

# 9

---

## Time Series Models: Basics

### 9.1 Time Series Data

A *time series* is a sequence of observations in chronological order, for example, daily log returns on a stock or monthly values of the Consumer Price Index (CPI). In this chapter, we study statistical models for time series. These models are widely used in econometrics, business forecasting, and many scientific applications.

A *stochastic process* is a sequence of random variables and can be viewed as the “theoretical” or “population” analog of a time series—conversely, a time series can be considered a sample from the stochastic process. “Stochastic” is a synonym for random.

One of the most useful methods for obtaining parsimony in a time series model is to assume *stationarity*, a property discussed next.

### 9.2 Stationary Processes

When we observe a time series, the fluctuations appear random, but often with the same type of stochastic behavior from one time period to the next. For example, returns on stocks or changes in interest rates can be very different from the previous year, but the mean, standard deviation, and other statistical properties often are similar from one year to the next.<sup>1</sup> Similarly, the demand for many consumer products, such as sunscreen, winter coats, and electricity, has random as well as seasonal variation, but each summer is similar to past summers, each winter to past winters, at least over shorter time periods. *Stationary stochastic processes* are probability models for time series with time-invariant behavior.

<sup>1</sup> It is the returns, not the stock prices, that have time-invariant behavior. Stock prices themselves tend to increase over time, so this year’s stock prices tend to be higher and more variable than those a decade or two ago.

A process is said to be *strictly stationary* if all aspects of its behavior are unchanged by shifts in time. Mathematically, stationarity is defined as the requirement that for every  $m$  and  $n$ , the distributions of  $Y_1, \dots, Y_n$  and  $Y_{1+m}, \dots, Y_{n+m}$  are the same; that is, the probability distribution of a sequence of  $n$  observations does not depend on their time origin. Strict stationarity is a very strong assumption, because it requires that “all aspects” of behavior be constant in time. Often, we can get by assuming less, namely, weak stationarity. A process is *weakly stationary* if its mean, variance, and covariance are unchanged by time shifts. More precisely,  $Y_1, Y_2, \dots$  is a *weakly stationary process* if

- $E(Y_i) = \mu$  (a constant) for all  $i$ ;
- $\text{Var}(Y_i) = \sigma^2$  (a constant) for all  $i$ ; and
- $\text{Corr}(Y_i, Y_j) = \rho(|i - j|)$  for all  $i$  and  $j$  for some function  $\rho(h)$ .

Thus, the mean and variance do not change with time and the correlation between two observations depends only on the *lag*, the time distance between them. For example, if the process is stationary, then the correlation between  $Y_2$  and  $Y_5$  is the same as the correlation between  $Y_7$  and  $Y_{10}$ , since each pair is separated by three units of time. The adjective “weakly” in “weakly stationary” refers to the fact that we are only assuming that means, variance, and covariances, not other distributional characteristics such as quantiles, skewness, and kurtosis, are stationary. The term *stationary* will sometimes be used as a shorthand for strictly stationary.

The function  $\rho$  is called the *autocorrelation function* of the process. Note that  $\rho(h) = \rho(-h)$ . Why?

The covariance between  $Y_t$  and  $Y_{t+h}$  is denoted by  $\gamma(h)$  and  $\gamma(\cdot)$  is called the *autocovariance function*. Note that  $\gamma(h) = \sigma^2 \rho(h)$  and that  $\gamma(0) = \sigma^2$ . Also,  $\rho(h) = \gamma(h)/\sigma^2 = \gamma(h)/\gamma(0)$ .

As mentioned, many financial time series are not stationary, but often the *changes* in them, perhaps after they have been log transformed, are stationary. For this reason, stationary time series models are far more applicable than they might appear. From the viewpoint of statistical modeling, it is not important whether it is the time series itself or changes in the time series that are stationary, because either way we get a parsimonious model.

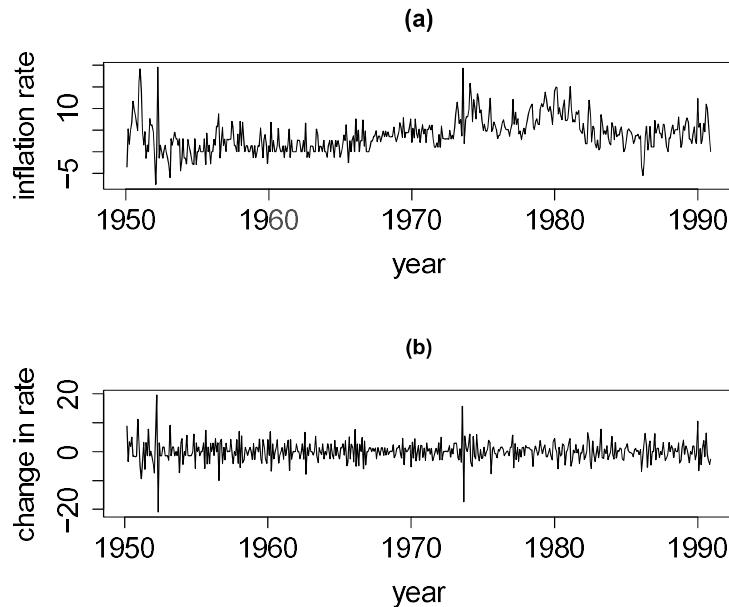
The beauty of a stationary process is that it can be modeled with relatively few parameters. For example, we do not need a different expectation for each  $Y_t$ ; rather they all have a common expectation,  $\mu$ . This implies that  $\mu$  can be estimated accurately by  $\bar{Y}$ . If instead we did not assume stationarity and each  $Y_t$  had its own unique expectation,  $\mu_t$ , then it would not be possible to estimate  $\mu_t$  accurately— $\mu_t$  could only be estimated by the single observation  $Y_t$  itself.

When a time series is observed, a natural question is whether it appears to be stationary. This is not an easy question to address, and we can never be absolutely certain of the answer. However, visual inspection of the time series and changes in the time series can be helpful. A *time series plot* is a plot of

the series in chronological order. This very basic plot is useful for assessing stationary behavior, though it can be supplemented with other plots, such as the plot of the sample autocorrelation function that will be introduced later. In addition, there are statistical tests of stationarity—these are discussed in Section 9.10.

A time series plot of a stationary series should show oscillation around some fixed level, a phenomenon called *mean-reversion*. If the series wanders without returning repeatedly to some fixed level, then the series should not be modeled as a stationary process.

*Example 9.1. Inflation rates and changes in inflation rates—Time series plots*

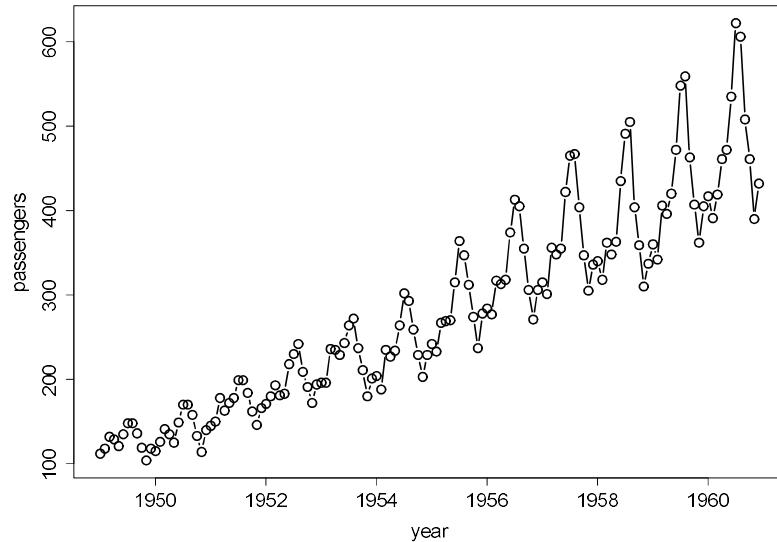


**Fig. 9.1.** Time series plots of (a) one-month (in percent, annual rate) inflation rate and (b) first differences in the rate. It is unclear if the series in (a) is stationary, but the differenced series in (b) seems suitable for modeling as stationary.

The one-month inflation rate (in percent, annual rate) is plotted in [Figure 9.1\(a\)](#). The data come from the Mishkin data set in R's Ecdat package. The series may be wandering without reverting to a fixed mean, as would be

expected with a stationary time series, or it may be slowly reverting to a mean of approximately 4%. In panel (b), the first differences, that is, the changes from one month to the next, are plotted. In contrast to the original series, the differenced series certainly oscillate around a fixed mean that is 0%, or nearly so. The differenced series is clearly stationary, but whether or not the original series is stationary needs further investigation. We will return to this question later.  $\square$

*Example 9.2. Air passengers*



**Fig. 9.2.** Time series plot of monthly totals of air passengers (in thousands).

[Figure 9.2](#) is a plot of the monthly totals of international airline passengers for the years 1949 to 1960. The data are in the data set `AirPassengers` in R's `Datasets` package. There are three types of nonstationarity seen in the plot. First is the obvious upward trend, second is the seasonal variation, and third is the increase over time in the size of the seasonal oscillations.  $\square$

### 9.2.1 White Noise

White noise is the simplest example of a stationary process. We will define several types of white noise with increasingly restrictive assumptions.

The sequence  $Y_1, Y_2, \dots$  is a *weak white noise process* with mean  $\mu$  and variance  $\sigma^2$ , which will be shortened to “weak WN( $\mu, \sigma^2$ ),” if

- $E(Y_i) = \mu$  for all  $i$ ;
- $\text{Var}(Y_i) = \sigma^2$  (a constant) for all  $i$ ; and
- $\text{Corr}(Y_i, Y_j) = 0$  for all  $i \neq j$ .

If the mean is not specified, then it is assumed that  $\mu = 0$ .

$Y_1, Y_2, \dots$  is an i.i.d. process, then we call it an *i.i.d. white noise process* or simply *i.i.d. WN( $\mu, \sigma^2$ )*. An i.i.d. white noise process is also a weak white noise process, but not vice versa.

If, in addition,  $Y_1, Y_2, \dots$  is an i.i.d. process with a specific marginal distribution, then this might be noted. For example, if  $Y_1, Y_2, \dots$  are i.i.d. normal random variables, then the process is called a *Gaussian white noise process*. Similarly, if  $Y_1, Y_2, \dots$  are i.i.d.  $t$  random variables with  $\nu$  degrees of freedom, then it is called a  $t_\nu$  white noise process.

A weak white noise process is weakly stationary with

$$\begin{aligned}\rho(0) &= 1, \\ \rho(h) &= 0 \text{ if } h \neq 0,\end{aligned}$$

so that

$$\begin{aligned}\gamma(0) &= \sigma^2, \\ \gamma(h) &= 0 \text{ if } h \neq 0.\end{aligned}$$

I.i.d. white noise is strictly stationary and weak white noise is weakly stationary.

### 9.2.2 Predicting White Noise

Because of the lack of correlation, past values of a white noise process contain no information that can be used to predict future values. More precisely, suppose that  $\dots, Y_1, Y_2, \dots$  is an i.i.d. WN( $\mu, \sigma^2$ ) process. Then

$$E(Y_{i+t} | Y_1, \dots, Y_i) = \mu \text{ for all } t \geq 1. \quad (9.1)$$

What this equation is saying is that one cannot predict the future deviations of a white noise process from its mean, because its future is independent of its past and present. Therefore, the best predictor of any future value of the process is simply the mean  $\mu$ , what you would use even if  $Y_1, \dots, Y_i$  had not been observed. For weak white noise, (9.1) need not be true, but it is still true that the best linear predictor<sup>2</sup> of  $Y_{i+t}$  given  $Y_1, \dots, Y_i$  is  $\mu$ .

---

<sup>2</sup> Best linear prediction is discussed in Section 14.10.1.

### 9.3 Estimating Parameters of a Stationary Process

Suppose we observe  $Y_1, \dots, Y_n$  from a stationary process. To estimate the mean  $\mu$  and variance  $\sigma^2$  of the process, we can use the sample mean  $\bar{Y}$  and sample variance  $s^2$ .

To estimate the autocovariance function, we use the *sample autocovariance function*

$$\hat{\gamma}(h) = n^{-1} \sum_{j=1}^{n-h} (Y_{j+h} - \bar{Y})(Y_j - \bar{Y}). \quad (9.2)$$

Equation (9.2) is an example of the usefulness of parsimony induced by the stationarity assumption. Because the correlation between  $Y_t$  and  $Y_{t+h}$  is independent of  $t$ , all  $n-h$  pairs of data points that are separated by a lag of  $h$  time units can be used to estimate  $\gamma(h)$ . Some authors define  $\hat{\gamma}(h)$  with the factor  $n^{-1}$  in (9.2) replaced by  $(n-h)^{-1}$ , but this change has little effect if  $n$  is reasonably large and  $h$  is small relative to  $n$ , as is typically the case.

To estimate  $\rho(\cdot)$ , we use the *sample autocorrelation function (sample ACF)* defined as

$$\hat{\rho}(h) = \frac{\hat{\gamma}(h)}{\hat{\gamma}(0)}.$$

#### 9.3.1 ACF Plots and the Ljung–Box Test

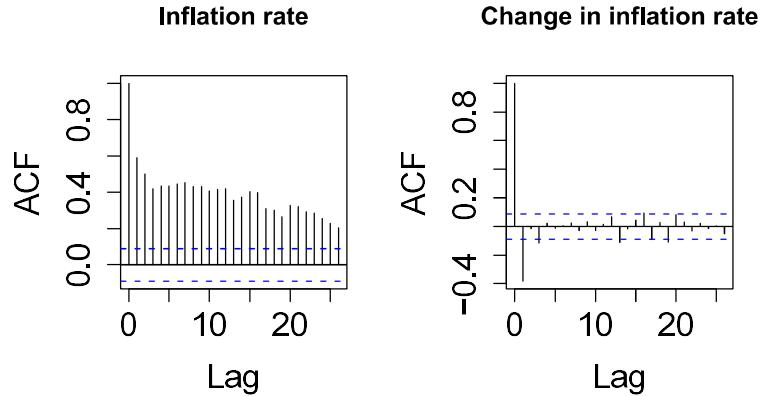
Most statistical software will plot a sample ACF with *test bounds*. These bounds are used to test the null hypothesis that an autocorrelation coefficient is 0. The null hypothesis is rejected if the sample autocorrelation is outside the bounds. The usual level of the test is 0.05, so one can expect to see about 1 out of 20 sample autocorrelations outside the test bounds simply by chance.

An alternative to using the bounds to test the autocorrelations one at a time is to use a simultaneous test. A *simultaneous test* is one that tests whether a group of null hypotheses are all true versus the alternative that at least one of them is false. The null hypothesis of the Ljung–Box test is  $H_0 : \rho(1) = \rho(2) = \dots = \rho(K) = 0$  for some  $K$ , say  $K = 5$  or 10. If the Ljung–Box test rejects, then we conclude that one or more of  $\rho(1) = \rho(2) = \dots = \rho(K)$  is nonzero.

If, in fact, the autocorrelations 1 to  $K$  are all zero, then there is only a 1 in 20 chance of falsely concluding that they are not all zero, assuming a level 0.05 test. In contrast, if the autocorrelations are tested one at a time, then there is a much higher chance of concluding that one or more is nonzero.

The Ljung–Box test is sometimes called simply the Box test, though the former name is preferable since the test is based on a joint paper of Ljung and Box.

*Example 9.3. Inflation rates and changes in the inflation rate—ACF plots and Ljung–Box test*



**Fig. 9.3.** Sample ACF plots of the one-month inflation rate (a) and changes in this rate (b).

We return to the inflation rate data used in Example 9.1. [Figure 9.3](#) contains plots of (a) the sample ACF of the one-month inflation rate and (b) the sample ACF of changes in the inflation rate. In (a) we see that the sample ACF decays to zero slowly. This is a sign of either nonstationarity or possibly of stationarity with long-memory dependence, which is discussed in Section 10.4. In contrast, the sample ACF in (b) decays to zero quickly, indicating clearly that the differenced series is stationary. Thus, the sample ACF plots agree with the conclusions reached by examining the time series plots in [Figure 9.1](#), specifically that the differenced series is stationary and the original series might not be. In Section 9.10 we will use hypothesis testing to further address the question of whether or not the original series is stationary.

Several of the autocorrelations of the rate changes series fall outside the test bounds, which suggests that the series is not white noise. To check, the Ljung–Box test was implemented using R’s `Box.test` function. The Ljung–Box test with  $K = 10$  has an extremely small  $p$ -value,  $6.665\text{e}{-13}$ , so the null hypothesis of white noise is strongly rejected. Other choices of  $K$  give similar results.  $K$  is called `lag` when `Box.test` is called and `df` in the output.

□

Although a stationary process is somewhat parsimonious with parameters, at least relative to a general nonstationary process, a stationary process is still

not sufficiently parsimonious for most purposes. The problem is that there are still an infinite number of parameters,  $\rho(1), \rho(2), \dots$ . What we need is a class of stationary time series models with only a finite, preferably small, number of parameters. The ARIMA models of this chapter are precisely such a class. The simplest ARIMA models are autoregressive (AR) models, and we turn to these first.

## 9.4 AR(1) Processes

Time series models with correlation can be built out of white noise. The simplest correlated stationary processes are *autoregressive processes*, where  $Y_t$  is modeled as a weighted average of past observations plus a white noise “error,” which is also called the “noise” or “disturbance.” We start with AR(1) processes, the simplest autoregressive processes.

Let  $\epsilon_1, \epsilon_2, \dots$  be  $\text{WN}(0, \sigma_\epsilon^2)$ . We say that  $Y_1, Y_2, \dots$  is an *AR(1) process* if for some constant parameters  $\mu$  and  $\phi$ ,

$$Y_t - \mu = \phi(Y_{t-1} - \mu) + \epsilon_t \quad (9.3)$$

for all  $t$ . The parameter  $\mu$  is the mean of the process. Think of the term  $\phi(Y_{t-1} - \mu)$  as representing “memory” or “feedback” of the past into the present value of the process. The process  $\{Y_t\}_{t=-\infty}^{+\infty}$  is correlated because the deviation of  $Y_{t-1}$  from its mean is fed back into  $Y_t$ . The parameter  $\phi$  determines the amount of feedback, with a larger absolute value of  $\phi$  resulting in more feedback and  $\phi = 0$  implying that  $Y_t = \mu + \epsilon_t$ , so that  $Y_t$  is  $\text{WN}(\mu, \sigma_\epsilon^2)$ . In applications in finance, one can think of  $\epsilon_t$  as representing the effect of “new information.” For example, if  $Y_t$  is the log return on an asset at time  $t$ , then  $\epsilon_t$  represents the effect on the asset’s price of business and economic information that is revealed at time  $t$ . Information that is truly new cannot be anticipated, so the effects of today’s new information should be independent of the effects of yesterday’s news. This is why we model new information as white noise.

If  $Y_1, \dots$  is a weakly stationary process, then  $|\phi| < 1$ . To see this, note that stationarity implies that the variances of  $(Y_t - \mu)$  and  $(Y_{t-1} - \mu)$  in (9.3) are equal, say, to  $\sigma_Y^2$ . Therefore,  $\sigma_Y^2 = \phi^2 \sigma_Y^2 + \sigma_\epsilon^2$ , which requires that  $|\phi| < 1$ . The mean of this process is  $\mu$ . Simple algebra shows that (9.3) can be rewritten as

$$Y_t = (1 - \phi)\mu + \phi Y_{t-1} + \epsilon_t. \quad (9.4)$$

Recall the linear regression model  $Y_t = \beta_0 + \beta_1 Y_t + \epsilon_t$  from your statistics courses or peek ahead to Chapter 12 for an introduction to regression analysis. Equation (9.4) is just a linear regression model with intercept  $\beta_0 = (1 - \phi)\mu$  and slope  $\beta_1 = \phi$ , since the model can be rewritten as

$$Y_t = (1 - \phi)\mu + \phi Y_{t-1} + \epsilon_t.$$

The term *autoregression* refers to the regression of the process on its own past values.

If  $|\phi| < 1$ , then repeated use of equation (9.3) shows that

$$Y_t = \mu + \epsilon_t + \phi\epsilon_{t-1} + \phi^2\epsilon_{t-2} + \dots = \mu + \sum_{h=0}^{\infty} \phi^h \epsilon_{t-h}, \quad (9.5)$$

and assumes that time parameter  $t$  of  $Y_t$  and  $\epsilon_t$  can be extended to negative values so that the white noise process is  $\dots, \epsilon_{-2}, \epsilon_{-1}, \epsilon_0, \epsilon_1, \dots$  and (9.3) is true for all integers  $t$ . Equation (9.5) is called *the infinite moving average* [MA( $\infty$ )] representation of the process. This equation shows that  $Y_t$  is a weighted average of *all* past values of the white noise process. This representation should be compared to the AR(1) representation that shows  $Y_t$  as depending only on  $Y_{t-1}$  and  $\epsilon_t$ . Since  $|\phi| < 1$ ,  $\phi^h \rightarrow 0$  as the lag  $h \rightarrow \infty$ . Thus, the weights given to the distant past are small. In fact, they are quite small. For example, if  $\phi = 0.5$ , then  $\phi^{10} = 0.00098$ , so  $\epsilon_{t-10}$  has virtually no effect on  $Y_t$ . For this reason, the sum in (9.5) could be truncated at a finite number of terms so there is no need to assume that the processes existed in the infinite past.

#### 9.4.1 Properties of a stationary AR(1) Process

When an AR(1) process is stationary, which implies that  $|\phi| < 1$ , then

$$E(Y_t) = \mu \quad \forall t, \quad (9.6)$$

$$\gamma(0) = \text{Var}(Y_t) = \frac{\sigma_\epsilon^2}{1 - \phi^2} \quad \forall t, \quad (9.7)$$

$$\gamma(h) = \text{Cov}(Y_t, Y_{t+h}) = \frac{\sigma_\epsilon^2 \phi^{|h|}}{1 - \phi^2} \quad \forall t \text{ and } \forall h, \quad (9.8)$$

and

$$\rho(h) = \text{Corr}(Y_t, Y_{t+h}) = \phi^{|h|} \quad \forall t \text{ and } \forall h. \quad (9.9)$$

It is important to remember that formulas (9.6) to (9.9) hold only if  $|\phi| < 1$  and only for AR(1) processes. Moreover, for  $Y_t$  to be stationary,  $Y_0$  must start in the stationary distribution so that  $E(Y_0) = \mu$  and  $\text{Var}(Y_0) = \sigma_\epsilon^2 / (1 - \phi^2)$ . Otherwise,  $Y_t$  is not stationary though it eventually converges to stationarity.

These formulas can be proved using (9.5). For example, using (7.11) in Section 7.3.2,

$$\text{Var}(Y_t) = \text{Var}\left(\sum_{h=0}^{\infty} \phi^h \epsilon_{t-h}\right) = \sigma_\epsilon^2 \sum_{h=0}^{\infty} \phi^{2h} = \frac{\sigma_\epsilon^2}{1 - \phi^2}, \quad (9.10)$$

which proves (9.7). In (9.10) the formula for summation of a geometric series was used. This formula is

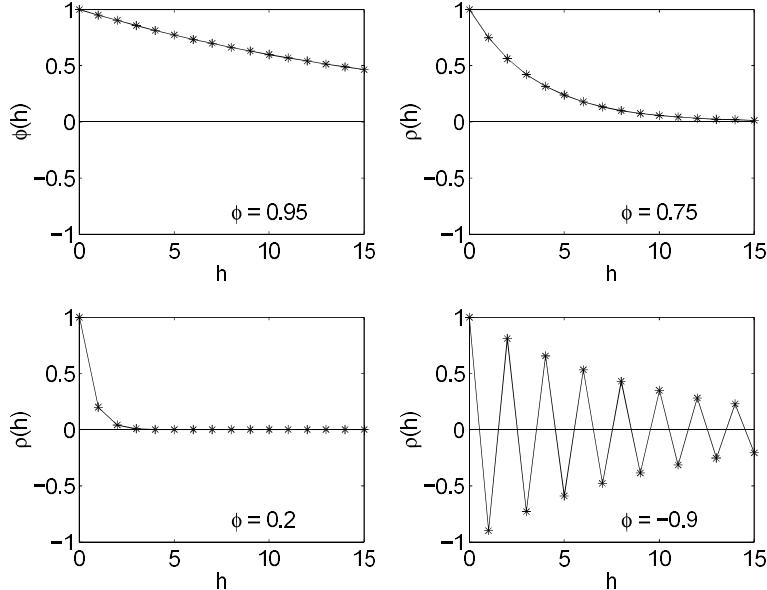
$$\sum_{i=0}^{\infty} r^i = \frac{1}{1-r} \text{ if } |r| < 1. \quad (9.11)$$

Also, for  $h > 0$ ,

$$\text{Cov} \left( \sum_{i=0}^{\infty} \epsilon_{t-i} \phi^i, \sum_{j=0}^{\infty} \epsilon_{t+h-j} \phi^j \right) = \frac{\sigma_{\epsilon}^2 \phi^{|h|}}{1 - \phi^2}, \quad (9.12)$$

thus verifying (9.8). Then (9.9) follows by dividing (9.8) by (9.7).

Be sure to distinguish between  $\sigma_{\epsilon}^2$ , which is the variance of the white noise process  $\epsilon_1, \epsilon_2, \dots$ , and  $\gamma(0)$ , which is the variance of the AR(1) process  $Y_1, Y_2, \dots$ . We can see from (9.7) that  $\gamma(0)$  is larger than  $\sigma_{\epsilon}^2$  unless  $\phi = 0$ , in which case  $Y_t = \mu + \epsilon_t$ , so that  $Y_t$  and  $\epsilon_t$  have the same variance.



**Fig. 9.4.** Autocorrelation functions of AR(1) processes with  $\phi$  equal to 0.95, 0.75, 0.2, and -0.9.

The ACF (autocorrelation function) of an AR(1) process depends upon only one parameter,  $\phi$ . This is a remarkable amount of parsimony, but it comes at a price. The ACF of an AR(1) process has only a limited range of shapes, as can be seen in [Figure 9.4](#). The magnitude of its ACF decays geometrically to zero, either slowly as when  $\phi = 0.95$ , moderately slowly as when  $\phi = 0.75$ , or rapidly as when  $\phi = 0.2$ . If  $\phi < 0$ , then the sign of the ACF alternates as its magnitude decays geometrically. If the sample ACF

of the data does not behave in one of these ways, then an AR(1) model is unsuitable. The remedy is to use more AR parameters, to switch to another class of models such as the moving average (MA) or autoregressive moving average (ARMA) models. We investigate these alternatives in this chapter.

#### 9.4.2 Convergence to the Stationary Distribution

Suppose that  $Y_0$  is an arbitrary starting value not chosen from the stationary distribution and that (9.3) holds for  $t = 1, \dots$ . Then the process is not stationary, but converges to the stationary distribution satisfying (9.6) to (9.9) as  $t \rightarrow \infty$ .<sup>3</sup> For example, since  $Y_t - \mu = \phi(Y_{t-1} - \mu) + \epsilon_t$ ,  $E(Y_1) - \mu = \phi\{E(Y_0) - \mu\}$ ,  $E(Y_2) - \mu = \phi^2\{E(Y_0) - \mu\}$ , and so forth, so that

$$E(Y_t) = \mu + \phi^t\{E(Y_0) - \mu\} \text{ for all } t > 0. \quad (9.13)$$

Since  $|\phi| < 1$ ,  $\phi^t \rightarrow 0$  and  $E(Y_t) \rightarrow \mu$  as  $t \rightarrow \infty$ . The convergence of  $\text{Var}(Y_t)$  to  $\sigma_\epsilon^2/(1-\phi^2)$  can be proved in a somewhat similar manner. The convergence to the stationary distribution can be very rapid when  $|\phi|$  is not too close to 1. For example, if  $\phi = 0.5$ , then  $\phi^{10} = 0.00097$ , so by (9.13)  $E(Y_{10})$  is very close to  $\mu$  unless  $E(Y_0)$  was extremely far from  $\mu$ .

#### 9.4.3 Nonstationary AR(1) Processes

If  $|\phi| \geq 1$ , then the AR(1) process is nonstationary, and the mean, variance, and correlation are not constant.

##### Random Walk ( $\phi = 1$ )

If  $\phi = 1$ , then

$$Y_t = Y_{t-1} + \epsilon_t$$

and the process is *not* stationary. This is the random walk process we saw in Chapter 2.

Suppose we start the process at an arbitrary point  $Y_0$ . It is easy to see that

$$Y_t = Y_0 + \epsilon_1 + \dots + \epsilon_t.$$

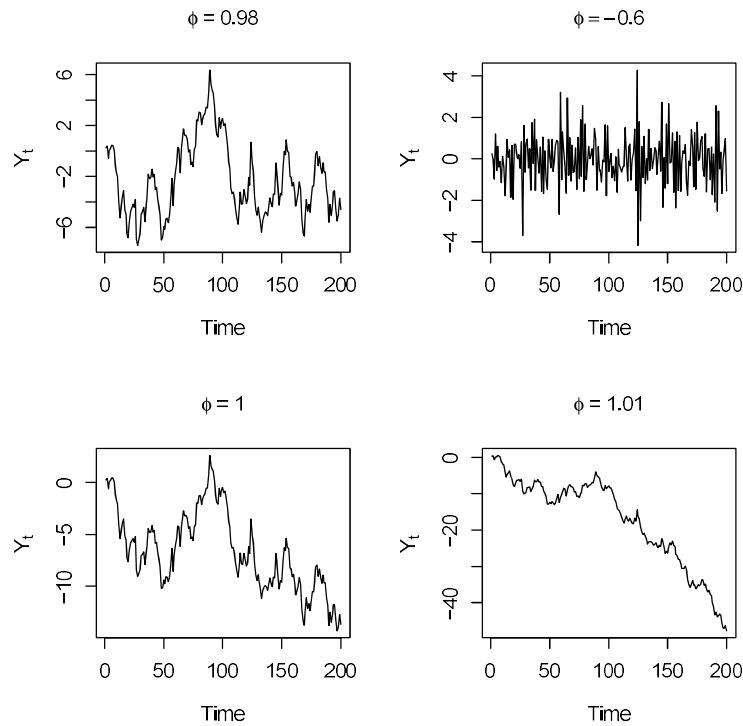
Then  $E(Y_t|Y_0) = Y_0$  for all  $t$ , which is constant but depends entirely on the arbitrary starting point. Moreover,  $\text{Var}(Y_t|Y_0) = t\sigma_\epsilon^2$ , which is not stationary but rather increases linearly with time. The increasing variance makes the random walk “wander” in that  $Y_t$  takes increasingly longer excursions away from its conditional mean of  $Y_0$  and therefore is not mean-reverting.

---

<sup>3</sup> However, there is a technical issue here. It must be assumed that  $Y_0$  has a finite mean and variance, since otherwise  $Y_t$  will not have a finite mean and variance for any  $t > 0$ .

### AR(1) Processes When $|\phi| > 1$

When  $|\phi| > 1$ , an AR(1) process has explosive behavior. This can be seen in [Figure 9.5](#). This figure shows simulations of 200 observations from AR(1) processes with various values of  $\phi$ . The explosive case where  $\phi = 1.01$  clearly is different from the other cases where  $|\phi| \leq 1$ . However, the case where  $\phi = 1$  is not that much different from  $\phi = 0.98$  even though the former is nonstationary while the latter is stationary. Longer time series would help distinguish between  $\phi = 0.98$  and  $\phi = 1$ .



**Fig. 9.5.** Simulations of 200 observations from AR(1) processes with various values of  $\phi$  and  $\mu = 0$ . The white noise process  $\epsilon_1, \epsilon_2, \dots, \epsilon_{200}$  is the same for all four AR(1) processes.

## 9.5 Estimation of AR(1) Processes

R has the function `arima` for fitting AR and other time series models. `arima` and similar functions in other software packages have two estimation meth-

ods, conditional least-squares and maximum likelihood. The two methods are explained in Section 9.5.2. They similar and generally give nearly the same estimates. In this book, we use the default method in R's `arima`, which is the MLE with the conditional least-squares estimate as the starting value for computing the MLE by nonlinear optimization.

### 9.5.1 Residuals and Model Checking

Once  $\mu$  and  $\phi$  have been estimated, one can estimate the white noise process  $\epsilon_1, \dots, \epsilon_n$ . Rearranging equation (9.3), we have

$$\epsilon_t = (Y_t - \mu) - \phi(Y_{t-1} - \mu). \quad (9.14)$$

In analogy with (9.14), the residuals,  $\hat{\epsilon}_2, \hat{\epsilon}_3, \dots, \hat{\epsilon}_n$ , are defined as

$$\hat{\epsilon}_t = (Y_t - \hat{\mu}) - \hat{\phi}(Y_{t-1} - \hat{\mu}), \quad t \geq 2, \quad (9.15)$$

and estimate  $\epsilon_2, \dots, \epsilon_n$ . The first noise,  $\epsilon_1$ , cannot be estimated since it is assumed that the observations start at  $Y_1$  so that  $Y_0$  is not available. The residuals can be used to check the assumption that  $Y_1, Y_2, \dots, Y_n$  is an AR(1) process; any autocorrelation in the residuals is evidence against the assumption of an AR(1) process.

To appreciate why residual autocorrelation indicates a possible problem with the model, suppose that we are fitting an AR(1) model,  $Y_t = \mu + \phi(Y_{t-1} - \mu) + \epsilon_t$ , but the true model is an AR(2) process<sup>4</sup> given by

$$(Y_t - \mu) = \phi_1(Y_{t-1} - \mu) + \phi_2(Y_{t-2} - \mu) + \epsilon_t.$$

Since we are fitting the incorrect AR(1) model, there is no hope of estimating  $\phi_2$  since it is not in the model. Moreover,  $\hat{\phi}$  does not necessarily estimate  $\phi_1$  because of bias caused by model misspecification. Let  $\phi^*$  be the expected value of  $\hat{\phi}$ . For the purpose of illustration, assume that  $\hat{\mu} \approx \mu$  and  $\hat{\phi} \approx \phi^*$ . This is a sensible approximation if the sample size  $n$  is large enough. Then

$$\begin{aligned} \hat{\epsilon}_t &\approx (Y_t - \mu) - \phi^*(Y_{t-1} - \mu) \\ &= \phi_1(Y_{t-1} - \mu) + \phi_2(Y_{t-2} - \mu) + \epsilon_t - \phi^*(Y_{t-1} - \mu) \\ &= (\phi_1 - \phi^*)(Y_{t-1} - \mu) + \phi_2(Y_{t-2} - \mu) + \epsilon_t. \end{aligned}$$

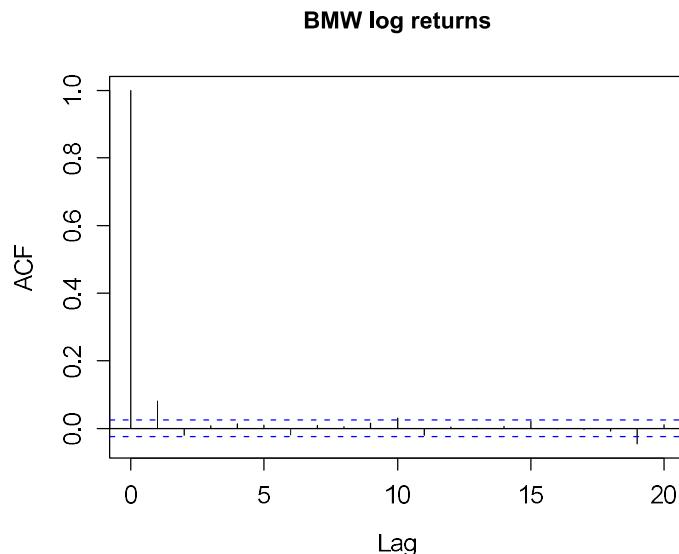
Thus, the residuals do not estimate the white noise process as they would if the correct AR(2) model were used. Even if there is no bias in the estimation of  $\phi_1$  by  $\hat{\phi}$  so that  $\phi_1 = \phi^*$  and the term  $(\phi_1 - \phi^*)(Y_{t-1} - \mu)$  drops out, the presence of  $\phi_2(Y_{t-2} - \mu)$  in the residuals causes them to be autocorrelated.

To check for residual autocorrelation, one can use the *test bounds* of ACF plots. Any residual ACF value outside the test bounds is significantly different

---

<sup>4</sup> We discuss higher-order AR models in more detail soon.

from 0 at the 0.05 level. As discussed earlier, the danger here is that some sample ACF values will be significant merely by chance, and to guard against this danger, one can use the Ljung–Box test that *simultaneously* tests that all autocorrelations up to a specified lag are zero. When the Ljung–Box test is applied to residuals, a correction is needed to account for the use of  $\hat{\phi}$  in place of the unknown  $\phi$ . Some software makes this correction automatically. In R the correction is not automatic but is done by setting the `fitdf` parameter in `Box.test` to the number of parameters that were estimated, so for an AR(1) model `fitdf` should be 1.



**Fig. 9.6.** Sample ACF of BMW log returns.

#### Example 9.4. BMW log returns—ACF plots and AR fit

Figure 9.6 is a sample ACF plot of the BMW log returns in the `bmw` data set in R's `evir` package. The autocorrelation coefficient at lag 1 is well outside the test bounds, so the series has some dependence. Also, the Ljung–Box test that the first `df` autocorrelations are 0 was performed using R's `Box.test` function. The parameter `df` specifies the number of autocorrelation coefficients to test was set equal to 5, though other choices give similar results. The output was

Box-Ljung test

```
data: bmw
X-squared = 44.987, df = 5, p-value = 1.460e-08
```

The  $p$ -value is very small, indicating that at least one of the first five autocorrelations is nonzero. Whether the amount of dependence is on any practical importance is debatable, but an AR(1) model to model the small amount of correlation might be appropriate.

Next, an AR(1) model was fit using the `arima` command in R. A summary of the results is below. The `order` parameter will be explained later, but for an AR(1) process it should be `c(1, 0, 0)`.

```
Call:
arima(x = bmw, order = c(1, 0, 0))

Coefficients:
ar1  intercept
0.081116  0.000340
s.e.  0.012722  0.000205

sigma^2 estimated as 0.000216260: log-likelihood = 17212.34,
aic = -34418.68
```

We see that  $\hat{\phi} = 0.081$  and  $\hat{\sigma}^2 = 0.000216$ . Although  $\hat{\phi}$  is small, it is statistically highly significant since it is 6.4 times its standard error so its  $p$ -value is near zero. As just mentioned, whether this small, but nonzero, value of  $\hat{\phi}$  is of practical significance is another matter. A positive value of  $\phi$  means that there is some information in today's return that could be used for prediction of tomorrow's return, but a small value of  $\phi$  means that the prediction will not be very accurate. The potential for profit might be negated by trading costs.

The sample ACF of the residuals is plotted in [Figure 9.7\(a\)](#). None of the autocorrelations at low lags is outside the test bounds. A few at higher lags are outside the bounds, but this type of behavior is expected to occur by chance or because, with a large sample size, very small but nonzero true correlations can be detected. The Ljung–Box test was applied, with `df` equal to 5 and `fitdf=1`, to the residuals with these results:

```
Box-Ljung test

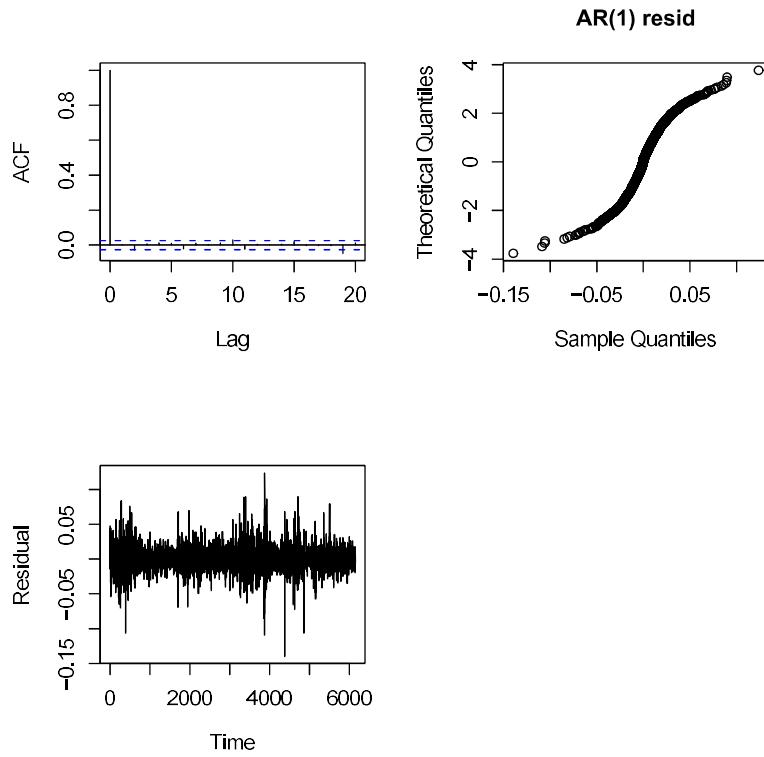
data: residuals(fitAR1)
X-squared = 6.8669, df = 5, p-value = 0.1431
```

The large  $p$ -value indicates that we should accept the null hypothesis that the residuals are uncorrelated, at least at small lags. This is a sign that the AR(1) model provides an adequate fit. However, the Ljung–Box test was repeated with `df` equal to 10, 15, and 20 and the  $p$ -values were 0.041, 0.045, and 0.040,

respectively. These values are “statistically significant” using the conventional cutoff of 0.05. The sample size is 6146, so it is not surprising that even a small amount of autocorrelation can be statistically significant. The practical significance of this autocorrelation is very doubtful.

We conclude that the AR(1) model is adequate for the BMW daily returns, but at longer lags some slight amount of autocorrelation appears to remain. However, the normal plot and time series plot of the AR(1) residuals in Figure 9.7(b) and (c) show heavy tails and volatility clustering. These are common features of economic data and will be modeled in subsequent chapters.

□



**Fig. 9.7.** ACF, normal plot, and time series plot of residuals from an AR(1) fit to the BMW log returns.

*Example 9.5. Inflation rate—AR(1) fit and checking residuals*

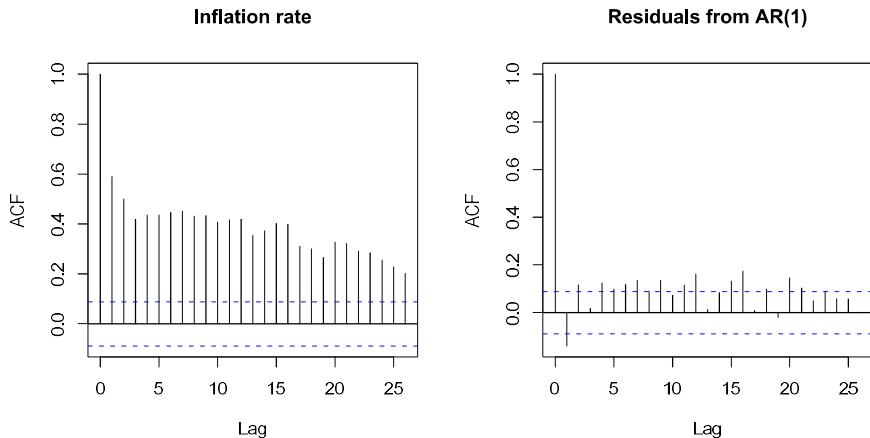
This example uses the inflation rate time series used earlier in Example 9.1. Although there is some doubt as to whether this series is stationary, we will fit an AR(1) model. The ACF of the residuals are shown in [Figure 9.8](#) and there is considerable residual autocorrelation, which indicates that the AR(1) model is not adequate. A Ljung–Box test confirms this result.

Box-Ljung test

```
data: fit$resid
X-squared = 46.1752, df = 12, p-value = 3.011e-06
```

One might try fitting an AR(1) to the changes in the inflation rate, since this series is clearly stationary. However, the AR(1) model also does not fit the changes in the inflation rate. We will return to this example when we have a larger collection of models in our statistics toolbox.

□



**Fig. 9.8.** ACF of the inflation rate time series and residuals from an AR(1) fit.

### 9.5.2 Maximum Likelihood and Conditional Least-Squares

Estimators for AR processes can be computed automatically by most statistical software packages, and the user need not know what is “under the

hood” of the software. Nonetheless, for readers interested in the estimation methodology, this section has been provided.

To find the likelihood for  $Y_1, \dots, Y_n$ , we use (A.41) and the fact that

$$f_{Y_k|Y_1, \dots, Y_{k-1}}(y_k|y_1, \dots, y_{k-1}) = f_{Y_k|Y_{k-1}}(y_k|y_{k-1}) \quad (9.16)$$

for  $k = 2, 3, \dots, n$ . A stochastic process with property (9.16) is called a *Markov process*. By (A.41) and (9.16), we have

$$f_{Y_1, \dots, Y_n}(y_1, \dots, y_n) = f_{Y_1}(y_1) \prod_{i=2}^n f_{Y_i|Y_{i-1}}(y_i|y_{i-1}). \quad (9.17)$$

By (9.7) and (9.8), we know that  $Y_1$  is  $N\{\mu, \sigma_\epsilon^2/(1-\phi^2)\}$ . Given  $Y_{i-1}$ , the only random component of  $Y_i$  is  $\epsilon_i$ , so that  $Y_i$  given  $Y_{i-1}$  is  $N\{\mu + \phi(Y_{i-1} - \mu), \sigma_\epsilon^2\}$ . It then follows that the likelihood for  $Y_1, \dots, Y_n$  is

$$\left(\frac{1}{\sqrt{2\pi}\sigma_\epsilon^n}\right) \exp\left\{-\frac{(Y_1 - \mu)^2}{2\sigma_\epsilon^2(1-\phi^2)}\right\} \prod_{i=2}^n \exp\left(-\frac{[Y_i - \{\mu + \phi(Y_{i-1} - \mu)\}]^2}{2\sigma_\epsilon^2}\right). \quad (9.18)$$

The maximum likelihood estimator maximizes the logarithm of (9.18) over  $(\mu, \phi, \sigma_\epsilon)$ . A somewhat simpler estimator deletes the marginal density of  $Y_1$  from the likelihood and maximizes the logarithm of

$$\left(\frac{1}{\sqrt{2\pi}\sigma_\epsilon^{n-1}}\right) \prod_{i=2}^n \exp\left(-\frac{[Y_i - \{\mu + \phi(Y_{i-1} - \mu)\}]^2}{2\sigma_\epsilon^2}\right). \quad (9.19)$$

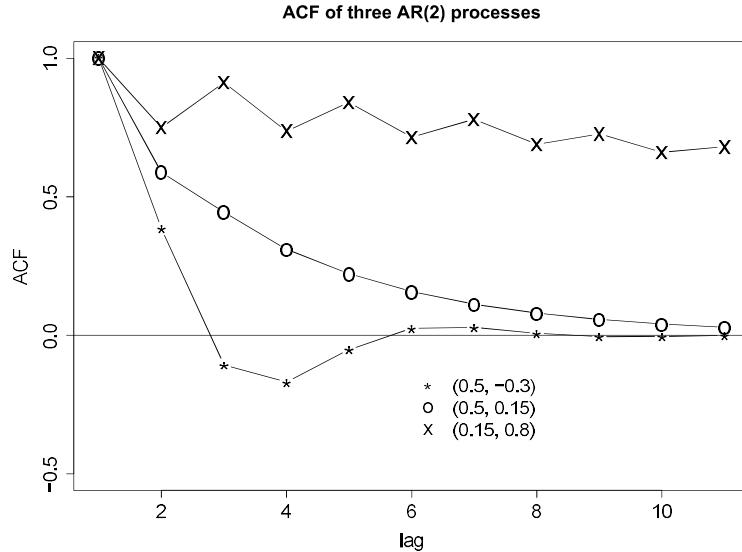
This estimator is called the conditional least-squares estimator. It is “conditional” because it uses the conditional density of  $Y_2, \dots, Y_n$  given  $Y_1$ . It is a least-squares estimator because the estimates of  $\mu$  and  $\phi$  minimize

$$\sum_{i=2}^n [Y_i - \{\mu + \phi(Y_{i-1} - \mu)\}]^2. \quad (9.20)$$

The default method for the function `arima` in R is to use the conditional least-squares estimates as starting values for maximum likelihood. The MLE is returned. The default option is used in the examples in this book.

## 9.6 AR( $p$ ) Models

We have seen that the ACF of an AR(1) process decays geometrically to zero and also alternates in sign if  $\phi < 0$ . This is a limited range of behavior and many time series do not behave in this way. To get a more flexible class of



**Fig. 9.9.** ACF of three AR(2) processes. The legend gives the values of  $\phi_1$  and  $\phi_2$ .

models, but one that still is parsimonious, we can use a model that regresses the current value of the process on several of the recent past values, not just the most recent. Thus, we let the last  $p$  values of the process,  $Y_{t-1}, \dots, Y_{t-p}$ , feed back into the current value  $Y_t$ .

Here's a formal definition. The stochastic process  $Y_t$  is an *AR( $p$ ) process* if

$$Y_t - \mu = \phi_1(Y_{t-1} - \mu) + \phi_2(Y_{t-2} - \mu) + \dots + \phi_p(Y_{t-p} - \mu) + \epsilon_t,$$

where  $\epsilon_1, \dots, \epsilon_n$  is  $\text{WN}(0, \sigma_\epsilon^2)$ .

This is a multiple linear regression<sup>5</sup> model with lagged values of the time series as the “ $x$ -variables.” The model can be reexpressed as

$$Y_t = \beta_0 + \phi_1 Y_{t-1} + \dots + \phi_p Y_{t-p} + \epsilon_t,$$

where  $\beta_0 = \{1 - (\phi_1 + \dots + \phi_p)\}\mu$ . The parameter  $\beta_0$  is called the “constant” or “intercept” as in an AR(1) model. It can be shown that  $\{1 - (\phi_1 + \dots + \phi_p)\} > 0$  for a stationary process, so  $\mu = 0$  if and only if  $\beta_0$  is zero.

Formulas for the ACFs of AR( $p$ ) processes with  $p > 1$  are more complicated than for an AR(1) process and can be found in the time series textbooks listed in Section 9.15. However, software is available for computing and plotting the

<sup>5</sup> See Chapter 12 for an introduction to multiple regression.

ACF of any AR processes, as well as for the MA and ARMA processes to be introduced soon. [Figure 9.9](#) is a plot of the ACFs of three AR(2) process. The ACFs were computed using R's `ARMAacf` function. Notice the wide variety of ACFs that are possible with two AR parameters.

Most of the concepts we have discussed for AR(1) models generalize easily to AR( $p$ ) models. The conditional least squares or maximum likelihood estimators can be calculated using software such as R's `arima` function. The residuals are defined by

$$\hat{\epsilon}_t = Y_t - \{\hat{\beta}_0 + \hat{\phi}_1 Y_{t-1} + \cdots + \hat{\phi}_{t-p} Y_{t-p}\}, \quad t \geq p+1.$$

If the AR( $p$ ) model fits the time series well, then the residuals should look like white noise. Residual autocorrelation can be detected by examining the sample ACF of the residuals and using the Ljung–Box test. Any significant residual autocorrelation is a sign that the AR( $p$ ) model does not fit well.

One problem with AR models is that they often need a rather large value of  $p$  to fit a data set. The problem is illustrated by the following two examples.

*Example 9.6. Changes in the inflation rate—AR( $p$ ) models*

[Figure 9.10](#) is a plot of AIC and BIC versus  $p$  for AR( $p$ ) fits to the changes in the inflation rate. Both criteria suggest that  $p$  should be large. AIC decreases steadily as  $p$  increases from 1 to 19, though there is a local minimum at 8. Even the conservative BIC criterion indicates that  $p$  should be as large as 6. Thus, AR models are not parsimonious for this example. The remedy is to use an MA or ARMA model, which are the next topics of the next sections.

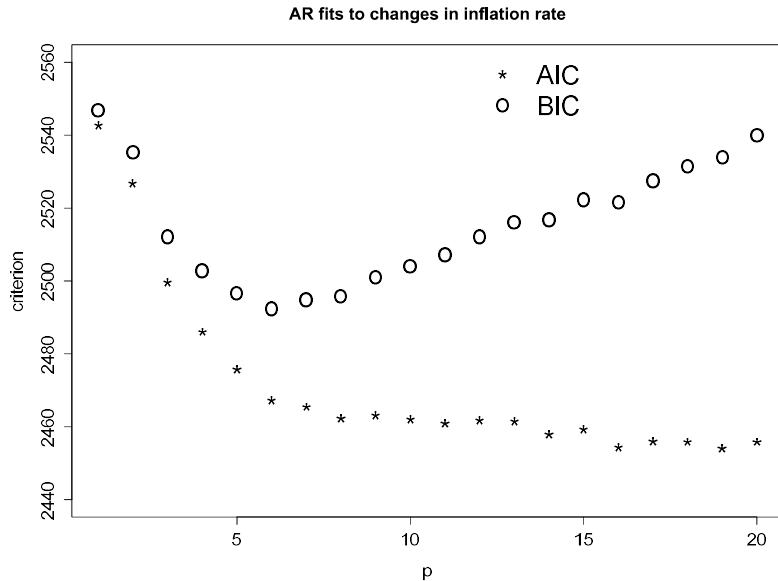
Many statistical software packages have functions to automate the search for the AR model that optimizes AIC or other criteria. The `auto.arima` function in R's `forecast` package found that  $p = 8$  is the first local minimum of AIC:

```
> auto.arima(diff(x),max.p=20,max.q=0,ic="aic")
Series: diff(x)
ARIMA(8,0,0) with zero mean

Coefficients:
      ar1      ar2      ar3      ar4      ar5
    -0.6274  -0.4977  -0.5158  -0.4155  -0.3443
  s.e.   0.0456   0.0536   0.0576   0.0606   0.0610

      ar6      ar7      ar8
    -0.2560  -0.1557  -0.1051
     0.0581   0.0543   0.0459

sigma^2 estimated as 8.539: log-likelihood = -1221.2
AIC = 2460.4   AICc = 2460.7   BIC = 2493.96
```



**Fig. 9.10.** Fitting AR( $p$ ) models to changes in the one-month inflation rate. AIC and BIC plotted against  $p$ .

The first local minimum of BIC is at 6:

```
> auto.arima(diff(x),max.p=10,max.q=0,ic="bic")
Series: diff(x)
ARIMA(6,0,0) with zero mean

Coefficients:
      ar1      ar2      ar3      ar4      ar5      ar6 
    -0.6057  -0.4554  -0.4558  -0.3345  -0.2496  -0.1481 
  s.e.   0.0454   0.0522   0.0544   0.0546   0.0526   0.0457 

sigma^2 estimated as 8.699: log-likelihood = -1225.67
AIC = 2465.33  AICc = 2465.51  BIC = 2490.5
```

We will see later that a more parsimonious fit can be obtained by going beyond AR models. □

#### Example 9.7. Inflation rates—AR( $p$ ) models

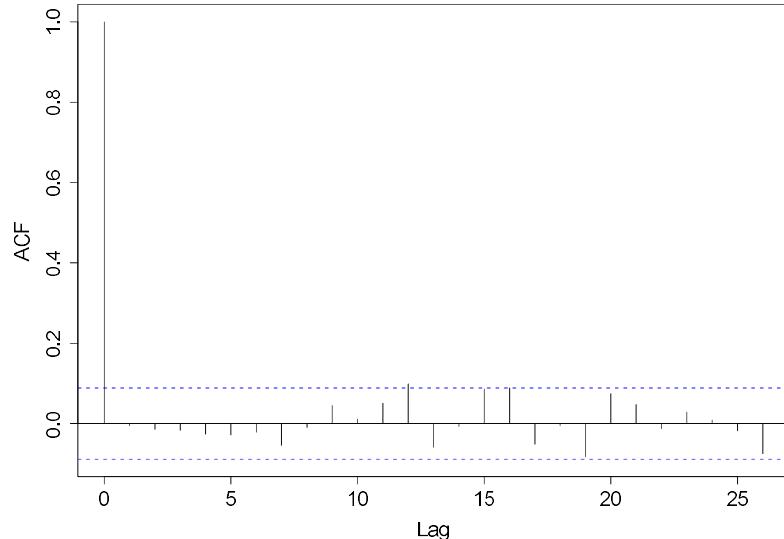
Since it is uncertain whether or not the inflation rates are stationary, one might fit an AR model to the inflation rates themselves, rather than their differences. An AR( $p$ ) models was fit to the inflation rates with  $p$  determined

automatically by `auto.arima`. The BIC criterion chose  $p = 2$  and AIC selected  $p = 7$ . Here are the results for  $p = 7$ .

```
Series: x
ARIMA(7,0,0) with non-zero mean

Coefficients:
            ar1      ar2      ar3      ar4      ar5      ar6      ar7  intercept
            0.366   0.129   -0.020   0.099   0.065   0.080   0.119       3.99
s.e.        0.045   0.048    0.048   0.048   0.049   0.048   0.046       0.78

sigma^2 estimated as 8.47:  log-likelihood = -1222
AIC = 2462    AICc = 2522    BIC = 2467
```



**Fig. 9.11.** ACF of residuals from an AR(7) fit to the inflation rates.

The ACF of the residuals is shown in Figure 9.11. □

## 9.7 Moving Average (MA) Processes

As we saw in Example 9.6, there is a potential need for large values of  $p$  when fitting AR processes. A remedy for this problem is to add a moving average component to an AR( $p$ ) process. The result is an *autoregressive-moving*

*average process*, often called an *ARMA process*. Before introducing ARMA processes, we start with pure moving average (MA) processes.

### 9.7.1 MA(1) Processes

The idea behind AR processes is to feed past data back into the current value of the process. This induces correlation between the past and present. The effect is to have at least some correlation at *all* lags. Sometimes data show correlation at only short lags, for example, only at lag 1 or only at lags 1 and 2. See, for example, [Figure 9.3\(b\)](#) where the sample ACF of changes in the inflation rate is approximately  $-0.4$  at lag 1, but then is approximately  $0.1$  or less in magnitude after one lag. AR processes do not behave this way and, as already seen in Example 9.6, do not provide a parsimonious fit. In such situations, a useful alternative to an AR model is a moving average (MA) model. A process  $Y_t$  is a *moving average process* if  $Y_t$  can be expressed as a weighted average (moving average) of the past values of the white noise process  $\epsilon_t$ .

The **MA(1)** (moving average of order 1) process is

$$Y_t - \mu = \epsilon_t + \theta\epsilon_{t-1}, \quad (9.21)$$

where as before the  $\epsilon_t$  are  $\text{WN}(0, \sigma_\epsilon^2)$ .<sup>6</sup>

One can show that

$$\begin{aligned} E(Y_t) &= \mu, \\ \text{Var}(Y_t) &= \sigma_\epsilon^2(1 + \theta^2), \\ \gamma(1) &= \theta\sigma_\epsilon^2, \\ \gamma(h) &= 0 \text{ if } |h| > 1, \\ \rho(1) &= \frac{\theta}{1 + \theta^2}, \end{aligned} \quad (9.22)$$

$$\rho(h) = 0 \text{ if } |h| > 1. \quad (9.23)$$

Notice the implication of (9.22) and (9.23)—an MA(1) model has zero correlation at all lags except lag 1 (and of course lag 0). It is relatively easy to derive these formulas and this is left as an exercise for the reader.

### 9.7.2 General MA Processes

The **MA( $q$ )** process is

$$Y_t = \mu + \epsilon_t + \theta_1\epsilon_{t-1} + \cdots + \theta_q\epsilon_{t-q}. \quad (9.24)$$

---

<sup>6</sup> Some textbooks and some software write MA models with the signs reversed so that model (9.21) is written as  $Y_t - \mu = \epsilon_t - \theta\epsilon_{t-1}$ . We have adopted the same form of MA models as R's `arima` function. These remarks apply as well to the general MA and ARMA models given by equations (9.24) and (9.25).

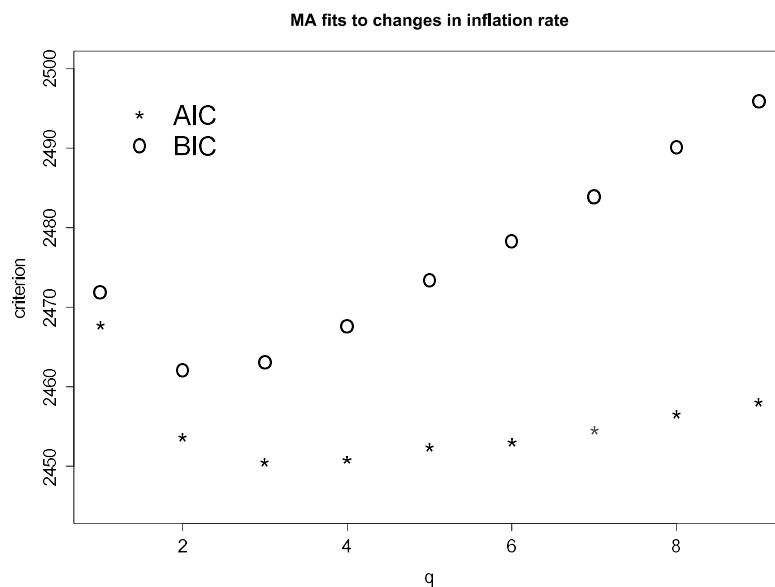
One can show that  $\gamma(h) = 0$  and  $\rho(h) = 0$  if  $|h| > q$ . Formulas for  $\gamma(h)$  and  $\rho(h)$  when  $|h| \leq q$  are given in time series textbooks and these functions can be computed in R by the function `armaACF`.

Unlike AR( $p$ ) models where the “constant” in the model is not the same as the mean, in an MA( $q$ ) model  $\mu$ , the mean of the process, is the same as  $\beta_0$ , the “constant” in the model. This fact can be appreciated by examining the right-hand side of equation (9.24), where  $\mu$  is the “intercept” or “constant” in the model and is also the mean of  $Y_t$  because  $\epsilon_t, \dots, \epsilon_{t-q}$  have mean zero.

MA( $q$ ) models can be fit easily using, for example, the `arima` function in R.

*Example 9.8. Changes in the inflation rate—MA models*

MA( $q$ ) models were fit to the changes in the inflation rate. Figure 9.12 shows plots of AIC and BIC versus  $q$ . BIC suggests that an MA(2) model is adequate, while AIC suggests an MA(3) model. We fit the MA(3) model. The Ljung–Box test was applied to the residuals with df equal to 5, 10, 15, and 20 and gave  $p$ -values of 0.97, 0.93, 0.54, and 0.15, respectively. The MA(2) also provided an adequate fit with the  $p$ -values from the Ljung–Box test all above 0.07. The output for the MA(3) model was



**Fig. 9.12.** Fitting MA( $q$ ) models to changes in the one-month inflation rate. AIC and BIC plotted against  $q$ .

```

Call:
arima(x = diff(x), order = c(0, 0, 3))

Coefficients:
      ma1        ma2        ma3  intercept
-0.632950 -0.102734 -0.108172 -0.000156
  s.e.   0.046017  0.051399  0.046985  0.020892

```

Thus, if an MA model is used, then only two or three MA parameters are needed. This is a strong contrast with AR models, which require far more parameters, perhaps as many as six.

## 9.8 ARMA Processes

Stationary time series with complex autocorrelation behavior often are more parsimoniously modeled by mixed autoregressive and moving average (ARMA) processes than by either a pure AR or pure MA process. For example, it is sometimes the case that a model with one AR and one MA parameter, called an ARMA(1,1) model, will provide a more parsimonious fit than a pure AR or pure MA model. This section introduces ARMA processes.

### 9.8.1 The Backwards Operator

The *backwards operator*  $B$  is a simple notation with a fancy name. It is useful for describing ARMA and ARIMA models. The backwards operator is defined by

$$B Y_t = Y_{t-1}$$

and, more generally,

$$B^k Y_t = Y_{t-k}.$$

Thus,  $B$  backs up time one unit while  $B^k$  does this repeatedly so that time is backed up  $k$  time units. Note that  $B c = c$  for any constant  $c$ , since a constant does not change with time. The backwards operator is sometimes called the *lag operator*.

### 9.8.2 The ARMA Model

An  $ARMA(p, q)$  model combines both AR and MA terms and is defined by the equation

$$(Y_t - \mu) = \phi_1(Y_{t-1} - \mu) + \cdots + \phi_p(Y_{t-p} - \mu) + \epsilon_t + \theta_1\epsilon_{t-1} + \cdots + \theta_q\epsilon_{t-q}, \quad (9.25)$$

which shows how  $Y_t$  depends on lagged values of itself and lagged values of the white noise process. Equation (9.25) can be written more succinctly with the backwards operator as

$$(1 - \phi_1 B - \cdots - \phi_p B^p)(Y_t - \mu) = (1 + \theta_1 B + \cdots + \theta_q B^q)\epsilon_t. \quad (9.26)$$

A white noise process is ARMA(0,0) since if  $p = q = 0$ , then (9.26) reduces to

$$(Y_t - \mu) = \epsilon_t.$$

### 9.8.3 ARMA(1,1) Processes

The ARMA(1,1) model is commonly used in practice and is simple enough to study theoretically. In the section, formulas for its variance and ACF will be derived. Without loss of generality, one can assume that  $\mu = 0$  when computing the variance and ACF. Multiplying the model

$$Y_t = \phi Y_{t-1} + \theta \epsilon_{t-1} + \epsilon_t \quad (9.27)$$

by  $\epsilon_t$  and taking expectations, one has

$$\text{Cov}(Y_t, \epsilon_t) = E(Y_t \epsilon_t) = \sigma_\epsilon^2, \quad (9.28)$$

since  $\epsilon_t$  is independent of  $\epsilon_{t-1}$  and  $Y_{t-1}$ . From (9.27) and (9.28),

$$\gamma(0) = \phi^2 \gamma(0) + (1 + \theta^2) \sigma_\epsilon^2 + 2\phi\theta\sigma_\epsilon^2, \quad (9.29)$$

and then solving (9.29) for  $\gamma(0)$  gives us the formula

$$\gamma(0) = \frac{(1 + \theta^2 + 2\phi\theta)\sigma_\epsilon^2}{1 - \phi^2}. \quad (9.30)$$

By similar calculations, multiplying (9.27) by  $Y_{t-1}$  and taking expectations yields a formula for  $\gamma(1)$ . Dividing this formula by the right-hand side of (9.29) gives us

$$\rho(1) = \frac{(1 + \phi\theta)(\phi + \theta)}{1 + \theta^2 + 2\phi\theta}. \quad (9.31)$$

For  $k \geq 2$ , multiplying (9.27) by  $Y_{t-k}$  and taking expectations results in the formula

$$\rho(k) = \phi\rho(k-1), \quad k \geq 2. \quad (9.32)$$

By (9.32), after one lag the ACF of an ARMA(1,1) process decays in the same way as the ACF of an AR(1) process with the same  $\phi$ .

**Table 9.1.** *AIC and BIC for ARMA models fit to the monthly changes in the risk-free interest returns. The minimum values of both criteria are shown in boldface. To improve the appearance of the table, 1290 was added to all AIC and BIC values.*

<i>p</i>	<i>q</i>	AIC	BIC
0	0	29.45	37.8
0	1	9.21	21.8
0	2	3.00	19.8
1	0	14.86	27.5
1	1	<b>2.67</b>	<b>19.5</b>
1	2	4.67	25.7
2	0	5.61	22.4
2	1	6.98	28.0
2	2	4.89	30.1

#### 9.8.4 Estimation of ARMA Parameters

The parameters of ARMA models can be estimated by maximum likelihood or conditional least-squares. These methods were introduced for AR(1) processes in Section 9.5. The estimation methods for AR(*p*) models are very similar to those for AR(1) models. For MA and ARMA, because the noise terms  $\epsilon_1, \dots, \epsilon_n$  are unobserved, there are complications that are best left for advanced time series texts.

#### Example 9.9. Changes in risk-free returns: ARMA models

This example uses the monthly changes in the risk-free returns shown in Figure 4.3. In [Table 9.1](#) AIC and BIC are shown for ARMA models with  $p, q = 0, 1, 2$ . We see that AIC and BIC are both minimized by the ARMA(1,1) model, though the MA(2) model is a very close second. The ARMA(1,1) and MA(2) fit nearly equally well, and it is difficult to decide between them.

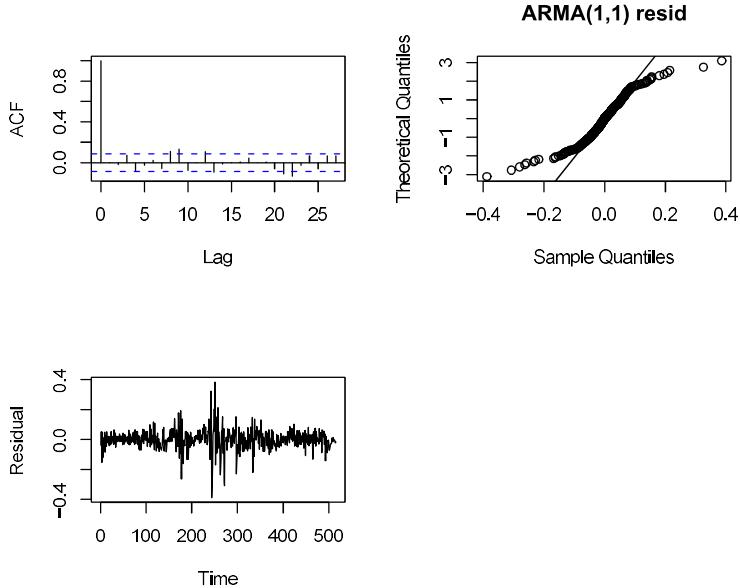
ACF, normal, and time series plots of the residuals from the ARMA(1,1) model are shown in [Figure 9.13](#). The ACF plot shows no short-term autocorrelation, which is another sign that the ARMA(1,1) model is satisfactory. However, the normal plot shows heavy tails and the time series plot shows volatility clustering. These problems will be addressed in later chapters.

□

#### 9.8.5 The Differencing Operator

The *differencing operator* is another useful notation and is defined as  $\Delta = 1 - B$ , where  $B$  is the backwards operator, so that

$$\Delta Y_t = Y_t - B Y_t = Y_t - Y_{t-1}.$$



**Fig. 9.13.** Residual plots for the ARMA(1,1) fit to the monthly changes in the risk-free returns.

For example, if  $p_t = \log(P_t)$  is the log price, then the log return is

$$r_t = \Delta p_t.$$

Differencing can be iterated. For example,

$$\begin{aligned} \Delta^2 Y_t &= \Delta(\Delta Y_t) = \Delta(Y_t - Y_{t-1}) = (Y_t - Y_{t-1}) - (Y_{t-1} - Y_{t-2}) \\ &= Y_t - 2Y_{t-1} + Y_{t-2}. \end{aligned}$$

$\Delta^k$  is called the  $k$ th-order differencing operator.

A general formula for  $\Delta^k$  can be derived from a binomial expansion:

$$\Delta^k Y_t = (1 - B)^k Y_t = \sum_{\ell=0}^k \binom{k}{\ell} (-1)^\ell Y_{t-\ell}. \quad (9.33)$$

## 9.9 ARIMA Processes

Often the first or perhaps second differences of nonstationary time series are stationary. For example, the first differences of a random walk (nonstationary) are white noise (stationary). In the section, *autoregressive integrated moving*

*average* (ARIMA) processes are introduced. They include stationary as well as nonstationary processes.

A time series  $Y_t$  is said to be an  $ARIMA(p, d, q)$  process if  $\Delta^d Y_t$  is ARMA( $p, q$ ). For example, if log returns on an asset are ARMA( $p, q$ ), then the log prices are ARIMA( $p, 1, q$ ). An ARIMA( $p, d, q$ ) is stationary only if  $d = 0$ . Otherwise, only its differences of order  $d$  or above are stationary.

Notice that an ARIMA( $p, 0, q$ ) model is the same as an ARMA( $p, q$ ) model. ARIMA( $p, 0, 0$ ), ARMA( $p, 0$ ), and AR( $p$ ) models are the same. Similarly, ARIMA( $0, 0, q$ ), ARMA( $0, q$ ), and MA( $q$ ) models are the same. A random walk is an ARIMA( $0, 1, 0$ ) model.

The inverse of differencing is “integrating.” The integral of a process  $Y_t$  is the process  $w_t$ , where

$$w_t = w_{t_0} + Y_{t_0} + Y_{t_0+1} + \cdots + Y_t. \quad (9.34)$$

Here  $t_0$  is an arbitrary starting time point and  $w_{t_0}$  is the starting value of the  $w_t$  process. It is easy to check that

$$\Delta w_t = Y_t, \quad (9.35)$$

so integrating and differencing are inverse processes.<sup>7</sup>

We will say that a process is I( $d$ ) if it is stationary after being differenced  $d$  times. For example, a stationary process is I(0). An ARIMA( $p, d, q$ ) process is I( $d$ ). An I( $d$ ) process is said to be “integrated to order  $d$ .”

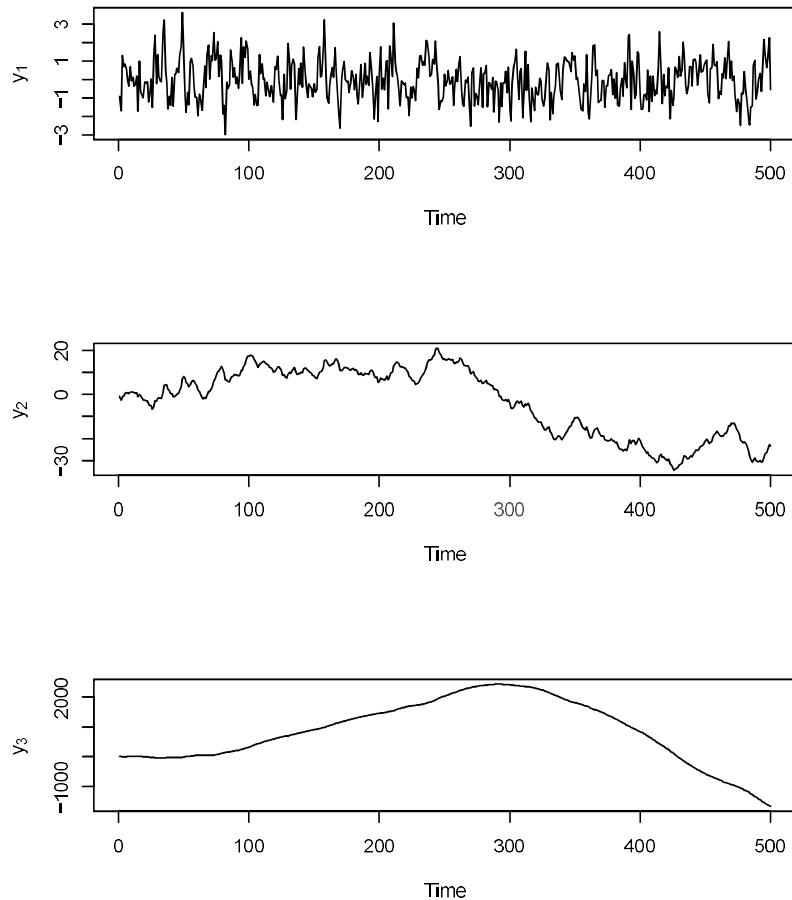
[Figure 9.14](#) shows an AR(1) process, its integral, and its second integral, meaning the integral of its integral. These three processes are I(0), I(1), and I(2), respectively. The three processes behave in entirely different ways. The AR(1) process is stationary and varies randomly about its mean, which is 0; one says that the process *reverts* to its mean. The integral of this process behaves much like a random walk in having no fixed level to which it reverts. The second integral has *momentum*. Once the process starts moving upward or downward, it tends to continue in that direction. If data show momentum like this, then the momentum is an indication that  $d = 2$ . The AR(1) process was generated by the R function `arima.sim`. This process was integrated twice with R’s `cumsum` function.

#### *Example 9.10. Fitting an ARIMA model to CPI data*

This example uses the `CPI.dat` data set in R’s `fEcofin` package. CPI is a seasonally adjusted U.S. Consumer Price Index. The data are monthly. Only data from January 1977 to December 1987 are used in this example. [Figure 9.15](#) shows time series plots of  $\log(\text{CPI})$  and the first and second differences of this series. The original series shows the type of momentum that

---

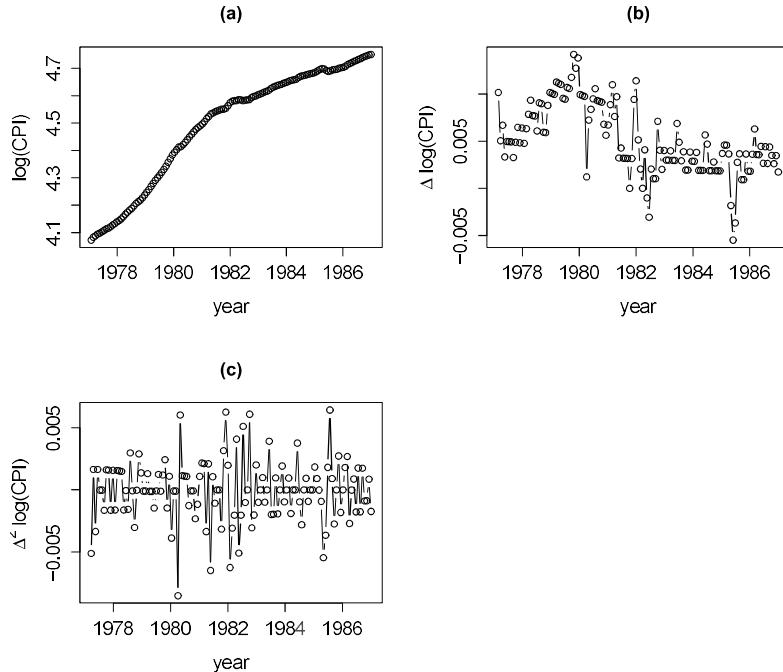
<sup>7</sup> An analog is, of course, differentiation and integration in calculus, which are inverses of each other.



**Fig. 9.14.** The top plot is of an AR(1) process with  $\mu = 0$  and  $\phi = 0.4$ . The middle and bottom plots are, respectively, the integral and second integral of this AR(1) process. Thus, from top to bottom, the series are I(0), I(1), and I(2), respectively.

is characteristic of an I(2) series. The first differences show no momentum, but they do not appear to be mean-reverting and so they may be I(1). The second differences appear to be mean-reverting and therefore seem to be I(0). ACF plots in Figures 9.16(a), (b), and (c) provide additional evidence that the log(CPI) is I(2).

Notice that the ACF of  $\Delta^2 \log(\text{CPI})$  has large correlations at the first two lags and then small autocorrelations after that. This suggests using an MA(2) for  $\Delta^2 \log(\text{CPI})$  or, equivalently, an ARIMA(0,2,2) model for  $\log(\text{CPI})$ . The

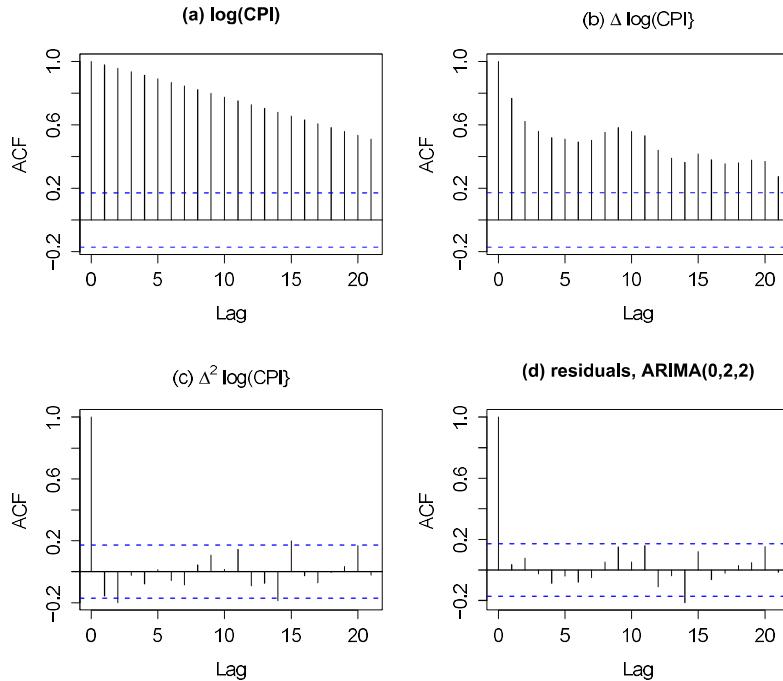


**Fig. 9.15.** (a)  $\log(\text{CPI})$ . (b) First differences of  $\log(\text{CPI})$ . (c) Second differences of  $\log(\text{CPI})$ .

ACF of the residuals from this fit is shown in Figure 9.16(d). The residual ACF has small correlations at short lags, which is an indication that the ARIMA(0,2,2) model fits well. Also, the residuals pass Ljung–Box tests for various choices of  $K$ , for example, with a  $p$ -value of 0.17 at  $K = 15$ .  $\square$

#### Example 9.11. Fitting an ARIMA model to industrial production (IP) data

This example uses the `IP.dat` data set in R's `fEcofin` package. The variable, `IP`, is a seasonally adjusted U.S. industrial production index. Figure 9.17 panels (a) and (b) show time series plots of `IP` and  $\Delta \text{IP}$  and panel (c) has the ACF of  $\Delta \text{IP}$ . `IP` appears to be I(1), implying that we should fit an ARMA model to  $\Delta \text{IP}$ . AR(1), AR(2), and ARMA(1,1) each fit  $\Delta \text{IP}$  reasonably well and the ARMA(1,1) model is selected using the BIC criterion with R's `auto.arima` function. The ACF of the residuals in Figure 9.17(d) indicates a satisfactory fit to the ARMA(1,1) model since it shows virtually no short-term autocorrelation. In summary, `IP` is well fit by an ARIMA(1,1,1) model.  $\square$



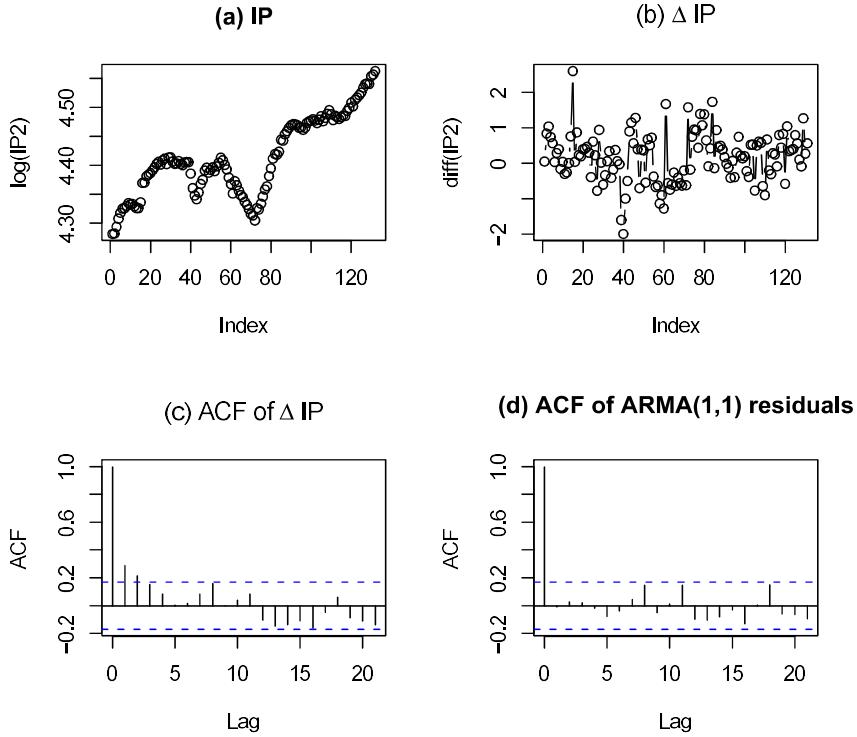
**Fig. 9.16.** ACF of (a)  $\log(\text{CPI})$ , (b) first differences of  $\log(\text{CPI})$ , (c) second differences of  $\log(\text{CPI})$ , and (d) residuals from an ARIMA(0,2,2) model fit to  $\log(\text{CPI})$ .

### 9.9.1 Drifts in ARIMA Processes

If a nonstationary process has a constant mean, then the first differences of this process have mean zero. For this reason, it is often assumed that a differenced process has mean zero. The `arima` function in R makes this assumption.

Instead of a constant mean, sometimes a nonstationary process has a mean with a deterministic linear trend, e.g.,  $E(Y_t) = \beta_0 + \beta_1 t$ . Then,  $\beta_1$  is called the *drift* of  $Y_t$ . Note that  $E(\Delta Y_t) = \beta_1$ , so if  $Y_t$  has a nonzero drift then  $\Delta Y_t$  has a nonzero mean. The R function `auto.arima` discussed in Section 9.11 allows a differenced process to have a nonzero mean, which is called the *drift* in the output.

These ideas can be extended to higher-degree polynomial trends and higher-order differencing. If  $E(Y_t)$  has an  $m$ th-degree polynomial trend, then the mean of  $E(\Delta^2 Y_t)$  has an  $(m - d)$ th-degree trend for  $d \leq m$ . For  $d > m$ ,  $E(\Delta^2 Y_t) = 0$ .



**Fig. 9.17.** (a) Time series plot of  $\text{IP}$ , (b) time series plot of  $\Delta \text{IP}$ , (c) ACF plot of  $\Delta \text{IP}$ , (d) ACF of residual from ARMA(1,1) fit to  $\Delta \text{IP}$ .

## 9.10 Unit Root Tests

We have seen that it can be difficult to tell whether a time series is best modeled as stationary or nonstationary. To help decide between these two possibilities, it can be helpful to use hypothesis testing.

What is meant by a unit root? Recall that an ARMA( $p, q$ ) process can be written as

$$(Y_t - \mu) = \phi_1(Y_{t-1} - \mu) + \cdots + \phi_p(Y_{t-p} - \mu) + \epsilon_t + \theta_1\epsilon_{t-1} + \cdots + \theta_q\epsilon_{t-q}. \quad (9.36)$$

The condition for  $\{Y_t\}$  to be stationary is that all roots of the polynomial

$$1 - \phi_1 x - \cdots - \phi_k x^k \quad (9.37)$$

have absolute values greater than one. (See Section A.21 for information about complex roots of polynomials and the absolute value of a complex number.) For example, when  $p = 1$ , then (9.37) is

$$1 - \phi x$$

and has one root,  $\phi^{-1}$ . We know that the process is stationary if  $|\phi| < 1$ , which, of course, is equivalent to  $|1/\phi| > 1$ .

If there is a unit root, that is, a root with an absolute value equal to 1, then the ARMA process is nonstationary and behaves much like a random walk. Not surprisingly, this is called the unit root case. The explosive case is when a root has an absolute value less than 1.

*Example 9.12. Inflation rates*

Recall from Examples 9.1 and 9.3 that we have had difficulty deciding whether the inflation rates are stationary or not.

If we fit stationary ARMA models to the inflation rates, then `auto.arima` selects an ARMA(2,1) model and the AR coefficients are  $\hat{\phi}_1 = 1.2074$  and  $\hat{\phi}_2 = -0.2237$ . The roots of

$$1 - \hat{\phi}_1 x - \hat{\phi}_2 x^2$$

can be found easily using R's `polyroot` function and are 1.022 and 4.377. Both roots have absolute values greater than 1, indicating possible stationarity, but the first is very close to 1 and since the roots are estimated with error, there is reason to suspect that this series may be nonstationary.  $\square$

Unit root tests are used to decide if an AR model has an absolute root equal to 1. One popular unit root test is the augmented Dickey–Fuller test, often called the ADF test. The null hypothesis is that there is a unit root. The usual alternative is that the process is stationary but one can instead use the alternative that the process is explosive.

Another unit root test is the Phillips–Perron test. It is similar to the Dickey–Fuller test but differs in some details.

A third test is the KPSS test. The null hypothesis for the KPSS test is stationarity and the alternative is a unit root, just the opposite of the hypotheses for the Dickey–Fuller and Phillips–Perron tests.

*Example 9.13. Inflation rates—Unit root tests*

Recall that we were undecided as to whether or not the inflation rate time series was stationary. The unit root tests might help resolve this issue, but unfortunately they do not provide unequivocal evidence in favor of stationarity. Both the augmented Dickey–Fuller and Phillips–Perron tests, which were implemented in R with the functions `adf.test` and `pp.test`, respectively, have small  $p$ -values, 0.016 for the former and less than 0.01 for the latter; see the

output below. The functions `pp.test`, `adf.test`, and `kpss.test` (used below) are in R's `tseries` package. Therefore, at level 0.05 the null hypothesis of a unit root is rejected by both tests in favor of the alternative of stationarity, the default alternative hypothesis for both `adf.test` and `pp.test`.

```
> adf.test(x)

Augmented Dickey--Fuller Test

data: x
Dickey-Fuller = -3.87, Lag order = 7, p-value = 0.01576
alternative hypothesis: stationary

> pp.test(x)

Phillips-Perron Unit Root Test

data: x
Dickey-Fuller Z(alpha) = -249, Truncation lag parameter = 5,
p-value = 0.01
alternative hypothesis: stationary

Warning message:
In pp.test(x) : p-value smaller than printed p-value
```

Although the augmented Dickey–Fuller and Phillips–Perron tests suggest that the inflation rate series is stationary since the null hypothesis of a unit root is rejected, the KPSS test leads one to the opposite conclusion. The null hypothesis for the KPSS is stationarity and it is rejected with a *p*-value smaller than 0.01. Here is the R output.

```
> kpss.test(x)

KPSS Test for Level Stationarity

data: x
KPSS Level = 2.51, Truncation lag parameter = 5, p-value = 0.01

Warning message:
In kpss.test(x) : p-value smaller than printed p-value
```

Thus, the unit root tests are somewhat contradictory. Perhaps the inflation rates are stationary with long-term memory. Long-memory processes will be introduced in Section 10.4.

□

### 9.10.1 How Do Unit Root Tests Work?

A full discussion of the theory behind unit root tests is beyond the scope of this book. Here, only the basic idea will be mentioned. See Section 9.14 for

more information. The Dickey–Fuller test is based on the AR(1) model

$$Y_t = \phi Y_{t-1} + \epsilon_t. \quad (9.38)$$

The null hypothesis ( $H_0$ ) is that there is a unit root, that is,  $\phi = 1$ , and the alternative ( $H_1$ ) is stationarity, which is equivalent to  $\phi < 1$ , assuming, as seems reasonable, that  $\phi > -1$ . Model (9.38) is equivalent to  $\Delta Y_t = (\phi - 1)Y_{t-1} + \epsilon_t$ , or

$$\Delta Y_t = \pi Y_{t-1} + \epsilon_t, \quad (9.39)$$

where  $\pi = \phi - 1$ . Stated in terms of  $\pi$ ,  $H_0$  is  $\pi = 0$  and  $H_1$  is  $\pi < 0$ . The Dickey–Fuller test regresses  $\Delta Y_t$  on  $Y_{t-1}$  and tests  $H_0$ . Because  $Y_{t-1}$  is nonstationary under  $H_0$ , the  $t$ -statistic for  $\pi$  has a nonstandard distribution so special tables need to be developed in order to compute  $p$ -values.

The augmented Dickey–Fuller test expands model (9.39) by adding a time trend and lagged values of  $\Delta Y_t$ . Typically, the time trend is linear so that the expanded model is

$$\Delta Y_t = \beta_0 + \beta_1 t + \pi Y_{t-1} + \sum_{j=1}^p \gamma_j \Delta Y_{t-j} + \epsilon_t. \quad (9.40)$$

The hypotheses are still  $H_0: \pi = 0$  and  $H_1: \pi < 0$ . There are several methods for selecting  $p$ . The `adf.test` function has a default value of  $p$  equal to `trunc((length(x)-1)^(1/3))`, where `x` is the input series ( $Y_t$  in our notation).

## 9.11 Automatic Selection of an ARIMA Model

It is useful to have an automatic method for selecting an ARIMA model. As always, an automatically selected model should not be accepted blindly, but it makes sense to start model selection with something chosen quickly and by objective criterion.

The R function `auto.arima` can select all three parameters,  $p$ ,  $d$ , and  $q$ , for an ARIMA model. The differencing parameter  $d$  is selected using the KPSS test. If the null hypothesis of stationarity is accepted when the KPSS is applied to the original time series, then  $d = 0$ . Otherwise, the series is differenced until the KPSS accepts the null hypothesis. After that,  $p$  and  $q$  are selected using either AIC or BIC.

*Example 9.14. Inflation rates—Automatic selection of an ARIMA model*

In this example, `auto.arima` is applied to the inflation rates. The ARIMA (1,1,1) model is selected by `auto.arima` using either AIC or BIC to select  $p$  and  $q$  after  $d = 1$  is selected by the KPSS test.

```

Series: x
ARIMA(1,1,1)

Coefficients:
ar1      ma1
0.238   -0.877
s.e.    0.055   0.027

sigma^2 estimated as 8.55:  log-likelihood = -1222
AIC = 2449  AICc = 2449  BIC = 2462

```

This is a very parsimonious model and residual diagnostics (not shown) show that it fits well.

$AIC_c$  in `auto.arima`'s output is the value of the corrected AIC criterion defined by (5.34). The sample size is 491 so, not surprisingly, corrected AIC is equal to AIC, at least after rounding to the nearest integer.  $\square$

## 9.12 Forecasting

Forecasting means predicting future values of a time series using the current *information set*, which is the set of present and past values of the time series. In some contexts, the information set could include other variables related to the time series, but in this section the information set contains only the past and present values of the time series that is being predicted.

ARIMA models are often used for forecasting. Consider forecasting using an AR(1) process. Suppose that we have data  $Y_1, \dots, Y_n$  and estimates  $\hat{\mu}$  and  $\hat{\phi}$ . We know that

$$Y_{n+1} = \mu + \phi(Y_n - \mu) + \epsilon_{n+1}. \quad (9.41)$$

Since  $\epsilon_{n+1}$  is independent of the past and present, by Result 14.10.1 in Section 14.10.2 the best predictor of  $\epsilon_{n+1}$  is its expected value, which is 0. We know, of course, that  $\epsilon_{n+1}$  is not 0, but 0 is our best guess at its value. On the other hand, we know or have estimates of all other quantities in (9.41). Therefore, we predict  $Y_{n+1}$  by

$$\hat{Y}_{n+1} = \hat{\mu} + \hat{\phi}(Y_n - \hat{\mu}).$$

By the same reasoning we forecast  $Y_{n+2}$  by

$$\hat{Y}_{n+2} = \hat{\mu} + \hat{\phi}(\hat{Y}_{n+1} - \hat{\mu}) = \hat{\mu} + \hat{\phi}\{\hat{\phi}(Y_n - \hat{\mu})\}, \quad (9.42)$$

and so forth. Notice that in (9.42) we do not use  $Y_{n+1}$ , which is unknown at time  $n$ , but rather the forecast  $\hat{Y}_{n+1}$ . Continuing in this way, we find the general formula for the  $k$ -step-ahead forecast:

$$\hat{Y}_{n+k} = \hat{\mu} + \hat{\phi}^k(Y_n - \hat{\mu}). \quad (9.43)$$

If  $|\hat{\phi}| < 1$ , as is true for a stationary series, then as  $k$  increases, the forecasts will converge exponentially fast to  $\hat{\mu}$ .

Formula (9.43) is valid only for AR(1) processes, but forecasting other AR( $p$ ) processes is similar. For an AR(2) process,

$$\hat{Y}_{n+1} = \hat{\mu} + \hat{\phi}_1(Y_n - \hat{\mu}) + \hat{\phi}_2(Y_{n-1} - \hat{\mu})$$

and

$$\hat{Y}_{n+2} = \hat{\mu} + \hat{\phi}_1(\hat{Y}_{n+1} - \hat{\mu}) + \hat{\phi}_2(Y_n - \hat{\mu}),$$

and so on.

Forecasting ARMA and ARIMA processes is similar to forecasting AR processes. Consider the MA(1) process,  $Y_t - \mu = \epsilon_t - \theta\epsilon_{t-1}$ . Then the next observation will be

$$Y_{n+1} = \mu + \epsilon_{n+1} - \theta\epsilon_n. \quad (9.44)$$

In the right-hand side of (9.44) we replace  $\mu$  and  $\theta$  by estimates and  $\epsilon_n$  by the residual  $\hat{\epsilon}_n$ . Also, since  $\epsilon_{n+1}$  is independent of the observed data, it is replaced by its mean 0. Then the forecast is

$$\hat{Y}_{n+1} = \hat{\mu} - \hat{\theta}\hat{\epsilon}_n.$$

The two-step-ahead forecast of  $Y_{n+2} = \mu + \epsilon_{n+2} - \theta\epsilon_{n+1}$  is simply  $\hat{Y}_{n+2} = \hat{\mu}$ , since  $\epsilon_{n+1}$  and  $\epsilon_{n+2}$  are independent of the observed data. Similarly,  $\hat{Y}_{n+k} = \hat{\mu}$  for all  $k > 2$ .

To forecast the ARMA(1,1) process

$$Y_t - \mu = \phi(Y_{t-1} - \mu) + \epsilon_t - \theta\epsilon_{t-1},$$

we use

$$\hat{Y}_{n+1} = \hat{\mu} + \hat{\phi}(Y_n - \hat{\mu}) - \hat{\theta}\hat{\epsilon}_n$$

as the one-step-ahead forecast and

$$\hat{Y}_{n+k} = \hat{\mu} + \hat{\phi}(\hat{Y}_{n+k-1} - \hat{\mu}), k \geq 2$$

for forecasting two or more steps ahead.

As a final example, suppose that  $Y_t$  is ARIMA(1,1,0), so that  $\Delta Y_t$  is AR(1). To forecast  $Y_{n+k}$ ,  $k \geq 1$ , one first fits an AR(1) model to the  $\Delta Y_t$  process and forecasts  $\Delta Y_{n+k}$ ,  $k \geq 1$ . Let the forecasts be denoted by  $\widehat{\Delta Y}_{n+k}$ ,  $k \geq 1$ . Then, since

$$Y_{n+1} = Y_n + \Delta Y_{n+1},$$

the forecast of  $Y_{n+1}$  is

$$\hat{Y}_{n+1} = Y_n + \widehat{\Delta Y}_{n+1},$$

and similarly

$$\hat{Y}_{n+2} = \hat{Y}_{n+1} + \widehat{\Delta Y}_{n+2} = Y_n + \widehat{\Delta Y}_{n+1} + \widehat{\Delta Y}_{n+2},$$

and so on.

Most time series software packages offer functions to automate forecasting. R's `predict` function forecasts using an “object” returned by the `arima` fitting function.

### 9.12.1 Forecast Errors and Prediction Intervals

When making forecasts, one would of course like to know the uncertainty of the predictions. To this end, one first computes the variance of the forecast error. Then a  $(1 - \alpha)100\%$  prediction interval is the forecast itself plus or minus the forecast error's standard deviation times  $z_{\alpha/2}$  (the normal upper quantile). The use of  $z_{\alpha/2}$  assume that  $\epsilon_1, \dots$  is Gaussian white noise. If the residuals are heavy-tailed, then we might be reluctant to make the Gaussian assumption. One way to avoid this assumption is discussed in Section 9.12.2.

Computation of the forecast error variance and the prediction interval is automated by modern statistical software, so we need not present general formulas for the forecast error variance. However, to gain some understanding of general principles, we will look at two special cases, one stationary and the other nonstationary.

#### Stationary AR(1) Forecast Errors

We will look first at the errors made when forecasting a stationary AR(1) process. The error in the first prediction is

$$Y_{n+1} - \hat{Y}_{n+1} = \{\mu + \phi(Y_n - \mu) + \epsilon_{n+1}\} - \{\hat{\mu} + \phi(Y_n - \hat{\mu})\} = (\mu - \hat{\mu}) + (\phi - \hat{\phi})Y_n - (\phi\mu - \hat{\phi}\hat{\mu}) + \epsilon_{n+1} \quad (9.45)$$

$$\approx \epsilon_{n+1}. \quad (9.46)$$

Here (9.45) is the exact error and (9.46) is a “large-sample” approximation. The basis for (9.46) is that as the sample size increases  $\hat{\mu} \rightarrow \mu$  and  $\hat{\phi} \rightarrow \phi$ , so the first three terms in (9.45) converge to 0 but the last term remains unchanged. The large-sample approximation simplifies formulas and helps us focus on the main components of the forecast error. Using the large-sample approximation again, so  $\hat{\mu}$  is replaced by  $\mu$  and  $\hat{\phi}$  by  $\phi$ , the error in the two-steps-ahead forecast is

$$\begin{aligned} Y_{n+2} - \hat{Y}_{n+2} &= \{\mu + \phi(Y_{n+1} - \mu) + \epsilon_{n+2}\} - \{\mu + \phi(\hat{Y}_{n+1} - \mu)\} \\ &= \phi(Y_{n+1} - \hat{Y}_{n+1}) + \epsilon_{n+1} \\ &= \phi\epsilon_{n+1} + \epsilon_{n+2}. \end{aligned} \quad (9.47)$$

Continuing in this manner, we find that the  $k$ -step-ahead forecasting error is

$$\begin{aligned} Y_{n+k} - \hat{Y}_{n+k} &\approx \{\mu + \phi(Y_{n+k-1} - \mu) + \epsilon_{n+k}\} - \{\mu + \phi(\hat{Y}_{n+k-1} - \mu)\} \\ &= \phi^{k-1}\epsilon_{n+1} + \phi^{k-2}\epsilon_{n+2} + \cdots + \phi\epsilon_{n+k-1} + \epsilon_{n+k}. \end{aligned} \quad (9.48)$$

By the formula for the sum of a finite geometric series, the variance of the right-hand side of (9.47) is

$$\begin{aligned} \left\{ \phi^{2(k-1)} + \phi^{2(k-2)} + \cdots + \phi^2 + 1 \right\} \sigma_\epsilon^2 &= \left( \frac{1 - \phi^{2k}}{1 - \phi^2} \right) \sigma_\epsilon^2 \\ &\rightarrow \frac{\sigma_\epsilon^2}{1 - \phi^2} \text{ as } k \rightarrow \infty. \quad (9.49) \end{aligned}$$

An important point here is that the variance of the forecast error does not diverge as  $k \rightarrow \infty$ , but rather the variance converges to  $\gamma(0)$ , the marginal covariance of the AR(1) process given by (9.7). This is an example of the general principle that for any stationary ARMA process, the variance of the forecast error converges to the marginal variance.

### Forecasting a Random Walk

For the random walk process,  $Y_{n+1} = \mu + Y_n + \epsilon_{n+1}$ , many of the formulas just derived for the AR(1) process still hold, but with  $\phi = 1$ . An exception is that the last result in (9.49) does not hold because the summation formula for a geometric series does not apply when  $\phi = 1$ . One result that does still hold is

$$Y_{n+k} - \hat{Y}_{n+k} = \epsilon_{n+1} + \epsilon_{n+2} + \cdots + \epsilon_{n+k-1} + \epsilon_{n+k}$$

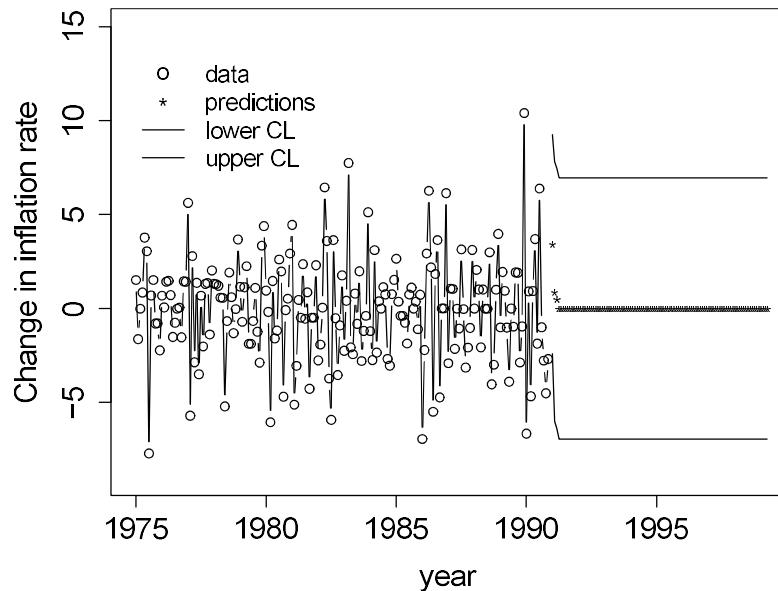
so the variance of the  $k$ -step-ahead forecast error is  $k\sigma_\epsilon^2$  and, unlike for the stationary AR(1) case, the forecast error variance diverges to  $\infty$  as  $k \rightarrow \infty$ .

### Forecasting ARIMA Processes

As mentioned before, in practice we do not need general formulas for the forecast error variance of ARIMA processes, since statistical software can compute the variance. However, it is worth repeating a general principle: For stationary ARMA processes, the variance of the  $k$ -step-ahead forecast error variance converges to a finite value as  $k \rightarrow \infty$ , but for a nonstationary ARIMA process this variance converges to  $\infty$ . The result of this principle is that for a nonstationary process, the forecast limits diverge away from each other as  $k \rightarrow \infty$ , but for a stationary process the forecast limits converge to parallel horizontal lines.

*Example 9.15. Forecasting the one-month inflation rate*

We saw in Example 9.8 that an MA(3) model provided a parsimonious fit to the changes in the one-month inflation rate. This implies that an ARIMA(0,1,3) model will be a good fit to the inflation rates themselves. The two models are, of course, equivalent, but they forecast different series. The first model gives forecasts and confidence limits for the changes in the inflation rate, while the second model provides forecasts and confidence limits for the inflation rate itself.



**Fig. 9.18.** Forecasts of changes in inflation rate.

Figures 9.18 and 9.19 plot forecasts and forecast limits from the two models out to 100 steps ahead. One can see that the forecast limits diverge for the second model and converge to parallel horizontal lines for the first model.  $\square$

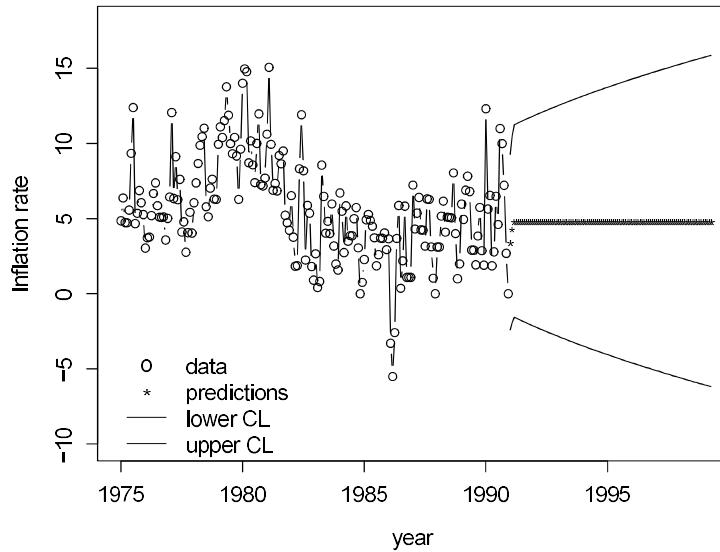
### 9.12.2 Computing Forecast Limits by Simulation

Simulation can be used to compute forecasts limits. This is done by simulating random forecasts and finding their  $\alpha/2$ -upper and -lower sample quantiles. A set of random forecasts up to  $m$  time units ahead is generated for an ARMA process by recursion:

$$\begin{aligned}\hat{Y}_{n+t} = & \hat{\mu} + \hat{\phi}_1(\hat{Y}_{n+t-1} - \hat{\mu}) + \cdots + \hat{\phi}_p(\hat{Y}_{n+t-p} - \hat{\mu}) \\ & + \hat{\epsilon}_{n+t} + \hat{\theta}_1\hat{\epsilon}_{n+t-1} + \cdots + \hat{\theta}_q\hat{\epsilon}_{n+t-q}, \quad t = 1, \dots, m,\end{aligned}\quad (9.50)$$

where

1.  $\hat{\epsilon}_k$  is the  $k$ th residual if  $k \leq n$ ,
2.  $\{\hat{\epsilon}_k : k = n+1, \dots, n+m\}$  is a resample from the residuals.



**Fig. 9.19.** Forecasts of inflation rate.

Thus,  $\hat{Y}_{n+1}$  is generated from  $Y_{n+1-p}, \dots, Y_n, \hat{\epsilon}_{n+1-q}, \dots, \hat{\epsilon}_{n+1}$ , then  $\hat{Y}_{n+2}$  is generated from  $Y_{n+2-p}, \dots, Y_n, \hat{Y}_{n+1}, \hat{\epsilon}_{n+2-q}, \dots, \hat{\epsilon}_{n+2}$ , then  $\hat{Y}_{n+3}$  is generated from  $Y_{n+3-p}, \dots, Y_n, \hat{Y}_{n+1}, \hat{Y}_{n+2}, \hat{\epsilon}_{n+3-q}, \dots, \hat{\epsilon}_{n+3}$ , and so forth.

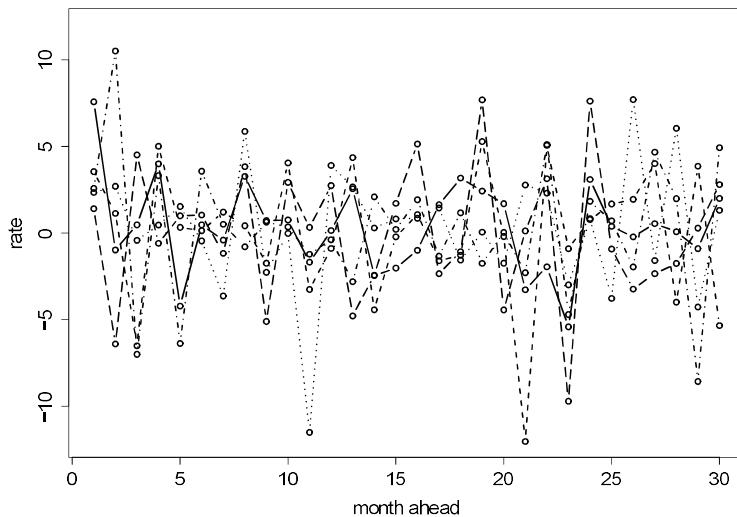
A large number, call it  $B$ , of sets of random forecasts are generated in this way. They differ because their sets of future noises generated in step 2 are mutually independent. For each  $t = 1, \dots, m$ , the  $\alpha/2$ -upper and -lower sample quantiles of the  $B$  random values of  $\hat{Y}_{n+t}$  are the forecast limits for  $Y_{n+t}$ .

To obtain forecasts, rather than forecast limits, one uses  $\hat{\epsilon}_k = 0$ ,  $k = n+1, \dots, n+m$ , in step 4. The forecasts are nonrandom, conditional given the data, and therefore need to be computed only once.

If  $Y_t = \Delta W_t$  for some nonstationary series  $\{W_1, \dots, W_n\}$ , then random forecasts of  $\{W_{n+1}, \dots\}$  can be obtained as partial sums of  $\{W_n, \hat{Y}_{n+1}, \dots\}$ . For example,

$$\begin{aligned}\hat{W}_{n+1} &= W_n + \hat{Y}_{n+1}, \\ \hat{W}_{n+2} &= \hat{W}_{n+1} + \hat{Y}_{n+2} = W_n + \hat{Y}_{n+1} + \hat{Y}_{n+2}, \\ \hat{W}_{n+3} &= \hat{W}_{n+2} + \hat{Y}_{n+3} = W_n + \hat{Y}_{n+1} + \hat{Y}_{n+2} + \hat{Y}_{n+3},\end{aligned}$$

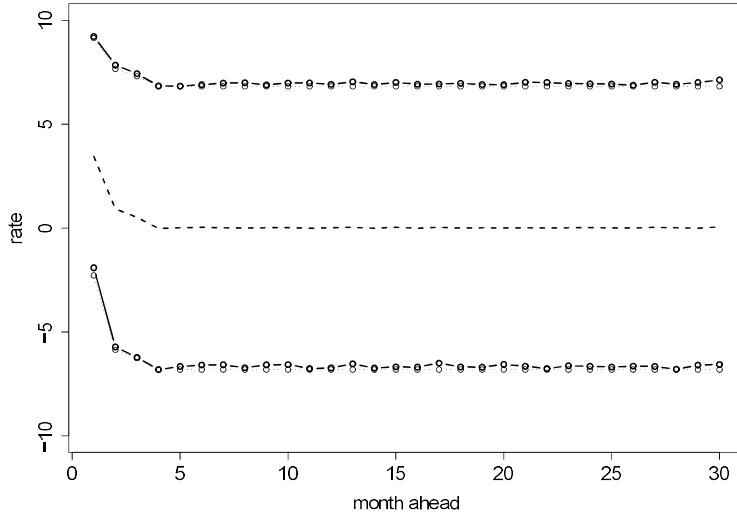
and so forth. Then, upper and lower quantiles of the randomly generated  $\widehat{W}_{n+k}$  can be used as forecast limits for  $W_{n+k}$ .



**Fig. 9.20.** Five random sets of forecasts of changes in the inflation rate computed by simulation.

*Example 9.16.* Forecasting the one-month inflation rate and changes in the inflation rate by simulation

To illustrate the amount of random variation in the forecasts, a small number (five) of sets of random forecasts of the changes in the inflation rate were generated out to 30 months ahead. These are plotted in Figure 9.20. Notice the substantial random variation between the random forecasts. Because of this large variation, to calculate forecasts limits a much larger number of random forecasts should be used. In this example,  $B = 50,000$  sets of random forecasts are generated. Figure 9.21 shows the forecast limits, which are the 2.5% upper and lower sample quantiles. For comparison, the forecast limits generated by R's function `ar` are also shown. The two sets of forecast limits are very similar even though the `ar` limits assume Gaussian noise but the residuals are heavy-tailed. Thus, the presence of heavy tails does not invalidate the Gaussian limits in this example with 95% forecast limits. If a larger confidence coefficient were used, that is, one very close to 1, then the forecast



**Fig. 9.21.** Forecast limits of changes in the inflation rate computed by simulation (solid), computed by `arima` (dotted), and the mean of the forecast (dashed). Notice that the two sets of future limits are very similar and nearly overprint each other, so they are difficult to distinguish visibly.

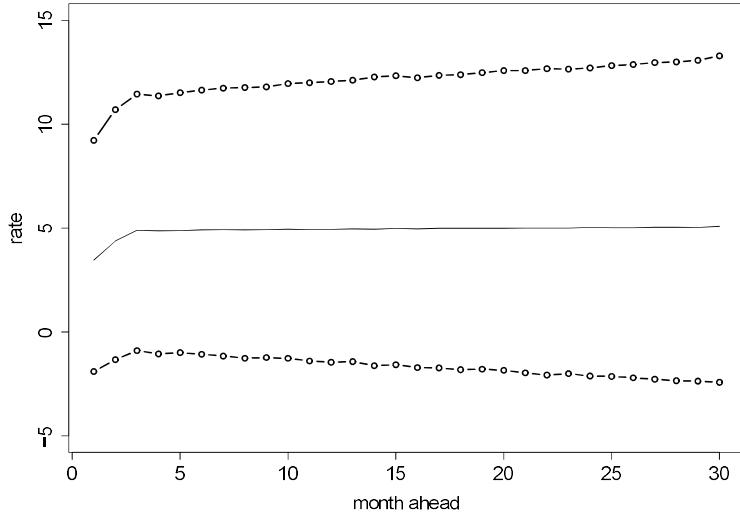
intervals based on sampling heavy-tailed residuals would be wider than those based on a Gaussian assumption.

As described above, forecasts for future inflation rates were obtained by taking partial sums of random forecasts of changes in the inflation rate and the forecast limits (upper and lower quantiles) are shown in Figure 9.22. As expected for a nonstationary process, the forecast limits diverge.  $\square$

There are two important advantages to using simulation for forecasting. They are

1. simulation can be used in situations where standard software does not compute forecast limits, and
2. simulation does not require that the noise series be Gaussian.

The first advantage will be important in some future examples, such as, multivariate AR processes fit by R's `ar` function. The second advantage is less important if one is generating 90% or 95% forecast limits, but if one wishes more extreme quantiles, say 99% forecast limits, then the second advantage could be more important since in most applications the noise series has heavier tails than Gaussian tails.



**Fig. 9.22.** Forecast limits for the inflation rate computed by simulation.

### 9.13 Partial Autocorrelation Coefficients

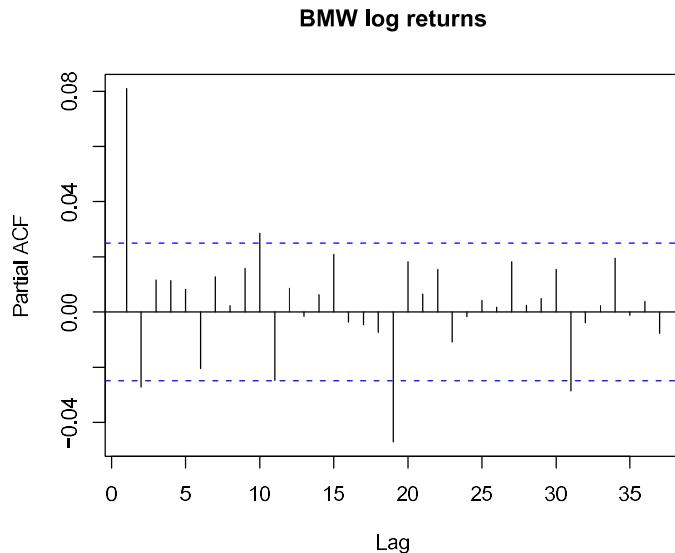
The partial autocorrelation function (PACF) can be useful for identifying the order of an AR process. The  $k$ th partial autocorrelation, denoted by  $\phi_{k,k}$ , for a stationary process  $Y_t$  is the correlation between  $Y_t$  and  $Y_{t+k}$ , conditional given  $Y_{t+1}, \dots, Y_{t+k-1}$ . For  $k = 1$ ,  $Y_{t+1}, \dots, Y_{t+k-1}$  is an empty set, so the partial autocorrelation coefficient is simply equal to the autocorrelation coefficient, that is,  $\phi_{1,1} = \rho(1)$ . Let  $\hat{\phi}_{k,k}$  denote the estimate of  $\phi_{k,k}$ .  $\hat{\phi}_{k,k}$  can be calculated by fitting the regression model

$$Y_t = \phi_{0,k} + \phi_{1,k}Y_{t-1} + \cdots + \phi_{k,k}Y_{t-k} + \epsilon_{k,t}.$$

If  $Y_t$  is an AR( $p$ ) process, then  $\phi_{k,k} = 0$  for  $k > p$ . Therefore, a sign that a time series can be fit by an AR( $p$ ) model is that the sample PACF will be nonzero up to  $p$  and then will be nearly zero for larger lags.

*Example 9.17.* PACF for BMW log returns

Figure 9.23 is the sample PACF for the BMW log returns. The large value of  $\hat{\phi}_{1,1}$  and the smaller values of  $\hat{\phi}_{k,k}$  for  $k = 2, \dots, 9$  are a sign that this time series can be fit by an AR(1) model, in agreement with the results in Example 9.4. Note that  $\hat{\phi}_{k,k}$  is outside the test bounds for some values of  $k > 9$ , particularly for  $k = 19$ . This is likely to be due to random variation.  $\square$

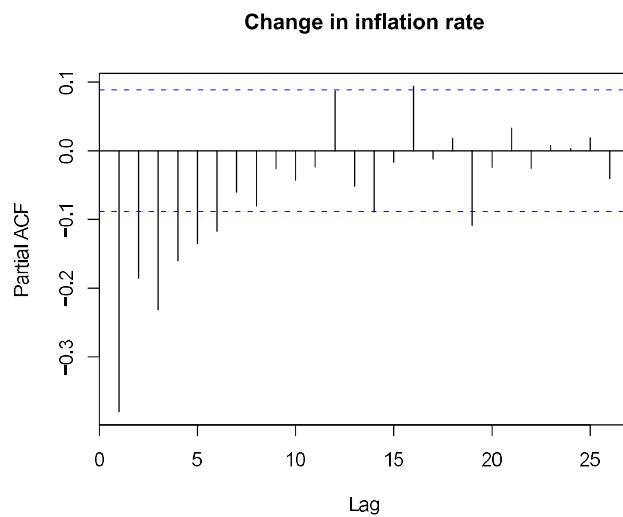


**Fig. 9.23.** Partial ACF for the BMW returns.

When computing resources were expensive, the standard practice was to identify a tentative ARMA model using the sample ACF and PACF, fit this model, and then check the ACF and PACF of the residuals. If the residual ACF and PACF revealed some lack of fit, then the model could be enlarged. As computing has become much cheaper and faster and the use of information-based model selection criteria has become popular, this practice has changed. Now many data analysts prefer to start with a relatively large set of models and compare them with selection criteria such as AIC and BIC. This can be done automatically by `auto.arima` in R or similar functions in other software packages.

*Example 9.18. PACF for changes in the inflation rate*

[Figure 9.24](#) is the sample PACF for the changes in the inflation rate. The sample PACF decays slowly to zero, rather than dropping abruptly to zero as for an AR process. This is an indication that this time series should not be fit by a pure AR process. An MA or ARMA process would be preferable. In fact, we saw previously that an MA(2) or MA(3) model provides a parsimonious fit.



**Fig. 9.24.** Sample PACF for changes in the inflation rate.

## 9.14 Bibliographic Notes

There are many books on time series analysis and only a few will be mentioned. Box, Jenkins, and Reinsel (2008) did so much to popularize ARIMA models that these are often called “Box–Jenkins models.” Hamilton (1994) is a comprehensive treatment of time series. Brockwell and Davis (1991) is particularly recommended for those with a strong mathematical preparation wishing to understand the theory of time series analysis. Brockwell and Davis (2003) is a gentler introduction to time series and is suited for those wishing to concentrate on applications. Enders (2004) and Tsay (2005) are time series textbooks concentrating on economic and financial applications; Tsay (2005)

is written at a somewhat more advanced level than Enders (2004). Gouriéroux and Jasiak (2001) has a chapter on the applications of univariate time series in financial econometrics, and Alexander (2001) has a chapter on time series models. Pfaff (2006) covers both the theory and application of unit root tests.

## 9.15 References

- Alexander, C. (2001) *Market Models: A Guide to Financial Data Analysis*, Wiley, Chichester.
- Box, G. E. P., Jenkins, G. M., and Reinsel, G. C. (2008) *Time Series Analysis: Forecasting and Control*, 4th ed., Wiley, Hoboken, NJ.
- Brockwell, P. J. and Davis, R. A. (1991) *Time Series: Theory and Methods*, 2nd ed., Springer, New York.
- Brockwell, P. J. and Davis, R. A. (2003) *Introduction to Time Series and Forecasting*, 2nd ed., Springer, New York.
- Enders, W. (2004) *Applied Econometric Time Series*, 2nd Ed., Wiley, New York.
- Gouriéroux, C., and Jasiak, J. (2001) *Financial Econometrics*, Princeton University Press, Princeton, NJ.
- Hamilton, J. D. (1994) *Time Series Analysis*, Princeton University Press, Princeton, NJ.
- Pfaff, B (2006) *Analysis of Integrated and Cointegrated Time Series with R*, Springer, New York.
- Tsay, R. S. (2005) *Analysis of Financial Time Series*, 2nd ed., Wiley, New York.

## 9.16 R Lab

### 9.16.1 T-bill Rates

Run the following code to input the `Tbrate` data set in the `Ecdat` package and plot the three quarterly time series in this data set as well as their auto- and cross-correlation functions. The last three lines of code run augmented Dickey–Fuller tests on the three series.

```
data(Tbrate, package="Ecdat")
library(tseries)
# r = the 91-day treasury bill rate
# y = the log of real GDP
# pi = the inflation rate
plot(Tbrate)
acf(Tbrate)
adf.test(Tbrate[,1])
adf.test(Tbrate[,2])
adf.test(Tbrate[,3])
```

**Problem 1**

- (a) *Describe the signs of nonstationarity seen in the time series and ACF plots.*
- (b) *Use the augmented Dickey–Fuller tests to decide which of the series are nonstationary. Do the tests corroborate the conclusions of the time series and ACF plots?*

Next run the augmented Dickey–Fuller test on the differenced series and plot the differenced series using the code below. Notice that the `pairs` function creates a scatterplot matrix, but the `plot` function applied to time series creates time series plots. [The `plot` function would create a scatterplot matrix if the data were in a `data.frame` rather than having “class” time series (`ts`). Check the class of `diff_rate` with `attr(diff_rate, "class")`.] Both types of plots are useful. The former shows cross-sectional associations, while the time series plots are helpful when deciding whether differencing once is enough to induce stationarity. You should see that the first-differenced data look stationary.

```
diff_rate = diff(Tbrate)
adf.test(diff_rate[,1])
adf.test(diff_rate[,2])
adf.test(diff_rate[,3])
pairs(diff_rate)           # scatterplot matrix
plot(diff_rate)           # time series plots
```

Next look at the autocorrelation functions of the differenced series. These will be on the diagonal of a  $3 \times 3$  matrix of plots. The off-diagonal plots are cross-correlation functions, which will be discussed in Chapter 10 and can be ignored for now.

```
acf(diff_rate)           # auto- and cross-correlations
```

**Problem 2**

1. *Do the differenced series appear stationary according to the augmented Dickey–Fuller tests?*
2. *Do you see evidence of autocorrelations in the differenced series? If so, describe these correlations.*

For the remainder of this lab, we will focus on the analysis of the 91-day T-bill rate. Since the time series are quarterly, it is good to see if the mean depends on the quarter. One way to check for such effects is to compare boxplots of the four quarters. The following code does this. Note the use of the `cycle` function to obtain the quarterly period of each observation; this information is embedded in the data and `cycle` simply extracts it.

```
par(mfrow=c(1,1))
boxplot(diff_rate[,1] ~ cycle(diff_rate))
```

**Problem 3** Do you see any seasonal differences in the boxplots? If so, describe them.

Regardless of whether seasonal variation is present, for now we will look at nonseasonal models. Seasonal models are introduced in Section 10.1. Next, use the `auto.arima` function in the `forecast` package to find a “best-fitting” nonseasonal arima model for the T-bill rates. The specifications `max.P=0` and `max.Q=0` force the model to be nonseasonal, since `max.P` and `max.Q` are the number of seasonal AR and MA components.

```
library(forecast)
auto.arima(Tbrate[,1],max.P=0,max.Q=0,ic="aic")
```

#### Problem 4

1. What order of differencing is chosen? Does this result agree with your previous conclusions?
2. What model was chosen by AIC?
3. Which goodness-of-fit criterion is being used here?
4. Change the criterion to BIC. Does the best-fitting model then change?

Finally, refit the best-fitting model with the following code, and check for any residual autocorrelation. You will need to replace the three question marks by the appropriate numerical values for the best-fitting model.

```
fit1 = arima(Tbrate[,1],order=c(?, ?, ?))
acf(residuals(fit1))
Box.test(residuals(fit1), lag = 10, type="Ljung")
```

**Problem 5** Do you think that there is residual autocorrelation? If so, describe this autocorrelation and suggest a more appropriate model for the T-bill series.

GARCH effects, that is, volatility clustering, can be detected by looking for auto-correlation in the mean-centered squared residuals. Another possibility is that some quarters are more variable than others. This can be detected for quarterly data by autocorrelation in the squared residuals at time lags that are a multiple of 4. Run the following code to look at autocorrelation in the mean-centered squared residuals.

```
resid2 = residuals(fit1)^2
acf(resid2)
Box.test(resid2, lag = 10, type="Ljung")
```

**Problem 6** Do you see evidence of GARCH effects?

### 9.16.2 Forecasting

This example shows how to forecast a time series using R. Run the following code to fit a nonseasonal ARIMA model to the quarterly inflation rate. The code also uses the `predict` function to forecast 36 quarters ahead. The standard errors of the forecasts are also returned by `predict` and can be used to create prediction intervals. Note the use of `col` to specify colors. Replace `c(?, ?, ?)` by the specification of the ARIMA model that minimizes BIC.

```
data(Tbrate, package="Ecdat")
# r = the 91-day Treasury bill rate
# y = the log of real GDP
# pi = the inflation rate
# fit the nonseasonal ARIMA model found by auto.arima
auto.arima(pi, max.P=0, max.Q=0, ic="bic")
fit = arima(pi, order=c(?, ?, ?))
forecasts = predict(fit, 36)
plot(pi, xlim=c(1980, 2006), ylim=c(-7, 12))
lines(seq(from=1997, by=.25, length=36),
      forecasts$pred, col="red")
lines(seq(from=1997, by=.25, length=36),
      forecasts$pred + 1.96*forecasts$se,
      col="blue")
lines(seq(from=1997, by=.25, length=36),
      forecasts$pred - 1.96*forecasts$se,
      col="blue")
```

**Problem 7** Include the plot with your work.

- (a) Why do the prediction intervals (blue curves) widen as one moves farther into the future?
- (b) What causes the predictions (red) and the prediction intervals to wiggle initially?

## 9.17 Exercises

1. This problem and the next use CRSP daily returns. First, get the data and plot the ACF in two ways:

```
library(Ecdat)
data(CRSPday)
crsp=CRSPday[,7]
```

```
acf(crsp)
acf(as.numeric(crsp))
```

- (a) Explain what “lag” means in the two ACF plots. Why does lag differ between the plots?
- (b) At what values of lag are there significant autocorrelations in the CRSP returns? For which of these values do you think the statistical significance might be due to chance?

2. Next, fit AR(1) and AR(p) models to the CRSP returns:

```
arima(crsp,order=c(1,0,0))
arima(crsp,order=c(2,0,0))
```

- (a) Would you prefer an AR(1) or an AR(2) model for this time series?  
Explain your answer.

- (b) Find a 95% confidence interval for  $\phi$  for the AR(1) model.

3. Consider the AR(1) model

$$Y_t = 5 - 0.55Y_{t-1} + \epsilon_t$$

and assume that  $\sigma_\epsilon^2 = 1.2$ .

- (a) Is this process stationary? Why or why not?  
(b) What is the mean of this process?  
(c) What is the variance of this process?  
(d) What is the covariance function of this process?

4. Suppose that  $Y_1, Y_2, \dots$  is an AR(1) process with  $\mu = 0.5$ ,  $\phi = 0.4$ , and  $\sigma_\epsilon^2 = 1.2$ .

- (a) What is the variance of  $Y_1$ ?  
(b) What are the covariances between  $Y_1$  and  $Y_2$  and between  $Y_1$  and  $Y_3$ ?  
(c) What is the variance of  $(Y_1 + Y_2 + Y_3)/2$ ?

5. An AR(3) model has been fit to a time series. The estimates are  $\hat{\mu} = 104$ ,  $\hat{\phi}_1 = 0.4$ ,  $\hat{\phi}_2 = 0.25$ , and  $\hat{\phi}_3 = 0.1$ . The last four observations were  $Y_{n-3} = 105$ ,  $Y_{n-2} = 102$ ,  $Y_{n-1} = 103$ , and  $Y_n = 99$ . Forecast  $Y_{n+1}$  and  $Y_{n+2}$  using these data and estimates.

6. Let  $Y_t$  be an MA(2) process,

$$Y_t = \mu + \epsilon_t + \theta_1\epsilon_{t-1} + \theta_2\epsilon_{t-2}.$$

Find formulas for the autocovariance and autocorrelation functions of  $Y_t$ .

7. Let  $Y_t$  be a stationary AR(2) process,

$$(Y_t - \mu) = \phi_1(Y_{t-1} - \mu) + \phi_2(Y_{t-2} - \mu) + \epsilon_t.$$

- (a) Show that the ACF of  $Y_t$  satisfies the equation

$$\rho(k) = \phi_1\rho(k-1) + \phi_2\rho(k-2)$$

for all values of  $k > 0$ . (These are a special case of the Yule–Walker equations.)

[Hint:  $\gamma(k) = \text{Cov}(Y_t, Y_{t-k}) = \text{Cov}\{\phi_1(Y_{t-1} - \mu) + \phi_2(Y_{t-2} - \mu) + \epsilon_t, Y_{t-k}\}$  and  $\epsilon_t$  and  $Y_{t-k}$  are independent if  $k > 0$ .]

- (b) Use part (a) to show that  $(\phi_1, \phi_2)$  solves the following system of equations:

$$\begin{pmatrix} \rho(1) \\ \rho(2) \end{pmatrix} = \begin{pmatrix} 1 & \rho(1) \\ \rho(1) & 1 \end{pmatrix} \begin{pmatrix} \phi_1 \\ \phi_2 \end{pmatrix}.$$

(c) Suppose that  $\rho(1) = 0.4$  and  $\rho(2) = 0.2$ . Find  $\phi_1$ ,  $\phi_2$ , and  $\rho(3)$ .

8. Use (9.11) to verify equation (9.12).
9. Show that if  $w_t$  is defined by (9.34) then (9.35) is true.
10. The time series in the middle and bottom panels of Figure 9.14 are both nonstationary, but they clearly behave in different manners. The time series in the bottom panel exhibits “momentum” in the sense that once it starts moving upward or downward, it often moves consistently in that direction for a large number of steps. In contrast, the series in the middle panel does not have this type of momentum and a step in one direction is quite likely to be followed by a step in the opposite direction. Do you think the time series model with momentum would be a good model for the price of a stock? Why or why not?
11. The MA(2) model  $Y_t = \mu + \epsilon_t + \theta_1 \epsilon_{t-1} + \theta_2 \epsilon_{t-2}$  was fit to data and the estimates are

Parameter	Estimate
$\mu$	45
$\theta_1$	0.3
$\theta_2$	-0.15

The last two values of the observed time series and residuals are

$t$	$Y_t$	$\hat{\epsilon}_t$
$n-1$	39.8	-4.3
$n$	42.7	1.5

Find the forecasts of  $Y_{n+1}$  and  $Y_{n+2}$ .

12. The ARMA(1,2) model  $Y_t = \mu + \phi_1 Y_{t-1} + \epsilon_t + \theta_1 \epsilon_{t-1} + \theta_2 \epsilon_{t-2}$  was fit to data and the estimates are

Parameter	Estimate
$\mu$	103
$\phi_1$	0.2
$\theta_1$	0.4
$\theta_2$	-0.25

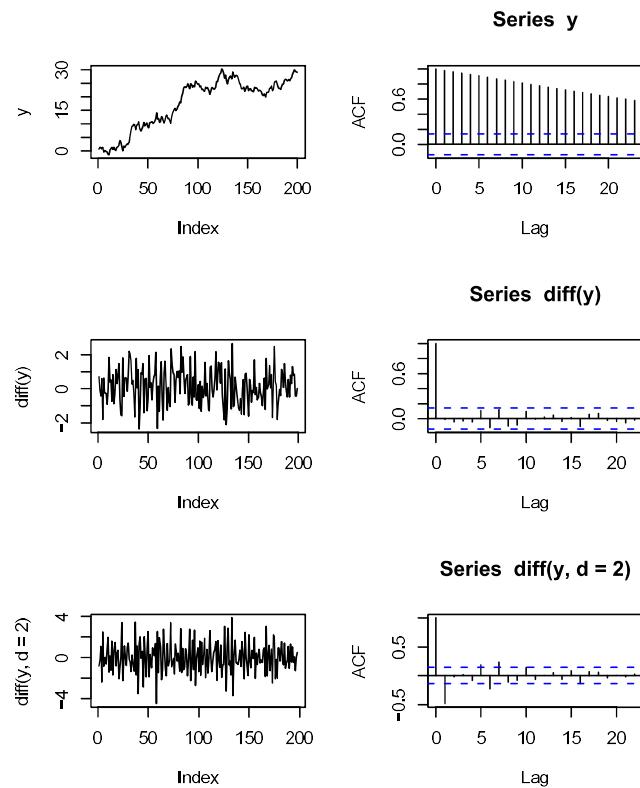
The last two values of the observed time series and residuals are

$t$	$Y_t$	$\hat{\epsilon}_t$
$n-1$	120.1	-2.3
$n$	118.3	2.6

- Find the forecasts of  $Y_{n+1}$  and  $Y_{n+2}$ .
13. To decide the value of  $d$  for an ARIMA( $p, d, q$ ) model for a time series  $y$ , plots were created using the R program:

```
par(mfrow=c(3,2))
plot(y,type="l")
acf(y)
plot(diff(y),type="l")
acf(diff(y))
plot(diff(y,d=2),type="l")
acf(diff(y,d=2))
```

The output was the following figure:



What value of  $d$  do you recommend? Why?

14. This problem fits an ARIMA model to the logarithms monthly one-month T-bill rates in the data set **Mishkin** in the **Ecdat** package. Run the following code to get the variable:

```
library(Ecdat)
data(Mishkin)
```

```
tb1 = log(Mishkin[,3])
```

- (a) Use time series and ACF plots to determine the amount of differencing needed to obtain a stationary series.
  - (b) Next use auto.arima to determine the best-fitting nonseasonal ARIMA models. Use both AIC and BIC and compare the results.
  - (c) Examine the ACF of the residuals for the model you selected. Do you see any problems?
15. Suppose you just fit an AR(2) model to a time series  $Y_t$ ,  $t = 1, \dots, n$ , and the estimates were  $\hat{\mu} = 100.1$ ,  $\hat{\phi}_1 = 0.5$ , and  $\hat{\phi}_2 = 0.1$ . The last three observations were  $Y_{n-2} = 101.0$ ,  $Y_{n-1} = 99.5$ , and  $Y_n = 102.3$ . What are the forecasts of  $Y_{n+1}$ ,  $Y_{n+2}$ , and  $Y_{n+3}$ ?
16. In Section 9.9.1, it was stated that “if  $E(Y_t)$  has an  $m$ th-degree polynomial trend, then the mean of  $E(\Delta^d Y_t)$  has an  $(m-d)$ th-degree trend for  $d \leq m$ . For  $d > m$ ,  $E(\Delta^d Y_t) = 0$ .” Prove these assertions.

