters ϕ and θ. The third list implicitly determines an ARMA(2,2) model by giving the starting values for the AR and MA parameters. The function `arima.mle` produces an object of class "`arima`" for which there are `print` and `plot` methods. Diagnostics from the fit can be created with the S-PLUS function `arima.diag,` and forecasts may be produced using `arima.forecast`.

Example 7 *Estimation of ARMA model for US/CA interest rate differential*

Consider estimating an ARMA(p, q) for the monthly US/CA interest rate differential data in the "`timeSeries`" `uscn.id` used in a previous example. To estimate an ARMA(1,1) model for the demeaned interest rate differential with starting values $\phi = 0.75$ and $\theta = 0$ use

```
> uscn.id.dm = uscn.id - mean(uscn.id)
> arma11.mod = list(ar=0.75,ma=0)
> arma11.fit = arima.mle(uscn.id.dm,model=arma11.mod)
> class(arma11.fit)
[1] "arima"
```

The components of `arma11.fit` are

```
> names(arma11.fit)
 [1] "model"     "var.coef"  "method"    "series"
 [5] "aic"       "loglik"    "sigma2"    "n.used"
 [9] "n.cond"    "converged" "conv.type" "call"
```

To see the basic fit simply type

```
> arma11.fit
Call: arima.mle(x = uscn.id.dm, model = arma11.mod)
Method:  Maximum Likelihood
Model :  1 0 1

Coefficients:
     AR : 0.82913
     MA : 0.11008

Variance-Covariance Matrix:
         ar(1)    ma(1)
ar(1) 0.002046 0.002224
ma(1) 0.002224 0.006467

Optimizer has  converged
Convergence Type: relative function convergence
AIC: -476.25563
```

The conditional MLEs are $\hat{\phi}_{cmle} = 0.829$ and $\hat{\theta}_{cmle} = -0.110$. Standard errors for these parameters are given by the square roots of the diagonal elements of variance-covariance matrix

```
> std.errs = sqrt(diag(arma11.fit$var.coef))
> names(std.errs) = colIds(arma11.fit$var.coef)
> std.errs
   ar(1)   ma(1)
 0.04523 0.08041
```

It appears that the $\hat{\theta}_{cmle}$ is not statistically different from zero.

To estimate the ARMA(1,1) for the interest rate differential data in regression form (3.12) with an intercept use

```
> arma11.fit2 = arima.mle(uscn.id,model=arma11.mod,xreg=1)
> arma11.fit2
Call: arima.mle(x = uscn.id, model = arma11.mod, xreg = 1)
Method:  Maximum Likelihood
Model :  1 0 1

Coefficients:
     AR : 0.82934
     MA : 0.11065

Variance-Covariance Matrix:
         ar(1)    ma(1)
ar(1) 0.002043 0.002222
ma(1) 0.002222 0.006465
Coeffficients for regressor(s): intercept
[1] -0.1347

Optimizer has  converged
Convergence Type: relative function convergence
AIC: -474.30852
```

The conditional MLEs for ϕ and θ are essentially the same as before, and the MLE for c is $\hat{c}_{cmle} = -0.1347$. Notice that the reported variance-covariance matrix only gives values for the estimated ARMA coefficients $\hat{\phi}_{cmle}$ and $\hat{\theta}_{cmle}$.

Graphical diagnostics of the fit produced using the `plot` method

```
> plot(arma11.fit)
```

are illustrated in Figure 3.9. There appears to be some high order serial correlation in the errors as well as heteroskedasticity.

The h-step ahead forecasts of future values may be produced with the S-PLUS function `arima.forecast`. For example, to produce monthly forecasts for the demeaned interest rate differential from July 1996 through June 1997 use

```
> fcst.dates = timeSeq("7/1/1996", "6/1/1997",
+ by="months", format="%b %Y")
> uscn.id.dm.fcst = arima.forecast(uscn.id.dm, n=12,
+ model=arma11.fit$model, future.positions=fcst.dates)
> names(uscn.id.dm.fcst)
[1] "mean"    "std.err"
```

The object `uscn.id.dm.fcst` is a list whose first component is a "`timeSeries`" containing the h-step forecasts, and the second component is a "`timeSeries`" containing the forecast standard errors:

```
> uscn.id.dm.fcst[[1]]
```

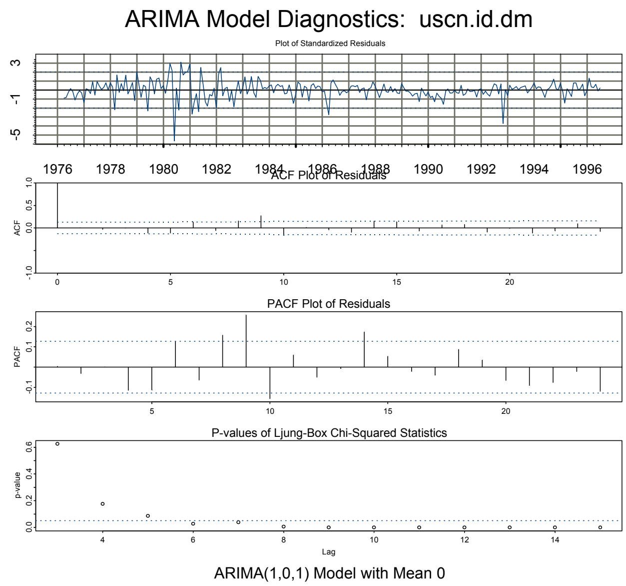

FIGURE 3.9. Residual diagnostics from ARMA(1,1) fit to US/CA interest rate differentials.

```
Positions        1
Jul 1996   0.09973
Aug 1996   0.08269
Sep 1996   0.06856
Oct 1996   0.05684
Nov 1996   0.04713
Dec 1996   0.03908
Jan 1997   0.03240
Feb 1997   0.02686
Mar 1997   0.02227
Apr 1997   0.01847
May 1997   0.01531
Jun 1997   0.01270
```

The data, forecasts and 95% forecast confidence intervals shown in Figure 3.10 are produced by

```
> smpl = positions(uscn.id.dm) >= timeDate("6/1/1995")
> plot(uscn.id.dm[smpl,],uscn.id.dm.fcst$mean,
+ uscn.id.dm.fcst$mean+2*uscn.id.dm.fcst$std.err,
+ uscn.id.dm.fcst$mean-2*uscn.id.dm.fcst$std.err,
+ plot.args=list(lty=c(1,4,3,3)))
```

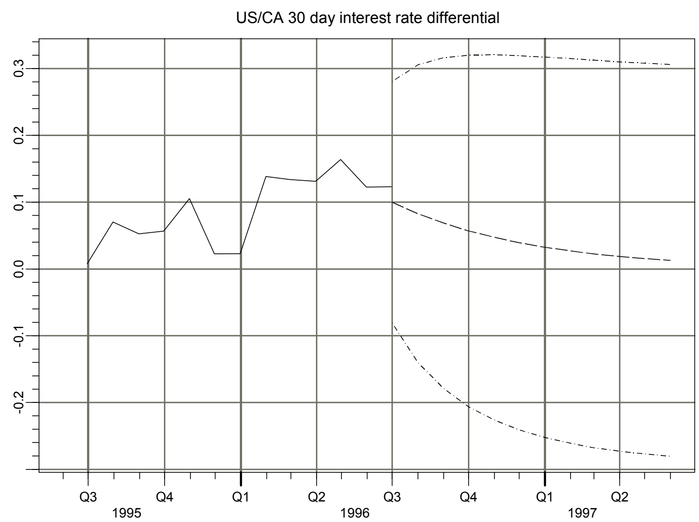

FIGURE 3.10. Forecasts for 12 months for the series `uscn.id.dm`.

Estimating AR(p) by Least Squares Using the `S+FinMetrics` Function `OLS`

As previously mentioned, the conditional MLEs for an AR(p) model may be computed using least squares. The `S+FinMetrics` function `OLS`, which extends the `S-PLUS` function `lm` to handle general time series regression, may be used to estimate an AR(p) in a particularly convenient way. The general use of OLS is discussed in Chapter 6, and its use for estimating an AR(p) is only mentioned here. For example, to estimate an AR(2) model for the US/CA interest rate differential use

```
> ar2.fit = OLS(USCNID~ar(2), data=uscn.id)
> ar2.fit

Call:
OLS(formula = USCNID ~ar(2), data = uscn.id)

Coefficients:
 (Intercept)    lag1    lag2
 -0.0265      0.7259  0.0758

Degrees of freedom: 243 total; 240 residual
Time period: from Apr 1976 to Jun 1996
Residual standard error: 0.09105
```

The least squares estimates of the AR coefficients are $\hat{\phi}_1 = 0.7259$ and $\hat{\phi}_2 = 0.0758$. Since $\hat{\phi}_1 + \hat{\phi}_2 < 1$ the estimated AR(2) model is stationary. To be sure, the roots of $\phi(z) = 1 - \hat{\phi}_1 z - \hat{\phi}_2 z^2 = 0$ are

```
> abs(polyroot(c(1,-ar2.fit$coef[2:3])))
[1]  1.222 10.798
```

are outside the complex unit circle.

3.2.7 Martingales and Martingale Difference Sequences

Let $\{y_t\}$ denote a sequence of random variables and let $I_t = \{y_{t,} y_{t-1}, \ldots\}$ denote a set of conditioning information or *information set* based on the past history of y_t. The sequence $\{y_t, I_t\}$ is called a *martingale* if

- $I_{t-1} \subset I_t$ (I_t is a filtration)
- $E[|y_t|] < \infty$
- $E[y_t|I_{t-1}] = y_{t-1}$ (martingale property)

The most common example of a martingale is the random walk model

$$y_t = y_{t-1} + \varepsilon_t,\ \varepsilon_t \sim WN(0, \sigma^2)$$

where y_0 is a fixed initial value. Letting $I_t = \{y_t, \ldots, y_0\}$ implies $E[y_t|I_{t-1}] = y_{t-1}$ since $E[\varepsilon_t|I_{t-1}] = 0$.

Let $\{\varepsilon_t\}$ be a sequence of random variables with an associated information set I_t. The sequence $\{\varepsilon_t, I_t\}$ is called a *martingale difference sequence* (MDS) if

- $I_{t-1} \subset I_t$
- $E[\varepsilon_t|I_{t-1}] = 0$ (MDS property)

If $\{y_t, I_t\}$ is a martingale, a MDS $\{\varepsilon_t, I_t\}$ may be constructed by defining

$$\varepsilon_t = y_t - E[y_t|I_{t-1}]$$

By construction, a MDS is an uncorrelated process. This follows from the *law of iterated expectations.* To see this, for any $k > 0$

$$\begin{aligned} E[\varepsilon_t \varepsilon_{t-k}] &= E[E[\varepsilon_t \varepsilon_{t-k}|I_{t-1}]] \\ &= E[\varepsilon_{t-k} E[\varepsilon_t|I_{t-1}]] \\ &= 0 \end{aligned}$$

In fact, if z_n is any function of the past history of ε_t so that $z_n \in I_{t-1}$ then

$$E[\varepsilon_t z_n] = 0$$

Although a MDS is an uncorrelated process, it does not have to be an independent process. That is, there can be dependencies in the higher order moments of ε_t. The *autoregressive conditional heteroskedasticity* (ARCH) process in the following example is a leading example in finance.

MDSs are particularly nice to work with because there are many useful convergence results (laws of large numbers, central limit theorems etc.). White (1984), Hamilton (1994) and Hayashi (2000) describe the most useful of these results for the analysis of financial time series.

Example 8 *ARCH process*

A well known stylized fact about high frequency financial asset returns is that volatility appears to be autocorrelated. A simple model to capture such volatility autocorrelation is the ARCH process due to Engle (1982). To illustrate, let r_t denote the daily return on an asset and assume that $E[r_t] = 0$. An ARCH(1) model for r_t is

$$\begin{aligned} r_t &= \sigma_t z_t & (3.15) \\ z_t &\sim iid\ N(0,1) \\ \sigma_t^2 &= \omega + \alpha r_{t-1}^2 & (3.16) \end{aligned}$$

where $\omega > 0$ and $0 < \alpha < 1$. Let $I_t = \{r_t, \ldots\}$. The `S+FinMetrics` function `simulate.garch` may be used to generate simulations from above ARCH(1) model. For example, to simulate 250 observations on r_t with $\omega = 0.1$ and $\alpha = 0.8$ use

```
> rt = simulate.garch(model=list(a.value=0.1, arch=0.8),
+      n=250, rseed=196)
> class(rt)
[1] "structure"
> names(rt)
[1] "et"      "sigma.t"
```

Notice that the function `simulate.garch` produces simulated values of both r_t and σ_t. These values are shown in Figure 3.11.

To see that $\{r_t, I_t\}$ is a MDS, note that

$$\begin{aligned} E[r_t|I_{t-1}] &= E[z_t\sigma_t|I_{t-1}] \\ &= \sigma_t E[z_t|I_{t-1}] \\ &= 0 \end{aligned}$$

Since r_t is a MDS, it is an uncorrelated process. Provided $|\alpha| < 1$, r_t is a mean zero covariance stationary process. The unconditional variance of r_t is given by

$$\begin{aligned} var(r_t) &= E[r_t^2] = E[E[z_t^2\sigma_t^2|I_{t-1}]] \\ &= E[\sigma_t^2 E[z_t^2|I_{t-1}] = E[\sigma_t^2] \end{aligned}$$

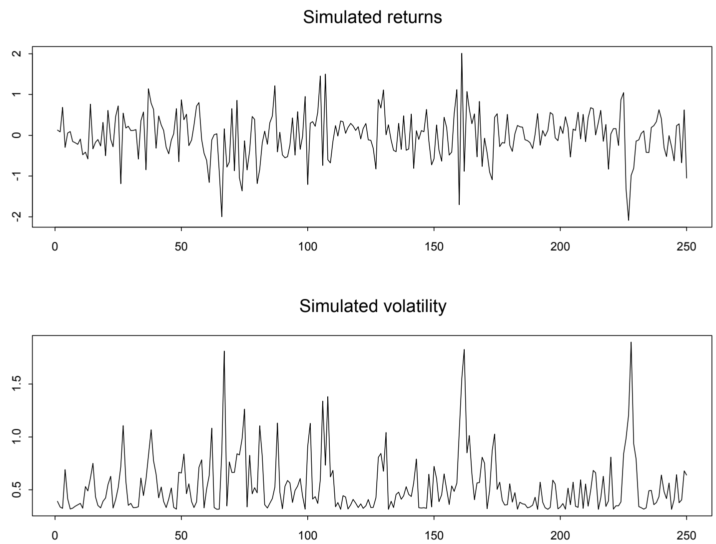

FIGURE 3.11. Simulated values from ARCH(1) process with $\omega = 1$ and $\alpha = 0.8$.

since $E[z_t^2|I_{t-1}] = 1$. Utilizing (3.16) and the stationarity of r_t, $E[\sigma_t^2]$ may be expressed as

$$E[\sigma_t^2] = \frac{\omega}{1-\alpha}$$

Furthermore, by adding ε_t^2 to both sides of (3.16) and rearranging it follows that r_t^2 has an AR(1) representation of the form

$$\varepsilon_t^2 = \omega + \alpha\varepsilon_{t-1}^2 + v_t$$

where $v_t = \varepsilon_t^2 - \sigma_t^2$ is a MDS.

3.2.8 Long-run Variance

Let y_t be a stationary and ergodic time series. Anderson's central limit theorem for stationary and ergodic processes (c.f. Hamilton (1994) pg. 195) states

$$\sqrt{T}(\bar{y} - \mu) \xrightarrow{d} N(0, \sum_{j=-\infty}^{\infty} \gamma_j)$$

or

$$\bar{y} \overset{A}{\sim} N\left(\mu, \frac{1}{T} \sum_{j=-\infty}^{\infty} \gamma_j\right)$$

The sample size, T, times the *asymptotic variance* of the sample mean is often called the *long-run variance* of y_t[6] :

$$lrv(y_t) = T \cdot avar(\bar{y}) = \sum_{j=-\infty}^{\infty} \gamma_j.$$

Since $\gamma_{-j} = \gamma_j$, $lrv(y_t)$ may be alternatively expressed as

$$lrv(y_t) = \gamma_0 + 2\sum_{j=1}^{\infty} \gamma_j.$$

Using the long-run variance, an asymptotic 95% confidence interval for μ takes the form

$$\bar{y} \pm 1.96 \cdot \sqrt{T^{-1}\widehat{lrv}(y_t)}$$

where $\widehat{lrv}(y_t)$ is a consistent estimate of $lrv(y_t)$.

Estimating the Long-Run Variance

If y_t is a linear process, it may be shown that

$$\sum_{j=-\infty}^{\infty} \gamma_j = \sigma^2 \left(\sum_{j=0}^{\infty} \psi_j \right)^2 = \sigma^2 \psi(1)^2$$

and so

$$lrv(y_t) = \sigma^2 \psi(1)^2 \tag{3.17}$$

Further, if $y_t \sim \text{ARMA}(p, q)$ then

$$\psi(1) = \frac{1 + \theta_1 + \cdots + \theta_q}{1 - \phi_1 - \cdots - \phi_p} = \frac{\theta(1)}{\phi(1)}$$

so that

$$lrv(y_t) = \frac{\sigma^2 \theta(1)^2}{\phi(1)^2}. \tag{3.18}$$

A consistent estimate of $lrv(y_t)$ may then be computed by estimating the parameters of the appropriate ARMA(p, q) model and substituting these estimates into (3.18). Alternatively, the ARMA(p, q) process may be approximated by a high order AR(p^*) process

$$y_t = c + \phi_1 y_{t-1} + \cdots + \phi_{p^*} y_{t-p^*} + \varepsilon_t$$

[6] Using spectral methods, $lrv(\bar{y})$ has the alternative representation

$$lrv(\bar{y}) = \frac{1}{T} 2\pi f(0)$$

where $f(0)$ denotes the spectral density of y_t evaluated at frequency 0.

where the lag length p^* is chosen such that ε_t is uncorrelated. This gives rise to the *autoregressive long-run variance* estimate

$$lrv_{AR}(y_t) = \frac{\sigma^2}{\phi^*(1)^2}. \tag{3.19}$$

A consistent estimate of $lrv(y_t)$ may also be computed using some non-parametric methods. An estimator made popular by Newey and West (1987) is the weighted autocovariance estimator

$$\widehat{lrv}_{NW}(y_t) = \hat{\gamma}_0 + 2\sum_{j=1}^{M_T} w_{j,T} \cdot \hat{\gamma}_j \tag{3.20}$$

where $w_{j,T}$ are weights which sum to unity and M_T is a truncation lag parameter that satisfies $M_T = O(T^{1/3})$. For MA(q) processes, $\gamma_j = 0$ for $j > q$ and Newey and West suggest using the *rectangular* weights $w_{j,T} = 1$ for $j \leq M_T = q$; 0 otherwise. For general linear processes, Newey and West suggest using the *Bartlett* weights $w_{j,T} = 1 - \frac{j}{M_T+1}$ with M_T equal to the integer part of $4(T/100)^{2/9}$.

Example 9 *Long-run variance of AR(1)*

Let y_t be an AR(1) process created using

```
> set.seed(101)
> e = rnorm(100,sd=1)
> y.ar1 = 1 + arima.sim(model=list(ar=0.75),innov=e)
```

Here $\psi(1) = \frac{1}{\phi(1)} = \frac{1}{1-\phi}$ and

$$lrv(y_t) = \frac{\sigma^2}{(1-\phi)^2}.$$

For $\phi = 0.75$, $\sigma^2 = 1$, $lrv(y_t) = 16$ implies for $T = 100$ an asymptotic standard error for $\bar{y}$ equal to $SE(\bar{y}) = 0.40$. If $y_t \sim WN(0,1)$, then the asymptotic standard error for $\bar{y}$ is $SE(\bar{y}) = 0.10$.

$lrv_{AR}(y_t)$ may be easily computed in S-PLUS using OLS to estimate the AR(1) parameters:

```
> ar1.fit = OLS(y.ar1~ar(1))
> rho.hat = coef(ar1.fit)[2]
> sig2.hat = sum(residuals(ar1.fit)^2)/ar1.fit$df.resid
> lrv.ar1 = sig2.hat/(1-rho.hat)^2
> as.numeric(lrv.ar1)
[1] 13.75
```

Here $lrv_{AR}(y_t) = 13.75$, and an estimate for $SE(\bar{y})$ is $\widehat{SE}_{AR}(\bar{y}) = 0.371$.

The S+FinMetrics function asymp.var may be used to compute the nonparameteric Newey-West estimate $lrv_{NW}(y_t)$. The arguments expected by asymp.var are

```
 > args(asymp.var)
function(x, bandwidth, window = "bartlett", na.rm = F)
```

where `x` is a "`timeSeries`", `bandwidth` sets the truncation lag M_T in (3.20) and `window` specifies the weight function. Newey and West suggest setting the `bandwidth` using the sample size dependent rule

$$M_T = 4(T/100)^{2/9}$$

which is equal to 4 in the present case. The Newey-West long-run variance estimate is then

```
> lrv.nw = asymp.var(y.ar1, bandwidth=4)
> lrv.nw
[1] 7.238
```

and the Newey-West estimate of $SE(\bar{y})$ is $\widehat{SE}_{NW}(\bar{y}) = 0.269$.

3.2.9 Variance Ratios

There has been considerable interest in testing the so-called *random walk* (RW) model for log stock prices (see chapter 2 in Campbell, Lo and MacKinlay (1997) for an extensive review). The RW model for log prices p_t has the form

$$p_t = \mu + p_{t-1} + \varepsilon_t, \ t = 1, \ldots, T$$

where ε_t is a random error term. Using $r_t = \Delta p_t$, the RW model may be rewritten as

$$r_t = \mu + \varepsilon_t$$

Campbell, Lo and MacKinlay distinguish three forms of the random walk model:

RW1 $\varepsilon_t \sim iid(0, \sigma^2)$

RW2 ε_t is an independent process (allows for heteroskedasticity)

RW3 ε_t is an uncorrelated process (allows for dependence in higher order moments)

For asset returns, RW1 and RW2 are not very realistic and, therefore, most attention has been placed on testing the model RW3.

Some commonly used tests for RW3 are based on constructing *variance ratios*. To illustrate, consider the simple two-period variance ratio

$$VR(2) = \frac{var(r_t(2))}{2 \cdot var(r_t)}$$

The numerator of the variance ratio is the variance of the two-period return, $r_t(2) = r_{t-1} + r_t$, and the deminator is two times the variance of the one-period return, r_t. Under RW1, is easy to see that $VR(2) = 1$. If $\{r_t\}$ is an ergodic-stationary process then

$$\begin{aligned} VR(2) &= \frac{var(r_{t-1}) + var(r_t) + 2 \cdot cov(r_t, r_{t-1})}{2 \cdot var(r_t)} \\ &= \frac{2\gamma_0 + 2\gamma_1}{2\gamma_0} = 1 + \rho_1 \end{aligned}$$

There are three cases of interest depending on the value of ρ_1. If $\rho_1 = 0$ then $VR(2) = 1$; if $\rho_1 > 1$ then $VR(2) > 1$; if $\rho_1 < 1$ then $VR(2) < 1$.

The general $q-$period variance ratio is

$$VR(q) = \frac{var(r_t(q))}{q \cdot var(r_t)} \tag{3.21}$$

where $r_t(q) = r_{t-q+1} + \cdots + r_t$. Under RW1, $VR(q) = 1$. For ergodic stationary returns, some algebra shows that

$$VR(q) = 1 + 2 \cdot \sum_{k=1}^{q} \left(1 - \frac{k}{q}\right) \rho_k$$

When the variance ratio is greater than one, returns are called *mean averting* due to the dominating presence of positive autocorrelations. When the variance ratio is less than one, returns are called *mean reverting* due to the dominating presence of negative autocorrelations. Using the Wold representation (3.6), it can be shown that

$$\lim_{q \to \infty} VR(q) = \frac{\sigma^2 \psi(1)^2}{\gamma_0} = \frac{lrv(r_t)}{var(r_t)}$$

That is, as q becomes large the variance ratio approaches the ratio of the long-run variance to the short-run variance. Furthermore, Under RW2 and RW3 it can be shown that $VR(q) \to 1$ as $q \to \infty$ provided

$$\frac{1}{T} \sum_{t=1}^{T} var(r_t) \to \bar{\sigma}^2 > 0$$

Test Statistics

Let $\{p_0, p_1, \ldots, p_{Tq}\}$ denote a sample of $Tq + 1$ log prices, which produces a sample of Tq one-period returns $\{r_1, \ldots, r_{Tq}\}$. Lo and MacKinlay (1988, 1989) develop a number of test statistics for testing the random walk hypothesis based on the estimated variance ratio

$$\widehat{VR}(q) = \frac{\widehat{var}(r_t(q))}{q \cdot \widehat{var}(r_t)} \tag{3.22}$$

The form of the statistic depends on the particular random walk model (RW1, RW2 or RW3) assumed under the null hypothesis.

Under RW1, (3.22) is computed using

$$\widehat{VR}(q) = \frac{\hat{\sigma}^2(q)}{\hat{\sigma}^2}$$

where

$$\begin{aligned}
\hat{\sigma}^2 &= \frac{1}{Tq}\sum_{k=1}^{Tq}(r_k - \hat{\mu})^2 \\
\hat{\sigma}^2(q) &= \frac{1}{Tq^2}\sum_{k=q}^{Tq}(r_k(q) - q\hat{\mu})^2 \\
\hat{\mu} &= \frac{1}{Tq}\sum_{k=1}^{Tq} r_k = \frac{1}{Tq}(p_{Tq} - p_0)
\end{aligned}$$

Lo and MacKinlay show that, under RW1,

$$\sqrt{Tq}(\widehat{VR}(q) - 1) \overset{A}{\sim} N(0, 2(q-1))$$

Therefore, the *variance ratio test statistic*

$$\hat{\psi}(q) = \left(\frac{Tq}{2(q-1)}\right)^{1/2}(\widehat{VR}(q) - 1) \tag{3.23}$$

has a limiting standard normal distribution under RW1.

Lo and MacKinlay also derive a modified version of (3.23) based on the following bias corrected estimates of σ^2 and $\sigma^2(q)$:

$$\begin{aligned}
\bar{\sigma}^2 &= \frac{1}{Tq - 1}\sum_{k=1}^{Tq}(r_k - \hat{\mu})^2 \\
\bar{\sigma}^2(q) &= \frac{1}{m}\sum_{k=q}^{Tq}(r_k(q) - q\hat{\mu})^2 \\
m &= q(Tq - q + 1)\left(1 - \frac{q}{Tq}\right)
\end{aligned}$$

Defining $\overline{VR}(q) = \bar{\sigma}^2(q)/\bar{\sigma}^2$, the biased corrected version of (3.23) has the form

$$\bar{\psi}(q) = \left(\frac{3Tq^2}{2(2q-1)(q-1)}\right)^{1/2}(\overline{VR}(q) - 1) \tag{3.24}$$

which has a limiting standard normal distribution under RW1.

The variance ratio statistics (3.23) and (3.24) are not valid under the empirically relevant RW3 and RW3 models. For this model, Lo and MacKinlay derived the heteroskedasticity robust variance ratio statistic

$$\psi^*(q) = \hat{\Omega}(q)^{-1/2}(\overline{VR}(q) - 1) \tag{3.25}$$

where

$$\begin{aligned}
\hat{\Omega}(q) &= \sum_{j=1}^{q-1} \left(\frac{2(q-j)}{j} \right) \hat{\delta}_j \\
\hat{\delta}_j &= \frac{\sum_{t=j+1}^{Tq} \hat{\alpha}_{0t}\hat{\alpha}_{jt}}{\left(\sum_{j=1}^{Tq} \hat{\alpha}_{0t} \right)^2} \\
\hat{\alpha}_{jt} &= (r_{t-j} - r_{t-j-1} - \hat{\mu})
\end{aligned}$$

Under RW2 or RW3, Lo and MacKinlay show that (3.25) has a limiting stardard normal distribution.

Example 10 *Testing the random walk hypothesis using variance ratios*

The variance ratio statistics (3.23), (3.24) and (3.25) may be computed using the `S+FinMetrics` function `varRatioTest`. The arguments for `varRatioTest` are

```
> args(varRatioTest)
function(x, n.periods, unbiased = T, hetero = F)
```

where `x` is the log return series (which may contain more than one series) and `n.periods` denotes the number of periods q in the variance ratio. If `unbiased=T` and `hetero=F` the bias corrected test statistic (3.24) is computed. If `unbiased=T` and `hetero=T` then the heteroskedasticity robust statistic (3.25) is computed. The function `varRatioTest` returns an object of class "`varRatioTest`" for which there are `print` and `plot` methods.

Consider testing the model RW3 for the daily log closing prices of the Dow Jones Industrial Average over the period 1/1/1960 through 1/1/1990. To compute the variance ratio (3.21) and the heteroskedasticity robust test (3.25) for $q = 1, \ldots, 60$ use

```
> VR.djia = varRatioTest(djia[timeEvent("1/1/1960","1/1/1990"),
+           "close"], n.periods=60, unbiased=T, hetero=T)
> class(VR.djia)
[1] "varRatioTest"
> names(VR.djia)
[1] "varRatio" "std.err"  "stat"     "hetero"
> VR.djia

Variance Ratio Test
```

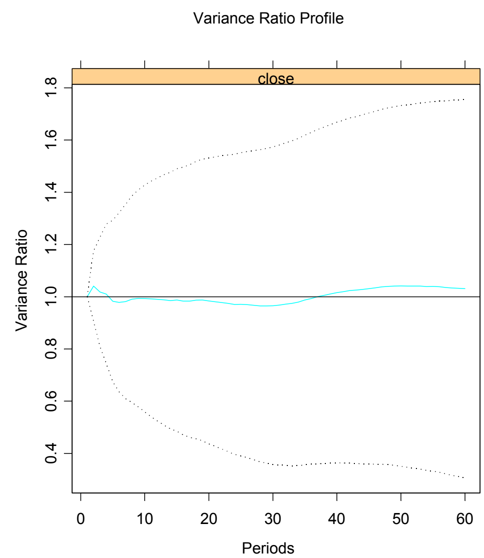

FIGURE 3.12. Variance ratios for the daily log prices of the Dow Jones Industrial Average.

```
Null Hypothesis: random walk with heteroskedastic errors

Variable: close
   var.ratio std.err       stat
 2    1.0403 0.06728  0.5994871
 3    1.0183 0.10527  0.1738146
...
60    1.0312 0.36227  0.0861747

 * : significant at 5% level
** : significant at 1% level
```

None of the variance ratios are statistically different from unity at the 5% level.

Figure 3.12 shows the results of the variance ratio tests based on `plot` method

```
> plot(VR.djia)
```

The variance ratios computed for different values of q hover around unity, and the $\pm$ 2 $\times$ standard error bands indicate that the model RW3 is not rejected at the 5% level.

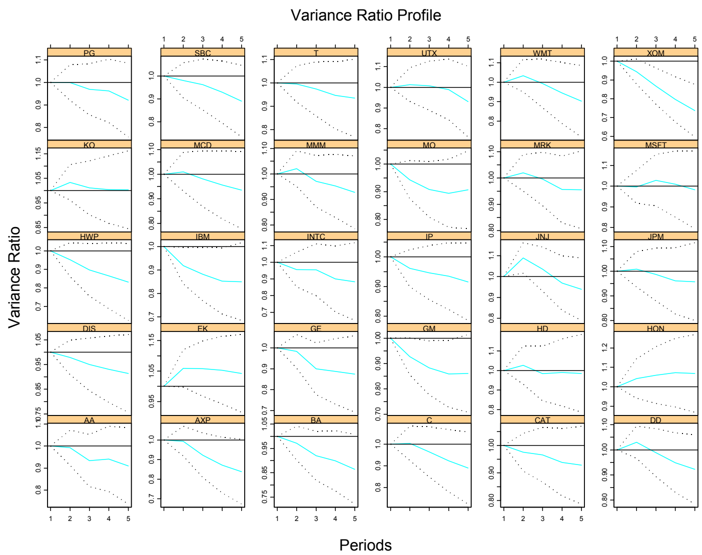

FIGURE 3.13. Variance ratio statistics for daily log prices on individual Dow Jones index stocks.

The RW3 model appears to hold for the Dow Jones index. To test the RW3 model for the top thirty stocks in the index individually, based on $q = 1, \ldots, 5$, use

```
> VR.DJ30 = varRatioTest(DowJones30, n.periods=5, unbiased=T,
+                         hetero=T)
> plot(VR.DJ30)
```

The results, illustrated in Figure 3.13, indicate that the RW3 model may not hold for some individual stocks.

3.3 Univariate Nonstationary Time Series

A univariate time series process $\{y_t\}$ is called *nonstationary* if it is not stationary. Since a stationary process has time invariant moments, a nonstationary process must have some time dependent moments. The most common forms of nonstationarity are caused by time dependence in the mean and variance.

Trend Stationary Process

$\{y_t\}$ is a *trend stationary* process if it has the form

$$y_t = TD_t + x_t$$

where TD_t are deterministic trend terms (constant, trend, seasonal dummies etc) that depend on t and $\{x_t\}$ is stationary. The series y_t is nonstationary because $E[TD_t] = TD_t$ which depends on t. Since x_t is stationary, y_t never deviates too far away from the deterministic trend TD_t. Hence, y_t exhibits *trend reversion*. If TD_t were known, y_t may be transformed to a stationary process by subtracting off the deterministic trend terms:

$$x_t = y_t - TD_t$$

Example 11 *Trend stationary AR(1)*

A trend stationary AR(1) process with $TD_t = \mu + \delta t$ may be expressed in three equivalent ways

$$\begin{aligned} y_t &= \mu + \delta t + u_t, u_t = \phi u_{t-1} + \varepsilon_t \\ y_t - \mu - \delta t &= \phi(y_{t-1} - \mu - \delta(t-1)) + \varepsilon_t \\ y_t &= c + \beta t + \phi y_{t-1} + \varepsilon_t \end{aligned}$$

where $|\phi| < 1$, $c = \mu(1-\phi) + \delta$, $\beta = \delta(1-\phi)t$ and $\varepsilon_t \sim WN(0, \sigma^2)$. Figure 3.14 shows $T = 100$ observations from a trend stationary AR(1) with $\mu = 1$, $\delta = 0.25$, $\phi = 0.75$ and $\sigma^2 = 1$ created with the S-PLUS commands

```
> set.seed(101)
> y.tsar1 = 1 + 0.25*seq(100) +
+ arima.sim(model=list(ar=0.75),n=100)
> tsplot(y.tsar1,ylab="y")
> abline(a=1,b=0.25)
```

The simulated data show clear trend reversion.

Integrated Processes

$\{y_t\}$ is an *integrated process* of order 1, denoted $y_t \sim I(1)$, if it has the form

$$y_t = y_{t-1} + u_t \tag{3.26}$$

where u_t is a stationary time series. Clearly, the first difference of y_t is stationary

$$\Delta y_t = u_t$$

Because of the above property, $I(1)$ processes are sometimes called *difference stationary* processes. Starting at y_0, by recursive substitution y_t has

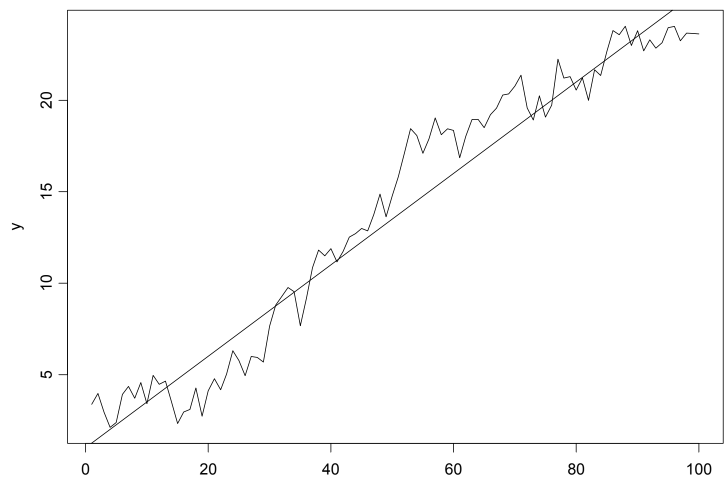

FIGURE 3.14. Simulated trend stationary process.

the representation of an *integrated sum* of stationary innovations

$$y_t = y_0 + \sum_{j=1}^{t} u_j. \tag{3.27}$$

The integrated sum $\sum_{j=1}^{t} u_j$ is called a *stochastic trend* and is denoted TS_t. Notice that

$$TS_t = TS_{t-1} + u_t$$

where $TS_0 = 0$. In contrast to a deterministic trend, changes in a stochastic trend are not perfectly predictable.

Since the stationary process u_t does not need to be differenced, it is called an integrated process of order zero and is denoted $u_t \sim I(0)$. Recall, from the Wold representation (3.6) a stationary process has an infinite order moving average representation where the moving average weights decline to zero at a geometric rate. From (3.27) it is seen that an $I(1)$ process has an infinite order moving average representation where all of the weights on the innovations are equal to 1.

If $u_t \sim IWN(0, \sigma^2)$ in (3.26) then y_t is called a *random walk*. In general, an $I(1)$ process can have serially correlated and heteroskedastic innovations u_t. If y_t is a random walk and assuming y_0 is fixed then it can be shown that

$$\begin{aligned}\gamma_0 &= \sigma^2 t \\ \gamma_j &= (t-j)\sigma^2 \\ \rho_j &= \sqrt{\frac{t-j}{t}}\end{aligned}$$

which clearly shows that y_t is nonstationary. Also, if t is large relative to j then $\rho_j \approx 1$. Hence, for an $I(1)$ process, the ACF does not decay at a geometric rate but at a linear rate as j increases.

An $I(1)$ process with drift has the form

$$y_t = \mu + y_{t-1} + u_t, \text{ where } u_t \sim I(0)$$

Starting at $t = 0$ an $I(1)$ process with drift μ may be expressed as

$$\begin{aligned}y_t &= y_0 + \mu t + \sum_{j=1}^{t} u_t \\ &= TD_t + TS_t\end{aligned}$$

so that it may be thought of as being composed of a deterministic linear trend $TD_t = y_0 + \mu t$ as well as a stochastic trend $TS_t = \sum_{j=1}^{t} u_j$.

An $I(d)$ process $\{y_t\}$ is one in which $\Delta^d y_t \sim I(0)$. In finance and economics data series are rarely modeled as $I(d)$ process with $d > 2$. Just as an $I(1)$ process with drift contains a linear deterministic trend, an $I(2)$ process with drift will contain a quadratic trend.

Example 12 *Simulated I(1) processes*

Consider the simulation of $T = 100$ observations from various $I(1)$ processes where the innovations u_t follow an AR(1) process $u_t = 0.75u_{t-1} + \varepsilon_t$ with $\varepsilon_t \sim GWN(0, 1)$.

```
> set.seed(101)
> u.ar1 = arima.sim(model=list(ar=0.75), n=100)
> y1 = cumsum(u.ar1)
> y1.d = 1 + 0.25*seq(100)+ y1
> y2 = rep(0,100)
> for (i in 3:100) {
+   y2[i] = 2*y2[i-1] - y2[i-2] + u.ar1[i]
+ }
```

The simulated data are illustrated in Figure 3.15 .

Example 13 *Financial time series*

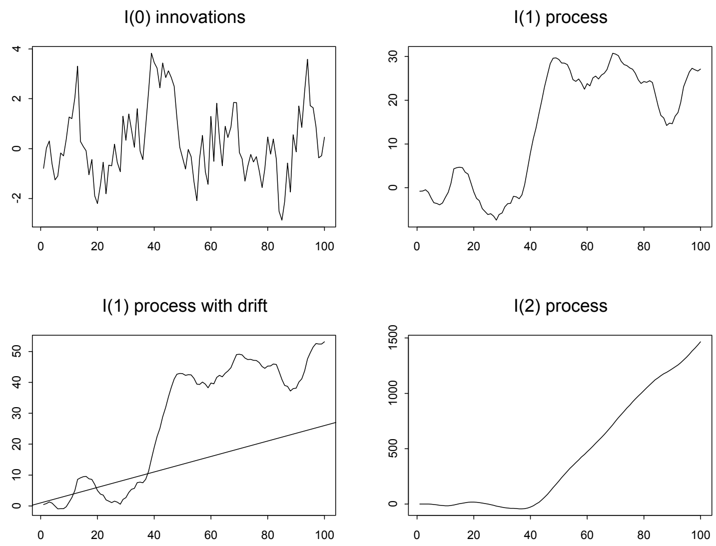

FIGURE 3.15. Simulated $I(d)$ processes for $d = 0$, 1 and 2.

Many financial time series are well characterized by $I(1)$ processes. The leading example of an $I(1)$ process with drift is the logarithm of an asset price. Common examples of $I(1)$ processes without drifts are the logarithms of exchange rates, nominal interest rates, and inflation rates. Notice that if inflation is constructed as the the difference in the logarithm of a price index and is an $I(1)$ process, then the logarithm of the price index is an $I(2)$ process. Examples of these data are illustrated in Figure 3.16. The exchange rate is the monthly log of the US/CA spot exchange rate taken from the `S+FinMetrics` "`timeSeries`" `lexrates.dat`, the asset price of the monthly S&P 500 index taken from the `S+FinMetrics` "`timeSeries`" object `singleIndex.dat`, the nominal interest rate is the 30 day T-bill rate taken from the `S+FinMetrics` "`timeSeries`" object `rf.30day`, and the monthly consumer price index is taken from the `S+FinMetrics` "`timeSeries`" object `CPI.dat`.

3.4 Long Memory Time Series

If a time series y_t is $I(0)$ then its ACF declines at a geometric rate. As a result, $I(0)$ process have *short memory* since observations far apart in time are essentially independent. Conversely, if y_t is $I(1)$ then its ACF declines at a linear rate and observations far apart in time are not independent. In

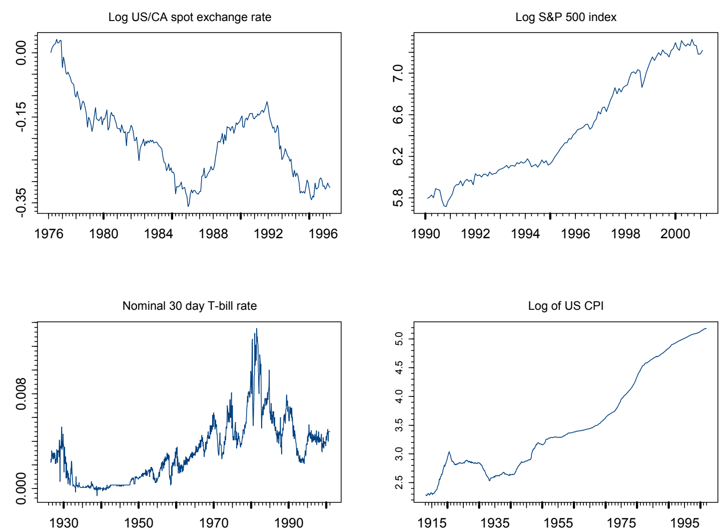

FIGURE 3.16. Monthly financial time series.

between $I(0)$ and $I(1)$ processes are so-called *fractionally integrated* $I(d)$ process where $0 < d < 1$. The ACF for a fractionally integrated processes declines at a polynomial (hyperbolic) rate, which implies that observations far apart in time may exhibit weak but non-zero correlation. This weak correlation between observations far apart is often referred to as *long memory*.

A fractionally integrated white noise process y_t has the form

$$(1-L)^d y_t = \varepsilon_t, \ \varepsilon_t \sim WN(0, \sigma^2) \tag{3.28}$$

where $(1-L)^d$ has the binomial series expansion representation (valid for any $d > -1$)

$$\begin{aligned}(1-L)^d &= \sum_{k=0}^{\infty} \begin{pmatrix} d \\ k \end{pmatrix} (-L)^k \\ &= 1 - dL + \frac{d(d-1)}{2!}L^2 - \frac{d(d-1)(d-2)}{3!}L^3 + \cdots\end{aligned}$$

If $d = 1$ then y_t is a random walk and if $d = 0$ then y_t is white noise. For $0 < d < 1$ it can be shown that

$$\rho_k \propto k^{2d-1}$$

as $k \to \infty$ so that the ACF for y_t declines hyperbolically to zero at a speed that depends on d. Further, it can be shown y_t is stationary and ergodic for $0 < d < 0.5$ and that the variance of y_t is infinite for $0.5 \leq d < 1$.

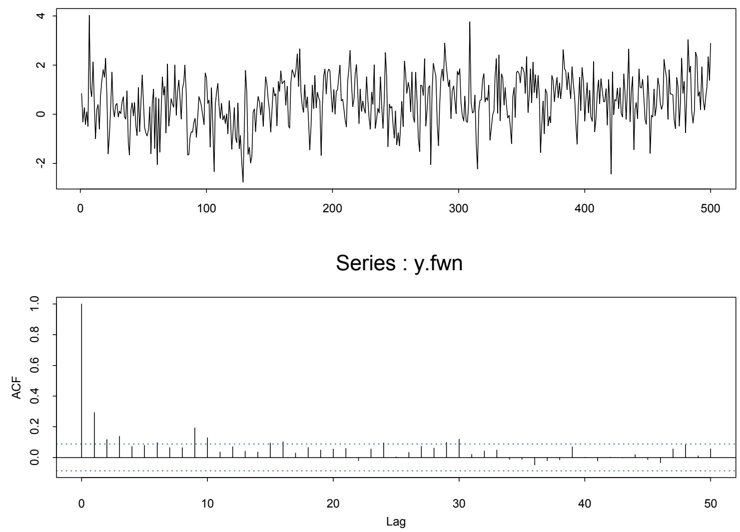

FIGURE 3.17. Simulated values from a fractional white noise process with $d = 0.3$ and $\sigma = 1$.

Example 14 *Simulated fractional white noise*

The S+FinMetrics function simulate.FARIMA may be used to generate simulated values from a fractional white noise process. To simulate 500 observations from (3.28) with $d = 0.3$ and $\sigma^2 = 1$ use

```
> set.seed(394)
> y.fwn = simulate.FARIMA(list(d=0.3), 500)
```

Figure 3.17 shows the simulated data along with the sample ACF created using

```
> par(mfrow=c(2,1))
> tsplot(y.fwn)
> tmp = acf(y.fwn,lag.max=50)
```

Notice how the sample ACF slowly decays to zero.

A fractionally integrated process with stationary and ergodic ARMA(p, q) errors

$$(1 - L)^d y_t = u_t, \ u_t \sim ARMA(p, q)$$

is called an *autoregressive fractionally integrated moving average* (ARFIMA) process. The modeling of long memory process is described in detail in Chapter 8.

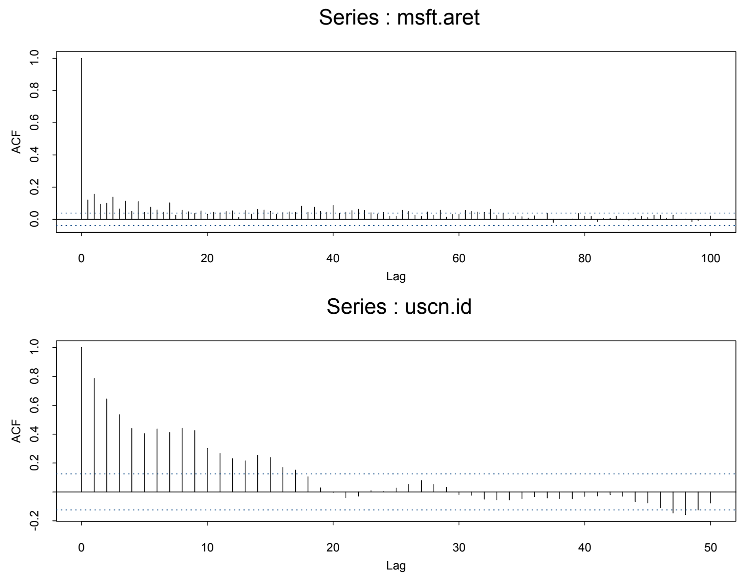

FIGURE 3.18. SACFs for the absolute value of daily returns on Microsoft and the monthly 30-day interest rate differential between U.S. bonds and Canadian bonds.

Example 15 *Long memory in financial time series*

Long memory behavior has been observed in certain types of financial time series. Ding, Granger and Engle (1993) find evidence of long memory in the absolute value of daily stock returns. Baillie and Bollerslev (1994) find evidence for long memory in the monthly interest rate differentials between short term U.S. government bonds and short term foreign government bonds. To illustrate, consider the absolute values of the daily returns on Microsoft over the 10 year period 1/2/1991 - 1/2/2001 taken from the `S+FinMetrics` "`timeSeries`" `DowJones30`

```
> msft.aret = abs(getReturns(DowJones30[,"MSFT"]))
```

Consider also the monthly US/CA 30-day interest rate differential over the period February 1976 through June 1996 in the "`timeSeries`" `uscn.id` constructed earlier and taken from the `S+FinMetrics` "`timeSeries`" object `lexrates.dat`. Figure 3.18 shows the SACFs these series create by

```
> par(mfrow=c(2,1))
> tmp = acf(msft.aret, lag.max=100)
> tmp = acf(uscn.id, lag.max=50)
```

For the absolute return series, notice the large number of small but apparently significant autocorrelations at very long lags. This is indicative of long memory. For the interest rate differential series, the ACF appears to decay fairly quickly, so the evidence for long memory is not as strong.

3.5 Multivariate Time Series

Consider n time series variables $\{y_{1t}\}, \ldots, \{y_{nt}\}$. A *multivariate time series* is the $(n \times 1)$ vector time series $\{\mathbf{Y}_t\}$ where the i^{th} row of $\{\mathbf{Y}_t\}$ is $\{y_{it}\}$. That is, for any time t, $\mathbf{Y}_t = (y_{1t}, \ldots, y_{nt})'$. Multivariate time series analysis is used when one wants to model and explain the interactions and co-movements among a group of time series variables. In finance, multivariate time series analysis is used to model systems of asset returns, asset prices and exchange rates, the term structure of interest rates, asset returns/prices, and economic variables etc. Many of the time series concepts described previously for univariate time series carry over to multivariate time series in a natural way. Additionally, there are some important time series concepts that are particular to multivariate time series. The following sections give the details of these extensions and provide examples using `S-PLUS` and `S+FinMetrics`.

3.5.1 *Stationary and Ergodic Multivariate Time Series*

A multivariate time series $\mathbf{Y}_t$ is covariance stationary and ergodic if all of its component time series are stationary and ergodic. The mean of $\mathbf{Y}_t$ is defined as the $(n \times 1)$ vector

$$E[\mathbf{Y}_t] = \boldsymbol{\mu} = (\mu_1, \ldots, \mu_n)'$$

where $\mu_i = E[y_{it}]$ for $i = 1, \ldots, n$. The variance/covariance matrix of $\mathbf{Y}_t$ is the $(n \times n)$ matrix

$$\begin{aligned} var(\mathbf{Y}_t) &= \Gamma_0 = E[(\mathbf{Y}_t - \boldsymbol{\mu})(\mathbf{Y}_t - \boldsymbol{\mu})'] \\ &= \begin{pmatrix} var(y_{1t}) & cov(y_{1t}, y_{2t}) & \cdots & cov(y_{1t}, y_{nt}) \\ cov(y_{2t}, y_{1t}) & var(y_{2t}) & \cdots & cov(y_{2t}, y_{nt}) \\ \vdots & \vdots & \ddots & \vdots \\ cov(y_{nt}, y_{1t}) & cov(y_{nt}, y_{2t}) & \cdots & var(y_{nt}) \end{pmatrix} \end{aligned}$$

The matrix $\mathbf{\Gamma}_0$ has elements $\gamma_{ij}^0 = cov(y_{it}, y_{jt})$. The correlation matrix of Y_t is the $(n \times n)$ matrix

$$corr(\mathbf{Y}_t) = \mathbf{R}_0 = \mathbf{D}^{-1}\mathbf{\Gamma}_0\mathbf{D}^{-1}$$

where $\mathbf{D}$ is an $(n \times n)$ diagonal matrix with j^{th} diagonal element $(\gamma_{jj}^0)^{1/2} = SD(y_{jt})$. The parameters μ, Γ_0 and $\mathbf{R}_0$ are estimated from data $(\mathbf{Y}_1, \ldots, \mathbf{Y}_T)$

using the sample moments

$$
\begin{aligned}
\bar{\mathbf{Y}} &= \frac{1}{T}\sum_{t=1}^{T}\mathbf{Y}_t \\
\hat{\mathbf{\Gamma}}_0 &= \frac{1}{T}\sum_{t=1}^{T}(\mathbf{Y}_t-\bar{\mathbf{Y}})(\mathbf{Y}_t-\bar{\mathbf{Y}})' \\
\hat{\mathbf{R}}_0 &= \hat{\mathbf{D}}^{-1}\hat{\mathbf{\Gamma}}_0\hat{\mathbf{D}}^{-1}
\end{aligned}
$$

where $\mathbf{D}$ is the $(n\times n)$ diagonal matrix with the sample standard deviations of y_{jt} along the diagonal. In order for the sample variance matrix $\hat{\mathbf{\Gamma}}_0$ and correlation matrix $\hat{\mathbf{R}}_0$ to be positive definite, the sample size T must be greater than the number of component time series n.

Example 16 *System of asset returns*

The `S+FinMetrics` "`timeSeries`" object `DowJones30` contains daily closing prices on the 30 assets in the Dow Jones index. An example of a stationary and ergodic multivariate time series is the continuously compounded returns on the first four assets in this index:

```
> Y = getReturns(DowJones30[,1:4],type="continuous")
> colIds(Y)
[1] "AA"  "AXP" "T"   "BA"
```

The `S-PLUS` function `colMeans` may be used to efficiently compute the mean vector of $\mathbf{Y}$

```
> colMeans(seriesData(Y))
       AA       AXP           T        BA
 0.0006661 0.0009478 -0.00002873 0.0004108
```

The function `colMeans` does not have a method for "`timeSeries`" objects so the extractor function `seriesData` is used to extract the data slot of the variable `Y`. The `S-PLUS` functions `var` and `cor`, which do have methods for "`timeSeries`" objects, may be used to compute $\hat{\mathbf{\Gamma}}_0$ and $\hat{\mathbf{R}}_0$

```
> var(Y)
            AA        AXP          T         BA
 AA 0.00041096 0.00009260 0.00005040 0.00007301
AXP 0.00009260 0.00044336 0.00008947 0.00009546
  T 0.00005040 0.00008947 0.00040441 0.00004548
 BA 0.00007301 0.00009546 0.00004548 0.00036829
> cor(Y)
        AA    AXP      T     BA
 AA 1.0000 0.2169 0.1236 0.1877
AXP 0.2169 1.0000 0.2113 0.2362
  T 0.1236 0.2113 1.0000 0.1179
```

```
 BA 0.1877 0.2362 0.1179 1.0000
```

If only the variances or standard deviations of $\mathbf{Y}_t$ are needed the S-PLUS functions colVars and colStdevs may be used

```
> colVars(seriesData(Y))
       AA       AXP         T        BA
 0.000411 0.0004434 0.0004044 0.0003683
> colStdevs(seriesData(Y))
       AA      AXP       T       BA
 0.020272 0.021056 0.02011 0.019191
```

Cross Covariance and Correlation Matrices

For a univariate time series y_t the autocovariances γ_k and autocorrelations ρ_k summarize the linear time dependence in the data. With a multivariate time series $\mathbf{Y}_t$ each component has autocovariances and autocorrelations but there are also cross lead-lag covariances and correlations between all possible pairs of components. The autocovariances and autocorrelations of y_{jt} for $j = 1, \ldots, n$ are defined as

$$
\begin{aligned}
\gamma_{jj}^k &= cov(y_{jt}, y_{jt-k}), \\
\rho_{jj}^k &= corr(y_{jt}, y_{jt-k}) = \frac{\gamma_{jj}^k}{\gamma_{jj}^0}
\end{aligned}
$$

and these are symmetric in k: $\gamma_{jj}^k = \gamma_{jj}^{-k}$, $\rho_{jj}^k = \rho_{jj}^{-k}$. The *cross lag covariances* and *cross lag correlations* between y_{it} and y_{jt} are defined as

$$
\begin{aligned}
\gamma_{ij}^k &= cov(y_{it}, y_{jt-k}), \\
\rho_{ij}^k &= corr(y_{jt}, y_{jt-k}) = \frac{\gamma_{ij}^k}{\sqrt{\gamma_{ii}^0 \gamma_{jj}^0}}
\end{aligned}
$$

and they are not necessarily symmetric in k. In general,

$$
\gamma_{ij}^k = cov(y_{it}, y_{jt-k}) \neq cov(y_{it}, y_{jt+k}) = cov(y_{jt}, y_{it-k}) = \gamma_{ij}^{-k}
$$

If $\gamma_{ij}^k \neq 0$ for some $k > 0$ then y_{jt} is said to *lead* y_{it}. Similarly, if $\gamma_{ij}^{-k} \neq 0$ for some $k > 0$ then y_{it} is said to *lead* y_{jt}. It is possible that y_{it} leads y_{jt} and vice-versa. In this case, there is said to be *feedback* between the two series.

All of the lag k cross covariances and correlations are summarized in the $(n \times n)$ lag k cross covariance and lag k cross correlation matrices

$$\begin{aligned}
\mathbf{\Gamma}_k &= E[(\mathbf{Y}_t - \boldsymbol{\mu})(\mathbf{Y}_{t-k} - \boldsymbol{\mu})'] \\
&= \begin{pmatrix} cov(y_{1t}, y_{1t-k}) & cov(y_{1t}, y_{2t-k}) & \cdots & cov(y_{1t}, y_{nt-k}) \\ cov(y_{2t}, y_{1t-k}) & cov(y_{2t}, y_{2t-k}) & \cdots & cov(y_{2t}, y_{nt-k}) \\ \vdots & \vdots & \ddots & \vdots \\ cov(y_{nt}, y_{1t-k}) & cov(y_{nt}, y_{2t-k}) & \cdots & cov(y_{nt}, y_{nt-k}) \end{pmatrix} \\
\mathbf{R}_k &= \mathbf{D}^{-1}\mathbf{\Gamma}_k\mathbf{D}^{-1}
\end{aligned}$$

The matrices $\mathbf{\Gamma}_k$ and $\mathbf{R}_k$ are not symmetric in k but it is easy to show that $\mathbf{\Gamma}_{-k} = \mathbf{\Gamma}_k'$ and $\mathbf{R}_{-k} = \mathbf{R}_k'$. The matrices $\mathbf{\Gamma}_k$ and $\mathbf{R}_k$ are estimated from data $(\mathbf{Y}_1, \ldots, \mathbf{Y}_T)$ using

$$\begin{aligned}
\hat{\mathbf{\Gamma}}_k &= \frac{1}{T}\sum_{t=k+1}^{T}(\mathbf{Y}_t - \bar{\mathbf{Y}})(\mathbf{Y}_{t-k} - \bar{\mathbf{Y}})' \\
\hat{\mathbf{R}}_k &= \hat{\mathbf{D}}^{-1}\hat{\mathbf{\Gamma}}_k\hat{\mathbf{D}}^{-1}
\end{aligned}$$

Example 17 *Lead-lag covariances and correlations among asset returns*

Consider computing the cross lag covariances and correlations for $k = 0, \ldots, 5$ between the first two Dow Jones 30 asset returns in the "`timeSeries`" `Y`. These covariances and correlations may be computed using the `S-PLUS` function `acf`

```
> Ghat = acf(Y[,1:2],lag.max=5,type="covariance",plot=F)
> Rhat = acf(Y[,1:2],lag.max=5,plot=F)
```

`Ghat` and `Rhat` are objects of class "`acf`" for which there is only a `print` method. For example, the estimated cross lag autocorrelations are

```
> Rhat
Call: acf(x = Y[, 1:2], lag.max = 5, plot = F)

Autocorrelation matrix:
  lag   AA.AA  AA.AXP AXP.AXP
1   0  1.0000  0.2169  1.0000
2   1  0.0182  0.0604 -0.0101
3   2 -0.0556 -0.0080 -0.0710
4   3  0.0145 -0.0203 -0.0152
5   4 -0.0639  0.0090 -0.0235
6   5  0.0142 -0.0056 -0.0169

  lag  AXP.AA
1   0  0.2169
```

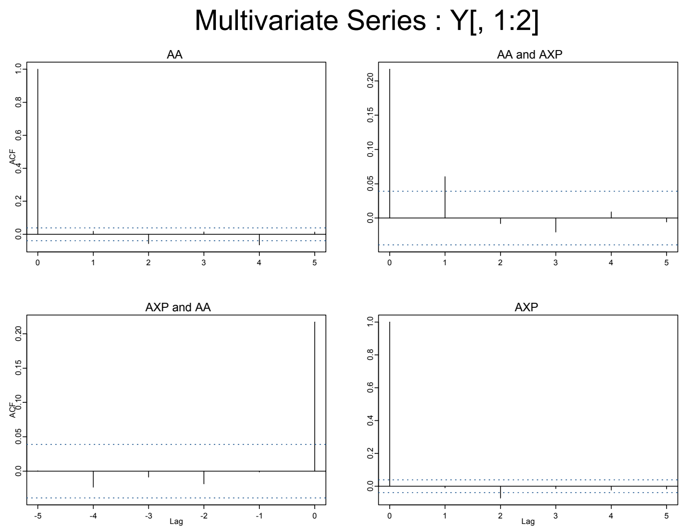

FIGURE 3.19. Cross lag correlations between the first two Dow Jones 30 asset returns.

```
2  -1 -0.0015
3  -2 -0.0187
4  -3 -0.0087
5  -4 -0.0233
6  -5  0.0003
```

The function `acf.plot` may be used to plot the cross lag covariances and correlations produced by `acf`.

```
> acf.plot(Rhat)
```

Figure 3.19 shows these cross lag correlations. The matrices $\hat{\mathbf{\Gamma}}_k$ and $\hat{\mathbf{R}}_k$ may be extracted from `acf` component of `Ghat` and `Rhat`, respectively. For example,

```
> Ghat$acf[1,,]
           [,1]       [,2]
[1,] 0.00041079 0.00009256
[2,] 0.00009256 0.00044318
> Rhat$acf[1,,]
       [,1]   [,2]
[1,] 1.0000 0.2169
[2,] 0.2169 1.0000
> Ghat$acf[2,,]
```

```
             [,1]       [,2]
[1,]  7.488e-006  2.578e-005
[2,] -6.537e-007 -4.486e-006
> Rhat$acf[2,,]
           [,1]     [,2]
[1,]  0.018229  0.06043
[2,] -0.001532 -0.01012
```

extracts $\hat{\mathbf{\Gamma}}_1$, $\hat{\mathbf{R}}_1$, $\hat{\mathbf{\Gamma}}_2$ and $\hat{\mathbf{R}}_2$.

3.5.2 Multivariate Wold Representation

Any $(n \times 1)$ covariance stationary multivariate time series $\mathbf{Y}_t$ has a Wold or linear process representation of the form

$$\begin{aligned}\mathbf{Y}_t &= \boldsymbol{\mu} + \boldsymbol{\varepsilon}_t + \mathbf{\Psi}_1 \boldsymbol{\varepsilon}_{t-1} + \mathbf{\Psi}_2 \boldsymbol{\varepsilon}_{t-2} + \cdots \\ &= \boldsymbol{\mu} + \sum_{k=0}^{\infty} \mathbf{\Psi}_k \boldsymbol{\varepsilon}_{t-k}\end{aligned} \tag{3.29}$$

where $\mathbf{\Psi}_0 = \mathbf{I}_n$ and $\boldsymbol{\varepsilon}_t$ is a multivariate white noise process with mean zero and variance matrix $E[\boldsymbol{\varepsilon}_t \boldsymbol{\varepsilon}_t'] = \mathbf{\Sigma}$. In (3.29), Ψ_k is an $(n \times n)$ matrix with (i, j)th element ψ_{ij}^k. In lag operator notation, the Wold form is

$$\begin{aligned}\mathbf{Y}_t &= \boldsymbol{\mu} + \mathbf{\Psi}(L) \boldsymbol{\varepsilon}_t \\ \mathbf{\Psi}(L) &= \sum_{k=0}^{\infty} \mathbf{\Psi}_k L^k\end{aligned}$$

The moments of $\mathbf{Y}_t$ are given by

$$\begin{aligned}E[\mathbf{Y}_t] &= \boldsymbol{\mu} \\ var(\mathbf{Y}_t) &= \sum_{k=0}^{\infty} \mathbf{\Psi}_k \mathbf{\Sigma} \mathbf{\Psi}_k'\end{aligned}$$

VAR Models

The most popular multivariate time series model is the *vector autoregressive* (VAR) model. The VAR model is a multivariate extension of the univariate autoregressive model. For example, a bivariate VAR(1) model has the form

$$\begin{pmatrix} y_{1t} \\ y_{2t} \end{pmatrix} = \begin{pmatrix} c_1 \\ c_2 \end{pmatrix} + \begin{pmatrix} \pi_{11}^1 & \pi_{12}^1 \\ \pi_{21}^1 & \pi_{22}^1 \end{pmatrix} \begin{pmatrix} y_{1t-1} \\ y_{2t-1} \end{pmatrix} + \begin{pmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{pmatrix}$$

or

$$\begin{aligned}y_{1t} &= c_1 + \pi_{11}^1 y_{1t-1} + \pi_{12}^1 y_{2t-1} + \varepsilon_{1t} \\ y_{2t} &= c_2 + \pi_{21}^1 y_{1t-1} + \pi_{22}^1 y_{2t-1} + \varepsilon_{2t}\end{aligned}$$

where

$$\begin{pmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{pmatrix} \sim iid \left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{12} & \sigma_{22} \end{pmatrix} \right)$$

In the equations for y_1 and y_2, the lagged values of both y_1 and y_2 are present.

The general VAR(p) model for $\mathbf{Y}_t = (y_{1t}, y_{2t}, \ldots, y_{nt})'$ has the form

$$\mathbf{Y}_t = \mathbf{c} + \mathbf{\Pi}_1 \mathbf{Y}_{t-1} + \mathbf{\Pi}_2 \mathbf{Y}_{t-2} + \cdots + \mathbf{\Pi}_p \mathbf{Y}_{t-p} + \boldsymbol{\varepsilon}_t, \; t = 1, \ldots, T \tag{3.30}$$

where $\mathbf{\Pi}_i$ are $(n \times n)$ coefficient matrices and $\boldsymbol{\varepsilon}_t$ is an $(n \times 1)$ unobservable zero mean white noise vector process with covariance matrix $\mathbf{\Sigma}$. VAR models are capable of capturing much of the complicated dynamics observed in stationary multivariate time series. Details about estimation, inference, and forecasting with VAR models are given in chapter eleven.

3.5.3 Long Run Variance

Let $\mathbf{Y}_t$ be an $(n \times 1)$ stationary and ergodic multivariate time series with $E[\mathbf{Y}_t] = \boldsymbol{\mu}$. Anderson's central limit theorem for stationary and ergodic process states

$$\sqrt{T}(\bar{\mathbf{Y}} - \boldsymbol{\mu}) \xrightarrow{d} N\left(\mathbf{0}, \sum_{j=-\infty}^{\infty} \mathbf{\Gamma}_j\right)$$

or

$$\bar{\mathbf{Y}} \overset{A}{\sim} N\left(\boldsymbol{\mu}, \frac{1}{T} \sum_{j=-\infty}^{\infty} \mathbf{\Gamma}_j\right)$$

Hence, the *long-run variance* of $\mathbf{Y}_t$ is T times the asymptotic variance of $\bar{\mathbf{Y}}$:

$$lrv(\mathbf{Y}_t) = T \cdot avar(\bar{\mathbf{Y}}) = \sum_{j=-\infty}^{\infty} \Gamma_j$$

Since $\mathbf{\Gamma}_{-j} = \mathbf{\Gamma}_j'$, $lrv(\mathbf{Y}_t)$ may be alternatively expressed as

$$lrv(\mathbf{Y}_t) = \mathbf{\Gamma}_0 + \sum_{j=1}^{\infty} (\mathbf{\Gamma}_j + \mathbf{\Gamma}_j')$$

Using the Wold representation of $\mathbf{Y}_t$ it can be shown that

$$lrv(\mathbf{Y}_t) = \mathbf{\Psi}(1)\mathbf{\Sigma}\mathbf{\Psi}(1)'$$

where $\mathbf{\Psi}(1) = \sum_{k=0}^{\infty} \mathbf{\Psi}_k$.

VAR Estimate of the Long-Run Variance

The Wold representation (3.29) may be approximated by high order VAR(p^*) model

$$\mathbf{Y}_t = \mathbf{c} + \mathbf{\Phi}_1 \mathbf{Y}_{t-1} + \cdots + \mathbf{\Phi}_{p^*} \mathbf{Y}_{t-p^*} + \varepsilon_t$$

where the lag length p^* is chosen such $p^* = O(T^{1/3})$. This gives rise to the *autoregressive long-run variance matrix* estimate

$$\widehat{lrv}_{AR}(\mathbf{Y}_t) = \hat{\mathbf{\Psi}}(1)\hat{\mathbf{\Sigma}}\hat{\mathbf{\Psi}}(1)' \tag{3.31}$$

$$\hat{\mathbf{\Psi}}(1) = (\mathbf{I}_n - \hat{\mathbf{\Phi}}_1 - \cdots - \hat{\mathbf{\Phi}}_p)^{-1} \tag{3.32}$$

$$\hat{\mathbf{\Sigma}} = \frac{1}{T}\sum_{t=1}^{T} \hat{\varepsilon}_t \hat{\varepsilon}_t' \tag{3.33}$$

where $\hat{\mathbf{\Phi}}_k$ $(k = 1, \ldots, p^*)$ are estimates of the VAR parameter matrices.

Non-parametric Estimate of the Long-Run Variance

A consistent estimate of $lrv(\mathbf{Y}_t)$ may be computed using non-parametric methods. A popular estimator is the Newey-West weighted autocovariance estimator

$$\widehat{lrv}_{NW}(\mathbf{Y}_t) = \hat{\mathbf{\Gamma}}_0 + \sum_{j=1}^{M_T} w_{j,T} \cdot \left(\hat{\mathbf{\Gamma}}_j + \hat{\mathbf{\Gamma}}_j'\right) \tag{3.34}$$

where $w_{j,T}$ are weights which sum to unity and M_T is a truncation lag parameter that satisfies $M_T = O(T^{1/3})$.

Example 18 *Newey-West estimate of long-run variance matrix for stock returns*

The `S+FinMetrics` function `asymp.var` may be used to compute the Newey-West long-run variance estimate (3.34) for a multivariate time series. The long-run variance matrix for the first four Dow Jones assets in the "`timeSeries`" `Y` is

```
> M.T = floor(4*(nrow(Y)/100)^(2/9))
> lrv.nw = asymp.var(Y,bandwidth=M.T)
> lrv.nw
            AA        AXP           T          BA
 AA 0.00037313 0.00008526 3.754e-005 6.685e-005
AXP 0.00008526 0.00034957 7.937e-005 1.051e-004
  T 0.00003754 0.00007937 3.707e-004 7.415e-006
 BA 0.00006685 0.00010506 7.415e-006 3.087e-004
```

3.6 References

[1] Alexander, C. (2001). *Market Models. A Guide to Financial Data Analysis.* John Wiley & Sons, Chichester, UK.

[2] Baille, R.T. and T. Bollerslev (1994). "The Long Memory of the Forward Premium," *Journal of International Money and Finance*, 13, 555-571.

[3] Box, G.E.P. and G.M. Jenkins (1976). *Time Series Analysis, Forecasting and Control. Revised Edition.* Holden Day, San Francisco.

[4] Campbell, J.Y., A.W. Lo, A.C. MacKinlay (1997). *The Econometrics of Financial Markets.* Princeton University Press, New Jersey.

[5] Box, G.E.P., and D.A. Pierce (1970). "Distribution of Residual Autocorrelations in Autoregressive-integrated Moving Average Time Series Models," *Journal of the American Statistical Association*, 65, 1509-1526.

[6] Chan, N.H. (2002). *Time Series: Applicatios to Finance.* John Wiley & Sons, New York.

[7] Ding, Z., C.W.J. Granger and R.F. Engle (1993). "A Long Memory Property of Stock Returns and a New Model," *Journal of Empirical Finance*, 1, 83-106.

[8] Engle, R.F. (1982). "Autoregressive Conditional Heteroskedasticity with Estimates of the Variance of United Kingdom Inflations," *Econometrica*, 50, 987-1097.

[9] Fuller, W.A. (1996). *Introduction to Statistical Time Series, Second Edition.* John Wiley & Sons, New York.

[10] Gourieroux, C and J. Jasiak (2001). *Financial Econometrics.* Princeton University Press, New Jersey.

[11] Granger, C.W.J. and M.J. Morris (1976). "Time Series Modeling and Interpretation," *Journal of the Royal Statistical Society*, Series A, 139, 246-257.

[12] Hamilton, J.D. (1994). *Time Series Analysis.* Princeton University Press, New Jersey.

[13] Harvey, A.C. (1993). *Time Series Models, Second Edition.* MIT Press, Massachusetts.

[14] Hayashi, F. (2000). *Econometrics.* Princeton University Press, New Jersey.

[15] Ljung, T. and G.E.P. Box (1979). "The Likelihood Function for a Stationary Autoregressive Moving Average Process," *Biometrika*, 66, 265-270.

[16] Mills, T.C. (1999). *The Econometric Modeling of Financial Time Series, Second Edition*. Cambridge University Press, Cambridge, UK.

[17] Newey, W.K. and K.D. West (1987). "A Simple Positive Semidefinite Heteroskedasticity and Autocorrelation Consistent Covariance Matrix," *Econometrica*, 55, 703-708.

[18] Tsay, R.S. (2001). *Analysis of Financial Time Series*, John Wiley & Sons, New York.

[19] Venables, W.N. and B.D. Ripley (1999). *Modern Applied Statistics with S-PLUS,* Third Edition. Springer-Verlag, New York.

[20] White, H. (1984). *Asymptotic Theory for Econometrians.* Academic Press, San Diego.