5

Additional Time Domain Topics

5.1 Introduction

In this chapter, we present material that may be considered special or advanced topics in the time domain. Chapter 6 is devoted to one of the most useful and interesting time domain topics, state-space models. Consequently, we do not cover state-space models or related topics—of which there are many—in this chapter. This chapter consists of sections of independent topics that may be read in any order. Most of the sections depend on a basic knowledge of ARMA models, forecasting and estimation, which is the material that is covered in Chapter 3, §3.1-§3.8. A few sections, for example the section on long memory models, require some knowledge of spectral analysis and related topics covered in Chapter 4. In addition to long memory, we discuss unit root testing, GARCH models, threshold models, regression with autocorrelated errors, lagged regression or transfer functions, and selected topics in multivariate ARMAX models.

5.2 Long Memory ARMA and Fractional Differencing

The conventional ARMA(p, q) process is often referred to as a short-memory process because the coefficients in the representation

$$x_t = \sum_{j=0}^{\infty} \psi_j w_{t-j},$$

obtained by solving

$$\phi(z)\psi(z) = \theta(z),$$

are dominated by exponential decay. As pointed out in §3.3, this result implies the ACF of the short memory process $\rho(h) \to 0$ exponentially fast as $h \to \infty$. When the sample ACF of a time series decays slowly, the advice given in

Chapter 3 has been to difference the series until it seems stationary. Following this advice with the glacial varve series first presented in Example 3.32 leads to the first difference of the logarithms of the data being represented as a first-order moving average. In Example 3.40, further analysis of the residuals leads to fitting an ARIMA$(1, 1, 1)$ model,

$$\nabla x_t = \phi \nabla x_{t-1} + w_t + \theta w_{t-1},$$

where we understand x_t is the log-transformed varve series. In particular, the estimates of the parameters (and the standard errors) were $\widehat{\phi} = .23(.05)$, $\widehat{\theta} = -.89(.03)$, and $\widehat{\sigma}_w^2 = .23$. The use of the first difference $\nabla x_t = (1 - B)x_t$ can be too severe a modification in the sense that the nonstationary model might represent an overdifferencing of the original process.

Long memory (or persistent) time series were considered in Hosking (1981) and Granger and Joyeux (1980) as intermediate compromises between the short memory ARMA type models and the fully integrated nonstationary processes in the Box–Jenkins class. The easiest way to generate a long memory series is to think of using the difference operator $(1 - B)^d$ for fractional values of d, say, $0 < d < .5$, so a basic long memory series gets generated as

$$(1 - B)^d x_t = w_t, \tag{5.1}$$

where w_t still denotes white noise with variance σ_w^2. The fractionally differenced series (5.1), for $|d| < .5$, is often called *fractional noise* (except when d is zero). Now, d becomes a parameter to be estimated along with σ_w^2. Differencing the original process, as in the Box–Jenkins approach, may be thought of as simply assigning a value of $d = 1$. This idea has been extended to the class of fractionally integrated ARMA, or ARFIMA models, where $-.5 < d < .5$; when d is negative, the term antipersistent is used. Long memory processes occur in hydrology (see Hurst, 1951, and McLeod and Hipel, 1978) and in environmental series, such as the varve data we have previously analyzed, to mention a few examples. Long memory time series data tend to exhibit sample autocorrelations that are not necessarily large (as in the case of $d = 1$), but persist for a long time. Figure 5.1 shows the sample ACF, to lag 100, of the log-transformed varve series, which exhibits classic long memory behavior:

```
u = acf(log(varve), 100, plot=FALSE)
plot(u[1:100], ylim=c(-.1,1), main="log(varve)") # get rid of lag 0
```

To investigate its properties, we can use the binomial expansion $(d > -1)$ to write

$$w_t = (1 - B)^d x_t = \sum_{j=0}^{\infty} \pi_j B^j x_t = \sum_{j=0}^{\infty} \pi_j x_{t-j} \tag{5.2}$$

where

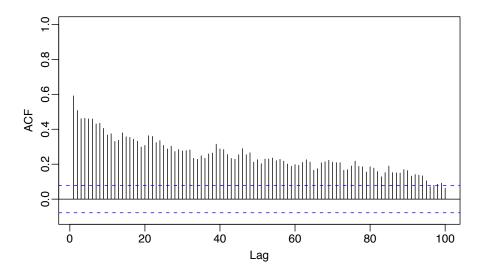$$\pi_j = \frac{\Gamma(j - d)}{\Gamma(j + 1)\Gamma(-d)} \tag{5.3}$$

log(varve)

Fig. 5.1. Sample ACF of the log transformed varve series.

with $\Gamma(x+1) = x\Gamma(x)$ being the gamma function. Similarly ($d < 1$), we can write

$$x_t = (1 - B)^{-d}w_t = \sum_{j=0}^{\infty} \psi_j B^j w_t = \sum_{j=0}^{\infty} \psi_j w_{t-j} \tag{5.4}$$

where

$$\psi_j = \frac{\Gamma(j + d)}{\Gamma(j + 1)\Gamma(d)}. \tag{5.5}$$

When $|d| < .5$, the processes (5.2) and (5.4) are well-defined stationary processes (see Brockwell and Davis, 1991, for details). In the case of fractional differencing, however, the coefficients satisfy $\sum \pi_j^2 < \infty$ and $\sum \psi_j^2 < \infty$ as opposed to the absolute summability of the coefficients in ARMA processes.

Using the representation (5.4)–(5.5), and after some nontrivial manipulations, it can be shown that the ACF of x_t is

$$\rho(h) = \frac{\Gamma(h + d)\Gamma(1 - d)}{\Gamma(h - d + 1)\Gamma(d)} \sim h^{2d-1} \tag{5.6}$$

for large h. From this we see that for $0 < d < .5$

$$\sum_{h=-\infty}^{\infty} |\rho(h)| = \infty$$

and hence the term *long memory*.

In order to examine a series such as the varve series for a possible long memory pattern, it is convenient to look at ways of estimating d. Using (5.3) it is easy to derive the recursions

$$\pi_{j+1}(d) = \frac{(j-d)\pi_j(d)}{(j+1)}, \tag{5.7}$$

for $j = 0, 1, \ldots$, with $\pi_0(d) = 1$. Maximizing the joint likelihood of the errors under normality, say, $w_t(d)$, will involve minimizing the sum of squared errors

$$Q(d) = \sum w_t^2(d).$$

The usual Gauss–Newton method, described in §3.6, leads to the expansion

$$w_t(d) = w_t(d_0) + w_t'(d_0)(d - d_0),$$

where

$$w_t'(d_0) = \left. \frac{\partial w_t}{\partial d} \right|_{d=d_0}$$

and d_0 is an initial estimate (guess) at to the value of d. Setting up the usual regression leads to

$$d = d_0 - \frac{\sum_t w_t'(d_0) w_t(d_0)}{\sum_t w_t'(d_0)^2}. \tag{5.8}$$

The derivatives are computed recursively by differentiating (5.7) successively with respect to d: $\pi_{j+1}'(d) = [(j-d)\pi_j'(d) - \pi_j(d)]/(j+1)$, where $\pi_0'(d) = 0$. The errors are computed from an approximation to (5.2), namely,

$$w_t(d) = \sum_{j=0}^{t} \pi_j(d) x_{t-j}. \tag{5.9}$$

It is advisable to omit a number of initial terms from the computation and start the sum, (5.8), at some fairly large value of t to have a reasonable approximation.

Example 5.1 Long Memory Fitting of the Glacial Varve Series

We consider analyzing the glacial varve series discussed in Example 2.6 and Example 3.32. Figure 2.6 shows the original and log-transformed series (which we denote by x_t). In Example 3.40, we noted that x_t could be modeled as an ARIMA$(1, 1, 1)$ process. We fit the fractionally differenced model, (5.1), to the mean-adjusted series, $x_t - \bar{x}$. Applying the Gauss–Newton iterative procedure previously described, starting with $d = .1$ and omitting the first 30 points from the computation, leads to a final value of $d = .384$, which implies the set of coefficients $\pi_j(.384)$, as given in Figure 5.2 with $\pi_0(.384) = 1$. We can compare roughly the performance of the fractional difference operator with the ARIMA model by examining the autocorrelation functions of the two residual series as shown in Figure 5.3. The ACFs of the two residual series are roughly comparable with the white noise model.

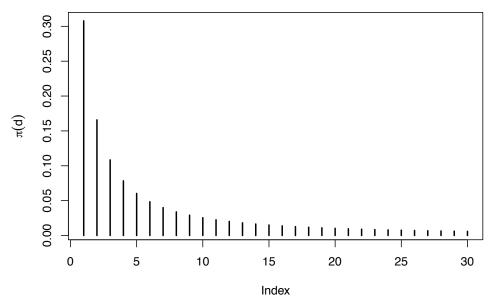To perform this analysis in R, first download and install the `fracdiff` package. Then use

Fig. 5.2. Coefficients $\pi_j(.384), j = 1, 2, \ldots, 30$ in the representation (5.7).

```
library(fracdiff)
lvarve = log(varve)-mean(log(varve))
varve.fd = fracdiff(lvarve, nar=0, nma=0, M=30)
varve.fd$d   # = 0.3841688
varve.fd$stderror.dpq   # = 4.589514e-06 (questionable result!!)
p = rep(1,31)
for (k in 1:30){ p[k+1] = (k-varve.fd$d)*p[k]/(k+1) }
plot(1:30, p[-1], ylab=expression(pi(d)), xlab="Index", type="h")
res.fd = diffseries(log(varve), varve.fd$d)   # frac diff resids
res.arima = resid(arima(log(varve), order=c(1,1,1))) # arima resids
par(mfrow=c(2,1))
acf(res.arima, 100, xlim=c(4,97), ylim=c(-.2,.2), main="")
acf(res.fd, 100, xlim=c(4,97), ylim=c(-.2,.2), main="")
```

The R package uses a truncated maximum likelihood procedure that was discussed in Haslett and Raftery (1989), which is a little more elaborate than simply zeroing out initial values. The default truncation value in R is $M = 100$. In the default case, the estimate is $\widehat{d} = .37$ with approximately the same (questionable) standard error. The standard error is (supposedly) obtained from the Hessian as described in Example 3.29.

Forecasting long memory processes is similar to forecasting ARMA models. That is, (5.2) and (5.7) can be used to obtain the truncated forecasts

$$\widetilde{x}_{n+m}^n = -\sum_{j=1}^n \pi_j(\widehat{d})\, \widetilde{x}_{n+m-j}^n, \tag{5.10}$$

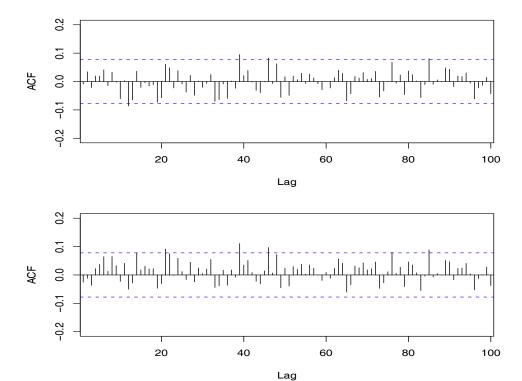for $m = 1, 2, \ldots$. Error bounds can be approximated by using

Fig. 5.3. ACF of residuals from the ARIMA$(1,1,1)$ fit to the logged varve series (top) and of the residuals from the long memory model fit, $(1 - B)^d x_t = w_t$, with $d = .384$ (bottom).

$$P_{n+m}^n = \widehat{\sigma}_w^2 \left(\sum_{j=0}^{m-1} \psi_j^2(\widehat{d}) \right) \tag{5.11}$$

where, as in (5.7),

$$\psi_j(\widehat{d}) = \frac{(j + \widehat{d})\psi_j(\widehat{d})}{(j + 1)}, \tag{5.12}$$

with $\psi_0(\widehat{d}) = 1$.

No obvious short memory ARMA-type component can be seen in the ACF of the residuals from the fractionally differenced varve series shown in Figure 5.3. It is natural, however, that cases will exist in which substantial short memory-type components will also be present in data that exhibits long memory. Hence, it is natural to define the general ARFIMA(p, d, q), $-.5 < d < .5$ process as

$$\phi(B)\nabla^d(x_t - \mu) = \theta(B)w_t, \tag{5.13}$$

where $\phi(B)$ and $\theta(B)$ are as given in Chapter 3. Writing the model in the form

$$\phi(B)\pi_d(B)(x_t - \mu) = \theta(B)w_t \tag{5.14}$$

makes it clear how we go about estimating the parameters for the more general model. Forecasting for the ARFIMA(p, d, q) series can be easily done, noting that we may equate coefficients in

$$\phi(z)\psi(z) = (1-z)^{-d}\theta(z) \tag{5.15}$$

and

$$\theta(z)\pi(z) = (1-z)^d\phi(z) \tag{5.16}$$

to obtain the representations

$$x_t = \mu + \sum_{j=0}^{\infty} \psi_j w_{t-j}$$

and

$$w_t = \sum_{j=0}^{\infty} \pi_j(x_{t-j} - \mu).$$

We then can proceed as discussed in (5.10) and (5.11).

Comprehensive treatments of long memory time series models are given in the texts by Beran (1994), Palma (2007), and Robinson (2003), and it should be noted that several other techniques for estimating the parameters, especially, the long memory parameter, can be developed in the frequency domain. In this case, we may think of the equations as generated by an infinite order autoregressive series with coefficients π_j given by (5.7) . Using the same approach as before, we obtain

$$
\begin{aligned}
f_x(\omega) &= \frac{\sigma_w^2}{|\sum_{k=0}^{\infty} \pi_k e^{-2\pi i k\omega}|^2} \\
&= \sigma_w^2|1 - e^{-2\pi i\omega}|^{-2d} = [4\sin^2(\pi\omega)]^{-d}\sigma_w^2
\end{aligned}
\tag{5.17}
$$

as equivalent representations of the spectrum of a long memory process. The long memory spectrum approaches infinity as the frequency $\omega \to 0$.

The main reason for defining the Whittle approximation to the log likelihood is to propose its use for estimating the parameter d in the long memory case as an alternative to the time domain method previously mentioned. The time domain approach is useful because of its simplicity and easily computed standard errors. One may also use an exact likelihood approach by developing an innovations form of the likelihood as in Brockwell and Davis (1991).

For the approximate approach using the Whittle likelihood (4.78), we consider using the approach of Fox and Taqqu (1986) who showed that maximizing the Whittle log likelihood leads to a consistent estimator with the usual asymptotic normal distribution that would be obtained by treating (4.78) as a conventional log likelihood (see also Dahlhaus, 1989; Robinson, 1995; Hurvich et al., 1998). Unfortunately, the periodogram ordinates are not asymptotically independent (Hurvich and Beltrao, 1993), although a quasi-likelihood in

the form of the Whittle approximation works well and has good asymptotic properties.

To see how this would work for the purely long memory case, write the long memory spectrum as

$$f_x(\omega_k; \ d, \sigma_w^2) = \sigma_w^2 g_k^{-d}, \tag{5.18}$$

where

$$g_k = 4\sin^2(\pi\omega_k). \tag{5.19}$$

Then, differentiating the log likelihood, say,

$$\ln L(\boldsymbol{x}; d, \sigma_w^2) \approx -m\ln\sigma_w^2 + d\sum_{k=1}^{m}\ln g_k - \frac{1}{\sigma_w^2}\sum_{k=1}^{m} g_k^d\, I(\omega_k) \tag{5.20}$$

at $m = n/2 - 1$ frequencies and solving for σ_w^2 yields

$$\sigma_w^2(d) = \frac{1}{m}\sum_{k=1}^{m} g_k^d\, I(\omega_k) \tag{5.21}$$

as the approximate maximum likelihood estimator for the variance parameter. To estimate d, we can use a grid search of the concentrated log likelihood

$$\ln L(\boldsymbol{x}; d) \approx -m\ln\sigma_w^2(d) + d\sum_{k=1}^{m}\ln g_k - m \tag{5.22}$$

over the interval $(-.5, .5)$, followed by a Newton–Raphson procedure to convergence.

Example 5.2 Long Memory Spectra for the Varve Series

In Example 5.1, we fit a long memory model to the glacial varve data via time domain methods. Fitting the same model using frequency domain methods and the Whittle approximation above gives $\widehat{d} = .380$, with an estimated standard error of .028. The earlier time domain method gave $\widehat{d} = .384$ with $M = 30$ and $\widehat{d} = .370$ with $M = 100$. Both estimates obtained via time domain methods had a standard error of about 4.6×10^{-6}, which seems implausible. The error variance estimate in this case is $\widehat{\sigma}_w^2 = .2293$; in Example 5.1, we could have used `var(res.fd)` as an estimate, in which case we obtain .2298. The R code to perform this analysis is

```
series = log(varve)  # specify series to be analyzed
d0 = .1  # initial value of d
n.per = nextn(length(series))
m = (n.per)/2  - 1
per = abs(fft(series-mean(series))[-1])^2  # remove 0 freq
per = per/n.per  # and scale the peridogram
g = 4*(sin(pi*((1:m)/n.per))^2)
```
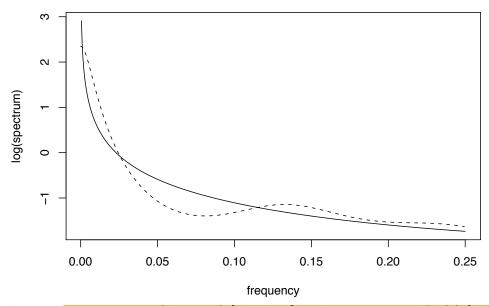
Fig. 5.4. Long Memory ($d = .380$) [solid line] and autoregressive AR(8) [dashed line] spectral estimators for the paleoclimatic glacial varve series.

```
# Function to calculate -log.likelihood
whit.like = function(d){
 g.d=g^d
 sig2 = (sum(g.d*per[1:m])/m)
 log.like = m*log(sig2) - d*sum(log(g)) + m
 return(log.like)    }
# Estimation (?optim for details - output not shown)
(est = optim(d0, whit.like, gr=NULL, method="L-BFGS-B",
    hessian=TRUE, lower=-.5, upper=.5,
    control=list(trace=1,REPORT=1)))
# Results:  d.hat = .380, se(dhat) = .028, and sig2hat = .229
cat("d.hat =", est$par, "se(dhat) = ",1/sqrt(est$hessian),"\n")
g.dhat = g^est$par;  sig2 = sum(g.dhat*per[1:m])/m
cat("sig2hat =",sig2,"\n")
```

One might also consider fitting an autoregressive model to these data using a procedure similar to that used in Example 4.15. Following this approach gave an autoregressive model with $p = 8$ and $\widehat{\phi} = (.34, .11, .04, .09, .08, .08, .02, .09)'$, with $\widehat{\sigma}_w^2 = .2267$ as the error variance. The two log spectra are plotted in Figure 5.4 for $\omega > 0$, and we note that long memory spectrum will eventually become infinite, whereas the AR(8) spectrum is finite at $\omega = 0$. The R code used for this part of the example (assuming the previous values have been retained) is

```
u = spec.ar(log(varve), plot=FALSE)   # produces AR(8)
g = 4*(sin(pi*((1:500)/2000))^2)
fhat = sig2*g^{-est$par}  # long memory spectral estimate
```

```
plot(1:500/2000, log(fhat), type="l", ylab="log(spectrum)",
    xlab="frequency")
lines(u$freq[1:250], log(u$spec[1:250]), lty="dashed")
ar.mle(log(varve))  # to get AR(8) estimates
```

Often, time series are not purely long memory. A common situation has the long memory component multiplied by a short memory component, leading to an alternate version of (5.18) of the form

$$f_x(\omega_k;\ d, \theta) = g_k^{-d} f_0(\omega_k; \boldsymbol{\theta}), \qquad (5.23)$$

where $f_0(\omega_k; \theta)$ might be the spectrum of an autoregressive moving average process with vector parameter $\boldsymbol{\theta}$, or it might be unspecified. If the spectrum has a parametric form, the Whittle likelihood can be used. However, there is a substantial amount of semiparametric literature that develops the estimators when the underlying spectrum $f_0(\omega; \boldsymbol{\theta})$ is unknown. A class of Gaussian semi-parametric estimators simply uses the same Whittle likelihood (5.22), evaluated over a sub-band of low frequencies, say $m' = \sqrt{n}$. There is some latitude in selecting a band that is relatively free from low frequency interference due to the short memory component in (5.23).

Geweke and Porter–Hudak (1983) developed an approximate method for estimating d based on a regression model, derived from (5.22). Note that we may write a simple equation for the logarithm of the spectrum as

$$\ln f_x(\omega_k; d) = \ln f_0(\omega_k; \boldsymbol{\theta}) - d \ln[4 \sin^2(\pi \omega_k)], \qquad (5.24)$$

with the frequencies $\omega_k = k/n$ restricted to a range $k = 1, 2, \ldots, m'$ near the zero frequency with $m' = \sqrt{n}$ as the recommended value. Relationship (5.24) suggests using a simple linear regression model of the form,

$$\ln I(\omega_k) = \beta_0 - d \ln[4 \sin^2(\pi \omega_k)] + e_k \qquad (5.25)$$

for the periodogram to estimate the parameters σ_w^2 and d. In this case, one performs least squares using $\ln I(\omega_k)$ as the dependent variable, and $\ln[4 \sin^2(\pi \omega_k)]$ as the independent variable for $k = 1, 2, \ldots, m$. The resulting slope estimate is then used as an estimate of $-d$. For a good discussion of various alternative methods for selecting m, see Hurvich and Deo (1999). The R package `fracdiff` also provides this method via the command `fdGPH()`; see the help file for further information.

One of the above two procedures works well for estimating the long memory component but there will be cases (such as ARFIMA) where there will be a parameterized short memory component $f_0(\omega_k; \boldsymbol{\theta})$ that needs to be estimated. If the spectrum is highly parameterized, one might estimate using the Whittle log likelihood (5.19) and

$$f_x(\omega_k; \boldsymbol{\theta}) = g_k^{-d} f_0(\omega_k; \boldsymbol{\theta})$$

and jointly estimating the parameters d and $\boldsymbol{\theta}$ using the Newton–Raphson method. If we are interested in a nonparametric estimator, using the conventional smoothed spectral estimator for the periodogram, adjusted for the long memory component, say $g_k^d\, I(\omega_k)$ might be a possible approach.

5.3 Unit Root Testing

As discussed in the previous section, the use of the first difference $\nabla x_t = (1-B)x_t$ can be too severe a modification in the sense that the nonstationary model might represent an overdifferencing of the original process. For example, consider a causal AR(1) process (we assume throughout this section that the noise is Gaussian),

$$x_t = \phi x_{t-1} + w_t. \tag{5.26}$$

Applying $(1-B)$ to both sides shows that differencing, $\nabla x_t = \phi \nabla x_{t-1} + \nabla w_t$, or

$$y_t = \phi y_{t-1} + w_t - w_{t-1},$$

where $y_t = \nabla x_t$, introduces extraneous correlation and invertibility problems. That is, while x_t is a causal AR(1) process, working with the differenced process y_t will be problematic because it is a non-invertible ARMA(1,1).

A unit root test provides a way to test whether (5.26) is a random walk (the null case) as opposed to a causal process (the alternative). That is, it provides a procedure for testing

$$H_0: \phi = 1 \quad \text{versus} \quad H_1: |\phi| < 1.$$

An obvious test statistic would be to consider $(\widehat{\phi} - 1)$, appropriately normalized, in the hope to develop an asymptotically normal test statistic, where $\widehat{\phi}$ is one of the optimal estimators discussed in Chapter 3, §3.6. Unfortunately, the theory of §3.6 will not work in the null case because the process is nonstationary. Moreover, as seen in Example 3.35, estimation near the boundary of stationarity produces highly skewed sample distributions (see Figure 3.11) and this is a good indication that the problem will be atypical.

To examine the behavior of $(\widehat{\phi} - 1)$ under the null hypothesis that $\phi = 1$, or more precisely that the model is a random walk, $x_t = \sum_{j=1}^{t} w_j$, or $x_t = x_{t-1} + w_t$ with $x_0 = 0$, consider the least squares estimator of ϕ. Noting that $\mu_x = 0$, the least squares estimator can be written as

$$\widehat{\phi} = \frac{\frac{1}{n}\sum_{t=1}^{n} x_t x_{t-1}}{\frac{1}{n}\sum_{t=1}^{n} x_{t-1}^2} = 1 + \frac{\frac{1}{n}\sum_{t=1}^{n} w_t x_{t-1}}{\frac{1}{n}\sum_{t=1}^{n} x_{t-1}^2}, \tag{5.27}$$

where we have written $x_t = x_{t-1} + w_t$ in the numerator; recall that $x_0 = 0$ and in the least squares setting, we are regressing x_t on x_{t-1} for $t = 1, \ldots, n$. Hence, under H_0, we have that

$$\widehat{\phi} - 1 = \frac{\frac{1}{n\sigma_w^2} \sum_{t=1}^n w_t x_{t-1}}{\frac{1}{n\sigma_w^2} \sum_{t=1}^n x_{t-1}^2}. \tag{5.28}$$

Consider the numerator of (5.28). Note first that by squaring both sides of $x_t = x_{t-1} + w_t$, we obtain $x_t^2 = x_{t-1}^2 + 2x_{t-1}w_t + w_t^2$ so that

$$x_{t-1}w_t = \frac{1}{2}(x_t^2 - x_{t-1}^2 - w_t^2),$$

and summing,

$$\frac{1}{n\sigma_w^2} \sum_{t=1}^n x_{t-1}w_t = \frac{1}{2}\left(\frac{x_n^2}{n\sigma_w^2} - \frac{\sum_{t=1}^n w_t^2}{n\sigma_w^2}\right).$$

Because $x_n = \sum_1^n w_t$, we have that $x_n \sim N(0, n\sigma_w^2)$, so that $\frac{1}{n\sigma_w^2} x_n^2 \sim \chi_1^2$, the chi-squared distribution with one degree of freedom. Moreover, because w_t is white Gaussian noise, $\frac{1}{n}\sum_1^n w_t^2 \to_p \sigma_w^2$, or $\frac{1}{n\sigma_w^2}\sum_1^n w_t^2 \to_p 1$. Consequently, $(n \to \infty)$

$$\frac{1}{n\sigma_w^2} \sum_{t=1}^n x_{t-1}w_t \xrightarrow{d} \tfrac{1}{2}(\chi_1^2 - 1). \tag{5.29}$$

Next we focus on the denominator of (5.28). First, we introduce standard Brownian motion.

Definition 5.1 *A continuous time process $\{W(t); t \geq 0\}$ is called **standard Brownian motion** if it satisfies the following conditions:*

(i) $W(0) = 0$;
(ii) $\{W(t_2) - W(t_1), W(t_3) - W(t_2), \ldots, W(t_n) - W(t_{n-1})\}$ are independent for any collection of points, $0 \leq t_1 < t_2 \cdots < t_n$, and integer $n > 2$;
(iii) $W(t + \Delta t) - W(t) \sim N(0, \Delta t)$ for $\Delta t > 0$.

The result for the denominator uses the functional central limit theorem, which can be found in Billlingsley (1999, §2.8). In particular, if $\xi_1, \ldots, \xi_n$ is a sequence of iid random variables with mean 0 and variance 1, then, for $0 \leq t \leq 1$, the continuous time process

$$S_n(t) = \frac{1}{\sqrt{n}} \sum_{j=1}^{[\![nt]\!]} \xi_j \xrightarrow{d} W(t), \tag{5.30}$$

as $n \to \infty$, where $[\![\ ]\!]$ is the greatest integer function and $W(t)$ is standard Brownian motion on $[0, 1]$. Note the under the null hypothesis, $x_s = w_1 + \cdots + w_s \sim N(0, s\sigma_w^2)$, and based on (5.30), we have $\frac{x_s}{\sigma_w\sqrt{n}} \to_d W(s)$. From this fact, we can show that $(n \to \infty)$

$$\sum_{t=1}^n \left(\frac{x_{t-1}}{\sigma_w\sqrt{n}}\right)^2 \frac{1}{n} \xrightarrow{d} \int_0^1 W^2(t)\, dt. \tag{5.31}$$

The denominator in (5.28) is off from the left side of (5.31) by a factor of n^{-1}, and we adjust accordingly to finally obtain $(n \to \infty)$,

$$n(\widehat{\phi} - 1) = \frac{\frac{1}{n\sigma_w^2} \sum_{t=1}^{n} w_t x_{t-1}}{\frac{1}{n^2\sigma_w^2} \sum_{t=1}^{n} x_{t-1}^2} \xrightarrow{d} \frac{\frac{1}{2}\left(\chi_1^2 - 1\right)}{\int_0^1 W^2(t)\,dt}. \tag{5.32}$$

The test statistic $n(\widehat{\phi} - 1)$ is known as the unit root or Dickey-Fuller (DF) statistic (see Fuller, 1996), although the actual DF test statistic is normalized a little differently. Because the distribution of the test statistic does not have a closed form, quantiles of the distribution must be computed by numerical approximation or by simulation. The R package `tseries` provides this test along with more general tests that we mention briefly.

Toward a more general model, we note that the DF test was established by noting that if $x_t = \phi x_{t-1} + w_t$, then $\nabla x_t = (\phi - 1)x_{t-1} + w_t = \gamma x_{t-1} + w_t$, and one could test $H_0 : \gamma = 0$ by regressing ∇x_t on x_{t-1}. They formed a Wald statistic and derived its limiting distribution [the previous derivation based on Brownian motion is due to Phillips (1987)]. The test was extended to accommodate AR(p) models, $x_t = \sum_{j=1}^{p} \phi_j x_{t-j} + w_t$, as follows. Subtract x_{t-1} from the model to obtain

$$\nabla x_t = \gamma x_{t-1} + \sum_{j=1}^{p-1} \psi_j \nabla x_{t-j} + w_t, \tag{5.33}$$

where $\gamma = \sum_{j=1}^{p} \phi_j - 1$ and $\psi_j = -\sum_{j=i}^{p} \phi_i$ for $j = 2, \ldots, p$. For a quick check of (5.33) when $p = 2$, note that $x_t = (\phi_1 + \phi_2)x_{t-1} - \phi_2(x_{t-1} - x_{t-2}) + w_t$; now subtract x_{t-1} from both sides. To test the hypothesis that the process has a unit root at 1 (i.e., the AR polynoimial $\phi(z) = 0$ when $z = 1$), we can test $H_0 : \gamma = 0$ by estimating γ in the regression of ∇x_t on $x_{t-1}, \nabla x_{t-1}, \ldots, \nabla x_{t-p+1}$, and forming a Wald test based on $t_\gamma = \widehat{\gamma}/\text{se}(\widehat{\gamma})$. This test leads to the so-called augmented Dickey-Fuller test (ADF). While the calculations for obtaining the asymptotic null distribution change, the basic ideas and machinery remain the same as in the simple case. The choice of p is crucial, and we will discuss some suggestions in the example. For ARMA(p, q) models, the ADF test can be used by assuming p is large enough to capture the essential correlation structure; another alternative is the Phillips-Perron (PP) test, which differs from the ADF tests mainly in how they deal with serial correlation and heteroskedasticity in the errors.

One can extend the model to include a constant, or even non-stochastic trend. For example, consider the model

$$x_t = \beta_0 + \beta_1 t + \phi x_{t-1} + w_t.$$

If we assume $\beta_1 = 0$, then under the null hypothesis, $\phi = 1$, the process is a random walk with drift β_0. Under the alternate hypothesis, the process is a causal AR(1) with mean $\mu_x = \beta_0(1-\phi)$. If we cannot assume $\beta_1 = 0$, then the

interest here is testing the null that $(\beta_1, \phi) = (0, 1)$, simultaneously, versus the alternative that $\beta_1 \neq 0$ and $|\phi| < 1$. In this case, the null hypothesis is that the process is a random walk with drift, versus the alternative hypothesis that the process is stationary around a global trend (consider the global temperature series examined in Example 2.1).

Example 5.3 Testing Unit Roots in the Glacial Varve Series

In this example we use the R package `tseries` to test the null hypothesis that the log of the glacial varve series has a unit root, versus the alternate hypothesis that the process is stationary. We test the null hypothesis using the available DF, ADF and PP tests; note that in each case, the general regression equation incorporates a constant and a linear trend. In the ADF test, the default number of AR components included in the model, say k, is $[\![(n-1)^{\frac{1}{3}}]\!]$, which corresponds to the suggested upper bound on the rate at which the number of lags, k, should be made to grow with the sample size for the general ARMA(p, q) setup. For the PP test, the default value of k is $[\![.04n^{\frac{1}{4}}]\!]$.

```
library(tseries)
adf.test(log(varve), k=0)              # DF test
   Dickey-Fuller = -12.8572, Lag order = 0, p-value < 0.01
   alternative hypothesis: stationary

adf.test(log(varve))                   # ADF test
   Dickey-Fuller = -3.5166, Lag order = 8, p-value = 0.04071
   alternative hypothesis: stationary

pp.test(log(varve))                    # PP test
   Dickey-Fuller Z(alpha) = -304.5376,
   Truncation lag parameter = 6, p-value < 0.01
   alternative hypothesis: stationary
```

In each test, we reject the null hypothesis that the logged varve series has a unit root. The conclusion of these tests supports the conclusion of the previous section that the logged varve series is long memory rather than integrated.

5.4 GARCH Models

Recent problems in finance have motivated the study of the volatility, or variability, of a time series. Although ARMA models assume a constant variance, models such as the autoregressive conditionally heteroscedastic or ARCH model, first introduced by Engle (1982), were developed to model changes in volatility. These models were later extended to generalized ARCH, or GARCH models by Bollerslev (1986).

In §3.8, we discussed the return or growth rate of a series. For example, if x_t is the value of a stock at time t, then the return or relative gain, y_t, of the stock at time t is

$$y_t = \frac{x_t - x_{t-1}}{x_{t-1}}. \tag{5.34}$$

Definition (5.34) implies that $x_t = (1+y_t)x_{t-1}$. Thus, based on the discussion in §3.8, if the return represents a small (in magnitude) percentage change then

$$\nabla[\log(x_t)] \approx y_t. \tag{5.35}$$

Either value, $\nabla[\log(x_t)]$ or $(x_t - x_{t-1})/x_{t-1}$, will be called the return, and will be denoted by y_t. It is the study of y_t that is the focus of ARCH, GARCH, and other volatility models. Recently there has been interest in stochastic volatility models and we will discuss these models in Chapter 6 because they are state-space models.

Typically, for financial series, the return y_t, does not have a constant conditional variance, and highly volatile periods tend to be clustered together. In other words, there is a strong dependence of sudden bursts of variability in a return on the series own past. For example, Figure 1.4 shows the daily returns of the New York Stock Exchange (NYSE) from February 2, 1984 to December 31, 1991. In this case, as is typical, the return y_t is fairly stable, except for short-term bursts of high volatility.

The simplest ARCH model, the ARCH(1), models the return as

$$y_t = \sigma_t \epsilon_t \tag{5.36}$$
$$\sigma_t^2 = \alpha_0 + \alpha_1 y_{t-1}^2, \tag{5.37}$$

where ϵ_t is standard Gaussian white noise; that is, $\epsilon_t \sim$ iid $N(0,1)$. As with ARMA models, we must impose some constraints on the model parameters to obtain desirable properties. One obvious constraint is that α_1 must not be negative, or else σ_t^2 may be negative.

As we shall see, the ARCH(1) models return as a white noise process with nonconstant conditional variance, and that conditional variance depends on the previous return. First, notice that the conditional distribution of y_t given y_{t-1} is Gaussian:

$$y_t \mid y_{t-1} \sim N(0, \alpha_0 + \alpha_1 y_{t-1}^2). \tag{5.38}$$

In addition, it is possible to write the ARCH(1) model as a non-Gaussian AR(1) model in the square of the returns y_t^2. First, rewrite (5.36)-(5.37) as

$$y_t^2 = \sigma_t^2 \epsilon_t^2$$
$$\alpha_0 + \alpha_1 y_{t-1}^2 = \sigma_t^2,$$

and subtract the two equations to obtain

$$y_t^2 - (\alpha_0 + \alpha_1 y_{t-1}^2) = \sigma_t^2 \epsilon_t^2 - \sigma_t^2.$$

Now, write this equation as

$$y_t^2 = \alpha_0 + \alpha_1 y_{t-1}^2 + v_t, \tag{5.39}$$

where $v_t = \sigma_t^2(\epsilon_t^2 - 1)$. Because ϵ_t^2 is the square of a N(0, 1) random variable, $\epsilon_t^2 - 1$ is a shifted (to have mean-zero), χ_1^2 random variable.

To explore the properties of ARCH, we define $Y_s = \{y_s, y_{s-1}, ...\}$. Then, using (5.38), we immediately see that y_t has a zero mean:

$$E(y_t) = EE(y_t \mid Y_{t-1}) = EE(y_t \mid y_{t-1}) = 0. \tag{5.40}$$

Because $E(y_t \mid Y_{t-1}) = 0$, the process y_t is said to be a *martingale difference*. Because y_t is a martingale difference, it is also an uncorrelated sequence. For example, with $h > 0$,

$$\begin{aligned} \text{cov}(y_{t+h}, y_t) &= E(y_t y_{t+h}) = EE(y_t y_{t+h} \mid Y_{t+h-1}) \\ &= E\{y_t E(y_{t+h} \mid Y_{t+h-1})\} = 0. \end{aligned} \tag{5.41}$$

The last line of (5.41) follows because y_t belongs to the information set Y_{t+h-1} for $h > 0$, and, $E(y_{t+h} \mid Y_{t+h-1}) = 0$, as determined in (5.40).

An argument similar to (5.40) and (5.41) will establish the fact that the error process v_t in (5.39) is also a martingale difference and, consequently, an uncorrelated sequence. If the variance of v_t is finite and constant with respect to time, and $0 \le \alpha_1 < 1$, then based on Property 3.1, (5.39) specifies a causal AR(1) process for y_t^2. Therefore, $E(y_t^2)$ and $\text{var}(y_t^2)$ must be constant with respect to time t. This, implies that

$$E(y_t^2) = \text{var}(y_t) = \frac{\alpha_0}{1 - \alpha_1} \tag{5.42}$$

and, after some manipulations,

$$E(y_t^4) = \frac{3\alpha_0^2}{(1 - \alpha_1)^2} \frac{1 - \alpha_1^2}{1 - 3\alpha_1^2}, \tag{5.43}$$

provided $3\alpha_1^2 < 1$. These results imply that the kurtosis, κ, of y_t is

$$\kappa = \frac{E(y_t^4)}{[E(y_t^2)]^2} = 3\frac{1 - \alpha_1^2}{1 - 3\alpha_1^2}, \tag{5.44}$$

which is always larger than 3 (unless $\alpha_1 = 0$), the kurtosis of the normal distribution. Thus, the marginal distribution of the returns, y_t, is leptokurtic, or has "fat tails."

In summary, an ARCH(1) process, y_t, as given by (5.36)-(5.37), or equivalently (5.38), is characterized by the following properties.

- If $0 \le \alpha_1 < 1$, the process y_t itself is white noise and its unconditional distribution is symmetrically distributed around zero; this distribution is leptokurtic.
- If, in addition, $3\alpha_1^2 < 1$, the square of the process, y_t^2, follows a causal AR(1) model with ACF given by $\rho_{y^2}(h) = \alpha_1^h \ge 0$, for all $h > 0$. If $3\alpha_1 \ge 1$, but $\alpha_1 < 1$, then y_t^2 is strictly stationary with infinite variance.

Estimation of the parameters α_0 and α_1 of the ARCH(1) model is typically accomplished by conditional MLE. The conditional likelihood of the data $y_2, ..., y_n$ given y_1, is given by

$$L(\alpha_0, \alpha_1 \mid y_1) = \prod_{t=2}^{n} f_{\alpha_0,\alpha_1}(y_t \mid y_{t-1}), \qquad (5.45)$$

where the density $f_{\alpha_0,\alpha_1}(y_t \mid y_{t-1})$ is the normal density specified in (5.38). Hence, the criterion function to be minimized, $l(\alpha_0, \alpha_1) \propto -\ln L(\alpha_0, \alpha_1 \mid y_1)$ is given by

$$l(\alpha_0, \alpha_1) = \frac{1}{2} \sum_{t=2}^{n} \ln(\alpha_0 + \alpha_1 y_{t-1}^2) + \frac{1}{2} \sum_{t=2}^{n} \left(\frac{y_t^2}{\alpha_0 + \alpha_1 y_{t-1}^2} \right). \qquad (5.46)$$

Estimation is accomplished by numerical methods, as described in §3.6. In this case, analytic expressions for the gradient vector, $l^{(1)}(\alpha_0, \alpha_1)$, and Hessian matrix, $l^{(2)}(\alpha_0, \alpha_1)$, as described in Example 3.29, can be obtained by straightforward calculations. For example, the 2×1 gradient vector, $l^{(1)}(\alpha_0, \alpha_1)$, is given by

$$\begin{pmatrix} \partial l/\partial\alpha_0 \\ \partial l/\partial\alpha_1 \end{pmatrix} = \sum_{t=2}^{n} \begin{pmatrix} 1 \\ y_{t-1}^2 \end{pmatrix} \times \frac{\alpha_0 + \alpha_1 y_{t-1}^2 - y_t^2}{2\left(\alpha_0 + \alpha_1 y_{t-1}^2\right)^2}.$$

The calculation of the Hessian matrix is left as an exercise (Problem Problem 5.9). The likelihood of the ARCH model tends to be flat unless n is very large. A discussion of this problem can be found in Shephard (1996).

It is also possible to combine a regression or an ARMA model for the mean with an ARCH model for the errors. For example, a regression with ARCH(1) errors model would have the observations x_t as linear function of p regressors, $\mathbf{z}_t = (z_{t1}, ..., z_{tp})'$, and ARCH(1) noise y_t, say,

$$x_t = \boldsymbol{\beta}' \mathbf{z}_t + y_t,$$

where y_t satisfies (5.36)-(5.37), but, in this case, is unobserved. Similarly, for example, an AR(1) model for data x_t exhibiting ARCH(1) errors would be

$$x_t = \phi_0 + \phi_1 x_{t-1} + y_t.$$

These types of models were explored by Weiss (1984).

Example 5.4 Analysis of U.S. GNP

In Example 3.38, we fit an MA(2) model and an AR(1) model to the U.S. GNP series and we concluded that the residuals from both fits appeared to behave like a white noise process. In Example 3.42 we concluded that the AR(1) is probably the better model in this case. It has been suggested that the U.S. GNP series has ARCH errors, and in this example, we will investigate this claim. If the GNP noise term is ARCH, the squares of
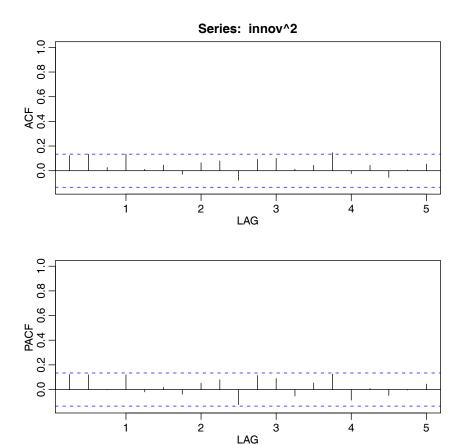
Fig. 5.5. ACF and PACF of the squares of the residuals from the AR(1) fit on U.S. GNP.

the residuals from the fit should behave like a non-Gaussian AR(1) process, as pointed out in (5.39). Figure 5.5 shows the ACF and PACF of the squared residuals it appears that there may be some dependence, albeit small, left in the residuals. The figure was generated in R as follows.

```
gnpgr = diff(log(gnp))   # get the growth rate
sarima(gnpgr, 1, 0, 0)   # fit an AR(1)
acf2(innov^2, 24)   # get (p)acf of the squared residuals
```

We used the R package **fGarch** to fit an AR(1)-ARCH(1) model to the U.S. GNP returns with the following results. A partial output is shown; we note that `garch(1,0)` specifies an `arch(1)` in the code below (details later).

```
library(fGarch)
summary(garchFit(~arma(1,0)+garch(1,0), gnpgr))
```

```
          Estimate  Std. Error  t value Pr(>|t|)
   mu      5.278e-03  8.996e-04    5.867 4.44e-09
   ar1     3.666e-01  7.514e-02    4.878 1.07e-06
   omega   7.331e-05  9.011e-06    8.135 4.44e-16
   alpha1  1.945e-01  9.554e-02    2.035   0.0418
```

```
Standardised Residuals Tests:  Statistic p-Value
 Jarque-Bera Test   R    Chi^2  9.118036  0.01047234
 Shapiro-Wilk Test  R    W      0.9842407 0.01433690
 Ljung-Box Test     R    Q(10)  9.874326  0.4515875
 Ljung-Box Test     R    Q(15)  17.55855  0.2865844
 Ljung-Box Test     R    Q(20)  23.41363  0.2689437
 Ljung-Box Test     R^2  Q(10)  19.2821   0.03682246
 Ljung-Box Test     R^2  Q(15)  33.23648  0.004352736
 Ljung-Box Test     R^2  Q(20)  37.74259  0.009518992
 LM Arch Test       R    TR^2   25.41625  0.01296901
```

In this example, we obtain $\widehat{\phi}_0 = .005$ (called `mu` in the output) and $\widehat{\phi}_1 = .367$ (called `ar1`) for the AR(1) parameter estimates; in Example 3.38 the values were .005 and .347, respectively. The ARCH(1) parameter estimates are $\widehat{\alpha}_0 = 0$ (called `omega`) for the constant and $\widehat{\alpha}_1 = .195$, which is significant with a p-value of about .04. There are a number of tests that are performed on the residuals [R] or the squared residuals [R^2]. For example, the Jarque–Bera statistic tests the residuals of the fit for normality based on the observed skewness and kurtosis, and it appears that the residuals have some non-normal skewness and kurtosis. The Shapiro–Wilk statistic tests the residuals of the fit for normality based on the empirical order statistics. The other tests, primarily based on the Q-statistic, are used on the residuals and their squares.

The ARCH(1) model can be extended to the general ARCH(m) model in an obvious way. That is, (5.36), $y_t = \sigma_t \epsilon_t$, is retained, but (5.37) is extended to

$$\sigma_t^2 = \alpha_0 + \alpha_1 y_{t-1}^2 + \cdots + \alpha_m y_{t-m}^2. \tag{5.47}$$

Estimation for ARCH(m) also follows in an obvious way from the discussion of estimation for ARCH(1) models. That is, the conditional likelihood of the data $y_{m+1}, \ldots, y_n$ given $y_1, \ldots, y_m$, is given by

$$L(\boldsymbol{\alpha} \mid y_1, \ldots, y_m) = \prod_{t=m+1}^{n} f_{\boldsymbol{\alpha}}(y_t \mid y_{t-1}, \ldots, y_{t-m}), \tag{5.48}$$

where $\boldsymbol{\alpha} = (\alpha_0, \alpha_1, \ldots, \alpha_m)$ and the conditional densities $f_{\boldsymbol{\alpha}}(\cdot \mid \cdot)$ in (5.48) are normal densities; that is, for $t > m$,

$$y_t \mid y_{t-1}, \ldots, y_{t-m} \sim \mathrm{N}(0, \alpha_0 + \alpha_1 y_{t-1}^2 + \cdots + \alpha_m y_{t-m}^2).$$

Another extension of ARCH is the generalized ARCH or GARCH model developed by Bollerslev (1986). For example, a GARCH(1, 1) model retains (5.36), $y_t = \sigma_t \epsilon_t$, but extends (5.37) as follows:

$$\sigma_t^2 = \alpha_0 + \alpha_1 y_{t-1}^2 + \beta_1 \sigma_{t-1}^2. \tag{5.49}$$

Under the condition that $\alpha_1 + \beta_1 < 1$, using similar manipulations as in (5.39), the GARCH(1, 1) model, (5.36) and (5.49), admits a non-Gaussian ARMA(1, 1) model for the squared process

$$y_t^2 = \alpha_0 + (\alpha_1 + \beta_1)y_{t-1}^2 + v_t - \beta_1 v_{t-1}, \tag{5.50}$$

where v_t is as defined in (5.39). Representation (5.50) follows by writing (5.36) as

$$y_t^2 - \sigma_t^2 = \sigma_t^2(\epsilon_t^2 - 1)$$
$$\beta_1(y_{t-1}^2 - \sigma_{t-1}^2) = \beta_1\sigma_{t-1}^2(\epsilon_{t-1}^2 - 1),$$

subtracting the second equation from the first, and using the fact that, from (5.49), $\sigma_t^2 - \beta_1\sigma_{t-1}^2 = \alpha_0 + \alpha_1 y_{t-1}^2$, on the left-hand side of the result. The GARCH(m, r) model retains (5.36) and extends (5.49) to

$$\sigma_t^2 = \alpha_0 + \sum_{j=1}^{m} \alpha_j y_{t-j}^2 + \sum_{j=1}^{r} \beta_j \sigma_{t-j}^2. \tag{5.51}$$

Conditional maximum likelihood estimation of the GARCH(m, r) model parameters is similar to the ARCH(m) case, wherein the conditional likelihood, (5.48), is the product of N$(0, \sigma_t^2)$ densities with σ_t^2 given by (5.51) and where the conditioning is on the first $\max(m, r)$ observations, with $\sigma_1^2 = \cdots = \sigma_r^2 = 0$. Once the parameter estimates are obtained, the model can be used to obtain one-step-ahead forecasts of the volatility, say $\widehat{\sigma}_{t+1}^2$, given by

$$\widehat{\sigma}_{t+1}^2 = \widehat{\alpha}_0 + \sum_{j=1}^{m} \widehat{\alpha}_j y_{t+1-j}^2 + \sum_{j=1}^{r} \widehat{\beta}_j \widehat{\sigma}_{t+1-j}^2. \tag{5.52}$$

We explore these concepts in the following example.

Example 5.5 GARCH Analysis of the NYSE Returns

As previously mentioned, the daily returns of the NYSE shown in Figure 1.4 exhibit classic GARCH features. We used the `fGarch` package to fit a GARCH$(1, 1)$ model to the series with an AR(1) conditional mean:

```
library(fGarch)
nyse = astsa::nyse
summary(nyse.g <- garchFit(~arma(1,0)+garch(1,1), nyse))
```

	Estimate	Std. Error	t value	Pr(>|t|)	
mu	6.548e-04	1.770e-04	3.699	0.000217	***
ar1	1.075e-01	2.516e-02	4.273	1.93e-05	***
omega	6.218e-06	1.381e-06	4.502	6.72e-06	***
alpha1	1.093e-01	1.538e-02	7.106	1.19e-12	***
beta1	8.138e-01	2.856e-02	28.490	< 2e-16	***

Standardised Residuals Tests:

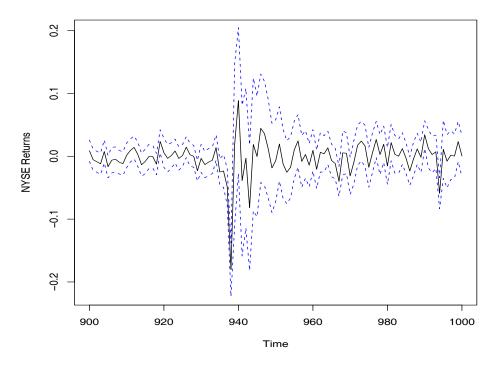|			Statistic	p-Value
Jarque-Bera Test	R	Chi^2	4029.819	0
Shapiro-Wilk Test	R	W	0.9507688	0
Ljung-Box Test	R	Q(10)	4.006516	0.9470526

Fig. 5.6. GARCH predictions of the NYSE volatility, $\pm 2\widehat{\sigma}_t$, displayed as dashed lines.

```
Ljung-Box Test     R    Q(15)   4.754317   0.9940101
Ljung-Box Test     R    Q(20)   6.748514   0.997432
Ljung-Box Test     R^2  Q(10)   3.439174   0.9691234
Ljung-Box Test     R^2  Q(15)   4.528756   0.9954202
Ljung-Box Test     R^2  Q(20)   7.352159   0.9953599
LM Arch Test       R    TR^2    3.936223   0.9845599
```

To explore the GARCH predictions of volatility, we calculated and plotted the 100 observations from the middle of the data (which includes the October 19, 1987 crash) along with the one-step-ahead predictions of the corresponding volatility, σ_t^2. The results are displayed as the data $\pm 2\widehat{\sigma}_t$ as a dashed line surrounding the data in Figure 5.6.

```r
u = nyse.g@sigma.t
plot(window(nyse, start=900, end=1000), ylim=c(-.22,.2), ylab="NYSE
    Returns")
lines(window(nyse-2*u, start=900, end=1000), lty=2, col=4)
lines(window(nyse+2*u, start=900, end=1000), lty=2, col=4)
```

Some key points can be gleaned from the examples of this section. First, it is apparent that the conditional distribution of the returns is rarely normal. `fGarch` allows for various distributions to be fit to the data; see the help file for information. Some drawbacks of the GARCH model are: (i) the model assumes positive and negative returns have the same effect because volatility depends on squared returns; (ii) the model is restrictive because of the tight constraints on the model parameters (e.g., for an ARCH(1), $0 \leq \alpha_1^2 < \frac{1}{3}$); (iii)